UNIVERSITY OF RIJEKA

FACULTY OF ENGINEERING

Daniel Štifanić

A DEEP LEARNING-BASED SYSTEM FOR ESTIMATING UNDERGROUND INFRASTRUCTURE PARAMETERS USING GPR DATA

DOCTORAL THESIS

Rijeka, 2026

UNIVERSITY OF RIJEKA

FACULTY OF ENGINEERING

Daniel Štifanić

A DEEP LEARNING-BASED SYSTEM FOR ESTIMATING UNDERGROUND INFRASTRUCTURE PARAMETERS USING GPR DATA

DOCTORAL THESIS

Thesis Supervisor: Prof. Zlatan Car, PhD

Thesis Co-supervisor: Assist. Prof. Nikola Anđelić, PhD

Rijeka, 2026

SVEUČILIŠTE U RIJECI

TEHNIČKI FAKULTET

Daniel Štifanić

SUSTAV ZA PROCJENU PARAMETARA PODZEMNE INFRASTRUKTURE IZ PODATAKA GPR-A TEMELJEN NA DUBOKOM UČENJU

DOKTORSKI RAD

Mentor: prof. dr. sc. Zlatan Car

Komentor: doc. dr. sc. Nikola Anđelić

Rijeka, 2026

Doctoral thesis supervisor: Prof. Zlatan Car, PhD, Catholic University of Croatia, Faculty of Engineering, Croatia

Doctoral thesis co-supervisor: Assist. Prof. Nikola Anđelić, PhD, University of Rijeka, Faculty of Engineering, Croatia

The doctoral thesis was defended on _____ at the University of Rijeka, Faculty of Engineering, Croatia, in front of the following Evaluation Committee:

1. Assoc. Prof. Ivan Volarić, PhD, University of Rijeka, Faculty of Engineering, Croatia – Committee Chair

2. Prof. Ivan Štajduhar, PhD, University of Rijeka, Faculty of Engineering, Croatia

3. Marko Budimir, PhD, INETEC – Institute for Nuclear Technology, Croatia

For my daughter – my motivation.

WORDS OF APPRECIATION

ABSTRACT

Non-destructive exploration and recording of underground infrastructure can be achieved by utilizing ground penetrating radar (GPR). Such infrastructure usually consists of cylindrical objects such as pipes, cables, rebar etc. which appear within the radar recording as hyperbola-shaped features. However, in real-world conditions, numerous factors can influence the shape of such features, further increasing the complexity of the analysis and interpretation. Radar recordings consist of a series of merged one-dimensional signals representing a forward time register of reflected events arriving at the radar receiver. Accordingly, a hybrid approach that combines time-frequency distributions (TFDs) with a deep convolutional neural network (CNN) is utilized in order to analyze hyperbola-shaped features with the purpose of estimating two-way travel time (TWTT), burial depth, cylindrical object diameter, and intersection angle (α).

In order to develop a deep learning (DL) based system, real-world GPR data is acquired utilizing a GSSI UtilityScan GPR system across multiple survey locations within the city of Rijeka, Croatia. A dedicated labeling protocol is developed to map precise geospatial records from a geographic information system (GIS) to the GPR survey lines, enabling the automated generation of bounding box annotations for detected hyperbola-shaped reflections. Additionally, synthetic GPR data generated using gprMax simulation software is incorporated into the training process to increase sample diversity and improve the generalization of the utilized models.

The proposed system is structured as a two-stage framework. In the first stage, the you only look once v11 (YOLOv11x) object detection model is employed for the detection and localization of hyperbola-shaped reflections within B-scans. In the second stage, the detected reflections are analyzed by the proposed cross-attention spatio-spectral fusion network, referred to as CASS-Fusion. The proposed network integrates spatial B-scan features extracted by an Xception backbone with localized spectral representations derived from Cohen's class TFDs, fused via a multi-head cross-attention mechanism. The experimental results show that the

proposed hybrid architecture consistently outperformed both the standalone spatial and standalone spectral models across all four descriptive parameters, validating the core assumption that spatial and spectral representations encode complementary and partially independent information about the underground infrastructure.

PROŠIRENI SAŽETAK

Nedestruktivno istraživanje i precizno mapiranje podzemne infrastrukture od apsolutne su važnosti za sigurno izvođenje građevinskih radova i napredno urbanističko planiranje. U tu se svrhu kao primarna tehnologija koristi georadar. Takva se infrastruktura uglavnom sastoji od cilindričnih objekata kao što su cijevi, kablovi, armaturne šipke itd., a koje se unutar radarskog zapisa (*eng. B-scan*) pojavljuju kao vizualne značajke hiperboličkog oblika. Međutim, u realnim terenskim uvjetima, brojni čimbenici poput heterogenosti tla, atenuacije signala te preklapanje refleksija značajno utječu na oblik takvih značajki, što eksponencijalno povećava kompleksnost analize i pravilne interpretacije. Često korišteni matematički modeli koji se temelje na preklapanju hiperbole i hiperboličkih značajki, inherentno su ograničeni nedostatkom prostornog zaključivanja potrebnog za rješavanje dvosmislenosti između brzine širenja elektromagnetskog vala i stvarne dubine objekata. Shodno navedenom, u ovom se istraživanju primjenjuje napredni hibridni pristup temeljen na dubokom učenju koji integrira vremensko-frekvencijske distribucije s dubokim konvolucijskim neuronskim mrežama kako bi se iz hiperboličkih značajki precizno procijenili ključni parametri, kao što su dvosmjerno vrijeme putovanja signala, dubina ukopa, promjer cilindričnog objekta te prilazni kut (α).

U svrhu razvoja sustava temeljenog na dubokom učenju, prikupljeni su georadarski podaci u stvarnim uvjetima pomoću visoko-rezolucijskog georadarskog sustava GSSI UtilityScan na više mjernih lokacija na području grada Rijeke, Hrvatska. Nadalje, razvijen je namjenski protokol za označavanje podataka kako bi se precizni geoprostorni zapisi iz geografskog informacijskog sustava preslikali na radarske zapise te samim time omogućili automatsko generiranje anotacija u obliku graničnih okvira koji obuhvaćaju refleksije hiperboličkog oblika. Dodatno, sintetički georadarski podaci generirani pomoću gprMax programske podrške za elektromagnetsku simulaciju uključeni su u proces treniranja kako bi se povećala raznolikost uzoraka i poboljšala generalizacija korištenih modela.

Predloženi sustav strukturiran je kao hijerarhijski razvojni okvir u dvije faze. U prvoj se fazi koristi YOLOv11x model za detekciju i lokalizaciju hiperboličkih refleksija unutar

radarskog zapisa. Eksperimentalna evaluacija dokazala je da YOLOv11x osigurava superiorne performanse i stabilnost u odnosu na RT-DETR-X model. U drugoj se fazi prethodno detektirane refleksije analiziraju pomoću predložene prostorno-spektralne fuzijske mreže s unakrsnom pažnjom, nazvane CASS-Fusion. Predložena mreža integrira prostorne značajke radarskog zapisa izdvojene pomoću Xception okosnice s lokaliziranim spektralnim reprezentacijama izvedenim iz vremensko-frekvencijskih distribucija iz Cohenove klase. Spajanje ovih potpuno različitih informacijskih domena vrši se putem mehanizma unakrsne pažnje. Eksperimentalni rezultati pokazuju da je predložena hibridna arhitektura nadmašila i samostalne prostorne i samostalne spektralne modele u procjeni sva četiri opisna parametra podzemne infrastrukture. To nadalje potvrđuje temeljnu pretpostavku da prostorne i spektralne reprezentacije sadrže komplementarne i djelomično neovisne informacije o podzemnoj infrastrukturi. Dodatno, istraživanje je otkrilo ključnu zakonitost: odabir optimalne spektralne reprezentacije izravno je uvjetovan fizičkom kompleksnošću ciljanog parametra. Drugim riječima, Wigner-Villeova distribucija se pokazala optimalnom za procjenu dvosmjernog vremena putovanja signala i promjera cjevovoda, dok je Born-Jordanova distribucija ostvarila optimalne rezultate pri procjeni dubine ukopa i prilaznog kuta. Naposljetku, cjelokupni sustav u potpunosti je validiran na terenskim georadarskim podacima, čime je uspješno uspostavljen dokaz koncepta u stvarnim radnim uvjetima.

Ključne riječi: Georadar, opisni parametri podzemne infrastrukture, detekcija i lokalizacija objekata, regresija temeljena na dubokom učenju, vremensko-frekvencijske distribucije, fuzija prostorno-spektralnih značajki.

TABLE OF CONTENTS

CHAPTER 5

CHAPTER 6

CHAPTER 7

CHAPTER 1

INTRODUCTION

Cities are increasingly using modern technologies to provide citizens with a better standard of living through information and communication technology (ICT) [120]. In the context of smart industries and communal services, specifically water supply and the energy sector, the implementation of modern technological solutions is essential to maximize efficiency and minimize losses [87]. Detailed documentation of the existing state of underground infrastructure serves as a fundamental prerequisite for the correct evaluation of these systems and the strategic planning of future investments [68]. However, a certain part of existing data, which mostly contains infrastructure built in the last century, is frequently characterized by low quality and incoherence [41]. Updating such data using classical methods presents a time-consuming and costly challenge, which can negatively impact the realization of new capital investments [32]. Furthermore, costs rise unnecessarily due to the difficulty of executing maintenance and intervention. Organized data on underground infrastructure is indeed critical for effective system management, as it directly impacts service reliability, operational efficiency, and long-term infrastructure sustainability.

Ground penetrating radar (GPR) is widely recognized as an efficient, non-destructive technique for subsurface surveying [51, 58, 70]. This method utilizes propagating electromagnetic (EM) waves to record changes in the EM characteristics of the subsurface [93]. Accordingly, radar scans consist of a series of 1D signals representing a time register of reflected events arriving at the radar receiver [61]. The relative contrast in permittivity between the background material and the underground infrastructure determines the propagation velocity of EM waves, which is the key factor in creating the radar reflection within the scan

[93]. The reliance on human-computer interaction (HCI) for interpreting GPR data is a well-documented bottleneck in geophysical signal processing [79]. While GPR data acquisition is physically standardized, the processing and interpretation stages remain highly operator-dependent, functioning as a quasi-manual signal processing loop where the operator's knowledge and experience directly shape the final output [95]. Additionally, manual analysis is demanding and inefficient, which further encourages the development of automatic systems for processing and analyzing this type of data. Due to the exponential increase in data volume, advancements in computational hardware, and progress in artificial intelligence (AI) algorithms, automated systems for the analysis and interpretation of GPR data are becoming increasingly common [94, 135]. Nevertheless, estimating the descriptive parameters of underground infrastructure remains a significant challenge [56]. Complex subsurface scattering and noise can obscure target features; therefore, traditional approaches often struggle with the non-stationary nature of GPR signals [19]. Such limitations can be particularly problematic for estimating subtle geometric parameters like pipeline diameter and orientation, which, unlike burial depth, are not directly encoded in simple signal attributes [110]. In order to address such challenge, new hybrid approach must combine the signal insights of TFDs with the adaptability of DL. Combining these methods can result in robust simultaneous estimation of multiple descriptive parameters of underground infrastructure in complex, real-world environments.

1.1 Related Work

Facilitated by the rapid development of AI technology, machine learning (ML) methods are progressively being utilized to process GPR data. Some of the commonly used algorithms are support vector machines (SVM), artificial neural networks (ANN), Naïve Bayes classifier, a genetic algorithm (GA) [5]. DL methods have also demonstrated broad applicability in the classification, detection, localization, and segmentation of GPR data. Therefore, such methods enable the direct learning of feature representations from the data [51].

In 1995, during a period of renewed interest in ANN, the first research paper using ANN to analyze GPR data was published by Molyneaux et al. [83]. The authors used three-layer, fully connected network to detect the presence and the size of a bar. The findings demonstrated the neural network's ability to identify the position of the rebar, but not its size. Overall, this preliminary research was a significant step forward in the use of ANN for automatic detection

of GPR object reflections. In the early 2000s, the popularity of ANNs declined significantly due to computational limitations, overfitting concerns, and the requirement for large training datasets. For that reason, Al-Nuaimy et al. [3] developed a series of signal and image processing steps before using a neural network in order to achieve high accuracy in GPR analysis. Around year 2004, SVM began to gain popularity for automatic detection in GPR images. Therefore, Zhang et al. [131] used SVM for landmine feature extraction and classification. A linear SVM was used to extract the important features to maximize the separation margin, reduce the number of misclassified samples, and provide better generalization. Following that, the SVM classifier's performance was evaluated using recursive classification and leave-one-out cross-validation. Pasolli et al. [90] proposed a novel pattern recognition system for the classification and identification of buried objects based on GPR imagery. Following the preprocessing and segmentation stages, the detection operation is performed iteratively using the proposed genetic optimized algorithm, while the classification of the material type is solved using a SVM classifier. Regardless of the complexity of GPR images, the experimental results show that the proposed system performs well in terms of detection and material recognition. In order to locate underground infrastructure, Birkenfeld [12] used a multilayer feed-forward network to detect hyperbola reflection in GPR data. Before defining the training process, radargrams (i.e., B-scans) were preprocessed to enhance relevant features, and then trapezoidal bounded images containing half of a hyperbolic reflection were used as the training data. Hundreds of radargrams were recorded at various locations to ensure that the detection system was as sensitive to a wide range of geological conditions as possible. Such an approach is proven to be appropriate for an automatic hyperbola detection system.

In 2015, deep convolutional neural networks (CNNs) started to be used for GPR data, despite their widespread use in other fields. Besaw and Stimac [11] in their research demonstrated deep CNNs' ability to extract meaningful features from 2-dimensional GPR B-scans and classify threats. In 2019, the application of DL method for real-time underground target detection using vehicle-borne GPR data was investigated by Zong et al. [137]. The research relied on a dataset of 489 labelled samples which included cables, rainwater wells, voids, metal and nonmetal pipes, as well as sparse and dense steel reinforcement. For feature extraction of the buried objects, the authors utilized DarkNet53, while YOLO algorithm was selected for classification and localization purposes. Experimental results indicated that the proposed method is effective for real-time detection and classification with an average precision (AP) of 0.869 and a recall of 0.892. Furthermore, Liu et al. [77] used a DL method for the

automatic detection and localization of reinforced steel bars in concrete. The dataset consisted of 13026 rebar targets in 3992 GPR images, which are collected on residential buildings under construction. The proposed AI method, single shot multibox detector (SSD), achieved an AP of 0.909 and area under the curve (AUC) of 0.94, which proves that the proposed method can be valuable for rebar detection with high accuracy in real-time. Feng et al. [35] in their research present a three-module system for the detection and reconstruction of underground targets. The first module, visual-inertial fusion is used to generate the pose information of GPR data. Then in the second module, the deep neural network is used to detect the B-scan of the GPR image, extract features to remove noise and predict dielectric to determine the depth of objects. The third module represents a 3D GPR migration module that reconstructs and visualizes the 3D underground targets. The used neural network, DepthNet, resulted in an average accuracy of 92.64% for B-scan feature detection. Based on YOLOv3, Li et al. [73] introduced a technique designed to identify voids in concrete or soil structures and parabolic targets of various sizes. The proposed approach yielded a mean average precision (mAP) of 83.17% for single-class pattern recognition, whereas the detection of metal bars within concrete structures achieved a mAP score of 76.10%. Yamaguchi et al. [127] demonstrated a novel detection model based on a 3D-CNN and Kirchhoff migration for subsurface pipe detection. The developed model's classification accuracy was up to 91% whereas in the case of transverse pipes, 3D-CNN improved classification accuracy by about 6% over 2D-CNN. Furthermore, using measurement and experimental field data examples it was observed that the two directions' pipes were successfully detected using a box-by-box search. By combining YOLO models and 3D GPR images, Liu et al. [80] in their research aimed to improve detection efficiency and increase the maintenance benefits of an asphalt road. According to the comparison results of various YOLO series models, the YOLOv5m model is chosen to conduct the rapid identification of road defects. The chosen model resulted in mAP values higher than 90%. Training reliable DL models requires large amounts of labelled data in order to achieve satisfactory performance. For that reason, Zhang et al. [132] proposed a generative adversarial nets (GANs)-based deep learning framework which consists of two components. Firstly, it generates new training data to address the scarcity of GPR data. Then, based on the DL approach it automatically detects hyperbola in the GPR B-scan image. According to the experimental results, the proposed method outperforms the existing state-of-the-art methods and achieves a mAP of 97%.

While the previously discussed techniques demonstrate the capability to classify or initially identify target regions, a significant limitation remains in their ability to adequately

characterize hyperbola-shaped reflections. The focus of research extends beyond the detection and localization of hyperbola-shaped reflections. The importance is also given to the estimation of critical parameters, including burial depth, cylindrical object diameter, EM wave velocity, and the intersection angle between the GPR survey line and the cylindrical object. Such parameters can potentially be extracted from the hyperbola-shaped reflections captured by GPR. Hyperbola fitting is a popular method for obtaining aforementioned information from these reflections. Nevertheless, a review of the existing literature indicates that only a limited number of researchers have successfully addressed some of these specific challenges. The research of Ristic et al. [96] offers a method for estimating both cylindrical object radius and EM wave propagation velocity from GPR data. The proposed method has an advantage in eliminating the measurement of velocity and its impact on radius estimation quality. Accurate estimation of velocity leads to better soil characterization. The suggested method outperforms existing methods in terms of accuracy and robustness, particularly when dealing with noise and raw data. In 2016, Dou et al. [31] proposed a real-time technique for the automatic recognition and fitting of hyperbolas in GPR data. The authors utilized a novel clustering algorithm to efficiently isolate and identify target signatures. Accordingly, this method successfully handles complex scenarios, including the detection of incomplete, distorted, or intersecting hyperbolic reflections. Furthermore, Lei et al. [72] in their research presented an automatic method of hyperbola detection and fitting. The proposed scheme for interpreting GPR images consists of faster R-CNN, preprocessing method, double cluster seeking estimate algorithm, and column-based transverse filter points method. The estimation of rebar diameter in facilities using GPR data was explored by Park et al. [89] using the YOLOv3 algorithm. A comparison of input data types revealed that migrated data provided superior accuracy over standard B-scan image data. Accordingly, the method achieved a mAP value of 93.89%, demonstrating that such a technique is valid for both rebar area detection and diameter estimation. Liu et al. [78] presented an automated YOLOv3 DL approach for underground infrastructure recognition and localization. Features corresponding to underground infrastructure were localized within the GPR data, after which migration and binarization were employed to estimate burial depth. Furthermore, Zhu and Ye [135] demonstrated a modular detection and quantitative inversion method for hyperbolic features in GPR B-scan imagery. In the first step, YOLOv7 is applied to automatically localize and identify hyperbolic features in the B-scans. The second step proposes a two-stage curve-fitting technique based on previously detected features. The approach performs quantitative inversion of depth and radius based on several critical points of the hyperbolic pattern and the GPR system characteristics. Research results indicate the success of

the developed method in estimating the depth and radius of underground infrastructure, provided that the radar's trajectory is perpendicular to the alignment of the target. However, such geometric assumptions are frequently invalid in real-world survey conditions. Barkataki et al. [8] proposed a model for estimating the diameter of underground targets from GPR data. By applying the fast Fourier transform (FFT), input data initially collected in the time domain were transformed into the frequency domain and utilized to develop an ANN. Experimental testing confirmed the effectiveness of the proposed approach in estimating object diameter. Xu et al. [125] proposed a methodology based on DL for estimating the burial depth and radius of underground pipelines. A parabola model was used instead of a hyperbola model to, according to the authors, effectively obtain the symmetry axis of the pipeline image, which, along with a series of 1D data, was used as input to long short-term memory (LSTM) neural networks for burial depth estimation. Radius estimation, on the other hand, was performed using CNNs based on the aforementioned depth estimate and additional GPR derived features. Synthetic test samples were generated using gprMax software in order to validate the effectiveness of the proposed methodology. He and Lai [43] in their research provided a thorough comparison of two hyperbolic and three non-hyperbolic fitting models by mathematically expressing them as a single optimization problem. The authors examined the influence of the distance between the transmitter and receiver within the antenna, object diameter, burial depth, and relative soil permittivity on the performance of mathematical models describing hyperbolic features. To quantitatively estimate the performance of the models, a novel cost function (the C-value) was proposed, and the evaluation was accordingly structured as an optimization problem. The results reveal that various parameters have distinct influences on the performance of the models relative to reality. Jin et al. [60] introduced GPR-Former, a transformer-based neural network architecture designed for the automatic extraction and fitting of hyperbolic signatures in GPR B-scans. The proposed methodology localizes subsurface target responses in order to estimate the mathematical shape parameters associated with hyperbolic features. Such an approach enables the determination of key parameters such as size, depth, and material. Experimental results indicate that GPR-Former automatically and efficiently extracts hyperbolic patterns from B-scan imagery with a significant improvement in performance compared to state-of-the-art methods.

In order to address structural integrity, Hou et al. [52] proposed a method for estimating the width and depth of hidden cracks in tunnel linings. The authors combined features from the time-frequency (TF) domain with a neural network. Experimental results indicate higher

accuracy and generalization capability for models trained on features from the TF domain compared to models trained only with features from the time domain. Considering that in practical applications information about the radar's approach angle relative to underground infrastructure is often unknown, Lai et al. [69] conducted an analysis on that topic. Experimental results revealed that an approach angle of 45˚ can produce velocity estimation errors of up to 30%. Consequently, the authors proposed a reliable method for correcting such errors.

Considering the aforementioned literature review, only He and Lai [44] presented a new technique based on hyperbolic patterns found within GPR data. The technique performs estimation of EM wave propagation velocity, pipeline orientation, and burial depth. The innovation of this method focuses on the inclusion of angle correction index in classical mathematical models describing hyperbolic features. Such capability represents significant advancement, as orientation is typically treated as a known prior in standard workflows despite the difficulty of accurately determining it in real-world field settings.

1.2 Research Motivation and Hypotheses

The exhaustive literature review presented above reveals a critical gap in the current body of knowledge: at the time of this research, no prior work has developed a DL based system that leverages TFDs to simultaneously estimate the complex multi-parameter characteristics of underground infrastructure. Specifically, existing methodologies fail to comprehensively address the simultaneous estimation of burial depth, pipeline diameter, EM wave velocity, and the intersection angle between a GPR survey line and a pipeline.

In order to address this gap, the main goal of the proposed research is to develop a DL based system capable of estimating the descriptive parameters of underground infrastructure and the surrounding medium by integrating TFDs with deep CNNs. In accordance with the aforementioned goal, the following research hypotheses are proposed:

- Utilizing deep learning techniques, it is possible to achieve high performance in the detection and localization of underground infrastructure objects.

- Through a customized deep neural network architecture, it is possible to achieve the estimation of underground infrastructure parameters (burial depth, cylindrical object diameter, and the intersection angle between the GPR survey line and the target object), from which the EM wave velocity of the surrounding medium can be derived.

Based on the aforementioned hypotheses, the scientific contributions of the thesis are defined as follows:

- A protocol for labeling GPR data based on precise geospatial records from GIS.

- A DL-based model for the detection and localization of underground infrastructure objects.

- A hybrid model for estimating the descriptive parameters of underground infrastructure and the surrounding medium based on TFDs and CNNs.

1.3 Research Methodology

The main goal of the proposed research is to develop a two-stage framework in which the first stage performs detection and localization of underground infrastructure objects, and the second stage estimates descriptive parameters of the infrastructure and the surrounding medium. The first step in developing the aforementioned framework is acquiring a dataset of real-world GPR data utilizing a Geophysical Survey Systems Inc. (GSSI, Nashua, NH, USA) UtilityScan GPR system. Afterwards, a dedicated labeling protocol will be developed to link precise geospatial records from a GIS to the GPR survey line. Such an approach will identify intersections with mapped utilities and generate bounding boxes in the corresponding GPR B-scans. As an additional layer of quality control, each annotation will be manually validated and refined by domain experts. In parallel, synthetic GPR data will be generated with gprMax simulation software in order to increase sample diversity and enable controlled variation of physical parameters. GPR data will then be preprocessed before the training of the models. In the first stage, the labeled data will be used to train object detection algorithms to identify hyperbola-shaped features associated with underground infrastructure. In the second stage, detected and localized hyperbola-shaped features will be used to train a hybrid DL model that estimates key descriptive parameters, including burial depth, cylindrical object diameter, EM

wave velocity, and the intersection angle between the GPR survey line and the target object. The hybrid model will be trained on sequential features obtained by combining detection outputs with TFDs from Cohen's class. Performance evaluation will be performed using k-fold cross-validation, using detection- and regression-oriented metrics as appropriate for each stage. In addition, a traditional hyperbola-fitting model will be implemented as a baseline for parameter estimation to support a direct comparison against the proposed DL based system.

1.4 Organization of the Thesis

The thesis is structured into seven chapters which progressively introduce the physical and methodological foundation of the research.

Chapter 1 introduces the research problem and establishes broader motivation for reliable characterization of underground infrastructure in modern utility management. Based on the research gap identified in the literature review, the chapter defines the main research focus and formulates the corresponding hypotheses, while also outlining the expected scientific contributions.

Chapter 2 provides the theoretical background required to interpret GPR data in the context of subsurface imaging. In other words, it describes the principles of EM wave propagation and the formation of hyperbola-shaped reflections associated with cylindrical targets. Such reflections can be mathematically characterized; therefore, the hyperbola-fitting model is presented. Furthermore, the chapter covers the acquisition of real-world GPR data, as well as the generation of synthetic data.

In Chapter 3, the DL concepts and architectures utilized within the proposed two-stage framework are presented. More precisely, it includes object detection approaches for detecting and localizing hyperbola-shaped features in the first stage, as well as regression-oriented models for estimating the descriptive parameters of underground infrastructure in the second stage.

Chapter 4 introduces TFDs and TF analysis using Cohen's class distributions as the signal processing component required for the hybrid estimation approach. At the end of the chapter, TF representations of GPR A-scan samples are demonstrated.

In Chapter 5, the proposed system for parameter estimation of underground infrastructure is presented. The chapter describes the end-to-end workflow and methodology of the two-stage framework, which includes the integration of detected hyperbola-shaped features with Cohen's class TFDs. In addition, the performance evaluation metrics are described, which includes the use of k-fold cross-validation along with the selection of detection- and regression-oriented metrics for each stage of the framework.

Experimental results are presented in Chapter 6 along with a corresponding discussion of the obtained performance for both stages of the framework. The performance of each stage is presented, compared, and discussed, together with the performance of a traditional hyperbola-fitting model used as a baseline to support direct comparison against the proposed DL-based system.

As the last section of the thesis, Chapter 7 concludes the research by summarizing the main findings, discussing the potential limitations, and outlining directions for future work.

CHAPTER 2

PRINCIPLES OF GROUND PENETRATING RADAR WAVE PROPAGATION AND SUBSURFACE IMAGING

GPR, often referred to as georadar or ground probing radar, is a versatile geophysical tool with a wide range of applications. Although GPR was initially applied primarily to natural geological materials, it has since proven equally effective for a variety of other media such as wood, concrete, and asphalt [61]. By utilizing propagating EM waves, GPR interacts with subsurface structures and produces reflections in response to contrast in electromagnetic properties [6]. Typically, a GPR unit consists of a transmitting and receiving antenna. The acquisition of subsurface data can be described as follows: an EM pulse emitted by the transmitting antenna penetrates the ground, where it is either reflected off an interface or scattered off point sources (both induced by contrast in relative permittivity). Afterwards, the returning reflected or scattered energy then propagates back to the ground surface and is captured by the receiving antenna [6]. Relative permittivity characterizes a material's ability to store electromagnetic energy in response to an applied field and influences the propagation of EM waves. The contrast in relative permittivity between the background material and the object of interest affects the propagation velocity of EM waves and is the primary controlling factor

in reflection generation [6]. A larger contrast in relative permittivity leads to a greater impedance mismatch, resulting in stronger reflection. The described behavior is quantitatively governed by the Fresnel equations [105]. In GPR data, an A-scan is considered as the fundamental unit of measurement, representing a one-dimensional radar trace that captures reflected or scattered EM waves from subsurface structures at a single location [123]. In other words, an A-scan is a one-dimensional representation of EM signal amplitude versus the TWTT of the radar pulse. A sequence of A-scans acquired along a survey line forms a B-scan, also known as a radargram. A B-scan creation process utilizing a GPR illustrated in the x-z coordinate system is shown in Figure 2.1.
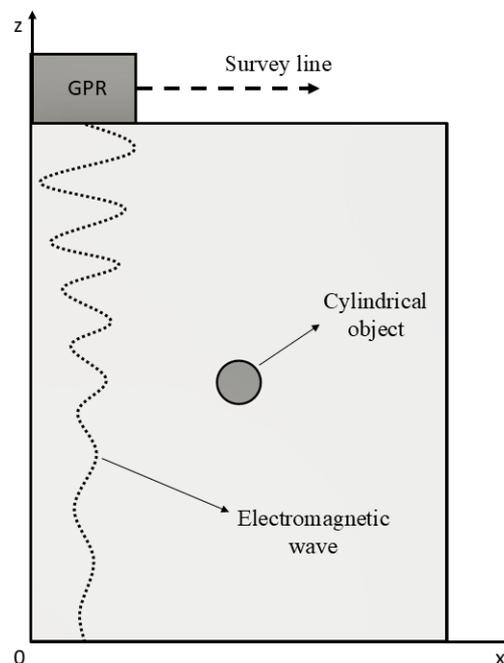
Figure 2.1 A B-scan creation process utilizing a GPR in x-z coordinate system.

According to Figure 2.1, the GPR survey line is represented by the x-axis, while the TWTT of the EM wave is represented by the z-axis. GPR data are typically characterized by weak, deep reflection echoes, which require the use of preprocessing techniques in order to enhance signal quality and interpretability [1]. Therefore, before further analysis, several standard preprocessing techniques need to be applied to all GPR data. First step is defining a time-zero position which is a reference point in the recorded radar trace (A-scan) that marks the moment when the transmitted EM wave leaves the transmitting antenna and begins propagating into the medium [130]. In theory, time-zero should coincide with the instant of transmission; however, in practice, various factors can influence the initial time pick. According to Yelf [129],

there are numerous approaches to determining a time-zero position, and not one correct method, which means that the choice of method usually depends on factors such as radar configuration, antenna design, internal delays in the electronics, subsurface conditions, etc. In practice, the time-zero position is determined either by adopting heuristic or calibration-based techniques. After the time-zero is accurately determined, the second step is usually applying a band-pass filter in order to attenuate both low-frequency drift and high-frequency random noise outside the effective bandwidth of the antenna system [9]. This way, the signal is restricted to the designed frequency range in which the antenna operates most efficiently. Third, an exponential gain function needs to be applied in order to compensate for signal attenuation with increasing depth and to improve the visibility of deeper echoes [28]. In other words, by utilizing such an approach, the weaker signals originating from greater depths are not overshadowed by stronger shallow reflections.

The following samples are generated within a homogeneous medium in order to isolate the case from the influence of a complex environment and noise effects. Samples of A-scans as well as the corresponding B-scan are shown in Figure 2.2. The left column presents individual A-scans obtained at distances of 0.0, 0.5, and 1.0 meters. Each A-scan represents the recorded amplitude of reflected EM wave as a function of time at a single location. Consecutive A-scans are recorded along the survey line of the GPR and aligned in sequence. By stacking A-scans, a B-scan is created as shown in the right column of Figure 2.2. The red ellipses and arrows indicate the position within the B-scan that corresponds to the individual A-scan shown in the left column. The cylindrical object utilized to generate the sample data has a diameter of 0.1 meters and is placed within a dielectric half-space with a relative permittivity of 9. The burial depth is 0.75 meters, and the GPR survey line is oriented perpendicular to the cylindrical object. These parameters directly influence the characteristics of the resulting reflection. Additionally, when the survey line is perpendicular to a cylindrical object, i.e. a subsurface pipeline, hyperbolic features appear within the B-scan. Moreover, the transformation from individual A-scans to a B-scan enables the visualization of subsurface structures as well as the identification and interpretation of buried objects of interest. In the presented sample, time-zero correction is not performed in order to visualize the influence of both the direct wave and the air wave. This way, the early-arriving signals remain visible in the B-scan and provide insight into how they may affect the subsequent interpretation. The direct wave represents the part of the EM signal that travels directly from the transmitting antenna to the receiving antenna [4].
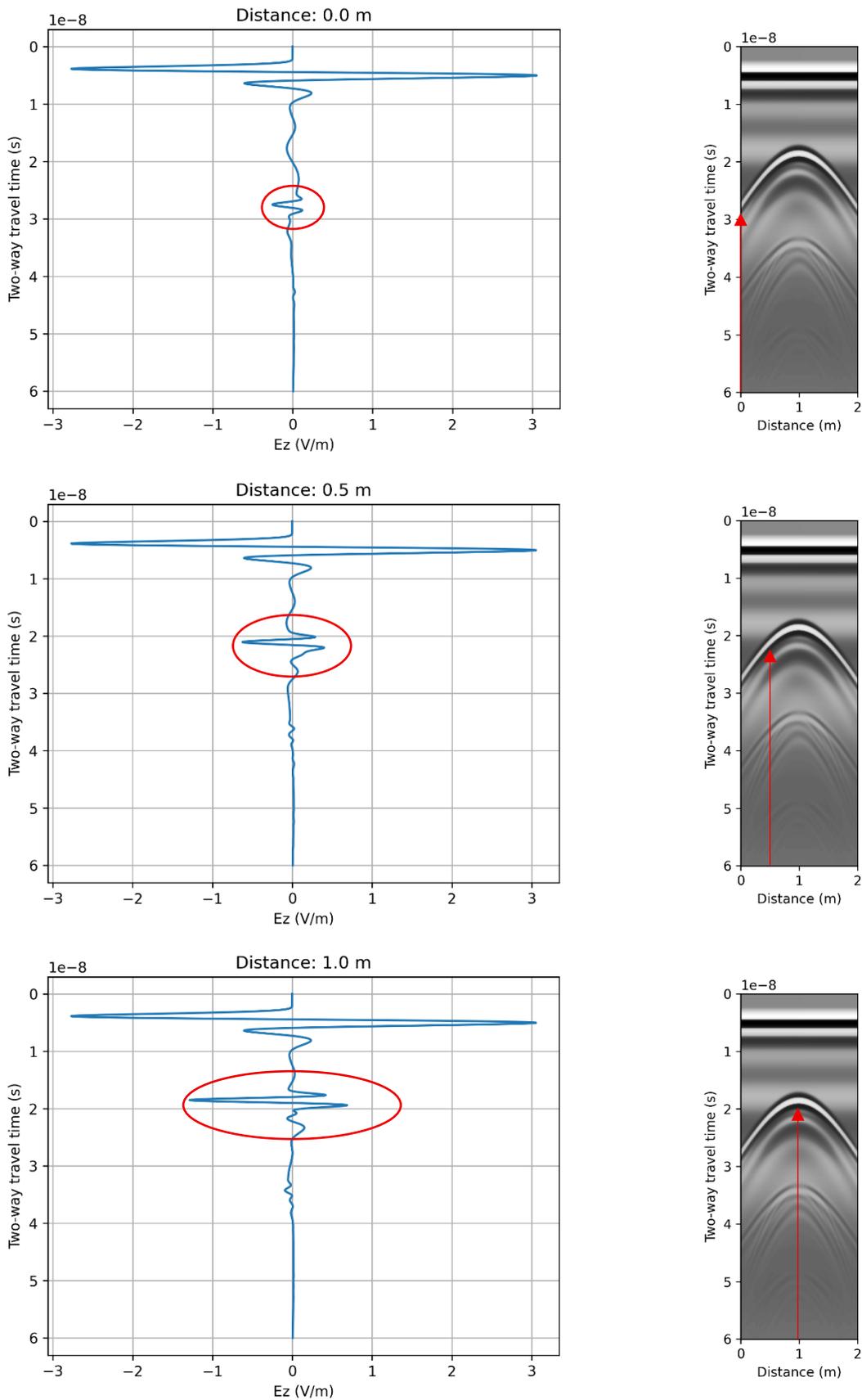
Figure 2.2 B-scan creation process. The left column displays A-scans recorded at distances of 0.0, 0.5, and 1.0 meters, while the right column shows the resulting B-scan, with arrows indicating the locations of the corresponding A-scans.

Such a signal arrives almost immediately since it travels through the air and does not contain subsurface reflection. However, the detection of direct wave is important since it can interfere with weaker reflected signals from shallow buried targets. On the other hand, an air wave refers to the part of the EM signal that travels through the air instead of propagating into the ground [4]. Such signals can reflect off the ground surface or nearby objects and produce artificial reflections that can be misinterpreted as genuine subsurface targets.

Sandy or dry soils typically have low relative permittivity, which leads to higher wave velocities, whereas clay-rich or moist soils slow down wave propagation. In GPR applications, the dependence of velocity on soil composition is crucial since directly influences the accuracy of depth estimation. Therefore, if the characteristics of the soil are known, the depth of the cylindrical objects in low-loss condition can be estimated utilizing the following equation [4]:

$$D = \frac{TWTT\ c}{2\sqrt{\varepsilon\mu}},$$
(2.1)

in which, the EM wave velocity is defined as

$$v = \frac{c}{\sqrt{\varepsilon\mu}},$$
(2.2)

where TWTT represents the two-way travel time of the EM wave, c denotes the velocity of EM waves in free space, ε is the relative permittivity, and μ stands for the relative permeability.

When only the burial depth parameter is observed with all other parameters fixed, a deeper cylindrical object will produce a wider (i.e., more open) hyperbolic reflection, whereas a shallower object will yield a narrower (i.e., more curved) reflection. This behavior is visually demonstrated in Figure 2.3, where in (a) the burial depth is 0.35 meters, and in (b) the burial depth is 1.75 meters. The samples are generated within a homogeneous medium with a relative permittivity of 9, using a cylindrical object with a diameter of 0.5 meters, and a GPR survey line oriented perpendicular to the object. The time-zero correction is performed; therefore, the grayish upper section of the B-scan is removed, and the TWTT scale is adjusted accordingly. The observed variation in the shape of the hyperbolic reflections is primarily caused by changes in geometric path length, where the increased distance to a deeper object causes the radar wave's travel time to change more slowly with horizontal movement. The widening of the antenna radiation pattern with depth, as energy disperses in the subsurface, further affects the amplitude and detectable extent of the hyperbolic feature and is thus considered a secondary factor [25].
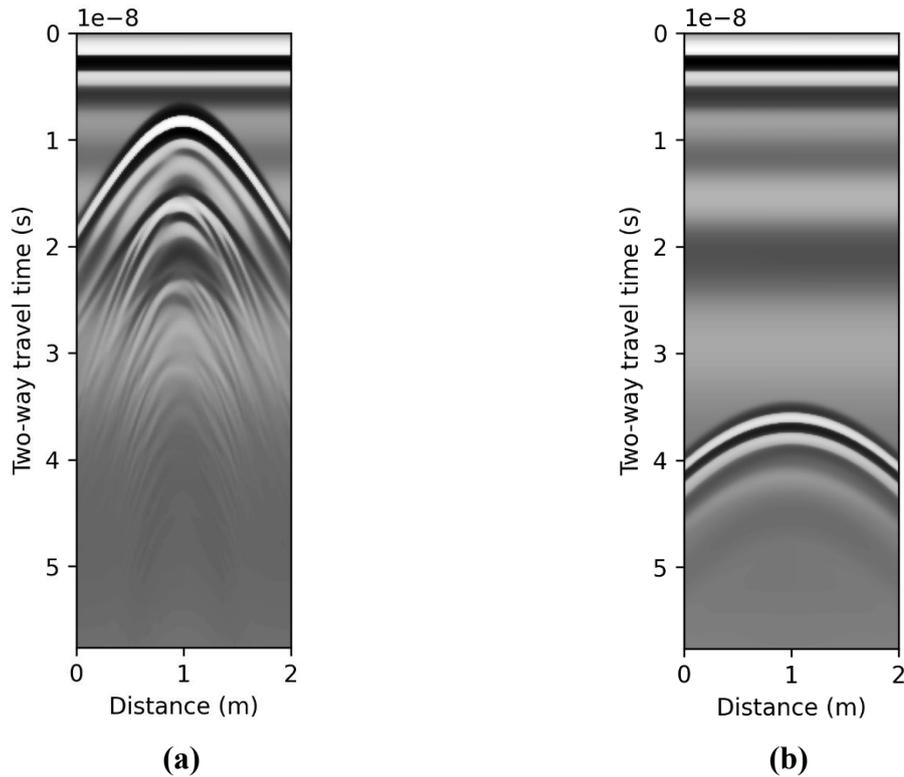
Figure 2.3 Comparison of hyperbolic reflections at different burial depths within a homogeneous medium with a relative permittivity of 9. The cylindrical object has a diameter of 0.5 meters, and the GPR survey line is oriented perpendicular to the object. In (a), the burial depth is 0.35 meters (measured from the ground surface to the top of the object), whereas in (b), the burial depth is 1.75 meters.

The relative permittivity of the subsurface medium directly influences the propagation velocity of EM waves as seen in (2.1) and (2.2). In other words, it affects the shape of hyperbolic reflections found within the B-scans. The wave velocity and relative permittivity are inversely related, therefore, a lower value of relative permittivity will result in wider hyperbolic reflection and vice versa. Such behavior is demonstrated in Figure 2.4, where the burial depth is 0.55 meters, the cylindrical object diameter is 0.5 meters, and the GPR survey line is oriented perpendicular to the object. Moreover, the relative permittivity parameter in sample (a) is 3, while in sample (b) is 21. A low relative permittivity parameter, which corresponds to a medium like dry sand, allows the signal to propagate faster, causing smaller changes in travel time as the antenna moves horizontally. In contrast, a material with high relative permittivity, which corresponds to a medium like wet clay, will cause the radar signal's travel time to change sharply with horizontal movement, therefore resulting in narrower hyperbolic reflection. Within Table 2.1, the typical ranges of relative permittivity and the corresponding EM wave velocities are provided for various geologic materials. Given values serve as a reference for interpreting
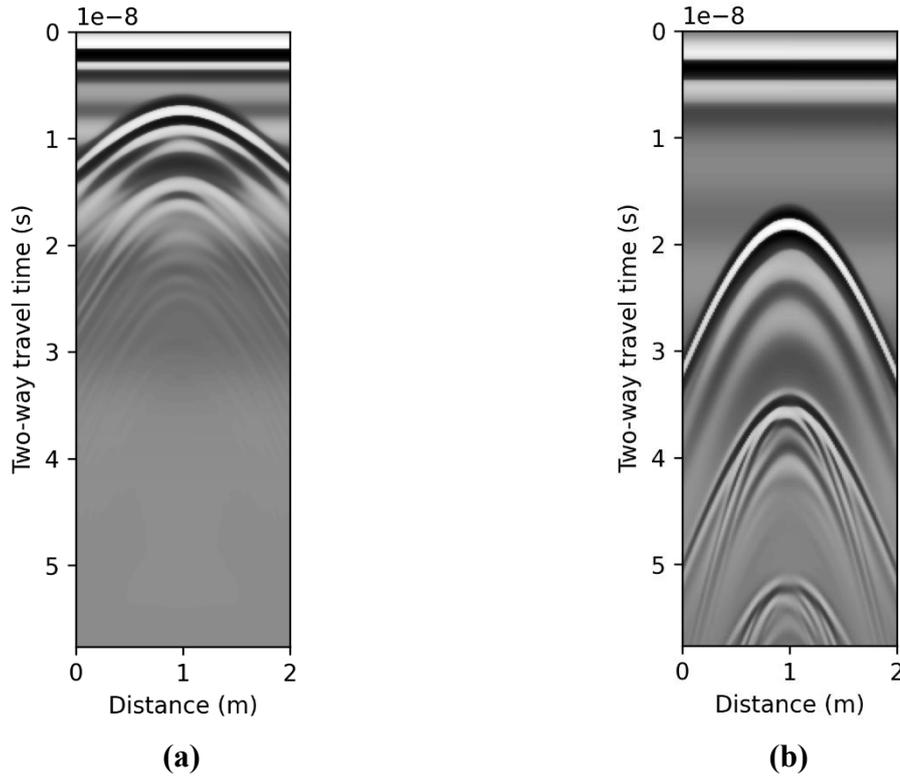
Figure 2.4 Comparison of hyperbolic reflections for different relative permittivities in a homogeneous medium. The burial depth is 0.55 meters (measured from the ground surface to the top of the object), the cylindrical object diameter is 0.5 meters, and the GPR survey line is oriented perpendicular to the object. In (a) relative permittivity is 3, whereas in (b), the relative permittivity is 21.

subsurface conditions and estimating material composition based on GPR survey data. Moreover, from (2.1) can be seen that parameters such as the cylindrical object diameter and the angle between the GPR survey line and the cylindrical object are not considered despite having an impact on the shape of the resulting hyperbolic reflection. Ignoring such parameters introduces inaccuracies and makes the interpretation of B-scan data more challenging. The effect of cylindrical object diameter on the resulting hyperbolic reflection is demonstrated in Figure 2.5. The samples are generated utilizing the following parameters: a burial depth of 1.175 meters, a homogeneous medium with a relative permittivity of 21, and a GPR survey line oriented perpendicular to the object. In sample (a), the object diameter is 0.15 meters, while in sample (b), the object diameter is 0.65 meters. It can be observed that a larger diameter produces a slightly wider hyperbolic reflection, and vice versa. By introducing a more complex surrounding environment it becomes evident that determining such a parameter from a B-scans represents a significantly more challenging task.

Table 2.1 Typical ranges of relative permittivity and corresponding EM wave velocities in various geologic materials [7].

Material	ε_r [29]	ε_r [27]	Velocity (m/ns)
Air	1	1	0.3
Distilled water	80	/	0.03
Fresh water	80	81	0.03
Sea water	80	/	0.03
Fresh water ice	3 – 4	4	0.15 – 0.17
Sea water ice	/	4 – 8	0.11 – 0.15
Snow	/	8 – 12	0.09 – 0.11
Permafrost	/	4 – 8	0.11 – 0.16
Sand, dry	3 – 5	4 – 6	0.12 – 0.17
Sand, wet	20 – 30	10 – 30	0.05 – 0.09
Sandstone, dry	/	2 – 3	0.17 – 0.21
Sandstone, wet	/	5 – 10	0.09 – 0.13
Limestones	4 – 8	/	0.11 – 0.15
Limestone, dry	/	7	0.11
Limestone, wet	/	8	0.11
Shales	5 – 15	/	0.08 – 0.13
Shale, wet	/	6 – 9	0.10 – 0.12
Silts	3 – 30	/	0.05 – 0.13
Clays	5 – 40	/	0.05 – 0.13
Clay, dry	/	2 – 6	0.12 – 0.21
Clay, wet	/	15 – 40	0.05 – 0.08
Soil, sandy dry	/	4 – 6	0.12 – 0.15
Soil, sandy wet	/	15 – 30	0.05 – 0.08
Soil, loamy dry	/	4 – 6	0.05 – 0.08
Soil, loamy wet	/	15 – 30	0.07 – 0.09
Soil, clayey dry	/	4 – 6	0.12 – 0.15
Soil, clayey wet	/	10 – 15	0.08 – 0.09
Coal, dry	/	3.5	0.16
Coal, wet	/	8	0.11
Granites	4 – 6	/	0.12 – 0.15
Granite, dry	/	5	0.13
Granite, wet	/	7	0.11
Salt, dry	5 – 6	4 – 7	0.11 – 0.15
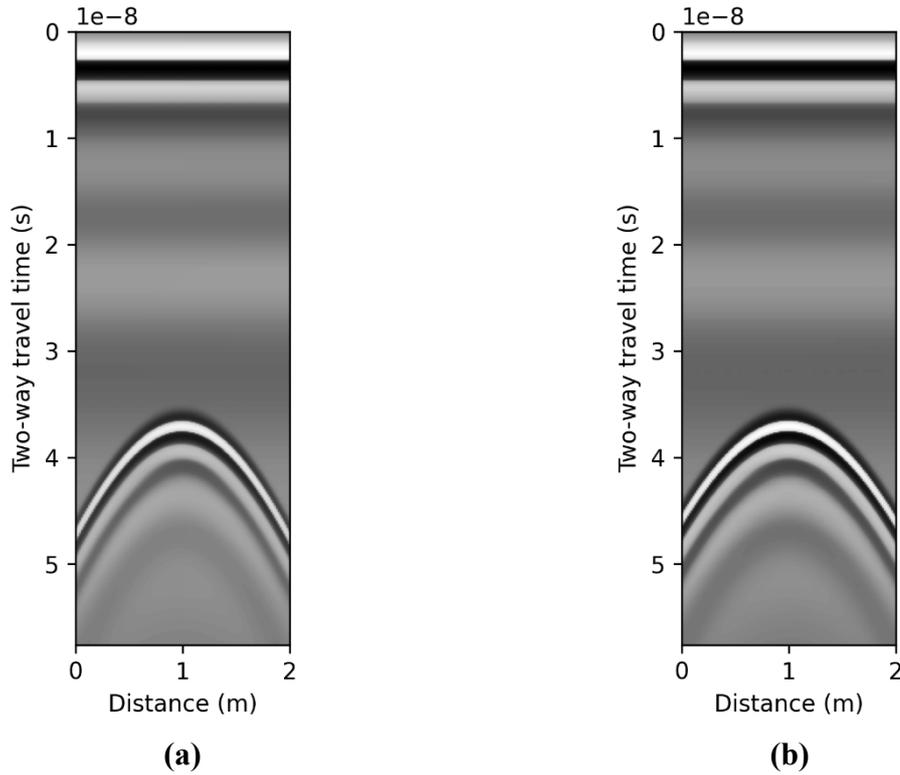
Figure 2.5 Comparison of hyperbolic reflections for different cylindrical object diameters. The burial depth is 1.175 meters (measured from the ground surface to the top of the object), the relative permittivity of the homogeneous medium is 21, and the GPR survey line is oriented perpendicular to the object. In (a), the diameter is 0.15 meters, whereas in (b), the diameter is 0.65 meters.

Within the 3D view of the B-scan creation process, the angle between the GPR survey line and the cylindrical object is represented by the parameter α, as shown in Figure 2.6. The values of the parameters, including relative permittivity, burial depth, and cylindrical object diameter, are 9, 0.75 meters, and 0.1 meters, respectively. The angle parameter is set to 90° in illustration (a) and the corresponding B-scan (c), whereas it is 30° in illustration (b) and the corresponding B-scan (d). The demonstrated example shows how different angles influence the resulting reflection within the B-scan. A smaller angle produces a wider hyperbola-shaped reflection with a flatter apex, while an angle closer to 90° results in a narrower reflection with a sharper peak. Additionally, when parameters such as cylindrical object diameter, angle, and characteristics of the real antenna system are introduced, the resulting reflection is no longer hyperbolic in the mathematical sense. Therefore, from this point onward, such reflections will be referred to as hyperbola-shaped reflections.

(a)

(b)

(c)

(d)
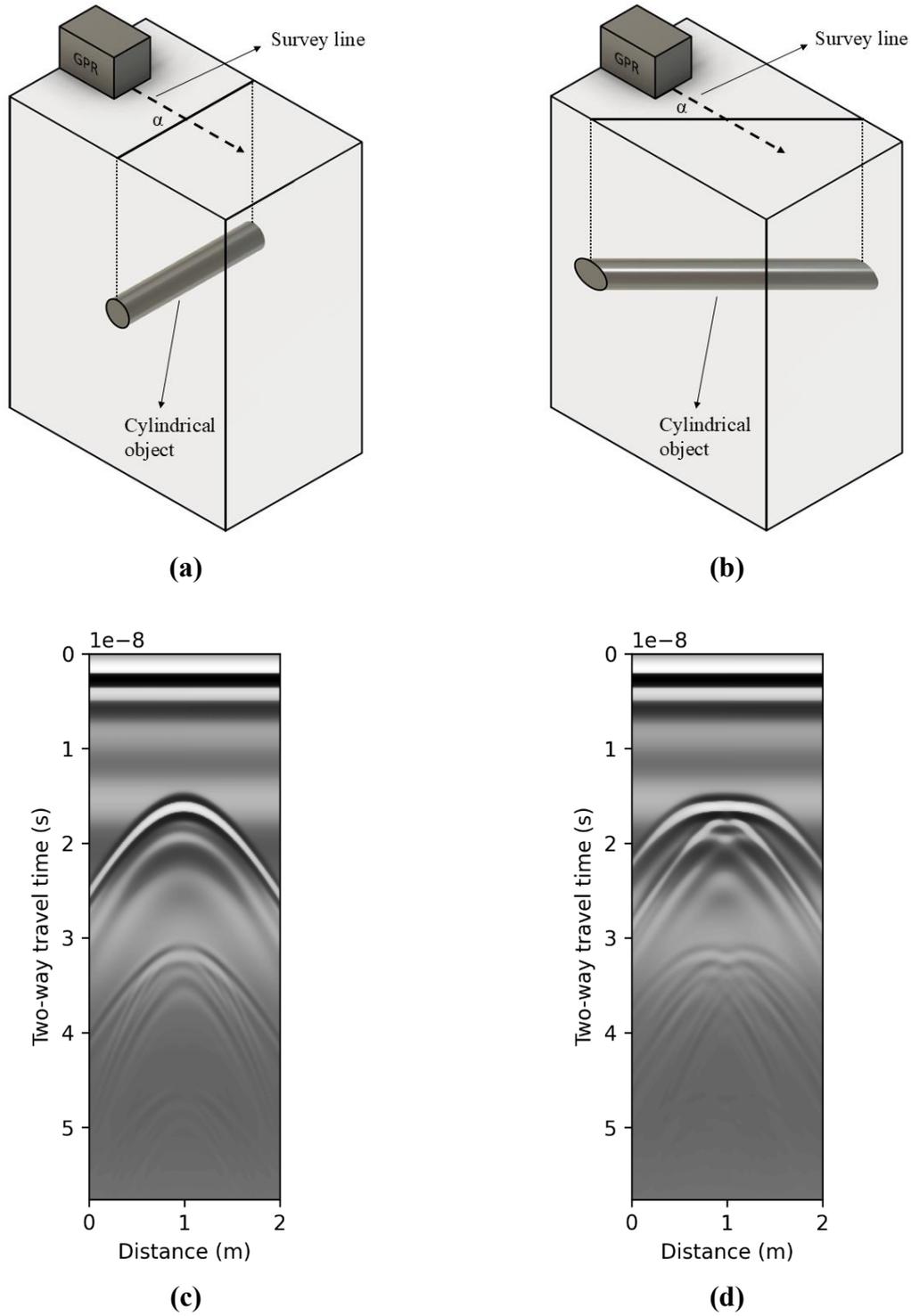
Figure 2.6 Illustration of the angle – α between the GPR survey line and the cylindrical object, along with the corresponding B-scans. The relative permittivity, burial depth, and cylindrical object diameter are defined as 9, 0.75 meters, and 0.1 meters, respectively. With the aforementioned parameters fixed, different values of the angles are used: (a) 90˚, and (b) 30˚.

2.1 Mathematical Model of Hyperbola-Shaped Reflections

The hyperbola-shaped reflections introduced in the previous section can be mathematically described using the mathematical model proposed by Sham and Lai [101]. The model includes all the parameters of interest, such as burial depth, cylindrical object diameter, angle between GPR survey line and a cylindrical object, as well as the separation between transmitter and receiver antennas. An illustration of the model is provided in Figure 2.7.
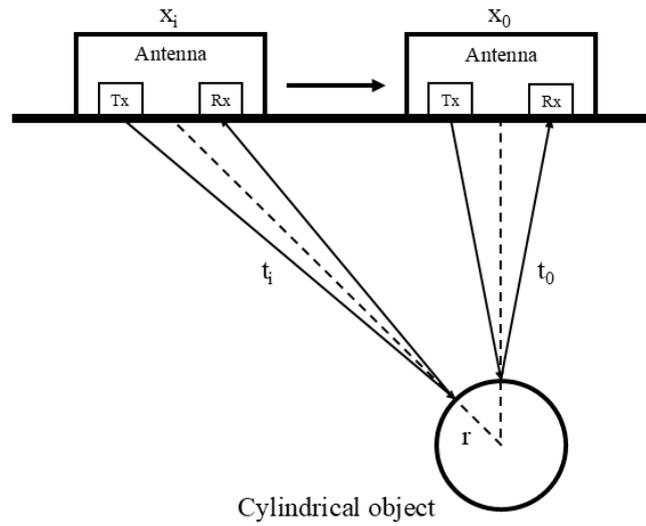
Figure 2.7 An illustration of the model employed to mathematically describe the hyperbola-shaped reflections.

From the illustration can be seen that the horizontal position of the antenna system is indicated as x_0 and x_i, while the vertical position expressed in terms of TWTT is denoted by t_0 and t_i. The parameter r represents the radius of the cylindrical object, while Tx and Rx correspond to the transmitter and receiver antennas, respectively. Based on Figure 2.6 and Figure 2.7, hyperbola-shaped reflection can be mathematically described as follows [44]:

$$t_i = \frac{D_1 + D_2}{v}, \tag{2.3}$$

$$D_1 = \sqrt{\left[(D_0 + r) - \frac{(D_0 + r)r}{\sqrt{(D_0 + r)^2 + (x_i \sin \alpha)^2}}\right]^2 + \left[x_i \sin \alpha - \frac{x_i \sin \alpha \, r}{\sqrt{(D_0 + r)^2 + (x_i \sin \alpha)^2}} - B_a\right]^2}, \tag{2.4}$$

and

$$D_2 = \sqrt{\left[(D_0 + r) - \frac{(D_0 + r)r}{\sqrt{(D_0 + r)^2 + (x_i \sin \alpha)^2}}\right]^2 + \left[x_i \sin \alpha - \frac{x_i \sin \alpha \, r}{\sqrt{(D_0 + r)^2 + (x_i \sin \alpha)^2}} + B_a\right]^2} \, , (2.5)$$

where D_1 represents the distance from the transmitter (Tx) to the cylindrical object, D_2 represents the distance from the cylindrical object to the receiver (Rx), v is the velocity of EM waves in the medium, D_0 stands for the burial depth (measured from the ground surface to the top of the cylindrical object), α defines the angle between a GPR survey line and a cylindrical object, and B_a is half of the antenna separation.

The presented mathematical model can accurately describe the hyperbola-shaped reflections, however, it comes at the cost of computational complexity. The model can be formulated as a hyperbola-fitting problem, where the objective is to estimate a set of parameters including v, D_o, r, and α, under the assumption that the antenna separation is known. Considering aforementioned, a mathematically defined hyperbola-shaped curve can be aligned with the hyperbola-shaped reflection located within the B-scan, by which the optimal parameters can be determined [72]. Therefore, the described formulation can be considered as an optimization task, in which the aim is to perform precise curve fitting. Optimization can be achieved by utilizing specific algorithms that iteratively adjust model parameters in order to reduce the error function [55].

Many iterative algorithms require careful selection of parameter values in order to achieve optimal performance [82]. In some cases, numerous parameters can be adjusted resulting in many possible combinations which increase computational cost and complexity. More traditional approaches to parameter value selection such as grid search and random search can be extremely costly (computationally) and sometimes infeasible [36]. Such approaches do not converge to the best solution as quickly since the set of parameter values for the upcoming iteration is not based on the evaluations from previous iterations. Therefore, a significant amount of time and computational resources can be spent evaluating models with a poor selection of parameter values. In contrast to the aforementioned approaches, the Bayesian approach for parameter value optimization selects the next set of parameter values based on previous evaluation results [108]. This way, a posterior distribution of functions that best describe the target function can be constructed. Therefore, Bayesian optimization (BO) can be described as an iterative algorithm with a probabilistic surrogate model and an acquisition

function [17]. As a result, the Bayesian approach for parameter values selection can find optimal values in fewer iterations compared to other, more traditional approaches. According to Wu et al. [122], BO is used to solve the hyperparameter tuning problem. Based on the obtained results, the proposed method proved to be efficient in hyperparameter search for ML models. Borgli et al. [15] in their research utilized the BO for automatic hyperparameter optimization of CNN models. With such an approach, the results were 10% better than other related work. Victoria and Maragatham [113] in their research used the Bayesian hyperparameter optimization algorithm in order to enhance the performance of the CNN model trained on the CIFAR-10 dataset. The authors concluded that BO for hyperparameter selection saves time and improves performance.

Considering the aforementioned, it can be concluded that BO effectively combines statistical modeling and decision-making strategies in order to optimize complex functions. Therefore, in this research a BO is utilized to find optimal parameter values and ensure convergence during the process of fitting hyperbola-shaped reflections within the B-scans. The BO process can be described as follows: first, the objective function f is sampled at a few initial points. These points can be selected randomly or through systematic methods. Second, a Gaussian process (GP) is typically utilized as the surrogate model, which can be defined as [103]:

$$f(x) \sim GP\big(m(x), k(x, x')\big), \tag{2.6}$$

where $m(x)$ represents the mean function, and $k(x, x')$ stands for covariance function. The $m(x)$ is often assumed to be zero if no prior knowledge is available, while $k(x, x')$ is considered as the kernel function that defines the covariance between any two points in the input space. One commonly used kernel is the squared exponential kernel, or in other words, Gaussian kernel, which is mathematically written as [119]:

$$k(x, x') = exp\left(-\frac{1}{2l^2}\|x - x'\|^2\right), \tag{2.7}$$

where l is the parameter that determines the rate at which the correlation decays with distance, and $\|x - x'\|^2$ represents squared Euclidean distance measuring the dissimilarity between points.

Determining the next sampling point is achieved by maximizing an acquisition function that balances the trade-off between exploration and exploitation. Frequently used acquisition functions include expected improvement (EI), and upper confidence bound (UCB) [62, 106]. The expected increase in the objective function relative to the best current observation is measured by the EI, which is mathematically defined as [62]:

$$EI(x) = \mathbb{E}[max(f(x) - f(x^+), 0)], \tag{2.8}$$

where $f(x^+)$ represents the current best observed value of f. Furthermore, the UCB can be written as [106]:

$$UCB(x) = \mu_{ucb}(x) + k_{ucb}\sigma_{ucb}(x), \tag{2.9}$$

where $\mu_{ucb}(x)$ and $\sigma_{ucb}(x)$ represent the mean and standard deviation of the GP's estimations at point x, while k_{ucb} corresponds to a parameter that balances exploration and exploitation.

The evaluation of the objective function in BO begins with the maximization of the acquisition function, through which point x is selected. Afterwards, the objective function value $f(x)$ is calculated and the result is added to the existing dataset. The updated dataset is then utilized in order to refine the GP model [103]. The process continues iteratively, including the repeated updates of the acquisition function, selection of new sampling points, and refinement of the surrogate model. This way, the surrogate model becomes more accurate in approximating the objective function, allowing the optimization process to gradually converge toward the optimal solution [100]. Since BO is an iterative process, defining stopping criteria ensures that the optimization process terminates preventing unnecessary computational overhead. Typical stopping criteria include reaching the maximum number of evaluations or achieving a convergence threshold where further improvements become minimal [62].

On the other hand, scalability, computational overhead, and dependence on the choice of surrogate model and acquisition function are considered as limitations of BO [34]. While BO is effective for low- to moderate-dimensional problems, in high-dimensional spaces, the complexity of the surrogate model also increases significantly, which can negatively affect the optimization process [30]. According to Snoek et al. [104], by increasing the number of evaluations, the surrogate model fitting process and maximizing the acquisition function can also introduce computational overhead. The overall performance of BO heavily depends on the selection of the surrogate model and acquisition function, both of which require careful

consideration and parameter tuning in order to achieve optimal results. In summary, by intelligently balancing exploration of unexplored regions and the exploitation of promising areas, BO can significantly reduce the number of evaluations required to locate the optimum.

2.2 Acquisition of Real-World GPR Data

Utilizing Geophysical Survey Systems Inc. (GSSI) UtilityScan GPR, a real-world B-scans are acquired throughout the city of Rijeka, Croatia. The GPR is equipped with a 350 MHz antenna in which, according to the manufacturer specification, the separation between the Tx and Rx antenna is 0.16 meters. The spatial sampling interval is 60 samples-per-meter, indicating that 60 A-scans are acquired for each meter of survey line during the data acquisition process. Additionally, the GPR is also equipped with a global navigation satellite system (GNSS), which, in combination with the Croatian positioning system (CROPOS), provides centimeter-level accuracy. The GPR system utilized in this research is shown in Figure 2.8.

Figure 2.8 GSSI UtilityScan GPR system equipped with 350 MHz antenna [39].

The subsurface surveying is performed at 33 micro-locations, which resulted in a total of 1051 B-scans representing underground pipeline infrastructures with various characteristics.

In order to capture different perspectives of the features of interest, the GPR survey is performed in three main lines (directions): horizontal, vertical, and diagonal. Additionally, the survey is conducted in a mesh pattern, with multiple survey lines recorded along each direction. The spacing between adjacent survey lines ranged from approximately 0.5 to 1.5 meters due to differences in terrain and site conditions. B-scan data acquired in this manner includes a variety of heterogeneous and complex subsurface environments with variations in composition, soil type, and moisture content. These properties directly affect the propagation velocity of EM waves and, thereby, impact the estimation of burial depth, pipeline diameter, and the angle between the GPR survey line and the pipeline [124]. The underground infrastructure consists of metal and plastic pipelines used for various purposes such as water and gas supply, district heating, sewage, stormwater drainage, etc. The characteristics and positions of the pipelines are stored in city infrastructure records, which are documented during construction and installation. These records are legally mandated to be documented with high precision under construction regulations in order to obtain a use permit. Geospatial data, such as GPR survey lines and pipeline positions, can be analyzed and visualized using the open-source software quantum geographic information system (QGIS). An example showing the pipeline positions for a wider area is presented in Figure 2.9.

Figure 2.9 Pipeline positions of the district heating system (red) and water supply (blue) within a wider area, visualized using QGIS software and Google Maps.

The displayed area corresponds to the University of Rijeka's Campus, where the pipeline highlighted in red represents the district heating system, while the blue highlighted one represents the water supply system. It is important to note that numerous other pipelines of various purposes are also buried within the same location, however, for simplicity of demonstration, only these two pipelines are shown in the example presented in Figure 2.9. Additionally, as shown in Figure 2.10, QGIS is utilized to visualize the GPR surveying at the selected micro-location. Similar to the previous example, the thicker lines represent the pipelines, while the thinner lines arranged in a mesh pattern correspond to the GPR survey lines. Furthermore, the circular dots of various colors mark the exact intersection points between the GPR survey lines and the pipelines. Such an extensive surveying strategy is crucial to ensure that pipelines are intersected at multiple angles, thereby maximizing the diversity of reflection signatures.
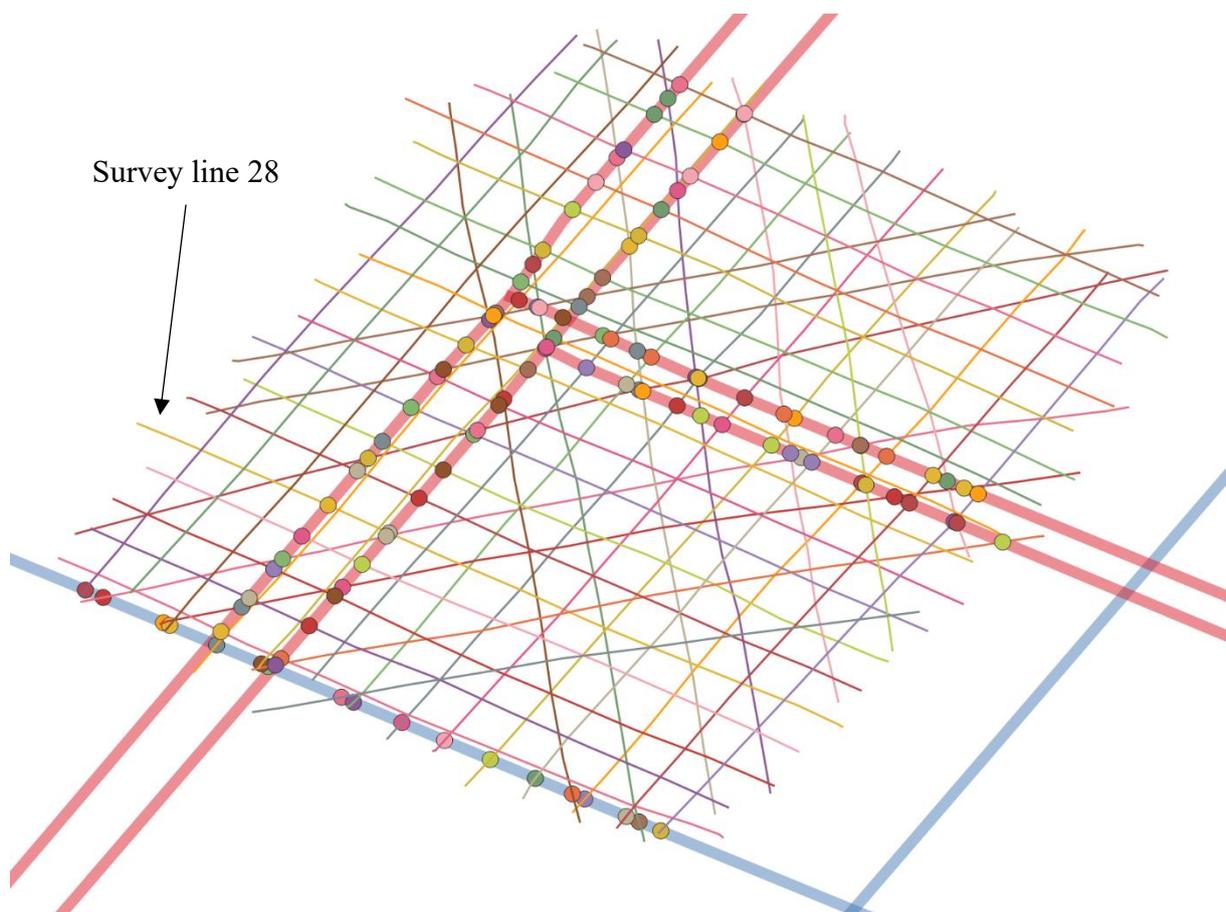
Figure 2.10 Visualization of GPR surveying at the selected micro-location. The thicker red line represents the district heating system, while the thicker blue line corresponds to the water supply network. The thinner lines arranged in a mesh pattern indicate the GPR survey lines, and the circular dots mark the intersection points between the GPR survey lines and the pipelines. Additionally, survey line 28 is highlighted with black arrow.

In addition to positional information provided in GPS coordinates, the pipeline records usually include parameters such as burial depth, pipeline diameter, and material composition. Therefore, by integrating GPR with GNSS and CROPOS data alongside city infrastructure records, precise pipeline annotations can be determined within each B-scan. Given this integration, a standardized protocol for annotating GPR data is required.

First, at the chosen location, the area of interest is surveyed using the GPR in a mesh pattern to ensure comprehensive spatial coverage. The acquired B-scans are georeferenced with the corresponding GPS coordinates, which are then compared with the city infrastructure records. Focusing on the area of interest, each pipeline segment is analyzed and compared with each GPR survey line in order to determine intersection points. If a crossing is detected, the GPS coordinates of the intersection point are calculated, and the angle between the GPR survey line and the pipeline is determined. Moreover, information such as burial depth and pipeline diameter is also stored. This process is iteratively repeated until all pipeline segments in the area of interest have been examined. In order to successfully annotate the hyperbola-shaped reflection within the B-scan, the vertical position of the reflection remains uncertain even when the burial depth is known from records. The uncertainty arises due to the unknown characteristics of the subsurface composition, more precisely, the relative permittivity of the soil. Therefore, the conversion from TWTT to actual burial depth cannot be directly established. In such cases, the relative permittivity value is measured or manually estimated based on expert experience with local soil conditions. Finally, as an additional quality control measure, four independent domain experts manually review, refine and validate all annotations. Such a protocol ensures that the annotations are fully aligned with the actual underground object of interest, i.e., the pipelines. The unprocessed B-scan acquired along survey line 28, highlighted in Figure 2.10 with black arrow, is presented in Figure 2.11. It can be seen that, without preprocessing, the hyperbola-shaped reflections are barely visible. Moreover, since time-zero correction has not yet been applied, the vertical positioning of the reflections is inaccurate. After performing the time-zero correction, the A-scans of the raw data at 1.5 and 2.0 meters are shown in Figure 2.12. The amplitude of the returning EM wave is a relative, not an absolute physical quantity, therefore the vertical axis is expressed in arbitrary units (a.u.). The use of unitless amplitude value arises from the high-resolution analog-to-digital converters used in advanced GPR systems [115]. These converters digitize the analog waveform into discrete amplitude levels, producing dimensionless numerical values that correspond to the relative voltage at the receiver's input.
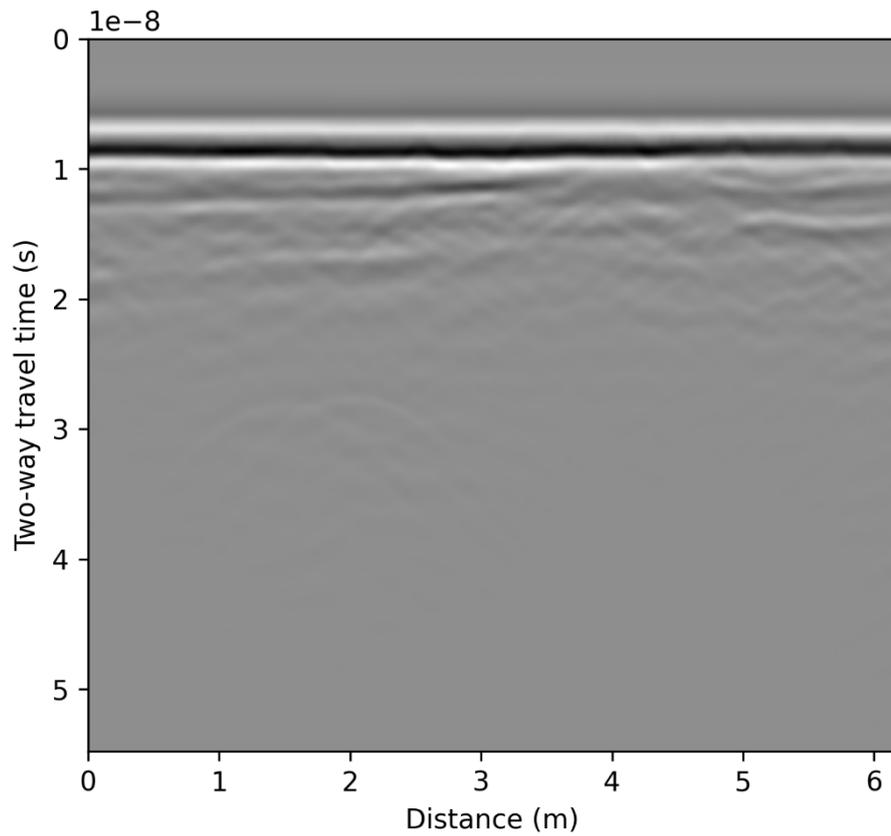
Figure 2.11 Unprocessed B-scan acquired using GSSI UtilityScan GPR equipped with a 350 MHz antenna, representing survey line 28 highlighted with black arrow in Figure 2.10. Two hyperbola-shaped reflections can be found between 1.0 and 2.5 meters along the x-axis, with their apexes located around 30 nanoseconds.
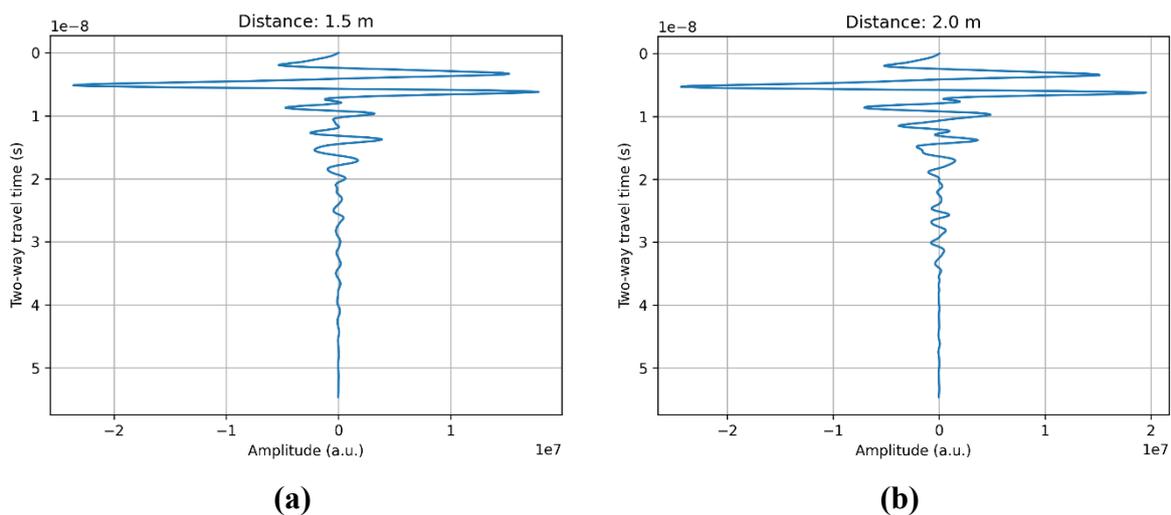

Figure 2.12 A-scans acquired at distances of (a) 1.5 meters and (b) 2.0 meters after time-zero correction. The apexes of the hyperbola-shaped reflections are visible within the signals at approximately 25 nanoseconds.

Since the digitization process scales the returning EM wave according to the converter's bit depth and dynamic range, the resulting amplitude values depend on the ADC quantization scale rather than any absolute physical magnitude. For aforementioned reasons, the amplitude expressed in arbitrary units allows consistent comparison of signal shapes and relative reflection strengths. Analyzing the A-scan data reveals that the returning EM waves are typically complex and non-stationary signals, characterized by a varying frequency spectrum [33]. In real-world applications, these signals often contain noise and multiple overlapping components. Therefore, it is methodologically justified to perform a TF analysis, as this approach will enable a more comprehensive characterization of the signal's temporal and spectral properties [88].

After applying the preprocessing techniques described in Chapter 2, the resulting, fully annotated B-scan is shown in Figure 2.13. The annotated hyperbola-shaped reflections correspond to the district heating system, featuring an outer pipeline diameter of 0.219 meters.
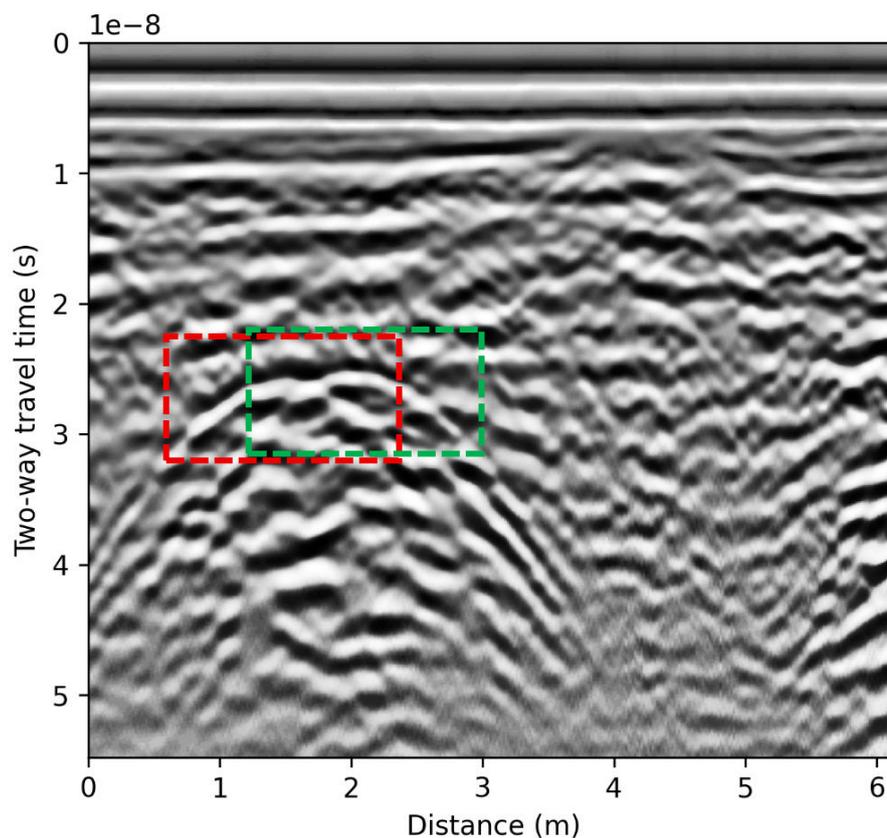
Figure 2.13 Preprocessed and annotated B-scan acquired using GSSI UtilityScan GPR equipped with a 350 MHz antenna, representing survey line 28 highlighted with black arrow in Figure 2.10. Two bounding box annotations are visible within the B-scan, each enclosing a hyperbola-shaped reflection corresponding to underground infrastructure, where both pipelines belong to the district heating system.

The reflection enclosed by the red bounding box appears at 25.07 nanoseconds which corresponds to a burial depth of 1.13 meters, while the reflection enclosed by the green bounding box appears at 24.40 nanoseconds, corresponding to a burial depth of 1.1 meters. The angle between the GPR survey line and the pipeline is 87° for the red-annotated pipeline and 89° for the green-annotated pipeline.

Utilizing the above-described standardized protocol for annotating GPR data, a total of 1596 annotations are generated across 1051 B-scans, yielding an average of approximately 1.52 annotations per one B-scan. These annotations can be used for the purpose of detecting and localizing underground infrastructure objects. However, for the estimation of underground infrastructure parameters such as burial depth, pipeline diameter, EM wave velocity and the angle between the GPR survey line and the pipeline, 732 annotations can be used. The remaining 864 annotations are excluded due to missing or imprecise parameter data, e.g., pipeline diameter information unavailable or burial depth insufficiency precise for the purpose of this research. TWTT is a directly observable parameter within the B-scan data, typically identifiable through mostly clear visual cues such as the apex of hyperbola-shaped reflections [124]. Therefore, the estimation of EM wave velocity can be achieved by incorporating TWTT together with the key parameters defined in (2.1) and (2.2).

A visual representation of the 732 parameter values describing underground infrastructure is shown in Figure 2.14. The vertical positions of the hyperbola-shaped reflections apexes within the B-scan data, expressed as TWTT, range from 9.2328 to 42.1902 nanoseconds, depending on the burial depth and the EM characteristics of the subsurface. The burial depths of the pipelines range from 0.4 to 1.95 meters, while their diameters vary between 0.025 to 0.6 meters. Additionally, the angle between the GPR survey line and the pipeline ranges from 19° to 90°, encompassing a wide range of intersection geometries. In order to provide a more straightforward overview of the data distribution and overall characteristics, a descriptive statistical analysis is performed with the results summarized in Table 2.2.
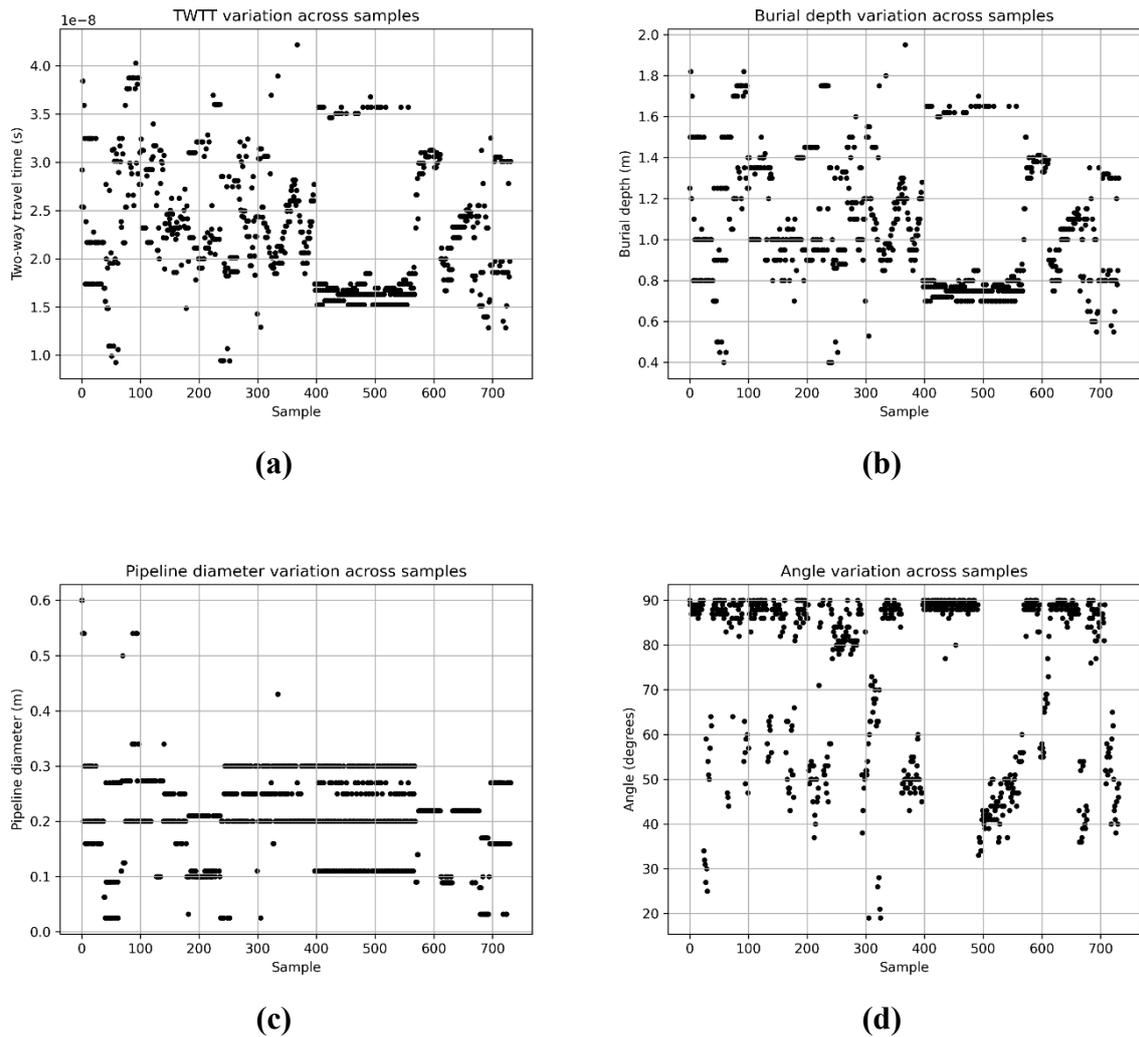
Figure 2.14 Variation of parameter values describing underground infrastructure across samples: (a) TWTT, (b) burial depth, (c) pipeline diameter, and (d) angle between the GPR survey line and the pipeline.

According to Table 2.2, the TWTT has a mean of 23.1606 nanoseconds and a standard deviation of 6.4493 nanoseconds, which indicates a moderate spread in reflection times. Most pipelines are located at shallow to moderate depths, with an average burial depth of 1.0555 meters and a low standard deviation of 0.3004 meters. Moreover, the majority of annotated objects are tightly clustered, as evidenced by the fact that 50% of samples fall within the 0.8 to 1.3 meter range. Pipeline diameters have a mean value of 0.2074 meters with an interquartile range of 0.11 meters. Such a narrow range suggests that the surveying area primarily contains pipelines of a consistent size, clustering around 0.16 to 0.27 meters in outer diameter. On the other hand, the angle between the GPR survey line and the pipeline varies widely, with a mean of 73.1636° and a standard deviation of 19.2149°. The high upper quartile (75% - 89.0076°)

strongly suggests that the majority of reflections are recorded with the survey line nearly perpendicular to the pipelines of interest. The heavy skew in the angle data represents critical methodological consideration, therefore, the system development should incorporate appropriate validation strategies to minimize errors when estimating parameters for pipelines recorded at angles smaller than 50˚.

Table 2.2 Descriptive statistics of the key parameters extracted from the GPR B-scan data, presenting the mean, standard deviation, minimum, 25th percentile, median (50%), 75th percentile, and maximum for TWTT, burial depth, pipeline diameter, and intersection angle.

n = 732	TWTT (ns)	Burial depth (m)	Pipeline diameter (m)	Angle – α (˚)
mean	23.1606	1.0555	0.2074	73.1636
std	6.4493	0.3004	0.0816	19.2149
min	9.2328	0.4000	0.0250	19.0000
25%	17.3826	0.8000	0.1600	53.0004
50%	22.1988	1.0000	0.2100	85.9967
75%	28.2043	1.3000	0.2700	89.0076
max	42.1902	1.9500	0.6000	90.0000

In order to analyze the relationships between individual pipeline parameters, a Pearson correlation analysis is performed. Such a statistical method quantifies both the strength and direction of linear associations between pairs of continuous variables [84]. As illustrated in Figure 2.15, the correlation analysis reveals an exceptionally high correlation between the TWTT and the burial depth, with a coefficient value of 0.9866. The relationship is mathematically described with (2.1) and (2.2). From a statistical perspective, including both TWTT and burial depth in a standard linear model can be problematic due to multicollinearity. However, from a physical and research standpoint, this strong correlation is invaluable. More precisely, the simultaneous estimation of TWTT and burial depth presents an opportunity to calculate the EM wave velocity for a given local environment. Moreover, moderate positive correlations are observed between TWTT and pipeline diameter, as well as between burial depth and pipeline diameter, both with a value of approximately 0.45. In other words, pipelines with larger outer diameters tend to be buried deeper than the smaller diameter pipelines.
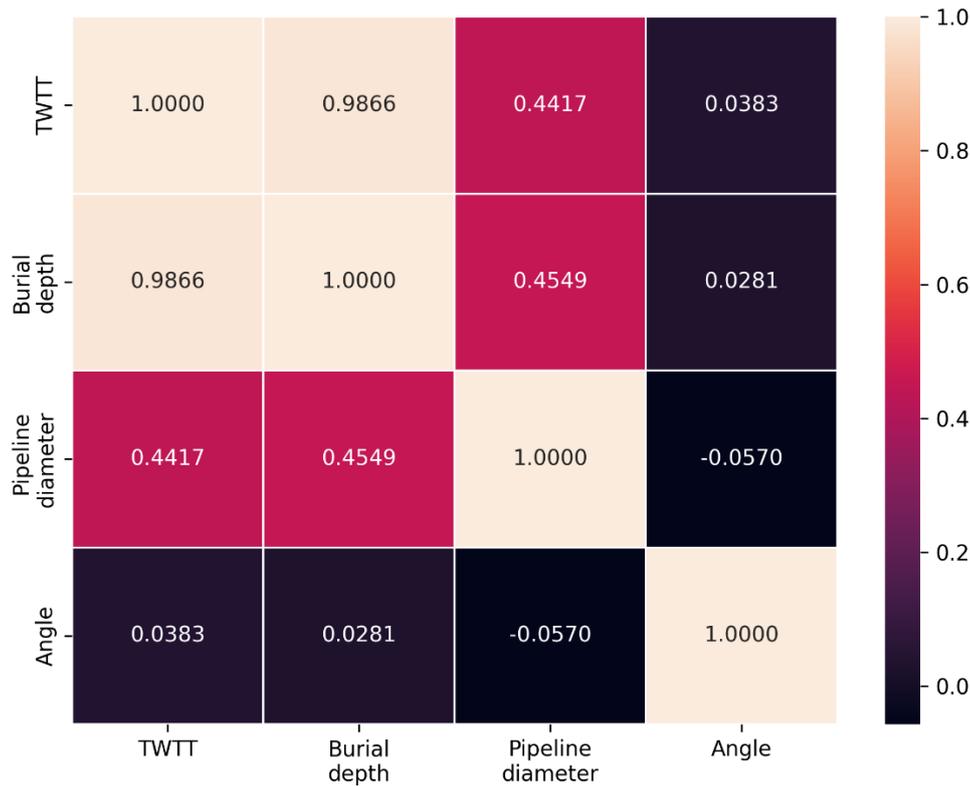
Figure 2.15 Pearson correlation matrix illustrating the relationships among underground pipeline characterization parameters such as TWTT, burial depth, pipeline diameter, and angle between the GPR survey line and the pipeline.

However, the observed pattern reflects the characteristics of the specific micro-locations surveyed and should not be interpreted as a universal rule. In contrast, the correlation coefficients involving the angle parameter (ranging from -0.0570 to 0.0383) confirm that the TWTT, burial depth, and pipeline diameter are not linearly affected by the angle at which the GPR survey line intersects the pipeline.

According to the aforementioned analyses, a key limitation identified is the insufficient dispersion of parameter values. In system development, where the main aim is parameter estimation of underground infrastructures, utilizing such data can potentially result in biased and poorly generalized estimations. The most challenging parameter is the angle – α, for which the descriptive statistics revealed a heavy skew. In order to address the critical data gap and ensure the robustness of the final system, synthetic data generation is required. Simulation software that can be utilized for the purpose of generating physically accurate and realistic GPR data is called gprMax [116]. Therefore, with particular attention to intersection angles below 50˚, the synthetic dataset can be specifically designed to augment the original data by incorporating the missing characteristics.

2.3 Generation of Synthetic GPR Data with GprMax

In order to address the challenges discussed in subsection 2.2 regarding data diversity and the underrepresentation of certain pipeline parameter values, the open-source simulation software gprMax is utilized. By using the finite-difference time-domain (FDTD) method to solve Maxwell's equations in 3D, gprMax simulates EM wave propagation [116]. Although the software was initially developed specifically for GPR modelling, it is applicable to a wide range of EM wave propagation problems.

All macroscopic EM phenomena are described by Maxwell's equations [99]. These represent a set of first-order partial differential equations which describe the relations between the fundamental EM field quantities and corresponding dependence on their sources. Therefore, Maxwell's macroscopic equations in partial differential equation form and a coherent system of units can be written as (top to bottom: Faraday's law, Ampère-Maxwell law, Gauss's law for magnetism, and Gauss's law for electricity) [40]:

$$\nabla \times E = -\frac{\partial B}{\partial t}, \tag{2.10}$$

$$\nabla \times H = \frac{\partial D}{\partial t} + J_c + J_s, \tag{2.11}$$

$$\nabla \cdot B = 0, \tag{2.12}$$

$$\nabla \cdot D = q_v, \tag{2.13}$$

where the variables are defined as follows: E represents the electric field (V/m), B denotes the magnetic field density (T), t is the time (s), H stands for the magnetizing field (A/m), D indicates the electric displacement field (C/m^2), J_c and J_s represent the conduction and impressed current densities (A/m^2), and q_v defines the volume electric charge density (C/m^3).

Maxwell's equations assume that field vectors are single-valued, bounded, and continuous functions of position and time. Therefore, to simulate the GPR response from a particular target, the aforementioned equations have to be solved subject to the geometry of the problem and the initial conditions [116]. The FDTD method solves Maxwell's equations

numerically by discretizing the continuous domains of space and time [111]. The chosen resolution, defined by the spatial Δx, Δy, Δz, and temporal Δt step sizes, is crucial since finer discretization will result in an FDTD model which is closer to a real representation of the problem. However, in real-world applications, the discretization steps cannot be infinitely small due to computational cost. Hence, the FDTD model is essentially a discretized representation of the real physical problem.

The simulation utilizes the built-in model of the commercial GSSI 400 MHz antenna to realistically implement the transmitter-receiver characteristics. The antenna model includes the specific physical geometry and shielded bow-tie structure of the actual hardware. In this way, the simulated wavelets and radiation patterns closely match real-world survey data. In order to enhance the realism of the simulation, the soil medium is modeled using the semi-empirical Peplinski mixing model, which is valid for the 0.3 – 1.3 GHz frequency range [91]. The model created in such a way derives dielectric properties from physical soil parameters, i.e., sand fraction, clay fraction, bulk density, and volumetric water content. The properties are spatially distributed using a fractal stochastic formulation to replicate the natural heterogeneity of subsurface environments. Based on the aforementioned setup, a dataset of synthetic GPR scans can be generated by varying the physical parameters of cylindrical objects and the surrounding soil medium. Simulations include two scenarios: homogeneous soil (dielectric half-space) – a model where the soil has a constant relative permittivity; and heterogeneous soil (Peplinski model) – a realistic model where soil properties vary stochastically using fractal distributions and the Peplinski mixing model. The computational domain is defined with dimensions of 2.38 m x 0.3 m x 2.0 m for the x, y, and z direction, respectively, and a spatial discretization of 0.002 m. Such high resolution is necessary in order to accurately model high-frequency wave propagation and small-scale soil heterogeneities. For the purpose of suppressing boundary reflections and simulating an unbounded domain, perfectly matched layers (PML) are applied to the edges of the computational grid [10]. The time window parameter determines how long the FDTD solver runs for each A-scan. Since the simulation is computationally expensive, an adaptive approach based on the input depth is employed. As a result, the target objects can be fully captured without running the simulation for unnecessarily long periods.

The cylindrical objects are modeled as perfect electric conductors (PEC), in other words, metallic objects with high reflectivity. The key input parameters defining such objects are depth, diameter, and angle between the GPR survey line and the object's longitudinal axis. By varying

these parameters, data diversity and coverage of underrepresented parameters can be achieved. In the case of homogeneous simulation scenario, the soil is modeled as a dielectric half-space with a fixed permittivity, which ranges from 3 to 21 across the generated B-scans. As for the heterogeneous simulation scenario, a fractal box is used to generate a stochastic distribution of soil properties using fractal correlated noise. Before each B-scan generation, values are randomly generated for sand fraction, clay fraction, bulk density, sand particle density, and volumetric water fraction. The ranges for these parameters are $0.05 - 0.95$, $0.05 - 0.60$, $1.1 - 2.2$ g/cm^3, $2.46 - 2.86$ g/cm^3, and $0.01 - 0.35$, respectively. Moreover, the stochastic distribution of materials within the soil is controlled by a fractal dimension parameter of 1.5. The number of materials used for the fractal distribution ranges from 12 to 43, allowing smooth transitions in soil properties.

The simulation process begins with A-scan acquisition where the GSSI 400 MHz antenna is moved across the surface in 120 steps. A step size of 0.0167 m is used to correspond with the spatial sampling interval of the real-world GPR data described in subsection 2.2. Accordingly, 60 A-scans are generated for each meter of survey line, resulting in a total B-scan length of 2 m. Experimental testing shows that a 2 m B-scan is sufficient for capturing the diverse range of hyperbola-shaped reflections similar to those found within the real-world GPR data. Accordingly, the simulation process iterated through a specific parameter space of depths $(0.35 - 1.95$ m), diameters $(0.1 - 0.5$ m), and angles $(15° - 90°)$ to generate the synthetic dataset. A total of 382 synthetic B-scans are generated, with two representative samples illustrating homogeneous and heterogeneous soil conditions presented in Figure 2.16. The characteristics of cylindrical object used in the samples are as follows: burial depth of 0.525 m (to the top), diameter of 0.15 m, and angle between the GPR survey line and the object's longitudinal axis of 45°. From Figure 2.16 (a), it can be seen that the background is a uniform gray, indicating a constant dielectric permittivity throughout the medium. The hyperbola-shaped reflection is clean and sharp without fading or broken curves, in other words, idealized response. Moreover, the shape of the reflection can be perfectly described by the mathematical model described in subsection 2.1. The aforementioned is possible since the hyperbola-shaped reflection is purely a function of the object's depth, diameter, intersection angle, and the soil's constant velocity. Such a simulation represents a perfect scenario; the closest real-world equivalent occurs when an object is located in a nearly homogeneous environment, for example, a pipeline buried in dry, sifted sand or a controlled concrete testbed. In contrast, the background of the heterogeneous soil B-scan in Figure 2.16 (b) is filled with smaller irregular reflections caused

by the fractal distribution of soil properties (sand, clay, water, etc.). In this case, the hyperbola-shaped reflection is not as clear as it is in the homogeneous soil. The reflection lacks the ideal geometry required for a perfect fit with standard mathematical models. The deviation is caused by the complexity of the materials and the variability of the surrounding medium. It is evident that the heterogeneous model successfully mimics the complex textures observed in real GPR data. Such complex reflections are expected in real-world situations where the environment consists of numerous materials mixed together, for example in typical underground utility surveys.

In cases where the hyperbola-shaped reflection differs from the ideal shapes described by analytical models, traditional curve-fitting methods may fail to accurately distinguish and estimate the parameters of the cylindrical objects of interest [99]. For this reason, the implementation of DL methods for the parameter estimation of underground infrastructure from GPR data is fully justified. DL-based models are capable of learning such complex, non-linear features that mathematical models cannot easily capture [71].

(a)   (b)
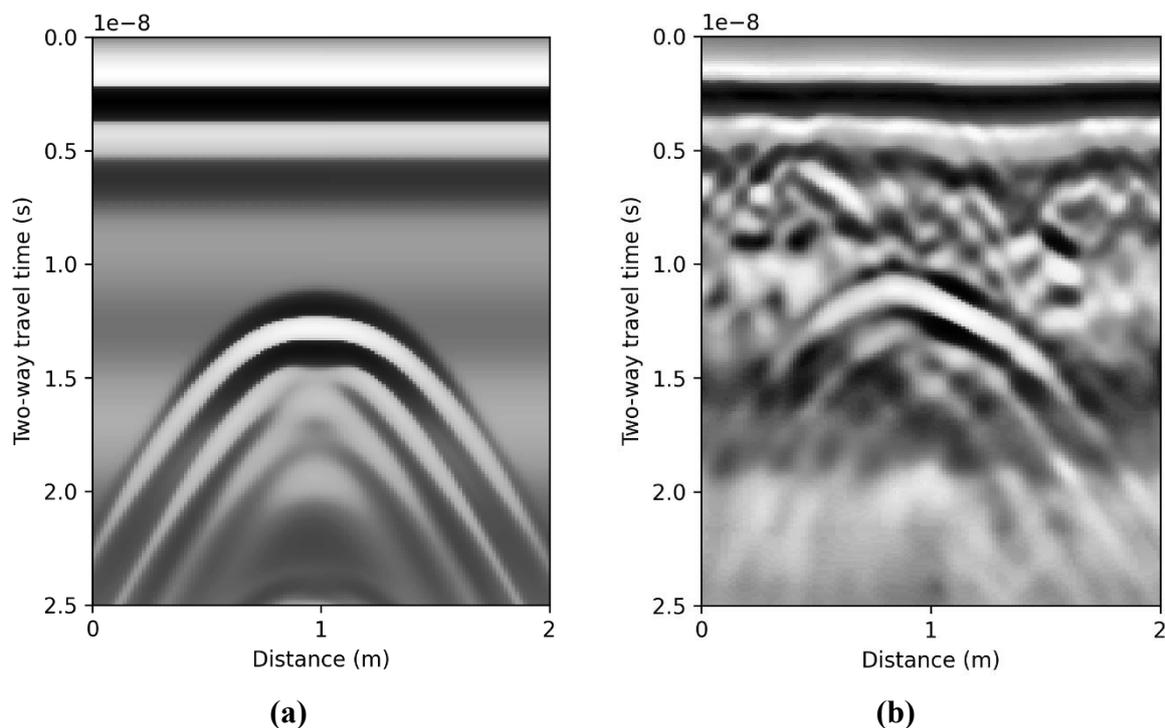
Figure 2.16 Samples of synthetic data generated using gprMax simulation software. The cylindrical object has a diameter of 0.15 m and is buried at a depth of 0.525 m (to the top), with a 45° orientation relative to the GPR survey line. The simulations represent two different soil environments: (a) homogeneous soil with a relative permittivity of 11, and (b) heterogeneous soil based on the Peplinski mixing model.

CHAPTER 3

DEEP LEARNING

The overall research workflow is structured as a comprehensive, two-stage automated framework. This framework is specifically designed to estimate the descriptive parameters of underground infrastructure located within the complex, heterogeneous soil environments. The illustration of the framework is given in Figure 3.1. Initially, a dataset containing real-world samples recorded utilizing a GSSI UtilityScan GPR is established. The system aims to achieve robust estimation of underground infrastructure descriptive parameters, therefore, relying solely on skewed real-world data can lead to biased and poorly generalizing models. In order to address this limitation, simulated synthetic data is used in parallel with real-world data. Before any model training procedure, both the acquired real-world GPR data and the synthetic data undergo necessary preprocessing and a ground-truth annotation process. In the first stage of the developed system, the processed and prepared datasets are utilized to train advanced object detection algorithms, which are described in subsection 3.1 of this chapter. The primary objective of this first stage is to automatically identify and localize hyperbola-shaped reflections within the B-scan data. The models' performances are evaluated in order to determine the best performing object detection model. Afterwards, the detected hyperbola-shaped reflections are isolated and further processed. The output of the first stage, i.e., the object detection stage, provided in the form of bounding box coordinates, is used to crop the corresponding regions of the B-scan. More specifically, for each detected hyperbola-shaped reflection, the relevant segment is extracted from the original B-scan based on the estimated bounding box dimensions. The absolute position along the vertical axis is preserved during the cropping procedure. This vertical area above the reflection significantly influences the resulting geometry and shape of

the hyperbola-shaped feature. In this manner, the extracted features retain both the spatial and structural information necessary for accurate parameter estimation.
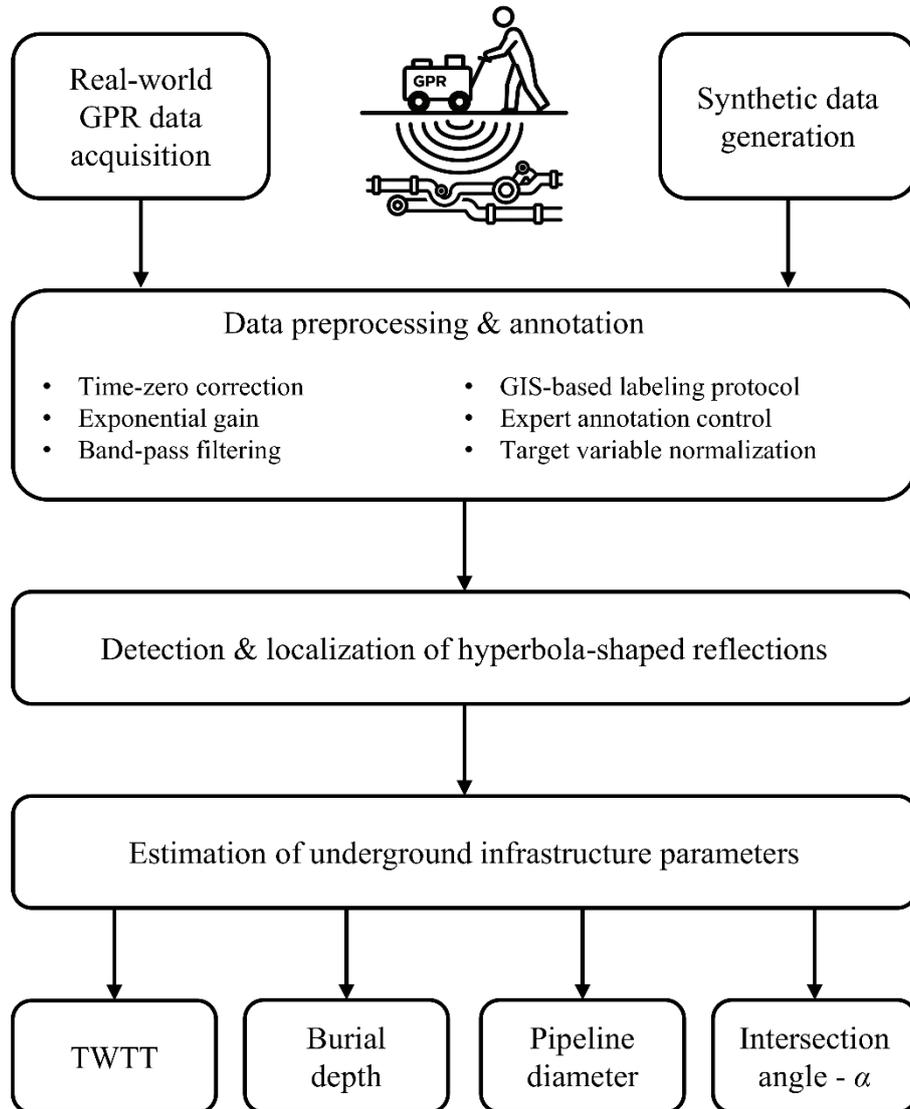
Figure 3.1 Framework of the DL-based system for the detection and localization of hyperbola-shaped reflections from GPR B-scan data, along with the estimation of underground infrastructure descriptive parameters.

Geometric signatures isolated in this way are then systematically transferred to the second stage of the framework, where they are utilized to train the parameter estimation model. The estimation of descriptive underground infrastructure parameters is enabled by leveraging a DL feature extraction backbone integrated with a custom regression head. Since the pipeline diameter and intersection angle represent a high-complexity regression problem, the model is optimized via a custom weighted MSE loss function. Such an approach guides the models to

allocate greater representational capacity toward the more challenging geometrical estimations, in this case, the pipeline diameter, and intersection angle. The comprehensive performance evaluation of the DL model architectures will be performed across multiple feature extraction backbones, as described in subsection 3.2. Furthermore, the impact of synthetic data inclusion will be investigated by directly comparing performances of models trained on real-world data with those trained on a combined dataset. Finally, the resulting performances will be analyzed in order to validate the optimal backbone configuration and input data composition.

3.1 Detection and Localization of Cylindrical Objects

Underground pipelines, when surveyed utilizing GPR, appear as hyperbola-shaped features within the B-scan profiles. The features appear due to the reflection and diffraction of EM waves from the curved surfaces of the pipelines; therefore, they represent the primary objects of interest in this research. Since accurate detection and localization of cylindrical objects are of critical importance, two state-of-the-art object detection algorithms are utilized: You Only Look Once (YOLO) version 11, and Real-Time Detection Transformer (RT-DETR). More precisely, YOLOv11 is employed due to its capability to perform end-to-end detection with high computational efficiency. The advanced convolutional backbones enable robust hierarchical feature extraction, while at the same time, the anchor-free design simplifies the detection framework [65]. On the other hand, RT-DETR utilizes self-attention mechanisms in order to capture long-range dependencies within the B-scan data [134]. Global context modeling can be effective in distinguishing the complex curvature of deep hyperbola-shaped features from surrounding geological layers. According to previous research, both algorithms have shown strong capability in identifying small, curved, and low-contrast targets [18, 74, 117, 128]. Therefore, they are considered as well-suited options for recognizing features of interest within the B-scans. The algorithms output hyperbola-shaped features precisely enclosed within bounding boxes. Each bounding box contains exactly one hyperbola-shaped feature, with coordinates representing the horizontal and vertical positions in terms of GPR travel distance and TWTT, respectively. The raw bounding box coordinate data will be processed afterwards in order to estimate the actual physical characteristics of the pipelines in question.

3.1.1 You Only Look Once v11

The first algorithm used is YOLOv11, which, at the time of its release, represented a significant step forward in the YOLO series for object detection tasks. Significant advancements have been made in both network architecture and training methodology. The key improvements are enhanced feature extraction, better optimization for computational efficiency, and increased detection accuracy [65]. The aforementioned improvements resulted in higher values of performance measures for the YOLOv11m model on the COCO benchmark dataset, while at the same time using 22% fewer parameters compared to YOLOv8m.

Several architectural innovations support these improvements, including the cross stage partial block with kernel size 2 (C3k2), spatial pyramid pooling-fast (SPPF), and the convolutional block with parallel spatial attention (C2PSA). The C2PSA component, for example, introduces significant advancement in spatial attention mechanisms. As a result, the model can focus more effectively on critical regions within an image. Together, all the aforementioned components enable more effective feature representation and information processing, thereby improving the model's ability to analyze and interpret complex visual information [42]. The innovations resulted in improved detection performance, especially in scenarios involving partially occluded or structurally complex objects [65]. Structure diagram of the YOLOv11 network is shown in Figure 3.2. In addition, YOLOv11 is also built with scalability in mind, providing a spectrum of models that range from smaller, resource-efficient nano versions to larger configurations intended for high-demand applications. YOLOv11 not only handles object detection but also proves useful in several related tasks, including image classification, instance segmentation, pose estimation, and oriented object detection [2, 59]. As a unified framework, it is capable of solving diverse computer vision tasks within a single architectural design.
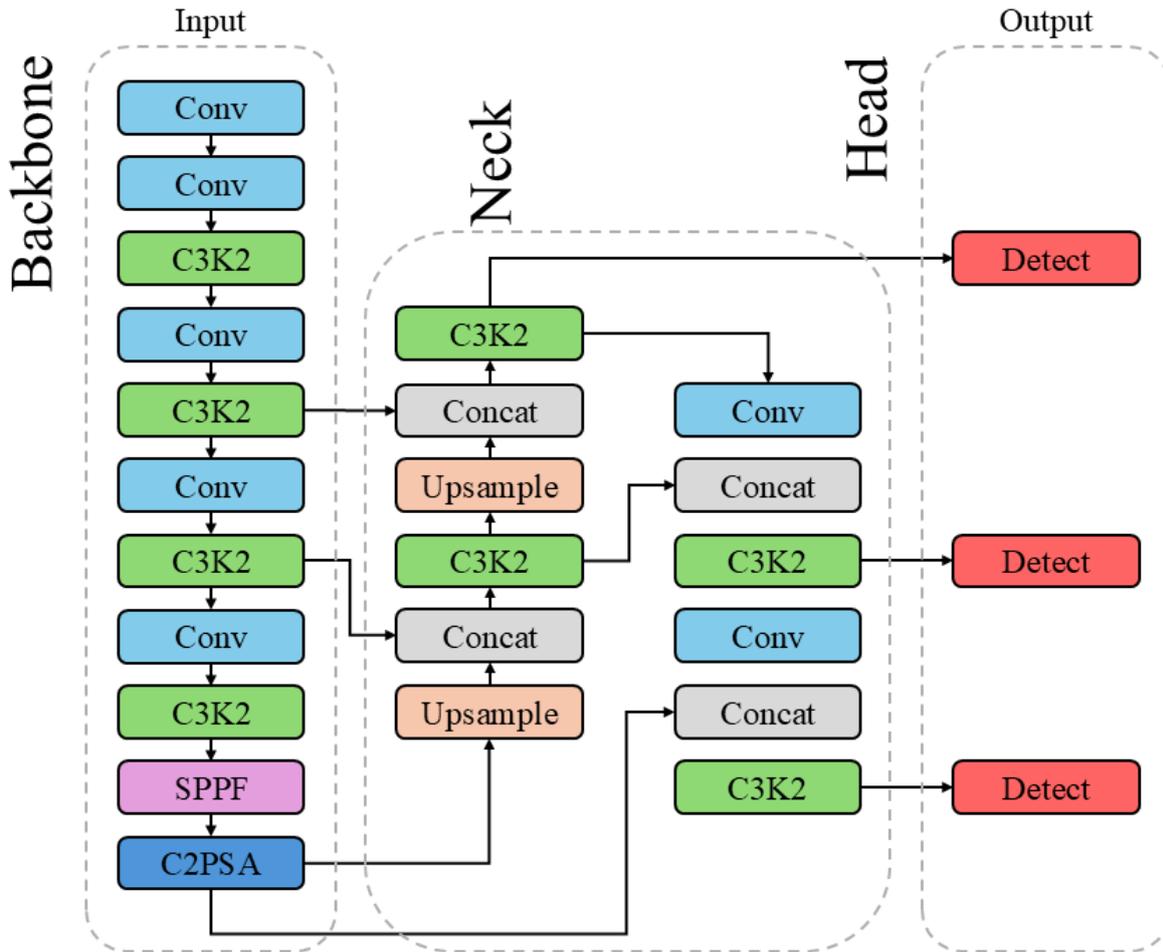
Figure 3.2 Structure diagram of the YOLOv11 network.

3.1.2 Real-Time Detection Transformer

RT-DETR can be seen as a rather different approach to object detection since it introduces a real-time end-to-end transformer-based framework. In this way, the computational limitations traditionally associated with DETR-based architectures can be effectively overcome. However, this benefit comes with design challenges that affect efficiency in practice. The original DETR architecture eliminates the need for hand-crafted anchor and non-maximum suppression (NMS) components by relying on bipartite matching to produce one-to-one predictions directly. On the other hand, in practice, slow training convergence, along with high computational cost, and hard-to-optimize queries can be expected [134]. These issues are especially restrictive when real-time deployment is required. Therefore, many DETR variants have been proposed to overcome the aforementioned limitations.

The development of RT-DETR follows a two-step process: initially focusing on maintaining detection accuracy while improving processing speed, followed by maintaining speed while improving accuracy [136]. The architecture itself consists of three primary components, including a convolutional backbone network, a hybrid encoder, and a transformer decoder with auxiliary prediction heads as shown in Figure 3.3. The hybrid encoder and the uncertainty-minimal query selection are considered as key innovations in RT-DETR. The hybrid encoder is designed to expeditiously process multi-scale features by decoupling intra-scale interactions and cross-scale fusion operations [45]. Considering above-mentioned, architectural separation enables efficient feature processing, while at the same time maintaining the model's ability to capture both local details and global contextual information. On the other hand, the uncertainty-minimal query selection improves the quality of the initial object queries provided to the decoder [67]. The transformer decoder is equipped with auxiliary prediction heads; therefore, it iteratively optimizes the object queries in order to output classification labels and bounding box coordinates [63]. By allowing the adjustment of decoder layers, the flexible speed-accuracy trade-offs are supported in RT-DETR, which enables adaptation to diverse computational constraints and deployment scenarios.
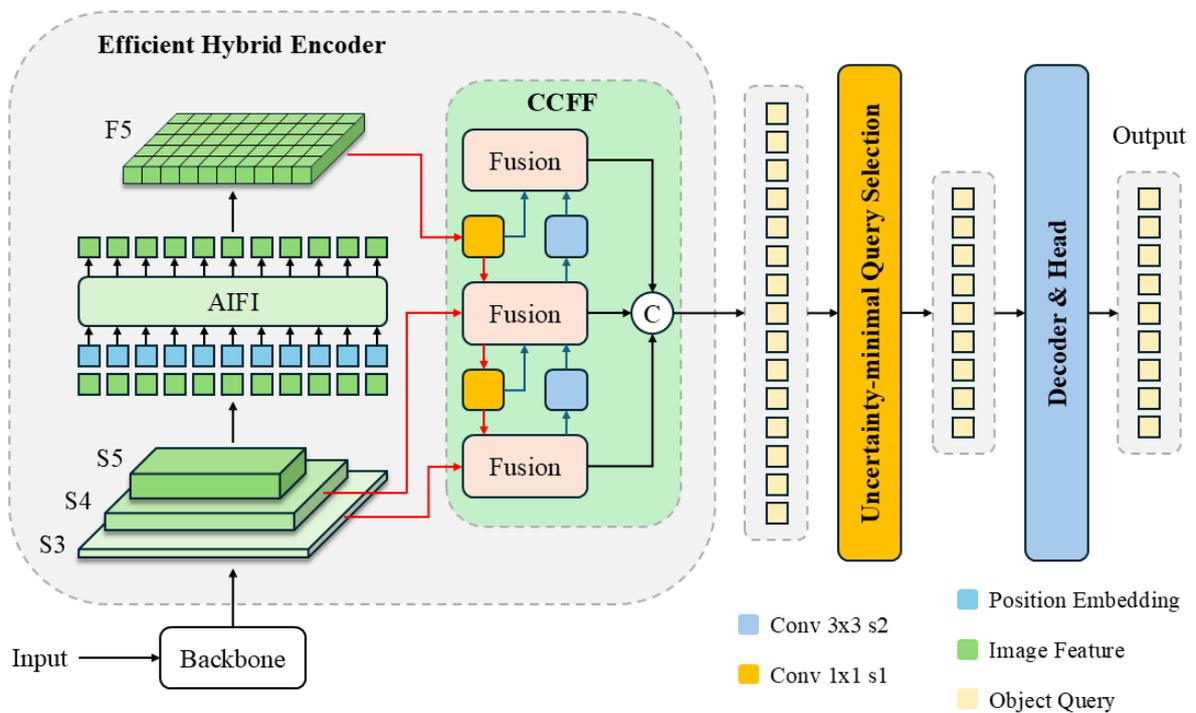
Figure 3.3 Structure diagram of the RT-DETR. AIFI denotes attention-based intra-scale feature interaction, while CCFF refers to cross-scale feature fusion.

3.2 Estimation of Underground Infrastructure Characteristics

In order to perform estimation of the underground infrastructure characteristics from B-scans, the output in the form of bounding box coordinates provided by the object detection algorithm is required. Using detection-derived regions of interest helps isolate the hyperbolic signature from surrounding reflections. By combining the aforementioned detected regions with a deep learning model architecture, the simultaneous estimation of parameters such as TWTT, burial depth, pipeline diameter, and the angle between the GPR survey line and a pipeline can be performed. In this setting, each cropped segment corresponds to a single hyperbolic reflection and is paired with continuous ground-truth labels for the four target variables. This estimation process relies on B-scan segments that are cropped according to the bounding box coordinates obtained during the object detection phase. For each identified hyperbola-shaped reflection, the corresponding segment is extracted from the original B-scan. Since the absolute position along the vertical axis significantly affects the shape and geometry of the hyperbola-shaped feature, such information is retained during the cropping. This way, the extracted hyperbola-shaped features preserve the spatial and structural information necessary for precise infrastructure characteristics estimation.

DL model architectures utilized for simultaneous parameter estimation of underground infrastructure are: DenseNet, Xception, NASNet, and EfficientNetV2. In all cases, ImageNet pretrained backbones are utilized as feature extractors. However, to enable multi-variable regression with the aforementioned models, a custom model head must be implemented since standard model heads are typically not designed to output continuous-valued parameters. Such a modification allows for the simultaneous estimation of multiple underground infrastructure characteristics [110]. The custom model head is composed of the following sequence of layers: a global average pooling 2D (GAP2D) layer, two fully connected (FC) layers, a dropout layer, an additional FC layer, and the output layer. The two initial FC layers each contain 128 neurons with the rectified linear unit (ReLU) activation functions. The dropout regularization is applied with a rate of 0.3 in order to minimize the probability of overfitting. The following FC layer consists of 64 neurons with a ReLU activation function. Finally, the output layer contains four neurons with a linear activation function, which enables the simultaneous regression of the four

target variables. The structure of the custom model head is shown in Figure 3.4, while an overview of the deep learning model architectures is provided in the following subsections.
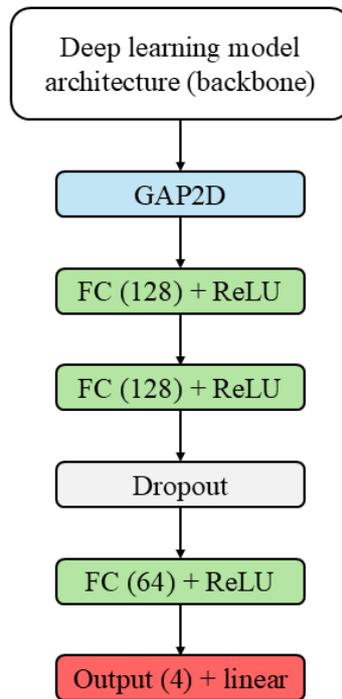
Figure 3.4 Structure of the custom model head implemented to extend standard deep learning architectures for multi-variable regression. This design enables the simultaneous estimation of underground infrastructure characteristics such as: TWTT, burial depth, pipeline diameter, and the angle between GPR survey line and the pipeline.

3.2.1 DenseNet

In 2016, Huang et al. [53] introduced the dense convolutional network (DenseNet). The network has shown various advantages, including alleviating the vanishing gradient problem, improving feature propagation, encouraging feature reuse, and reducing the number of parameters. The aforementioned is possible due to the dense connectivity among layers, which allows each layer to receive direct input from all preceding layers and pass its own feature maps to all subsequent layers in a feed-forward manner. The dense connection allows for direct access to gradients from the loss function and original signal throughout the neural network hierarchy [54]. This way, the training stability and convergence are improved. Due to the high efficiency and robustness to overfitting, architecture is commonly used in computer vision tasks. The structure of a deep DenseNet model with three dense blocks is shown in Figure 3.5.
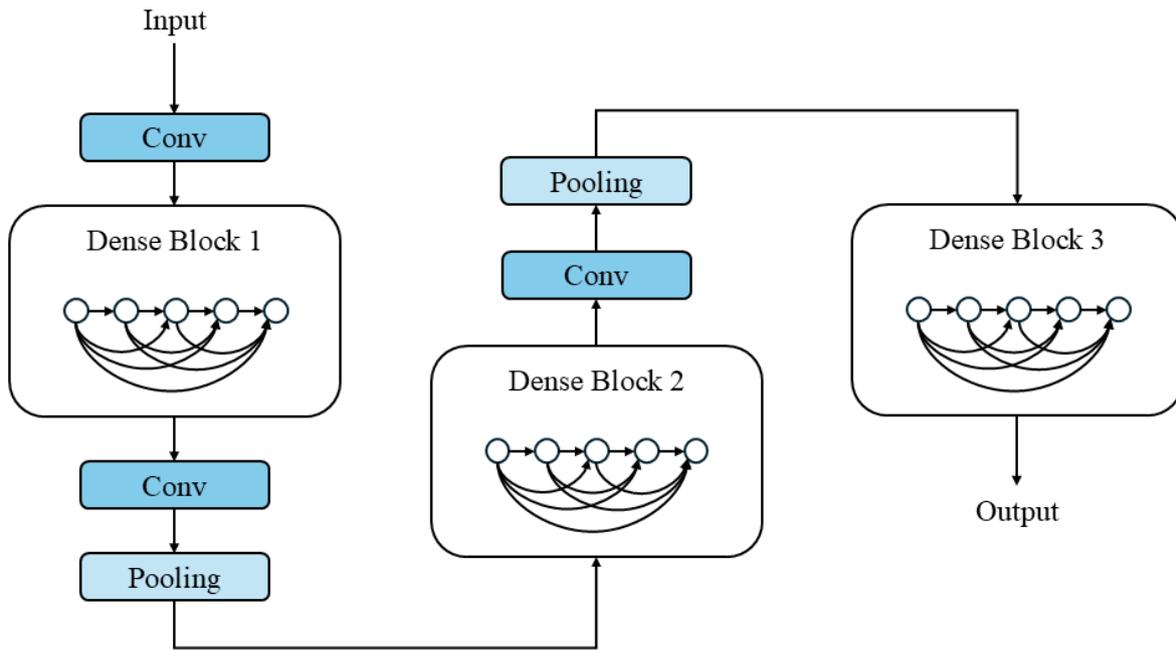
Figure 3.5 Structure of a deep DenseNet with three dense blocks. Transition layers, positioned between two adjacent blocks, change the size of feature maps through the application of convolution and pooling operators.

3.2.2 Xception

The Xception architecture proposed by Chollet [22] provides a substantial development in depthwise separable convolutions by reformulating the Inception hypothesis. The innovation utilizes depthwise separable convolutions as the fundamental building block, where Inception modules are replaced by depthwise convolutions followed by pointwise convolutions. The subsequent 1×1 pointwise convolution combines information across channels, whereas the depthwise convolution performs spatial filtering independently for each channel. The underlying hypothesis assumes that cross-channel and spatial correlations within feature maps can be fully decoupled. The depthwise separable convolution mechanism reduces parameter complexity, lowers computational cost, and improves model efficiency while maintaining equivalent or greater performance [64]. To facilitate gradient flow and convergence in deeper configurations, the architecture also incorporates linear residual connections around the convolutional blocks. Considering the aforementioned, the Xception model architecture is suitable for applications that require a balance between model complexity and performance. The structure of a Xception architecture is shown in Figure 3.6.
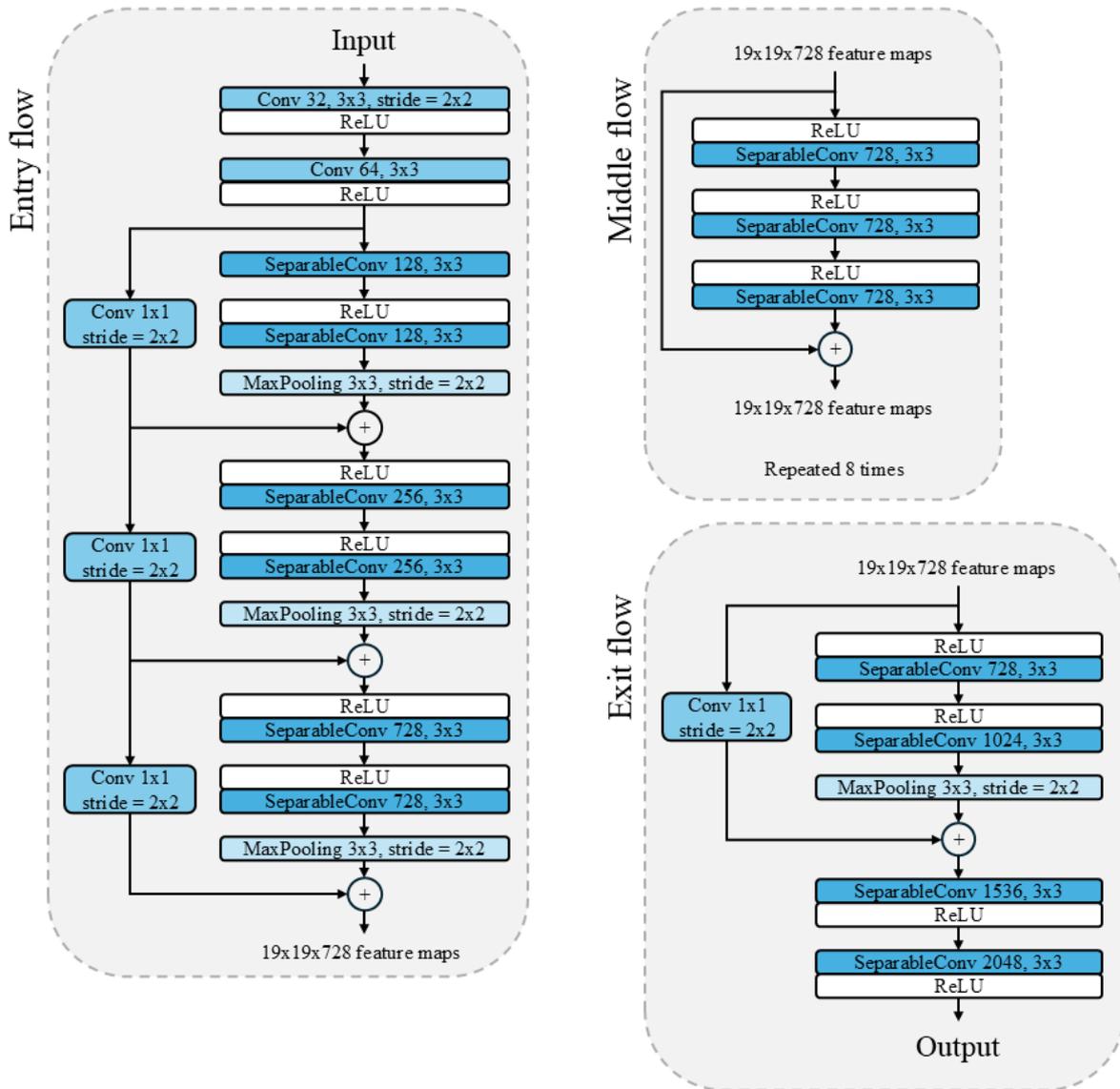
Figure 3.6 Structure of a Xception architecture. All convolution and separable convolution layers are followed by batch normalization layer.

3.2.3 NASNet

Zoph et al. [138] presented an approach to automatically design the model architectures directly on the dataset of interest by utilizing neural architecture search (NAS) techniques. Due to the high computational cost of searching on large-scale data, the authors employed a proxy task, identifying optimal cells on the CIFAR-10 dataset before transferring them to ImageNet. Reinforcement learning-based search algorithms are used in order to find the best convolutional cell structures within a predefined search space. The NASNet architecture used for ImageNet is presented in Figure 3.7.
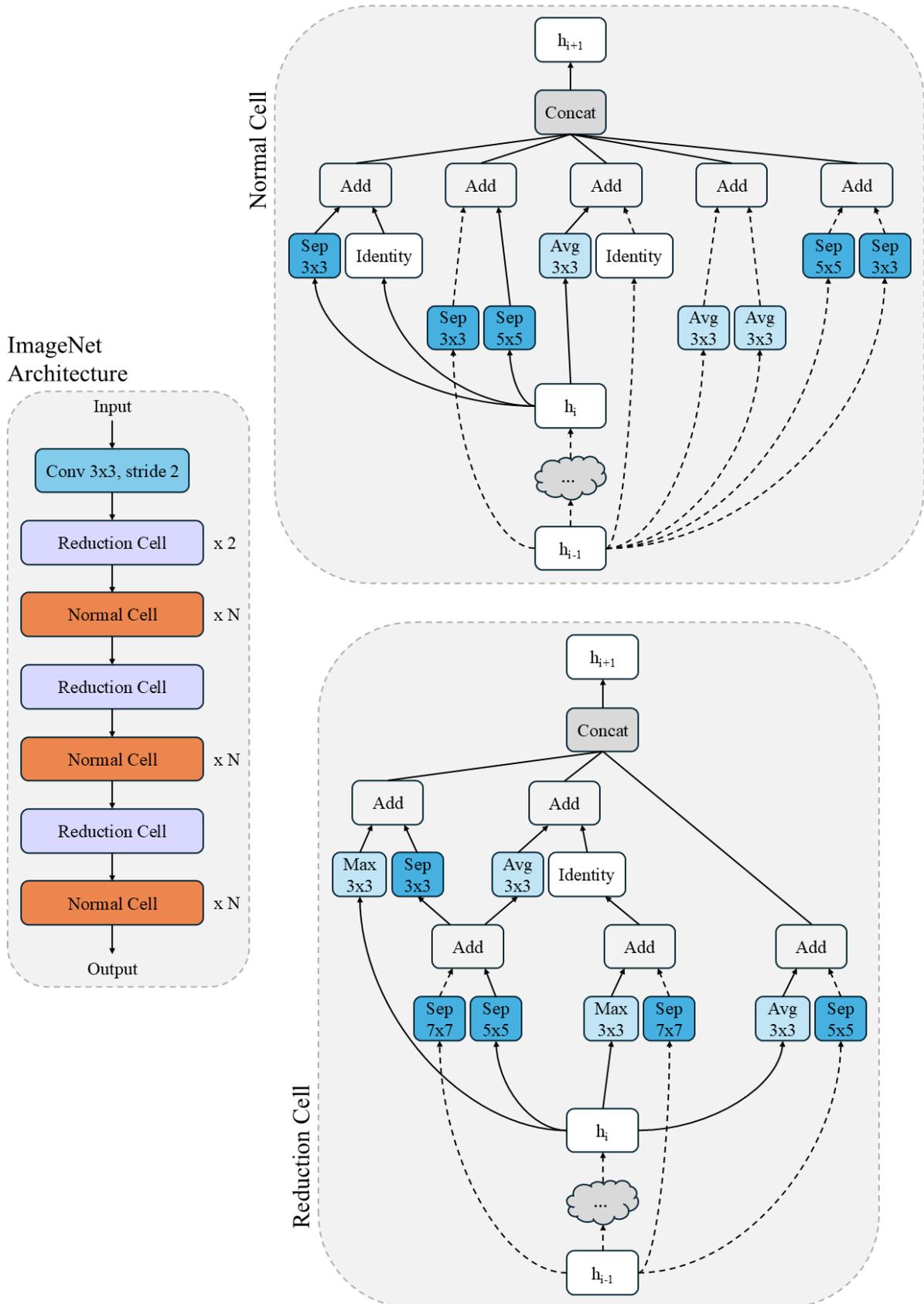
Figure 3.7 Overview of the NASNet scalable cell-based design. The diagram highlights the
ImageNet architecture configuration and the structure of the optimal NASNet-A convolutional
cells with B = 5 blocks originally identified on CIFAR-10.

To maximize performance and computational efficiency, the automated search mechanism systematically analyzes architectural configurations. The process identifies two types of modules: normal cells, which preserve spatial dimensions, and reduction cells, which perform downsampling [92]. Transferable convolutional cells can be stacked to construct neural networks with different depths. Therefore, applications requiring cutting-edge performance can benefit from the NASNet model design.

3.2.4 EfficientNetV2

Tan and Le [112] introduced the EfficientNetV2 family of models, achieving improved parameter efficiency and faster training by integrating training-aware NAS with optimized scaling strategies. By including Fused-MBConv (mobile inverted bottleneck convolution) blocks in addition to traditional MBConv blocks and implementing progressive resizing during training, the training efficiency can be increased. Fused-MBConv replaces the depthwise-pointwise decomposition with a single regular convolution in early stages, which tends to run more efficiently on modern accelerators when feature maps are large. The overall structure of the EfficientNetV2-L (large) architecture, together with the designs of the MBConv and Fused-MBConv blocks is shown in Figure 3.8. The progressive learning mechanism, which gradually increases image resolution and regularization strength during training, is utilized to enable the model to learn from smaller, less regularized images before progressing to full resolution inputs [107]. EfficientNetV2 is therefore appropriate for applications where high training efficiency and rapid model deployment are essential.
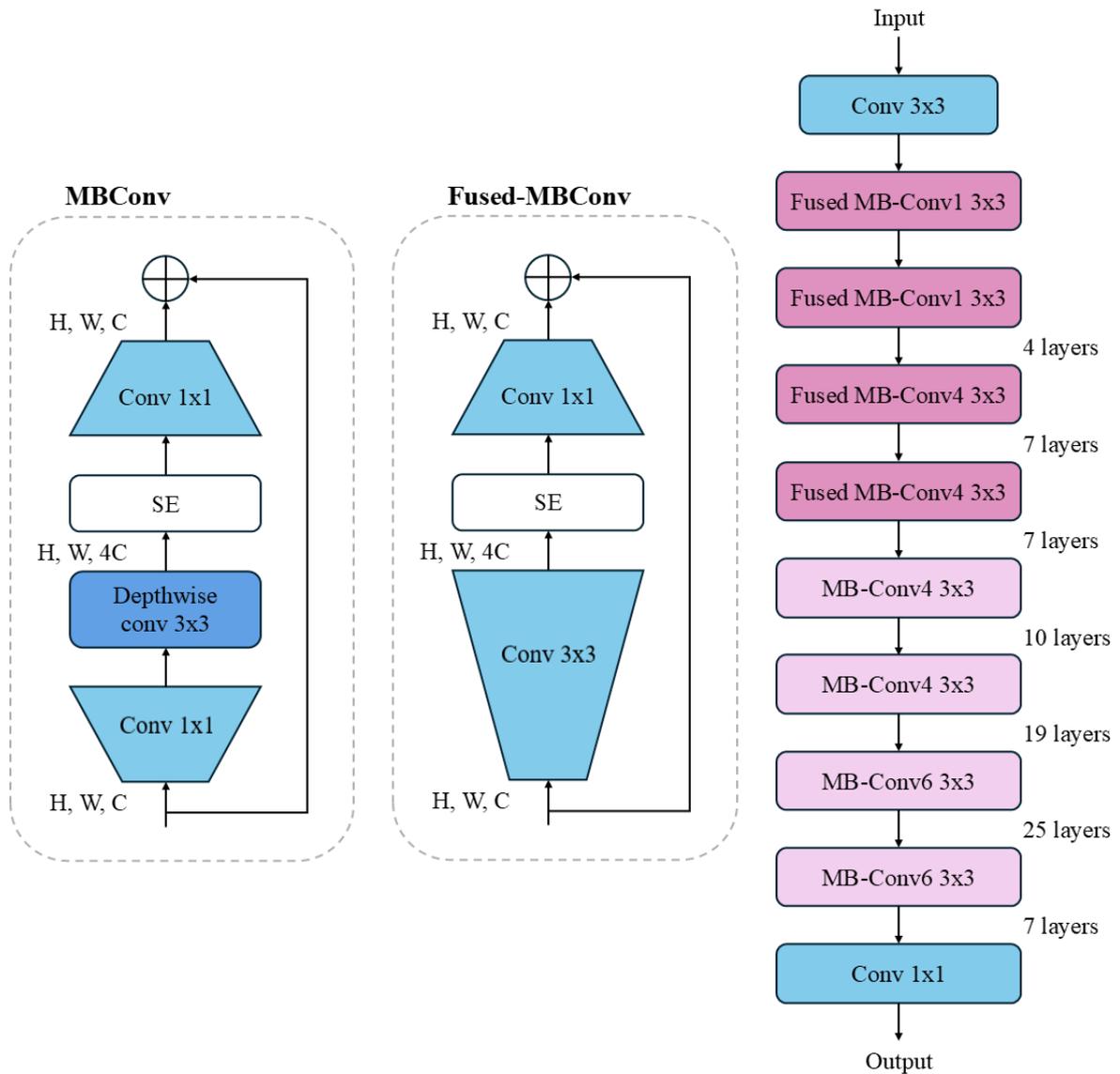
Figure 3.8 Structures of the EfficientNetV2-L (large) architecture, along with the designs of the MBConv and Fused-MBConv blocks.

CHAPTER 4

TIME-FREQUENCY ANALYSIS USING COHEN'S CLASS DISTRIBUTIONS

For stationary signals, the statistical and spectral properties remain constant throughout the entire observation period. In contrast, the term non-stationary signal refers to signals whose statistical properties change over time. The variations in statistical properties may be observed through different characteristics [14]. They may appear as amplitude fluctuations, modulations in instantaneous frequency, or the presence of spectral components. Therefore, most real-world signals, including GPR data, are non-stationary and can be either single- or multi-component [76]. A single-component signal is defined by one instantaneous frequency that changes over time, with chirp signals serving as a typical example. On the other hand, multi-component non-stationary signals consist of multiple simultaneous oscillatory modes, each with its own time-varying frequency. The aforementioned factors significantly increase the complexity of the signal analysis and interpretation. As a result, such signals often require methods capable of separating and tracking overlapping components over time. The main drawback of classical signal analysis, such as the Fourier transform, is the inherent trade-off between time and frequency resolution (known as the Heisenberg-Gabor limit) [23]. In other words, signal information is restricted to either the time domain or the frequency domain. The aforementioned limitation can be particularly problematic for multi-component non-stationary signals, where information on when a specific frequency component appears and how long it lasts is crucial

[14]. In order to overcome these challenges, TFDs have been introduced [24]. Such methods allow simultaneous localization of signal features in both time and frequency. In this chapter, the main concepts of non-stationary signal analysis are presented, with a focus on quadratic TFDs from Cohen's class.

4.1 Wigner-Ville Distribution

A one-dimensional time-domain signal $s(t)$ can be mapped into a two-dimensional time-frequency (t, f) representation using the Wigner-Ville distribution (WVD) [114]. It is considered as a fundamental TFD of Cohen's class. The WVD is obtained by applying the Fourier transform of the signal's instantaneous autocorrelation function (IAF) [13]. The mathematical definition of the standard Fourier transform $\mathcal{F}$ of a signal $s(t)$ is given by [16]:

$$S(f) = \mathop{\mathcal{F}}_{t \to f} \{s(t)\} = \int_{-\infty}^{\infty} s(t) e^{-j2\pi f t} \, dt.$$ (4.1)

The IAF of the signal $s(t)$, which correlates the signal with itself at two time instants centered around time t with a lag τ, is expressed as [14]:

$$K_s(t, \tau) = s\left(t + \frac{\tau}{2}\right) s^*\left(t - \frac{\tau}{2}\right),$$ (4.2)

where $s^*(t)$ represents the complex conjugate of the signal $s(t)$. Accordingly, the WVD is defined as the Fourier transform of the IAF with respect to the lag τ [13]:

$$WVD_s(t, f) = \int_{-\infty}^{\infty} s\left(t + \frac{\tau}{2}\right) s^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi f \tau} \, d\tau.$$ (4.3)

Mathematical properties satisfied by the WVD are realness, preservation of time and frequency marginals, time- and frequency-shift invariance, energy conservation, time and frequency support, instantaneous frequency, and group delay [47]. However, it is important to acknowledge that the WVD does not satisfy the non-negativity and cross-term suppression properties that are desirable for TFDs. Due to the quadratic nature of the WVD formulation, the interference terms, also known as cross-terms, are generated. Cross-terms occur centrally between the signal components, i.e., auto-terms, in the TF representation of the multi-

component signals and oscillate perpendicular to the line connecting them [48]. The rate of the aforementioned oscillation is proportional to the TF distance between the interacting auto-terms. Such artifacts manifest as "outer" cross-terms between distinct components or "inner" cross-terms caused by non-linear frequency modulation within a single component. The overall existence of cross-terms can create difficulties in the visual interpretation of complex TFD results. However, by smoothing the WVD utilizing the filtering function, i.e., kernel function, the unwanted cross-terms can be attenuated [14].

4.2 Smoothed and Pseudo Wigner-Ville Distributions

The pseudo Wigner-Ville distribution (PWVD) is derived by performing the same procedure used for the WVD, but with the additional step of applying a window function $h(\tau)$ to the lag variable of the IAF [75]. Since applying a window in the lag domain is equivalent to smoothing the distribution in the frequency domain, the PWVD attenuates cross-terms that oscillate along the frequency axis [46]. Accordingly, the PWVD is mathematically defined as [38]:

$$PWVD_s(t,f) = \int_{-\infty}^{\infty} h(\tau) s\left(t + \frac{\tau}{2}\right) s^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi f\tau}\, d\tau. \tag{4.4}$$

While the PWVD suppresses frequency-oscillating interference, it does not affect cross-terms oscillating in the time direction [23]. Furthermore, the application of the lag window decreases the frequency resolution of the auto-terms. Mathematical properties satisfied by the PWVD include realness, time- and frequency-shift invariance, preservation of the time marginal, energy conservation, time support, instantaneous frequency, and reduced interference [14].

In order to address the remaining cross-terms of the PWVD, the smoothed pseudo Wigner-Ville distribution (SPWVD) is utilized. By introducing an additional time-smoothing window $g(t)$, independent smoothing in both time and frequency directions is allowed. The SPWVD can be expressed as [37]:

$$SPWVD_s(t,f) = \int\limits_{-\infty}^{\infty} h(\tau) \int\limits_{-\infty}^{\infty} g(u-t)\, s\left(u+\frac{\tau}{2}\right) s^*\left(u-\frac{\tau}{2}\right) du\, e^{-j2\pi f\tau}\, d\tau. \qquad (4.5)$$

By adjusting the lengths of the windows $h(\tau)$ and $g(t)$, one is able to control the smoothing in the frequency and time domains, respectively [85]. Still, a trade-off remains between the level of the cross-terms and the resulting time-frequency resolution. Moreover, the SPWVD satisfies key mathematical properties such as realness, time- and frequency-shift invariance, and energy conservation, while providing reduced interference [49]. However, the time smoothing causes the loss of the time marginal and time support properties.

4.3 Reduced Interference Time-Frequency Distributions

Unlike the SPWVD, which applies separable smoothing windows, the Born-Jordan distribution (BJD) utilizes a signal-dependent kernel [26]. Such an approach changes how smoothing is introduced into the TF representation. In the ambiguity domain, the kernel is a sinc function, which is equivalent to a rectangular smoothing window in the time-lag domain. This operation effectively averages the IAF over the lag interval. The BJD is mathematically defined as [57]:

$$BJD_s(t,f) = \int\limits_{-\infty}^{\infty} \frac{1}{|\tau|} \int\limits_{t-\frac{|\tau|}{2}}^{t+\frac{|\tau|}{2}} s\left(u+\frac{\tau}{2}\right) s^*\left(u-\frac{\tau}{2}\right) du\, e^{-j2\pi f\tau}\, d\tau. \qquad (4.6)$$

The BJD attenuates the cross-terms significantly, especially the ones located between auto-terms [66]. The ability to reduce interference while preserving the signal's energy distribution boundaries is considered a key advantage of the BJD. Mathematical properties satisfied by the BJD are realness, time- and frequency-shift invariance, preservation of time and frequency marginals, energy conservation, instantaneous frequency, group delay, time support, frequency support, and reduced interference [14].

The next distribution utilized in this research is the Zhao-Atlas-Marks distribution (ZAMD), also known as the cone-shaped distribution (CSD). The ZAMD is obtained by

smoothing the BJD in the frequency direction. The distribution enhances the TF representation through a cone-shaped kernel defined in the time-lag domain [133]. Compared to the SPWVD, the ZAMD uses a non-separable kernel that limits the smoothing of the IAF to a triangular region where $|t| \leq |\tau|/2$. This way, the distribution effectively attenuates cross-terms while preserving the finite time support of the signal. The ZAMD is defined as [133]:

$$ZAMD_s(t,f) = \int_{-\infty}^{\infty} h(\tau) \int_{t-\frac{|\tau|}{2}}^{t+\frac{|\tau|}{2}} s\left(u + \frac{\tau}{2}\right) s^*\left(u - \frac{\tau}{2}\right) du \, e^{-j2\pi f\tau} \, d\tau. \qquad (4.7)$$

By restricting the inner integration limits, the ZAMD ensures that no signal energy is attributed to time instants where the signal does not exist, thereby eliminating the time-smearing effect observed in other smoothed distributions [48]. The ZAMD satisfies several useful mathematical properties, including realness, time- and frequency-shift invariance, time support, frequency support, and reduced interference [14]. Due to the specific constraints of the cone kernel, the ZAMD generally does not satisfy the time and frequency marginals, which represents a trade-off between precise time localization and total energy conservation.

The Choi-Williams distribution (CWD) is introduced in order to reduce WVD cross-terms while maintaining the marginal properties that are commonly sacrificed in smoothed distributions such as the SPWVD and ZAMD. Instead of applying a smoothing window in the time or frequency domain, the CWD utilizes an exponential kernel function in the ambiguity domain [21]. The kernel acts as a low-pass filter for the cross-terms, which are typically located far from the origin in the ambiguity domain. On the other hand, the auto-terms, which are concentrated near the origin, are preserved. The CWD can be written as follows [21]:

$$CWD_s(t,f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\sqrt{\pi\sigma}}{|\tau|} e^{\frac{-\pi^2\sigma(t-u)^2}{\tau^2}} s\left(u + \frac{\tau}{2}\right) s^*\left(u - \frac{\tau}{2}\right) du \, e^{-j2\pi f\tau} \, d\tau. \qquad (4.8)$$

Parameter σ controls the trade-off between cross-terms attenuation and auto-terms resolution [23]. In other words, a larger σ yields better resolution but less interference attenuation, and vice versa. Unlike the SPWVD and ZAMD, the CWD is designed to satisfy the marginal properties. Accordingly, some of the mathematical properties satisfied by the CWD are realness, time- and frequency-shift invariance, preservation of time and frequency marginals, energy conservation, instantaneous frequency, group delay, and reduced interference [46].

4.4 Time-Frequency Representations of GPR A-scans

In this section, Cohen's class TFDs are applied to GPR A-scan signals. Real-world data are considered non-stationary due to the frequency-dependent attenuation caused by complex subsurface media. Therefore, by transforming the one-dimensional A-scan into the TF domain, additional potentially useful spectral features (e.g. energy localization and bandwidth variation over time) can be extracted from the original data. These features can be critical in improving the estimation of underground infrastructure characteristics. The B-scan shown in Figure 2.11, containing the hyperbola-shaped reflections, is used for the TF representation. However, before further analysis, preprocessing is required to enhance signal quality and interpretability, as described in Chapter 2. Additionally, prior to TFD computation, the Hilbert transform is applied in order to obtain the corresponding analytic signal. This pre-processing step is essential for Wigner-type distributions since it suppresses artifacts introduced by negative frequency components. To further condition the data prior the transformation, the individual A-scans are zero-centered and tapered using a Tukey window. A 10% taper ratio is applied (smoothly attenuating the first and last 5% of the signal) to reduce spectral leakage and edge artifacts. Furthermore, the distributions are computed across N frequency bins, where N represents the total number of discrete points in the isolated A-scan. For the TFDs requiring independent smoothing (PWVD, SPWVD, BJD, ZAMD, and CWD), Hamming window is utilized. The temporal and spectral window sizes are dynamically scaled to approximately $N/4$ and $N/8$, respectively. Additionally, the kernel width parameter σ for the CWD is set to 3 to control cross-term suppression. The resulting TF matrices are afterwards normalized, mapping the minimum and maximum amplitude values to a range between 0 and 1 in order to ensure stable feature extraction by the hybrid architecture:

$$x_{TFD\_norm} = \frac{x_{TFD} - x_{TFD\_min}}{x_{TFD\_max} - x_{TFD\_min}}, \qquad (4.9)$$

where x_{TFD_norm} is the normalized value, x_{TFD} represents the original amplitude value of an individual data point, while x_{TFD_min} and x_{TFD_max} represent the global minimum and maximum amplitudes, respectively. The following analysis demonstrates the computation and comparative performance of six TFDs: WVD, PWVD, SPWVD, BJD, ZAMD, and CWD, as shown in Figures 4.1 and 4.2. As the fundamental distribution of Cohen's class, the WVD

(Figures 4.1a and 4.2a) provides the highest theoretical TF resolution. From the TF representation, it can be seen that the distribution provides sharp localization of the hyperbola-shaped reflection event at approximately 24 – 30 nanoseconds with the center frequency near 350 MHz. However, due to the WVD's quadratic nature, cross-terms are generated between components of the signal. By applying a window function to the lag variable of the IAF, the PWVD (Figures 4.1b and 4.2b) successfully attenuates cross-terms that oscillate along the frequency axis. Although the time resolution remains high, the frequency resolution of auto-terms is slightly degraded compared to the WVD. In other words, the PWVD preserves the sharp definition of the hyperbola-shaped apex's arrival time. Introducing an additional time-smoothing window in the SPWVD (Figures 4.1c and 4.2c) results in independent smoothing in both the time and frequency directions. Cross-terms are almost entirely suppressed in the TF representation. The smoothing comes at the cost of a slight blurring of the distribution. Accordingly, the sharp edge of the pipeline reflection is widened, potentially masking spectral features necessary for precise infrastructure parameter estimation. The BJD (Figures 4.1d and 4.2d) reveals a horizontal smoothing effect since it is designed to suppress cross-terms while preserving time support and the signal's energy distribution boundaries. The TF representation retains more vertical detail compared to the SPWVD. Due to a cone-shaped kernel defined in the time-lag domain, the ZAMD (Figures 4.1e and 4.2e) provides a balanced compromise between cross-term suppression and auto-term concentration. From the presented TF representation it can be seen that ZAMD eliminates the time-smearing effect observed in other smoothed distributions. Finally, by utilizing an exponential kernel function in the ambiguity domain controlled by a scaling factor σ, the CWD (Figures 4.1f and 4.2f) achieves a balanced result. The distribution effectively suppresses the dominant cross-terms, while at the same time preserving sharper auto-term localization than the SPWVD.
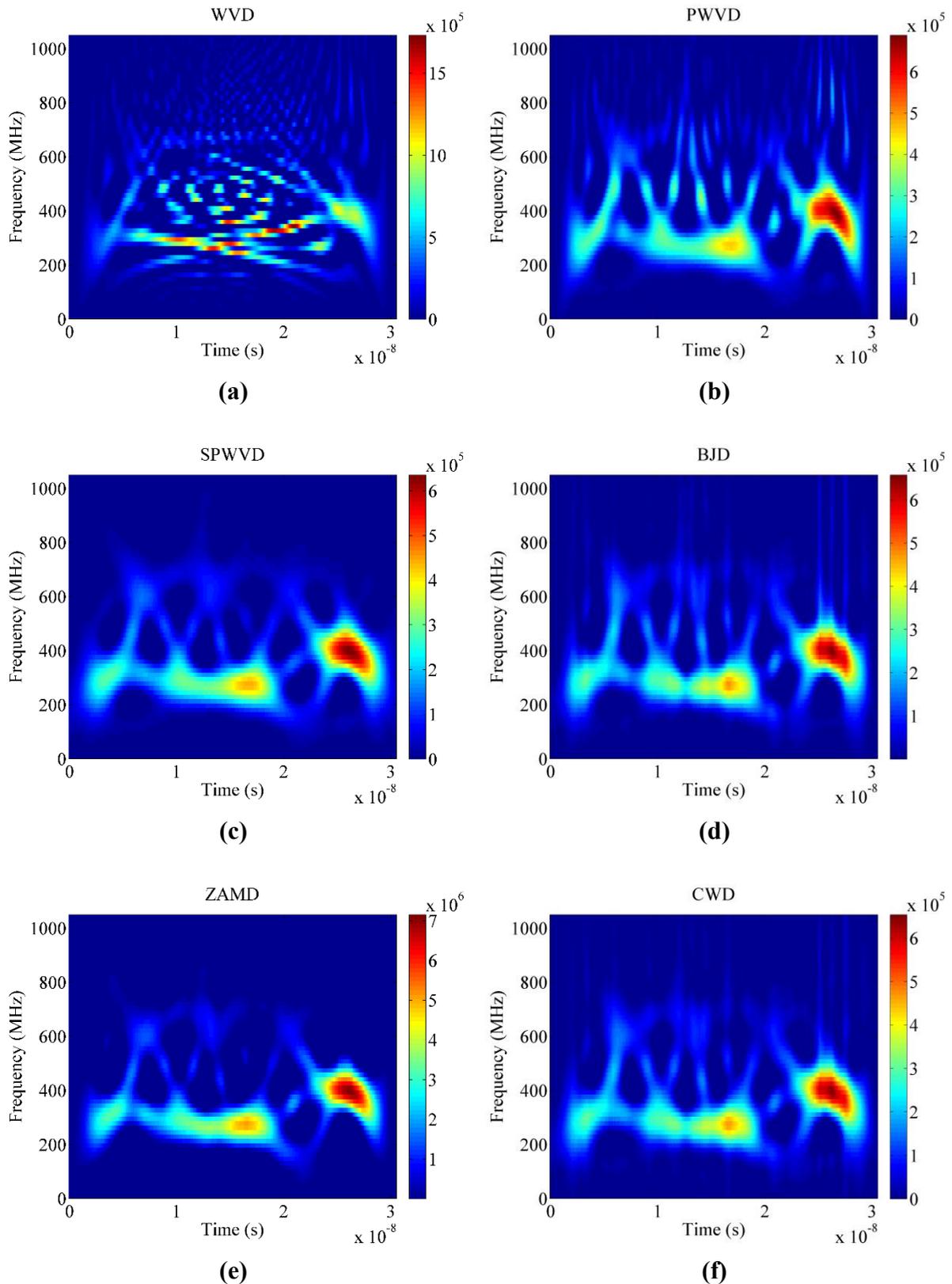
Figure 4.1 TFDs of the A-scan acquired at a distance of 2.0 meters: (a) WVD, (b) PWVD, (c) SPWVD, (d) BJD, (e) ZAMD, and (f) CWD. The hyperbola-shaped reflection (apex) appears within the A-scan at 24.40 nanoseconds, which corresponds to a burial depth of 1.1 meters.
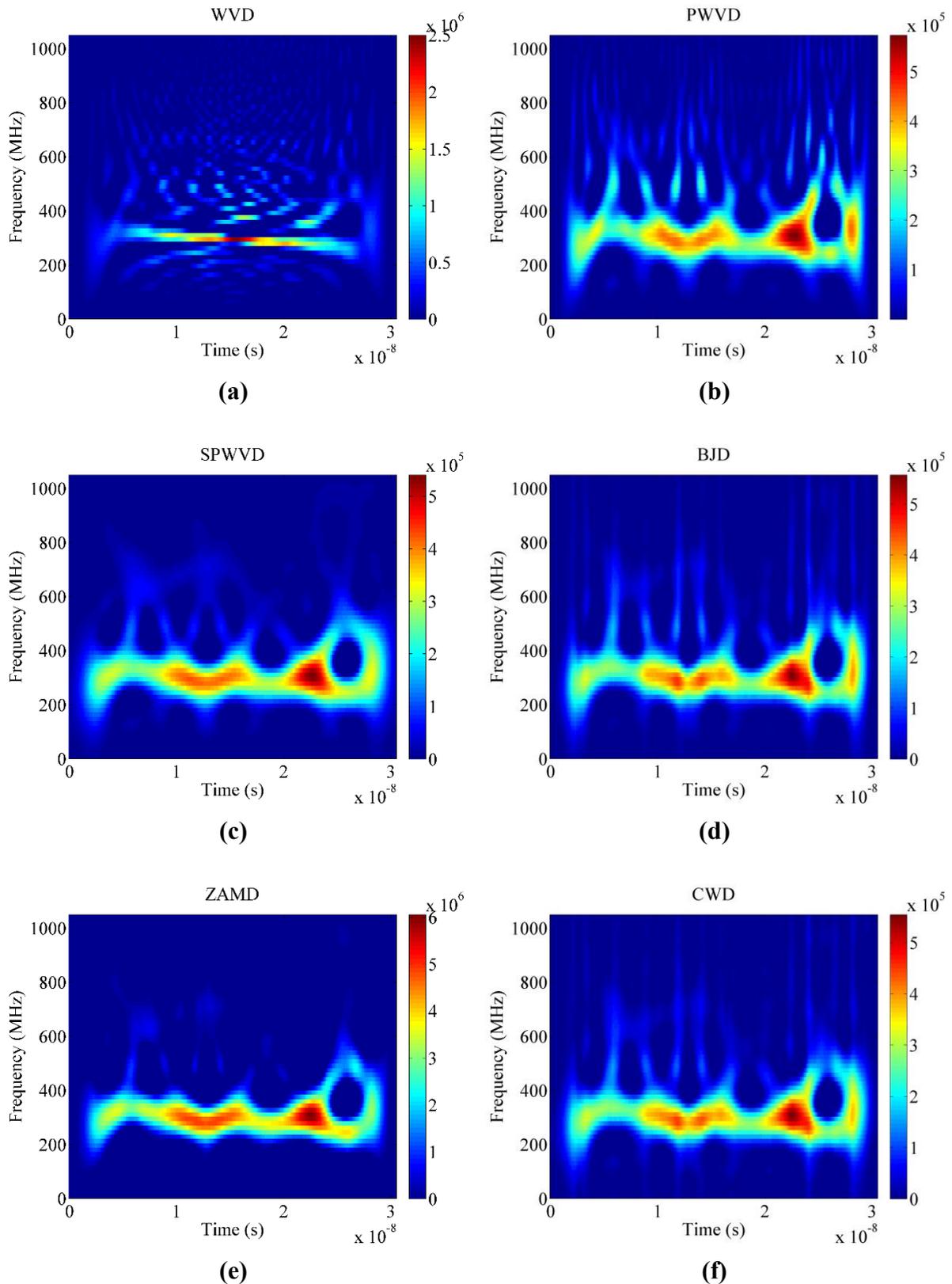
Figure 4.2 TFDs of the A-scan acquired at a distance of 2.5 meters: (a) WVD, (b) PWVD, (c) SPWVD, (d) BJD, (e) ZAMD, and (f) CWD. The right branch of the hyperbola-shaped reflection appears within the A-scan at 26.50 nanoseconds.

CHAPTER 5

DEEP LEARNING-BASED SYSTEM FOR ESTIMATING UNDERGROUND INFRASTRUCTURE PARAMETERS

This chapter describes the proposed DL-based framework designed to improve the baseline estimation of underground infrastructure parameters. The end-to-end pipeline relies on the preliminary detection and localization of hyperbola-shaped reflections from the B-scan GPR data. Afterwards, the isolated data is passed into the core architecture, referred to as the cross-attention spatio-spectral fusion network (CASS-Fusion). In order to thoroughly utilize the multi-domain characteristics of B-scan data, the network is structured as dual-branch architecture. Branch 1 only processes the spatial domain by passing cropped GPR B-scans, formatted to a structural shape of (H, W, 3), through a dedicated spatial feature extractor, which can utilize any suitable convolutional backbone. At the same time, branch 2 processes the spectral domain by passing TFDs with a dimensions of (120, 64, 64, 1) into a parallel spectral feature extractor. To synthesize the independent representations, the resulting high-level feature vectors are integrated via a cross-attention and dynamic gated fusion module. The fused multi-

domain features are afterwards processed by a custom regression head, which simultaneously estimates four descriptive parameters of underground infrastructure: TWTT, burial depth, pipeline diameter, and intersection angle. The proposed DL-framework is illustrated in Figure 5.1.
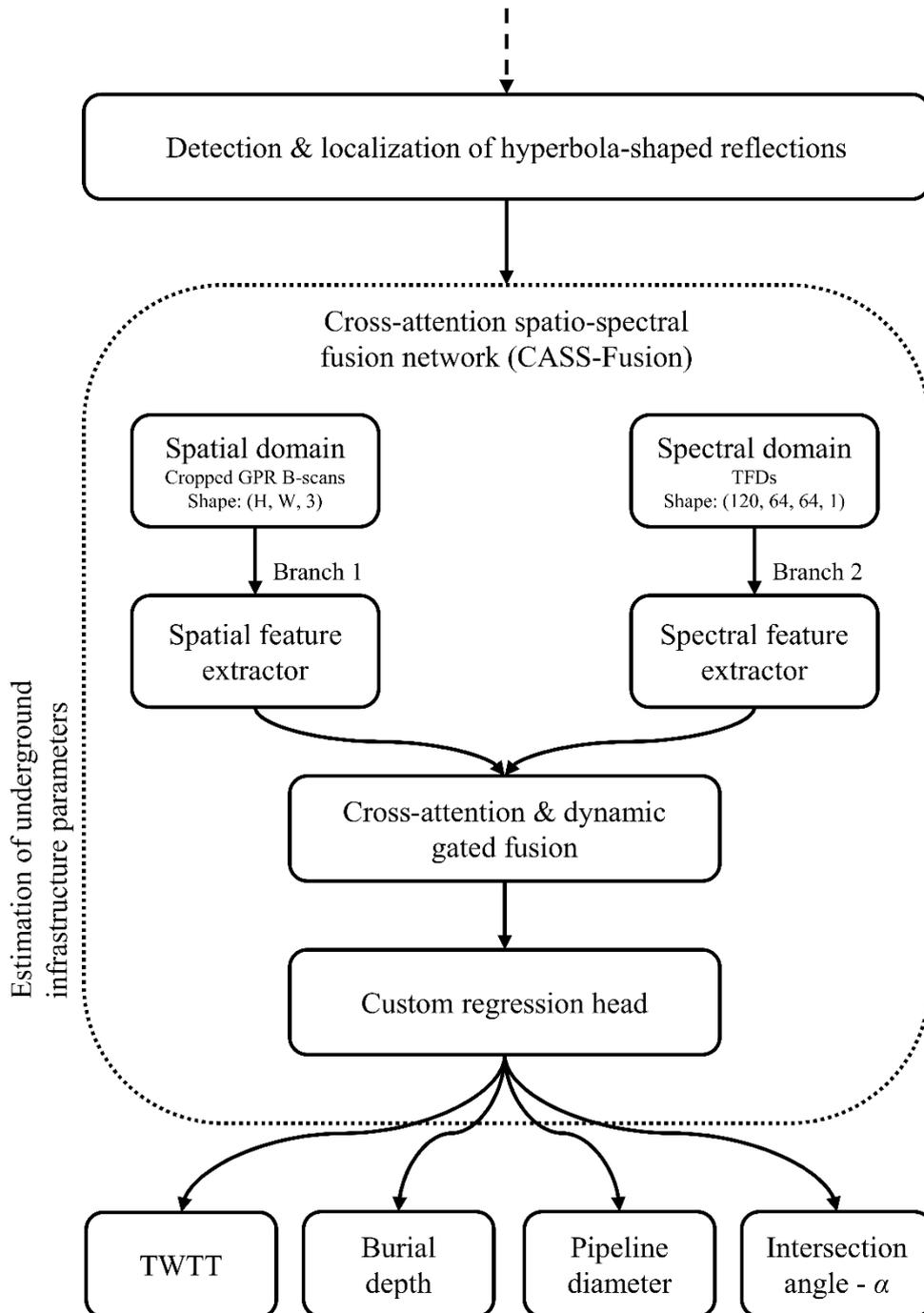
Figure 5.1 The proposed framework for underground infrastructure estimation, highlighting the multi-domain fusion of spatial B-scans and spectral TFDs via the CASS-Fusion network.

5.1 Proposed Cross-Attention Spatio-Spectral Fusion Network

The CASS-Fusion network is a multi-domain DL framework for the simultaneous estimation of four underground infrastructure descriptive parameters. The model fuses features from two complementary representations of the same GPR data: the spatial B-scan data, and a sequence of TFDs computed from the individual A-scans within the cropped B-scan. The motivation for dual-domain fusion is grounded in the physical properties of GPR hyperbola-shaped reflections [81]. The B-scan data encodes the geometric signature of the hyperbola-shaped reflection, from which descriptive parameters can be estimated by a spatial feature extractor. However, the estimation performance is limited by the spatial domain representation alone, since some, potentially important features depend on subtle frequency domain characteristics such as resonance patterns and spectral attenuation that are not directly visible in the B-scan pixel values [50]. TFDs, by representing the joint TF content of each A-scan, expose spectral features and provide information that is complementary to and partially independent of the spatial representation [98]. The CASS-Fusion network consists of a spatial feature extractor processing B-scans, a 3D convolutional and bidirectional LSTM (BiLSTM) spectral feature extractor processing TFD sequences, a multi-head cross-attention block that computes spectral-spatial attended representations by allowing TFD features to query the spatial map, and a dynamic gated fusion mechanism that adaptively combines the cross-attended representation with a direct spatial residual connection prior to the regression head [102, 109, 118, 126].

A visual representation of the data preparation required for the CASS-Fusion network is shown in Figure 5.2. The spatial input is a B-scan cropped according to the coordinates of the detected and localized hyperbola-shaped reflection as described in subsection 3.2, which is resized to the fixed spatial dimensions required by the utilized backbone. The magnifying glass within the illustration highlights the A-scan extraction step. Within the cropped B-scan, each vertical column of data corresponds to one A-scan, a single 1D time-domain waveform recording the amplitude of the reflected EM signal, several of which are illustrated in the magnified view. Each A-scan within the cropped B-scan is extracted as a 1D signal and processed independently to compute its TFD. The number of A-scans varies from sample to sample depending on the width of the detected hyperbola-shaped reflection.
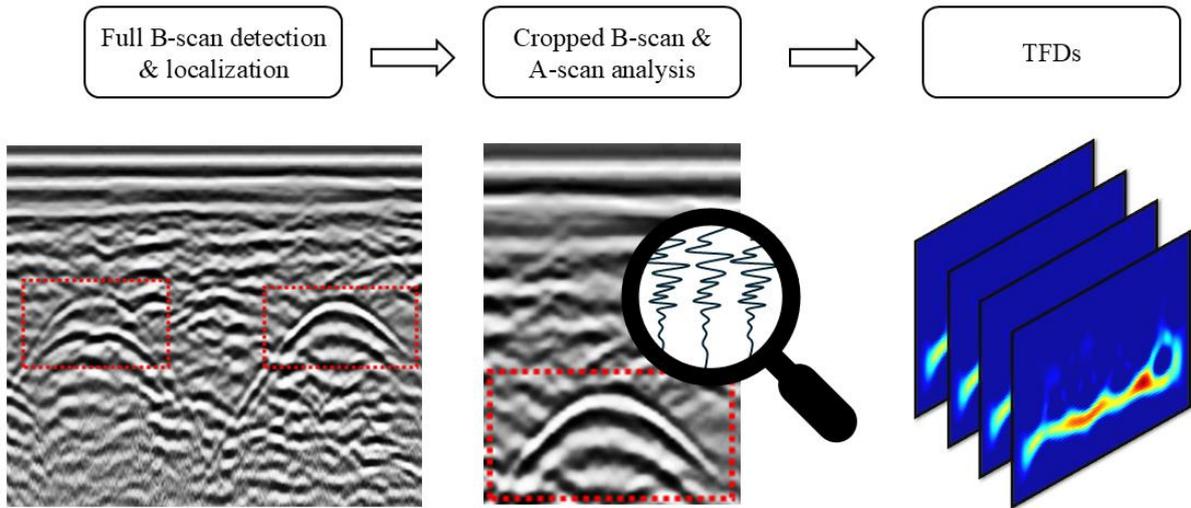
Figure 5.2 Visual representation of the data preparation required for the CASS-Fusion network. The workflow isolates hyperbola-shaped reflections to produce cropped spatial B-scans for branch 1, while at the same time extracting and transforming A-scans into TFDs for branch 2.

In order to produce a fixed length sequence compatible with the network requirement, the A-scan sequence is symmetrically center-cropped or zero-padded to a fixed length of 120. If more than 120 A-scans are available, the excess is removed equally from both the left and right ends of the sequence. If fewer than 120 A-scans are available, zeros are inserted symmetrically on both sides. In both cases, the apex of the hyperbola-shaped feature remains at approximately the center of the resulting 120-element sequence. The spatial sampling interval of the GPR used in this research is 60 samples-per-meter, which at the fixed input window of 120 A-scans corresponds to a physical survey extent of 2 meters. The fixed-window design provides flexibility for use with GPR devices operating at different spatial sampling intervals. Utilizing a device with a sampling interval of 30 samples per meter will, for example, result in a configuration where TFDs occupy every other position within the 120-slot window, with the remaining positions zero-padded. This configuration allows recordings from GPR systems with different spatial resolutions to be represented within the same fixed-length input structure. After the TFD for each of the 120 A-scans in the sequence is computed, they are stacked together as shown on the right side of the illustration. Each TFD image is resized to the fixed resolution of 64×64 using bilinear interpolation and stored as a single-channel array. Structurally, the 120 generated TFDs are stacked across a newly defined leading axis. The stacking process produces the $(120, 64, 64, 1)$ 4D tensor required by the spectral branch of the CASS-Fusion architecture.

Two outputs are extracted from the spatial backbone for use in the hybrid fusion model: a spatial feature map with the shape of (H, W, C), and a global pooled descriptor yielding a single vector that encodes a summary of the entire B-scan. The complete layer-wise structure of the proposed CASS-Fusion network is shown in Figure 5.3. The spectral branch processes the TFD sequence through a hierarchical architecture that first extracts 3D local patterns from the joint TF-position space using 3D convolutions and then models sequential dependencies across A-scan positions using stacked BiLSTMs. The input tensor of shape (120, 64, 64, 1) is accepted by the 3D CNN and processed through three convolutional blocks, each following the pattern Conv3D – BatchNormalization – ReLU – MaxPooling3D. The three dimensions of each 3D convolutional filter extend across the temporal A-scan axis and both spatial axes of the TFD image simultaneously. As a result, the network can successfully learn patterns that are coherent across neighboring A-scan positions and across the TF plane jointly. Following the standard practice of increasing representational depth as spatial resolution decreases, the filter count doubles with each block (32 – 64 – 128). Batch normalization after each convolutional layer normalizes the activation distributions across the batch dimension. The third pooling layer applies an asymmetric pool size of (1, 2, 2), thereby preserving the resolution along the A-scan position axis. This way, all 30 A-scan positions are retained, while the spatial dimensions of each TFD image are reduced from 16×16 to 8×8. Such an asymmetric design reflects the physical distinction between the TFD's spatial and sequential components. While the sequential axis must remain uncompressed to support deep sequential encoding via the BiLSTM layers, the spatial TFD axes undergo dimensionality reduction to extract compact spectral features. Following the 3D CNN, the output tensor of shape (30, 8, 8, 128) is processed using a time-distributed wrapper, which applies the same sub-layer independently to each of the 30 A-scan positions. First, a time-distributed flatten layer reshapes each spatial volume produced by the third convolutional block into a flat vector, concatenating all spatial positions and all feature channels into a single 1D representation per A-scan. In the context of TFDs, the spatial position within the image contains physical meaning, as different TF regions encode different signal characteristics. Therefore, flattening preserves the full spatial layout, which allows the projection to learn which specific TF regions and feature channels are jointly informative for the regression targets. The FC layer following the flatten can thereby learn spatially-aware weighted combinations of activations, rather than being constrained to treat all spatial positions as interchangeable, as global pooling assumes.
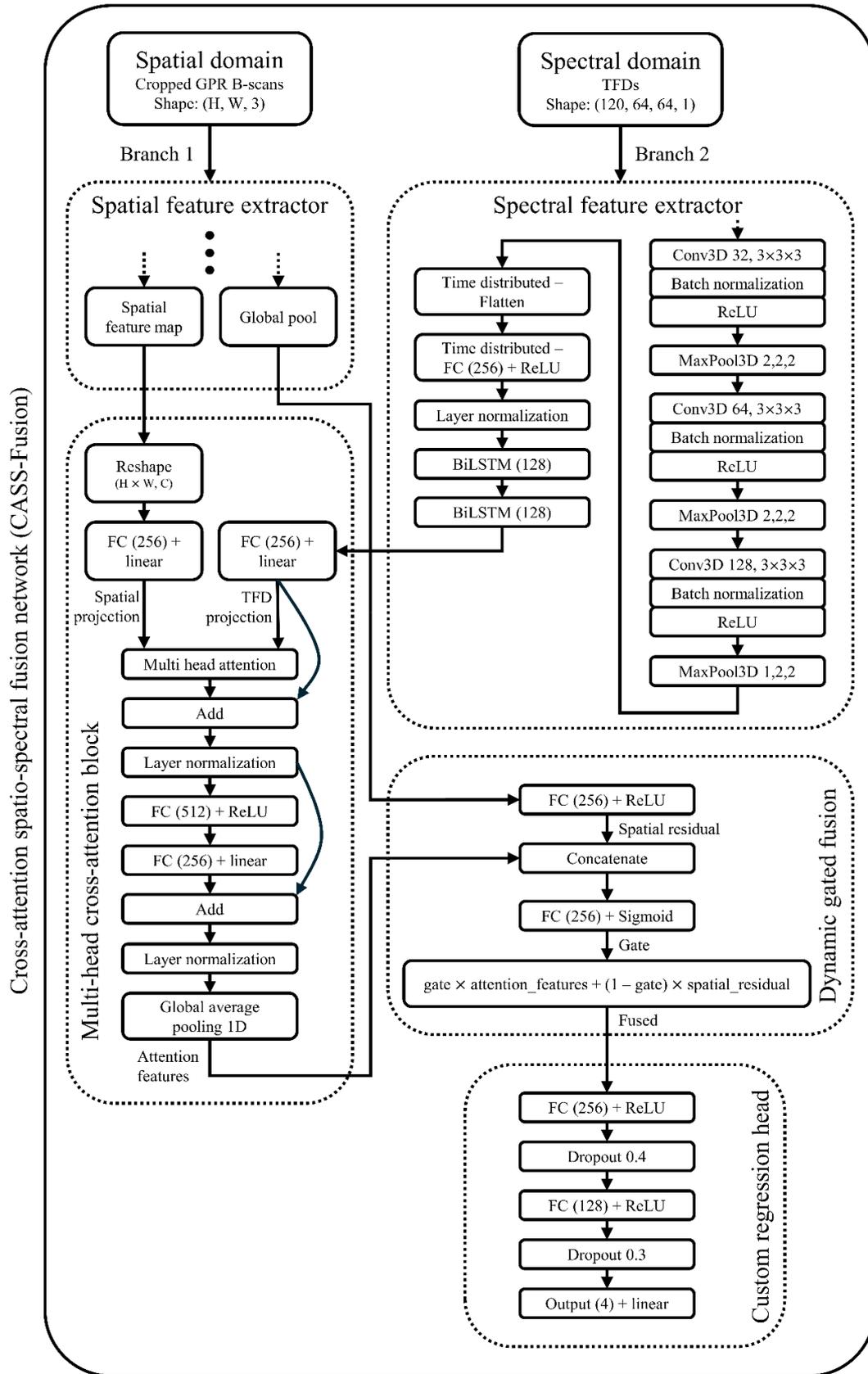
Figure 5.3 Structure of the proposed cross-attention spatio-spectral fusion (CASS-Fusion) network, illustrating the parallel spatial and spectral feature extractors, the multi-head cross-attention block, the dynamic gated fusion mechanism, and the custom regression head.

A time-distributed FC layer with 256 neurons and ReLU activation function follows, projecting each flattened vector into a 256-dimensional feature space. The FC projection serves two purposes simultaneously: dimensionality reduction to a compact representation and learning of a non-linear task-specific transformation that maps raw convolutional activations into a representation space suitable for sequential modelling. A layer normalization layer is applied across the 256 feature dimensions at each of the 30 spatial positions, preventing the BiLSTM input distribution from drifting during training. Two stacked BiLSTM layers, each with 128 hidden units per direction, are applied sequentially to the normalized (30, 256) sequence. Each bidirectional wrapper concatenates the outputs of a forward-direction LSTM and a backward-direction LSTM, which produces a 256-dimensional output vector at each of the 30 A-scan positions. Both layers preserve the full output sequence across all 30 spatial positions rather than returning only the final state. Processing the sequence in both directions allows the model to use future context (right-side information) when encoding left-side positions, and vice versa. This process enables the hidden states to encode the global geometry of hyperbola-shaped feature rather than only the local history. The output of the second BiLSTM is used as the query input to the cross-attention block in the hybrid model.

The cross-attention block implements a single transformer encoder layer in which the spectral TFD features are formulated as queries and the spatial features (from the spatial feature extractor) are formulated as keys and values. The asymmetric cross-domain attention enables each TFD sequence position to selectively aggregate relevant spatial information from the B-scan. Before the attention computations, both input sequences must be mapped into a common subspace of dimension 256, which is achieved by two independent linear projection layers. The projected spatial sequence serves as both the key sequence and the value sequence in the attention operation. The choice of a linear projection is consistent with the transformer formulation. The TFD projection learns a task-specific remapping of the LSTM hidden states into the query subspace, distinct from the representation used internally by the BiLSTM.

The multi-head attention mechanism is configured with 4 heads, a key dimension of 64, and an attention dropout rate of 0.1. The physical interpretation is that each of the 30 TFD query positions computes a soft weighted combination of all 100 spatial key positions, where the weights reflect the compatibility between the spectral features at that A-scan position and the spatial features at each location in the B-scan. This way, the model is able to correlate the frequency content at a particular side position of the hyperbola-shaped reflection with the

corresponding spatial shape of the reflection branch at that position in the B-scan. Following the attention computation, standard transformer post-processing is applied. The attention residual connection adds the original TFD query sequence back to the attention output, which ensures the original spectral features are preserved regardless of attention quality during early training. Layer normalization is then applied across the 256 feature dimensions. To expand and re-project the feature space, two FC layers, the first with 512 neurons and ReLU activation, and the second with 256 neurons and linear activation, are applied to each of the 30 spatial positions. The implementation of the two residual connections provides a direct gradient flow that stabilizes the training of the cross-attention. To conclude the module, a 1D global average pooling layer integrates the cross-attended spectral-spatial features, representing the final output of the block.

The dynamic gated fusion mechanism combines the cross-attended representation, referred to as attention features, with the spatial residual vector. Through this mechanism, the network adaptively modulates the contribution of each branch on a per-feature-dimension basis for each input sample. Moreover, the spatial residual connection ensures that the global spatial representation from the spatial backbone is always available to the fusion stage, even if the cross-attention produces suboptimal outputs. The gating vector is computed from the concatenation of both 256-dimensional representations, making the gate conditioned on both the attended spectral-spatial features and the spatial residual simultaneously.

The custom regression head maps the 256-dimensional fused representation to the four-dimensional parameter output through two FC layers with progressive dimensionality reduction and regularization. The use of a single shared regression head for all four descriptive parameters guides the network to learn a unified representation that captures the shared physical relationships among the parameters.

5.2 Evaluation Criteria

In order to evaluate the proposed two-stage framework, performance is estimated separately for the detection and localization stage and the parameter estimation stage, using metrics appropriate for each learning task. For first stage of the proposed framework, which includes object detection and localization, precision, recall, and mAP are utilized. Precision and

recall are calculated from the number of true positives (TP), false positives (FP), and false negatives (FN) as follows [86]:

$$Precision = \frac{TP}{TP + FP}, \tag{5.1}$$

and

$$Recall = \frac{TP}{TP + FN}. \tag{5.2}$$

Object detection accuracy is further quantified using the intersection over union (IoU) between the predicted bounding box A and the ground truth bounding box B, mathematically written as [86]:

$$IoU = \frac{|A_{bbox} \cap B_{bbox}|}{|A_{bbox} \cup B_{bbox}|}, \tag{5.3}$$

where $A_{bbox} \cap B_{bbox}$ represents the area where the estimated and ground truth boxes overlap, and $A_{bbox} \cup B_{bbox}$ represents the total area covered by both bounding boxes. The mAP score is acquired by calculating precision and recall at various IoU thresholds, followed by calculating the AP for each threshold, and afterwards averaging the AP values to obtain the mAP score. The metrics defined in (5.1-5.3) have a value range between 0.0 and 1.0, where higher values indicate better detection and localization performance.

For second stage of the proposed framework, which includes estimation of underground infrastructure parameters (TWTT, burial depth, cylindrical object diameter, and the intersection angle between the GPR survey line and the target object), the task is formulated as a regression problem. The coefficient of determination (R^2), mean absolute error (MAE), and root mean squared error (RMSE) are utilized. The mathematical representation of the aforementioned measures is given by [20]:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}, \tag{5.4}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|, \tag{5.5}$$

and

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5.6}$$

where y_i denotes the ground truth value, $\hat{y}_i$ the corresponding estimate, $\bar{y}$ the mean of all observed true values, and n the number of samples. The R^2 ranges from 0.0 to 1.0, where a value of 1.0 indicates that the model explains all the variability in the data. On the other hand, lower MAE and RMSE values correspond to more accurate estimations of the model.

In both stages, k-fold cross-validation is utilized to perform a robust estimation of models' performances. Recognized as a standard approach for evaluating generalization, this technique divides the dataset into k equal folds [97]. Afterwards, the model is trained and evaluated k times, each time using a different fold as the validation set while the remaining k-1 folds are used for training purposes. The metrics from each fold are then averaged to provide a less biased estimate of model performance. By employing the described technique, it is ensured that all samples from the dataset are used for both training and validation exactly once [121]. However, in practice, the selection of k represents a direct trade-off between computational cost and the stability of the estimated generalization performance. In the case of this research, 5-fold cross-validation ($k = 5$) is employed. When dividing the dataset into five folds, it is ensured that the training and validation splits preserve an approximately similar data distribution. Such strategy reduces the risk that the validation set contains samples which are systematically different from those used for training purposes. Ignoring the aforementioned balance could result in biased performance estimates.

CHAPTER 6

RESULTS AND DISCUSSION

This chapter presents the experimental results of the proposed two-stage DL framework for automated characterization of underground infrastructure from GPR data. The first stage addresses the detection and localization of hyperbola-shaped reflections within B-scans, while the second stage analyzes the detected reflections to estimate the descriptive parameters of underground infrastructure, namely TWTT, burial depth, pipeline diameter, and the intersection angle between the GPR survey line and the cylindrical object i.e., pipeline. Although the synthetic data is generated primarily to support the second stage, its impact on detection and localization performance is also experimentally investigated. Since hyperbola fitting is one of the most widely used approaches for estimating descriptive parameters of underground infrastructure, the performance of the mathematical fitting model is experimentally evaluated on real-world GPR data in order to establish the main limitations of the method and to motivate the DL-based solution. Afterwards, experiments with multiple DL backbone architectures are conducted on real-world and combined real and synthetic datasets to identify the optimal baseline model and training configuration. Building on the selected baseline, the proposed CASS-Fusion hybrid architecture is evaluated, which integrates spatial B-scan features with localized spectral representations derived from TFDs to improve the estimation of all four descriptive parameters. In order to ensure statistically reliable evaluation of the DL models, 5-fold cross-validation is employed throughout the experiments. The performance of the detection stage is quantified using precision, recall, and mAP, whereas the parameter estimation models are evaluated using R^2, MAE, and RMSE. Finally, the complete end-to-end framework is validated on field-acquired GPR data to establish a proof-of-concept under real-world operating conditions.

6.1 Performance of Cylindrical Object Detection and Localization

The subsection presents the first stage of the proposed system which includes the detection and localization of hyperbola-shaped reflections within the B-scans produced by buried cylindrical objects, i.e., pipelines. Accordingly, the YOLOv11x and RT-DETR-X models are trained for this task. In both model families, the 'x' variants are employed as they represent the largest and best performing model architecture. Although such variants have a high computational cost, the inference speed remained within acceptable limits for GPR analysis. Model training is performed using the AdamW optimization algorithm in both cases. As for the YOLOv11x, an initial learning rate of 0.002, a final learning rate factor of 0.01, momentum of 0.9, and weight decay of 0.0005 are utilized. Similarly, for RT-DETR-X, the initial learning rate, final learning rate factor, momentum, and weight decay are set to 0.0001, 0.01, 0.937, and 0.0005, respectively. For training purposes of both models, the maximum number of epochs is set to 300 with early stopping enabled in order to prevent overfitting. A batch size of 4 is used for YOLOv11x, along with an input image resolution of 640 × 640. In the case of RT-DETR-X, a batch size of 8 is used while the input image resolution is set to 720 × 720. Data augmentation is performed exclusively for the object detection and localization task in order to enhance models' generalization. The augmentation strategy includes horizontal flipping with a probability of 0.5, scaling with a factor of 0.5, translation with a factor of 0.1, mosaic, and random erasing with a probability of 0.4. However, the mosaic augmentation is disabled in the last 10 epochs of the training process to stabilize convergence. To investigate the impact of input data composition, experiments are conducted in two scenarios. In all cases, 5-fold cross-validation is utilized to ensure a consistent evaluation benchmark. In the first scenario, YOLOv11x and RT-DETR-X models are trained exclusively on real-world data. Although the synthetic data is generated primarily for improving the estimation of underground infrastructure descriptive parameters, in the second scenario, synthetic data is combined with real-world data in the training set to investigate the impact on detection and localization performance. It is important to emphasize that validation sets in the 5-fold cross-validation remain unchanged when including synthetic data; the synthetic data is added only to the training sets. The performances of YOLOv11x and RT-DETR-X models achieved on real data, as well as on the combination of real and synthetic data, are shown in Figure 6.1 and Figure 6.2, respectively, with comprehensive numerical details given in Tables A.1 and A.2.
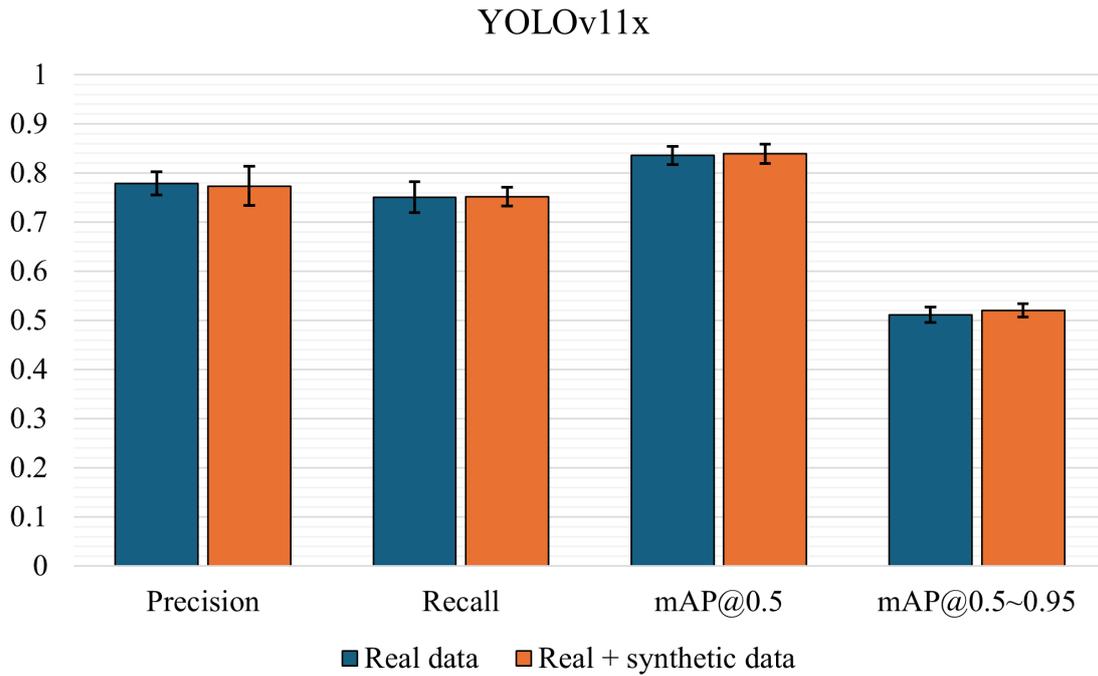
YOLOv11x

Figure 6.1 Performance comparison of YOLOv11x model trained on real data only versus real and synthetic data for detection and localization of hyperbola-shaped reflections within B-scan data.
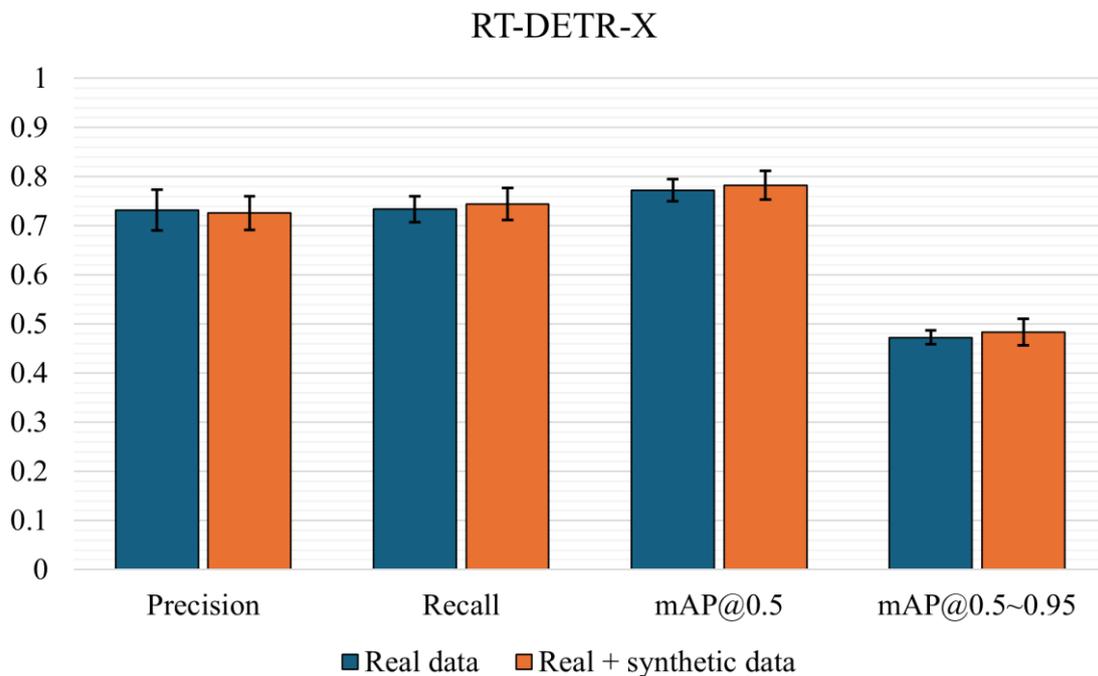
RT-DETR-X

Figure 6.2 Performance comparison of RT-DETR-X model trained on real data only versus real and synthetic data for detection and localization of hyperbola-shaped reflections within B-scan data.

According to the experimental results, the YOLOv11x trained only on real data achieved a precision of 0.7789 ± 0.0238, recall of 0.7502 ± 0.0317, mAP@0.5 of 0.8354 ± 0.0182, and mAP@0.5~0.95 of 0.5112 ± 0.0154. Additionally, the impact of synthetic data included in training set within each fold is experimentally analyzed, which resulted in precision, recall, mAP@0.5, and mAP@0.5~0.95 of 0.7732 ± 0.0400, 0.7517 ± 0.0190, 0.8388 ± 0.0193, and 0.5202 ± 0.0139, respectively. On the other hand, RT-DETR-X trained on real data only resulted in a precision of 0.7319 ± 0.0418, recall of 0.7335 ± 0.0262, mAP@0.5 of 0.7721 ± 0.0221, and mAP@0.5~0.95 of 0.4725 ± 0.0142. The inclusion of synthetic data along with real data in the training set yielded a precision, recall, mAP@0.5, and mAP@0.5~0.95 of 0.7256 ± 0.0340, 0.7443 ± 0.0328, 0.7822 ± 0.0295, and 0.4837 ± 0.0271, respectively.

Analyzing the results, it can be seen that the YOLOv11x consistently outperformed the RT-DETR-X model across all evaluated metrics. The result remained consistent regardless of whether the models are trained on real-world data only or on a combination of real and synthetic data. For the YOLOv11x, the use of the combined dataset resulted in mAP@0.5, and mAP@0.5~0.95 improvements of approximately 0.4% and 1.7%, respectively. The inclusion of synthetic data in the case of RT-DETR-X resulted in a mAP@0.5 improvement of approximately 1.29% and a mAP@0.5~0.95 improvement of approximately 2.3%. Although the inclusion of synthetic data led to an improvement in all performance measures except precision, where a slight decrease is observed for both YOLOv11x and RT-DETR-X, the overall impact on generalization is positive. The observed decrease in precision indicates that while synthetic data improves target detection (higher recall), it may also introduce a marginal increase in false positives. However, prioritizing recall may be crucial in underground infrastructure detection and localization tasks; high recall ensures the detection of all potential pipelines. The trade-off is justified since such an approach can prevent accidental damage during excavation.

In summary, YOLOv11x provided more stable and reliable detection and localization performance in the context of this research. Although competitive, RT-DETR-X demonstrated greater variability in performance measures. The reason behind such behavior can be transformer-based architecture which RT-DETR-X employs, that may require larger datasets or more complex fine-tuning to achieve comparable consistency and performance.

6.2 Quantitative Analysis of Estimated Infrastructure Parameters

This subsection presents the quantitative experimental results of the second stage of the proposed framework, focusing on the estimation of underground infrastructure descriptive parameters from GPR data. To enable a systematic comparison across fundamentally different methodological approaches, the evaluation progresses from the hyperbola fitting approach as a conventional baseline, through the DL regression baselines, to the proposed CASS-Fusion hybrid architecture.

To enhance numerical stability and improve the convergence of the employed methods, certain target variables are normalized. Specifically, TWTT parameter values are linearly scaled to the range between 0.0 and 1.0 for model training purposes. In order to ensure MAE and RMSE preserve physical meaning, the TWTT values are denormalized to their original range $(0 - 60$ ns$)$ prior to calculating the performance measures. For the intersection angle (α), a sine transformation – $\sin(\alpha)$, is applied instead of using raw angular values in degrees or radians, yielding values in the range from 0.0 to 1.0. Burial depth and pipeline diameter parameters remain unnormalized since their natural ranges are already sufficiently constrained.

6.2.1 Performance of the Hyperbola Fitting Approach

After the successful detection and localization of cylindrical objects i.e., pipelines from real-world GPR data in the first stage, the next step is to employ a hyperbola fitting approach to estimate the descriptive parameters of underground infrastructure. This represents a fundamental approach for translating the visual geometry of the hyperbola-shaped reflection into quantitative descriptive parameters of underground infrastructure. Hyperbola fitting is performed using the methodology described in Chapter 2, subsection 2.1, on real-world GPR data acquired throughout the city of Rijeka, Croatia, as described in subsection 2.2. Fitting of hyperbola-shaped features is performed on 732 samples which contain full and precise underground infrastructure parameter data. For each sample, a hyperbola-shaped curve is generated and fitted to the reflection localized by the detection algorithm from the first stage.

In order to ensure an optimal fit, BO is utilized to maximize a fitness score based on pixel intensity alignment. More precisely, pixel intensities along the generated hyperbola-

shaped curve are summed, where the higher cumulative values indicate a stronger alignment with the reflection of interest. Additionally, a GP surrogate model is used to guide the search of the optimization process. For each B-scan sample, the algorithm is initialized with 400 random exploration steps followed by 32 iterative optimization steps. A large number of exploration steps is utilized to minimize the probability of convergence to local minima. The search space is defined within the physical bounds expected for the surveyed environment:

- TWTT (ns) – Bounds are concentrated within the lower half of the bounding box produced by the detection algorithm from the first stage in the vertical direction.

- Burial depth (m) – Bounds are dynamically calculated based on previously defined TWTT and expected relative permittivity values for the surveying location (range 5 – 18).

- Pipeline diameter (m) – Limited to the range between 0.025 and 0.6.

- Intersection angle – α (˚) – Bounded between 19˚ and 90˚.

After the features are successfully fitted, the performance is quantified using the R^2, MAE, and RMSE. Since such an approach is not a learning-based method, the performance is evaluated across the full dataset, and the results are presented in Table 6.1.

Table 6.1 Performance of the BO hyperbola fitting method on real-world GPR data for estimating the descriptive parameters of underground infrastructure.

n = 732	R^2	MAE	RMSE
TWTT	0.9917	0.4582	0.5879
Burial depth	0.1171	0.2192	0.2821
Pipeline diameter	-0.8274	0.0872	0.1103
Angle – α	-0.1169	0.1088	0.1432

Analyzing the obtained results, a significant disparity in estimation performance across different parameters can be observed. High values of R^2 (0.9917) and low values of MAE (0.4582) and RMSE (0.5879) in the case of TWTT estimation indicate that by using such a mathematical model, the apex of the hyperbola-shaped reflection can be identified sufficiently. One of the reasons that helped the method achieve high values of performance measures is that

the bounds for TWTT are directly derived from the resulting bounding box coordinates from the first stage i.e. object detection and localization stage. This way, the optimization process has a precisely defined TWTT window in which the optimization is performed. Furthermore, the apex of the hyperbola-shaped reflection is usually the most visually distinct feature in a B-scan, characterized by the highest signal amplitude and strongest contrast compared to other parts of the hyperbola-shaped feature. Therefore, the fitness function, which maximizes pixel intensity, is able to converge to the global maximum within the defined search space.

The estimation of physical descriptive parameters proved to be more challenging. Estimates of burial depth yielded a weak correlation with the ground truth, with the values of R^2, MAE, and RMSE of 0.1171, 0.2192, and 0.2821, respectively. Such values indicate that the approach could not distinguish the surrounding medium velocity sufficiently. The estimation of the pipeline diameter and intersection angle resulted in negative R^2 values of -0.8274, and -0.1169, respectively. In other words, estimation of these parameters is less accurate than a simple mean-baseline prediction. The reason for such behavior arises from the velocity-depth ambiguity. As visually demonstrated in Chapter 2, a hyperbola-shaped reflection generated by a shallow object in a slow medium can physically appear nearly identical to one generated by a deeper object in a fast medium. This complexity is further increased when accounting for pipeline diameter and intersection angle, since such parameters primarily affect the "tails" of the hyperbola-shaped reflection. An additional complicating factor is that the reflections in real-world GPR data often fade due to signal attenuation or are obscured by noise. To visually represent the complexity of the task, four B-scans are shown in Figure 6.3 with the corresponding ground truth hyperbola-shaped curve (green), and an estimated curve (red). It can be seen that both hyperbola-shaped curves overlap almost perfectly, however, the values of parameters used to generate the estimated one differ from the ground truth by a significant margin, as shown in Table 6.2.
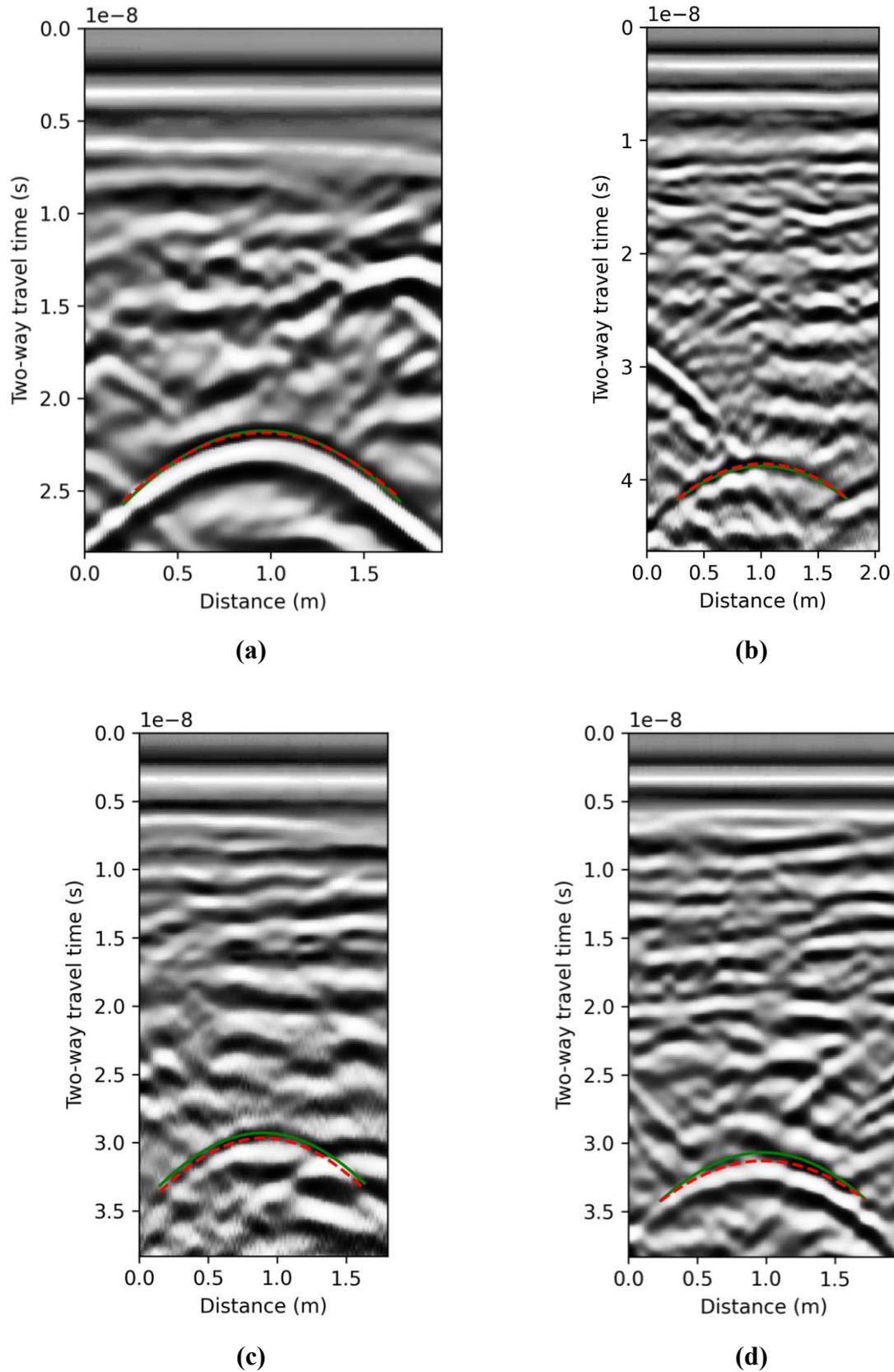
Figure 6.3 Qualitative performance of the mathematical model and the BO fitting algorithm. The ground truth hyperbola-shaped curves are shown in solid green, while the estimated curves are shown in dashed red.

Table 6.2 Quantitative comparison between ground truth and estimated parameters for the visual examples presented in Figure 6.3.

	TWTT (ns)			Burial depth (m)			Pipeline diameter (m)			Angle − α (°)		
	True	Est.	|Δ|	True	Est.	|Δ|	True	Est.	|Δ|	True	Est.	|Δ|
(a)	21.68	21.80	0.12	1.00	1.10	0.1	0.20	0.29	0.09	62	70	8
(b)	38.76	38.54	0.22	1.75	1.38	0.37	0.54	0.14	0.4	89	51	38
(c)	29.24	29.58	0.34	1.35	1.06	0.29	0.27	0.14	0.13	89	51	38
(d)	30.64	31.23	0.59	1.45	1.25	0.2	0.20	0.31	0.11	80	52	28

Since the BO adjusts all variables simultaneously, a good fit can also be produced by the incorrect combination of parameters. Therefore, multiple parameter configurations may produce a high fitness score, even though only one corresponds to the true physical state of the underground infrastructure. In such cases, the properties of the surrounding medium should also be taken into account when estimating these parameters. An overestimation of EM propagation velocity, which is a property of the surrounding medium, requires an overestimation of burial depth to maintain the same TWTT. Accordingly, the samples presented in Figure 6.3 demonstrate that the BO hyperbola fitting method finds a solution that satisfies the visual curve but misrepresents the physical reality of the subsurface. By using traditional, non-learning-based methods, it becomes very difficult to explicitly model all possible scenarios, which gives DL-based methods a clear advantage in such cases.

In summary, the hyperbola fitting approach lacked sufficient visual information to reliably distinguish between changes in burial depth, pipeline diameter, and intersection angle. The results suggest that, while the approach is a robust tool for localizing the target in time (when bounding box coordinates of the detected and localized hyperbola-shaped reflections are provided beforehand), it is insufficient for reliable characterization of physical properties when used as a standalone approach without additional constraints or prior knowledge of the medium's velocity.

6.2.2 Performance of Deep Learning Regression Baselines

The experimental results from the second stage of the proposed system are presented in this subsection. The main focus of this stage is the estimation of descriptive parameters of underground infrastructure situated within heterogeneous and complex soil environments. By utilizing the detected and localized hyperbola-shaped reflections from the first stage as input, the framework employs DL model architectures as feature extraction backbones to estimate aforementioned parameters. More precisely, DenseNet201, Xception, NASNetLarge, and EfficientNetV2L are implemented for the task, each pretrained on ImageNet and adapted with a custom designed regression head, as illustrated in Figure 3.4. To ensure robust regression performance, hyperparameter search is performed to identify the optimal training configuration. Accordingly, the Adam optimization algorithm is selected, with a training procedure consisting of 500 epochs, a batch size of 4, and an initial learning rate of 10^{-4}. A critical component of the training strategy is a custom weighted MSE loss function, which is implemented to prioritize the learning of more complex parameters. This way, the estimations from the output layer are scaled by predefined importance weights: 1.0 for TWTT and burial depth, 1.5 for pipeline diameter, and 3.0 for the intersection angle. The weighting approach guides the model to allocate greater representational capacity toward the more challenging estimations. The values of importance weights are determined empirically to achieve a balance between the model's overall performance and the estimation accuracy of each individual descriptive parameter. Furthermore, the training procedure also integrates early stopping mechanisms and an adaptive learning rate scheduler to balance training efficiency. Similar to the object detection and localization stage, experiments are conducted in two scenarios in order to investigate the impact of input data composition on the estimation of underground infrastructure parameters. In all cases, 5-fold cross-validation is utilized to ensure a consistent and robust evaluation benchmark. In the first scenario, the DL regression models are trained on real-world GPR data only. In the second scenario, synthetic data, which is specifically generated to capture diverse underground infrastructure parameters, is combined with the real-world data. More precisely, the synthetic data is added only to the training sets, which ensures that the validation sets within the 5-fold cross-validation remain unchanged. The quantitative results for TWTT estimation across different DL architectures on real-world and combined (real + synthetic) B-scan data, expressed as R^2, MAE, and RMSE, are illustrated in Figures 6.4, 6.5, and B.1, respectively. The comprehensive numerical details are given in Tables B.1 and B.2.
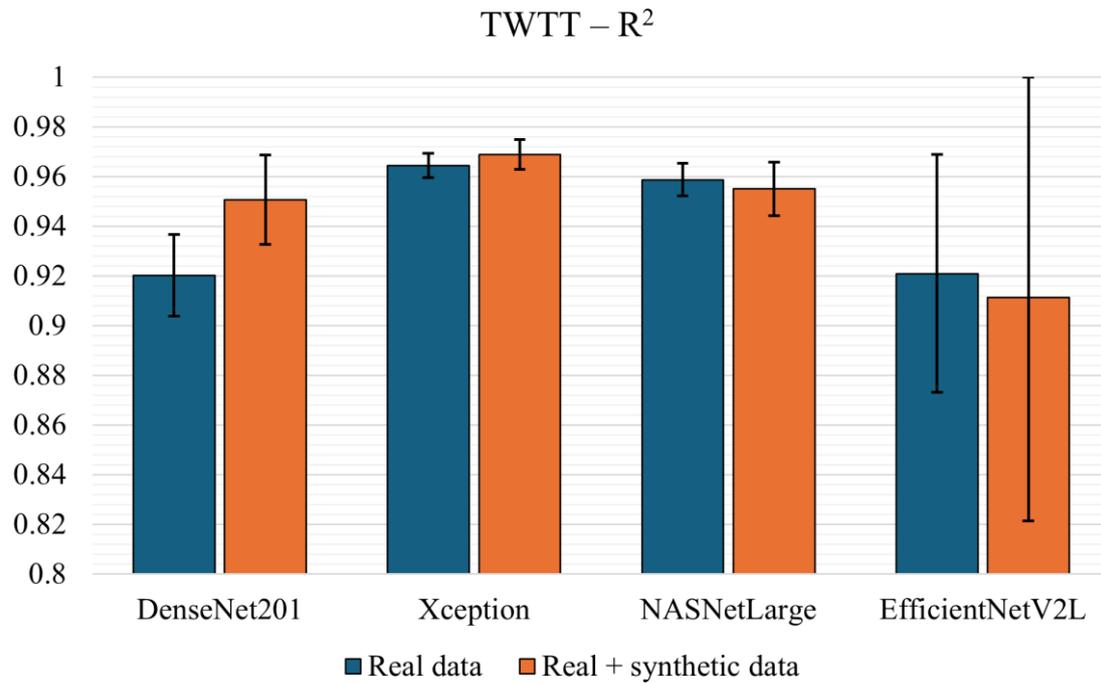
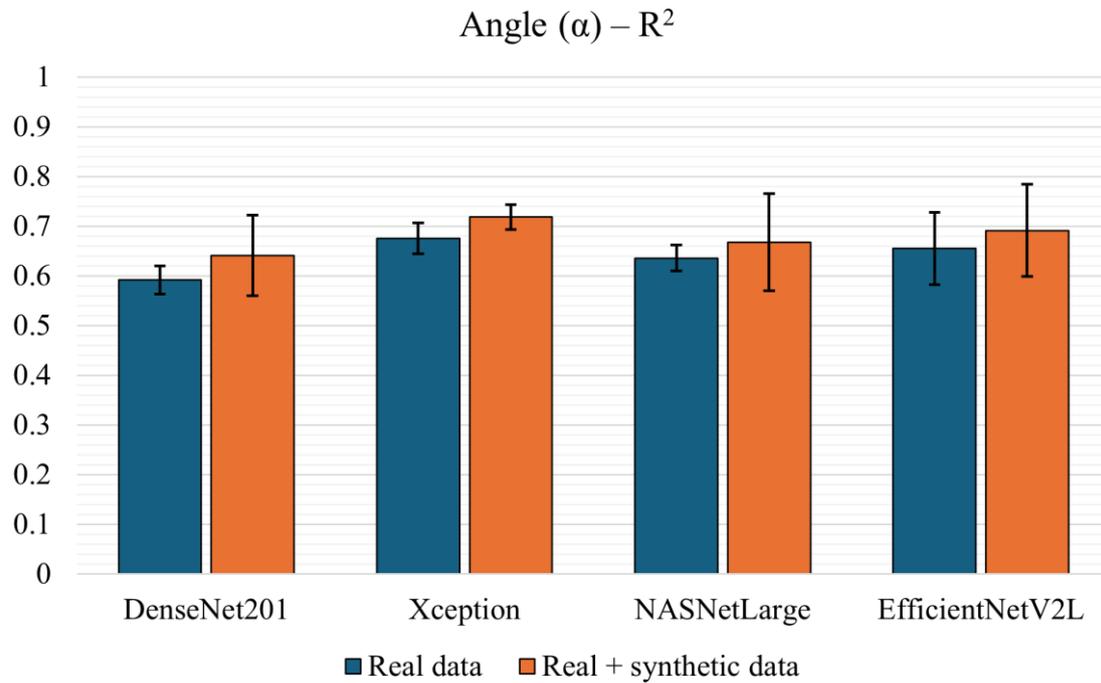**Figure 6.4** Performance comparison of different DL model architectures for TWTT estimation (expressed as R^2) across real-world and combined (real + synthetic) B-scan data.
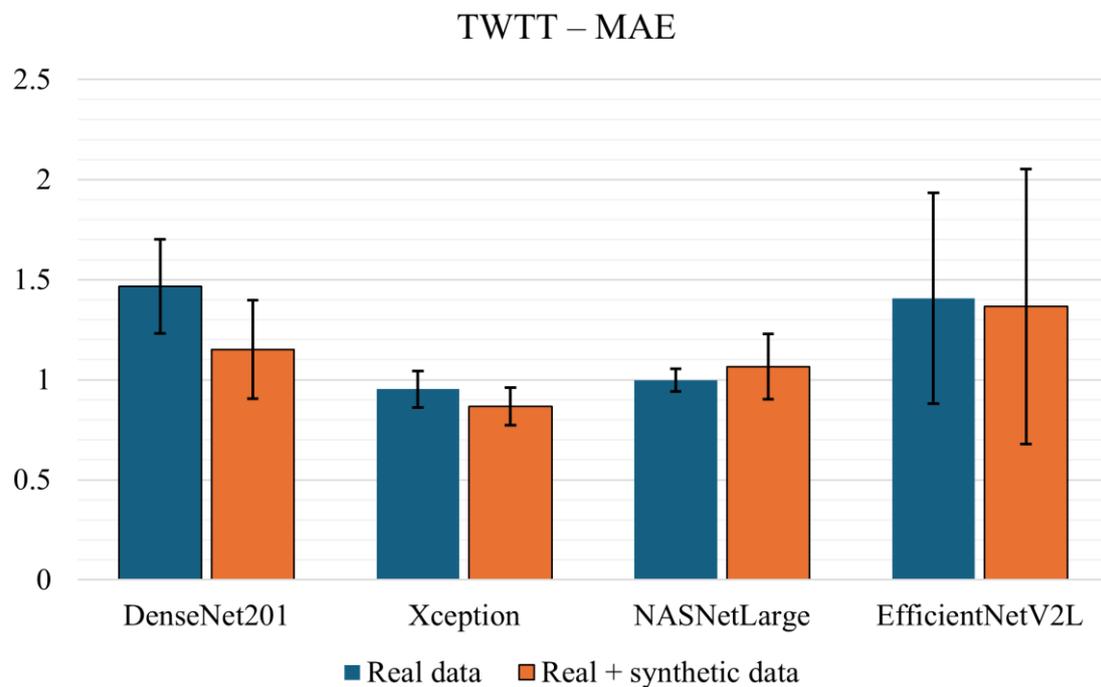
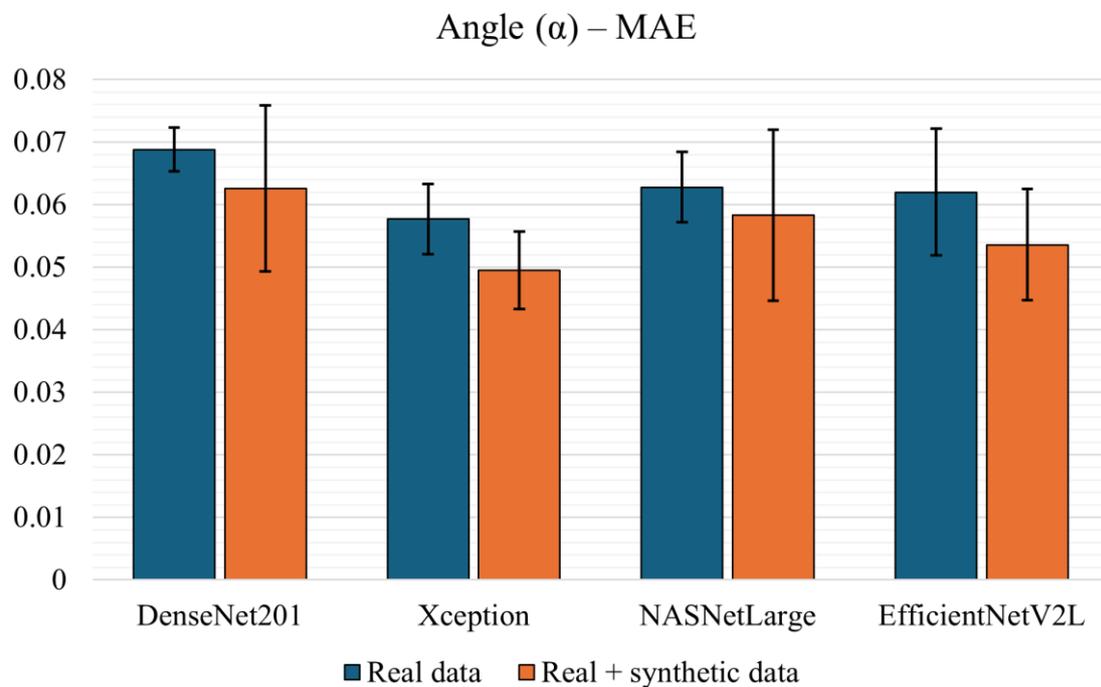**Figure 6.5** Performance comparison of different DL model architectures for TWTT estimation (expressed as MAE) across real-world and combined (real + synthetic) B-scan data.

Regardless of the data composition, the Xception consistently outperformed DenseNet201, NASNetLarge, and EfficientNetV2L in estimating the TWTT. Utilizing only real-world data for model training, Xception achieved an R^2 score of 0.9644 ± 0.0048, which is significantly higher than DenseNet201 and EfficientNetV2L R^2 scores of 0.9202 ± 0.0164 and 0.9209 ± 0.0479, respectively. In addition, the Xception also achieved the lowest MAE and RMSE with the values of 0.9527 ± 0.0911 and 1.2114 ± 0.1002, respectively. Furthermore, the introduction of synthetic data into the training set resulted in non-uniform effects across different DL models. The largest improvement is observed in DenseNet201, where such inclusion of synthetically generated data resulted in an R^2 improvement of ~3% (from 0.9202 to 0.9505) and a significant reduction in MAE and RMSE. Xception, despite already achieving high performance, reached a peak R^2 of 0.9689 ± 0.0060 along with the lowest MAE, and RMSE when trained on combined data. However, not all DL model architectures responded positively on the synthetic data inclusion. The NASNetLarge model, while showing competitive performance on real-world data (R^2 of 0.9587 ± 0.0065), training on combined data resulted in R^2 marginal decrease of ~0.4%. A similar trend is observed for EfficientNetV2L, where the standard deviation increased significantly when the model is trained on combined data, directly implying less stability and lower generalization.

The quantitative results for burial depth estimation across different DL architectures on real-world and combined (real + synthetic) B-scan data, expressed as R^2, MAE, and RMSE, are presented in Figures 6.6, 6.7, and B.2, respectively. The comprehensive numerical details are given in Tables B.3 and B.4. Evaluating burial depth estimation revealed similar trends to the TWTT analysis. In the case where the training is performed on real-world data only, the Xception architecture resulted in superior regression performances compared to the other models with R^2 of 0.9544 ± 0.0089, MAE of 0.0472 ± 0.0038, and RMSE of 0.0636 ± 0.0066. This consistent performance is expected, given that a target's physical burial depth is fundamentally dictated by the radar signal's TWTT. The achieved results establish a robust baseline for burial depth estimation prior to any synthetic data inclusion. In comparison, the initial R^2 values for DenseNet201, NASNetLarge, and EfficientNetV2L clustered around 0.94, with corresponding MAE values consistently remaining above the 0.05 threshold. Furthermore, the introduction of synthetic data into training process enabled further performance improvement for DenseNet201 and Xception. More precisely, DenseNet201 benefited the most from the combined data, achieving an R^2 improvement of ~0.9%. However, for NASNetLarge and EfficientNetV2L, synthetic data inclusion lowered overall performance.
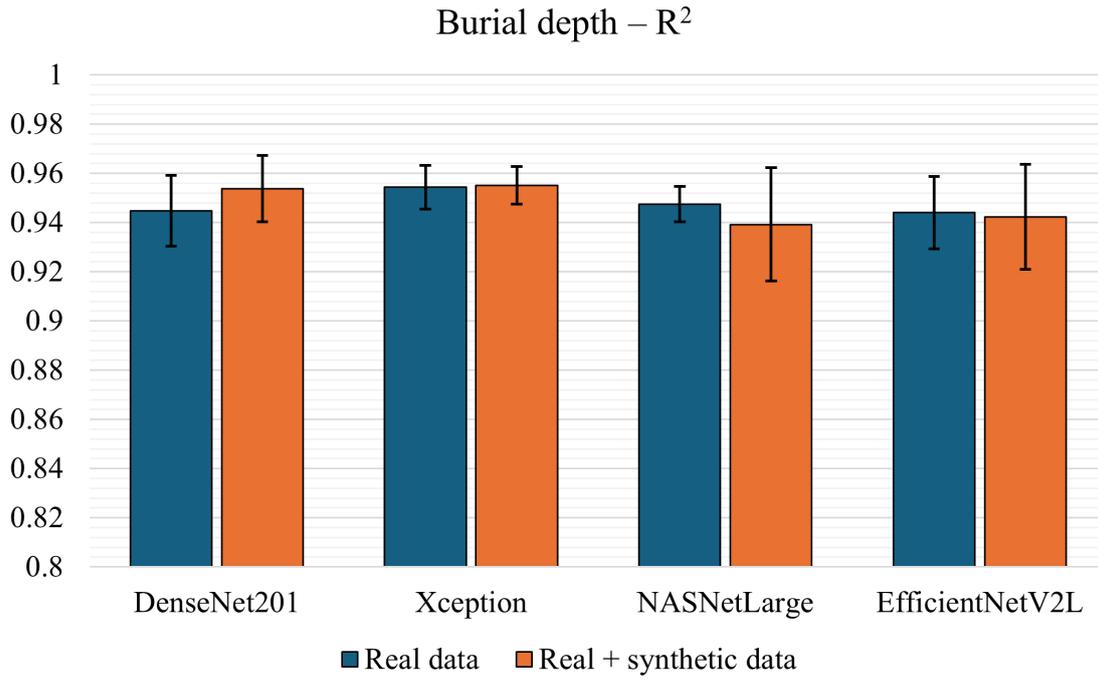
Figure 6.6 Performance comparison of different DL model architectures for burial depth estimation (expressed as R2) across real-world and combined (real + synthetic) B-scan data.
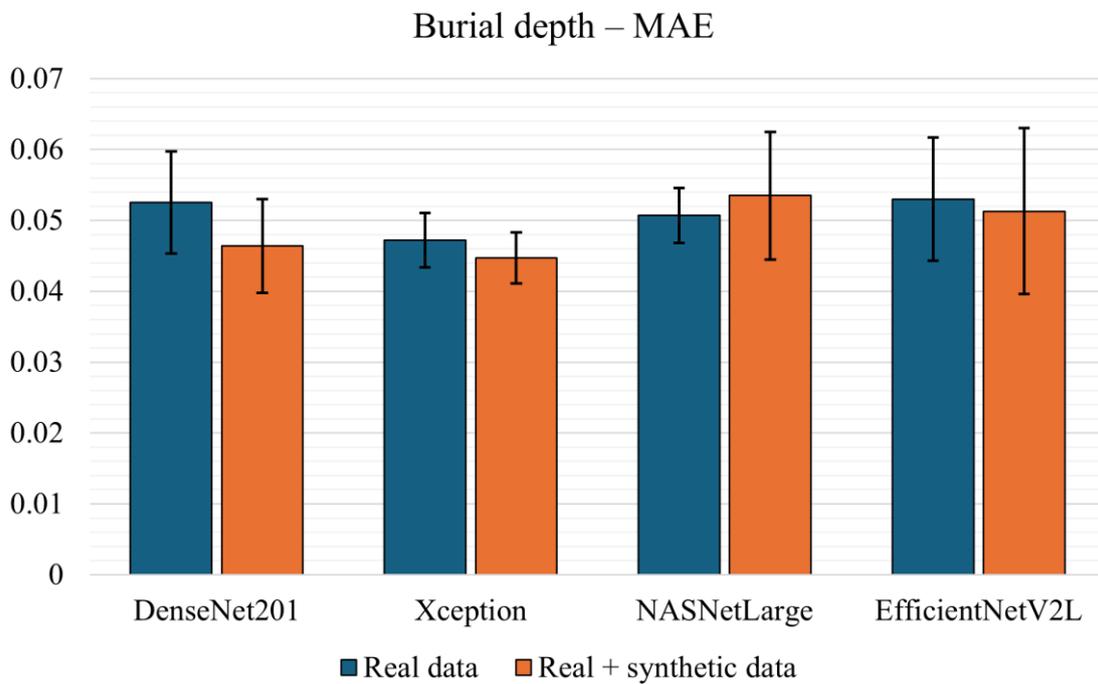
Figure 6.7 Performance comparison of different DL model architectures for burial depth estimation (expressed as MAE) across real-world and combined (real + synthetic) B-scan data.

The quantitative results for pipeline diameter estimation across different DL architectures on real-world and combined (real + synthetic) B-scan data, expressed as R^2, MAE, and RMSE, are presented in Figures 6.8, 6.9, and B.3, respectively. The comprehensive numerical details are given in Tables B.5 and B.6. While the general behavioral trends from TWTT and burial depth analyses persisted, estimating pipeline diameter proved more difficult, resulting in consistently lower R^2 values across all models. When trained exclusively on real-world data, NASNetLarge and Xception achieved the highest regression performance. NASNetLarge reached the highest R^2 score of 0.5802 ± 0.0875, marginally better than Xception (0.5789 ± 0.0716) and EfficientNetV2L (0.5753 ± 0.0527), while DenseNet201 underperformed at 0.5319 ± 0.0403. Additionally, NASNetLarge produced the lowest MAE (0.0358 ± 0.0052), whereas Xception achieved the best overall RMSE (0.0524 ± 0.0028). Introducing synthetic data again produced distinctly non-uniform effects. For DenseNet201 and Xception, the combined approach proved highly beneficial. DenseNet201 showed an R^2 improvement of ~3.8% (from 0.5319 to 0.5702) along with reductions in both MAE and RMSE. Xception maintained its consistent capability, achieving an overall best R^2 of 0.5910 ± 0.0430 on the combined dataset. However, incorporating synthetic data lowered NASNetLarge and EfficientNetV2L performances, indicating a severe loss of stability and generalization.
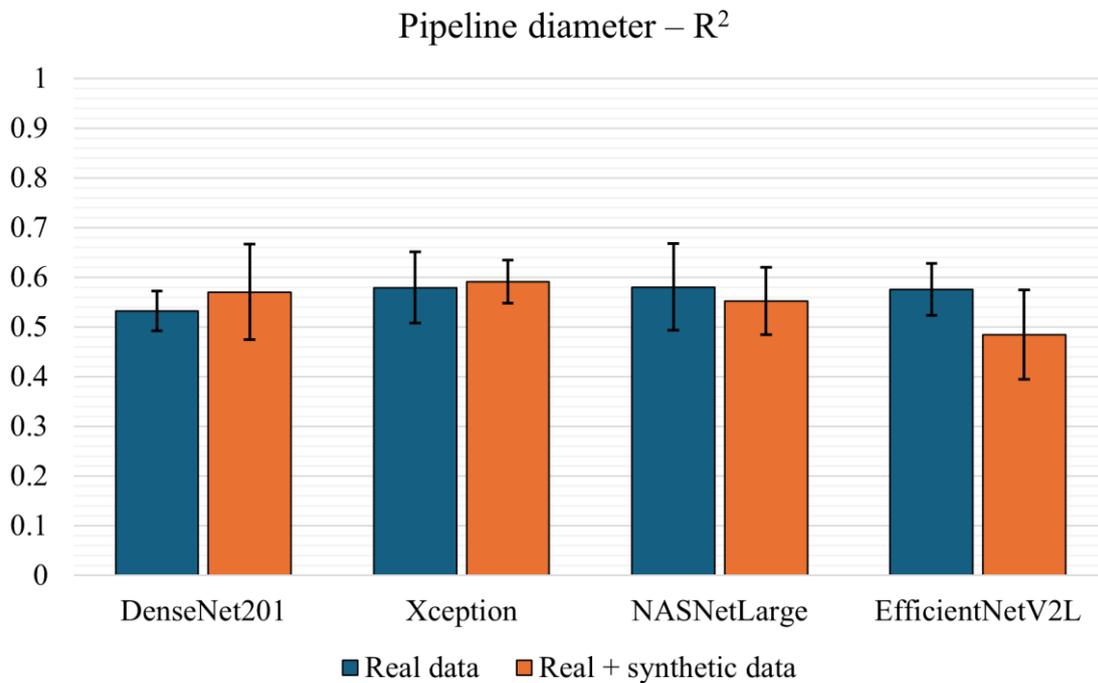
Figure 6.8 Performance comparison of different DL model architectures for pipeline diameter estimation (expressed as R2) across real-world and combined (real + synthetic) B-scan data.
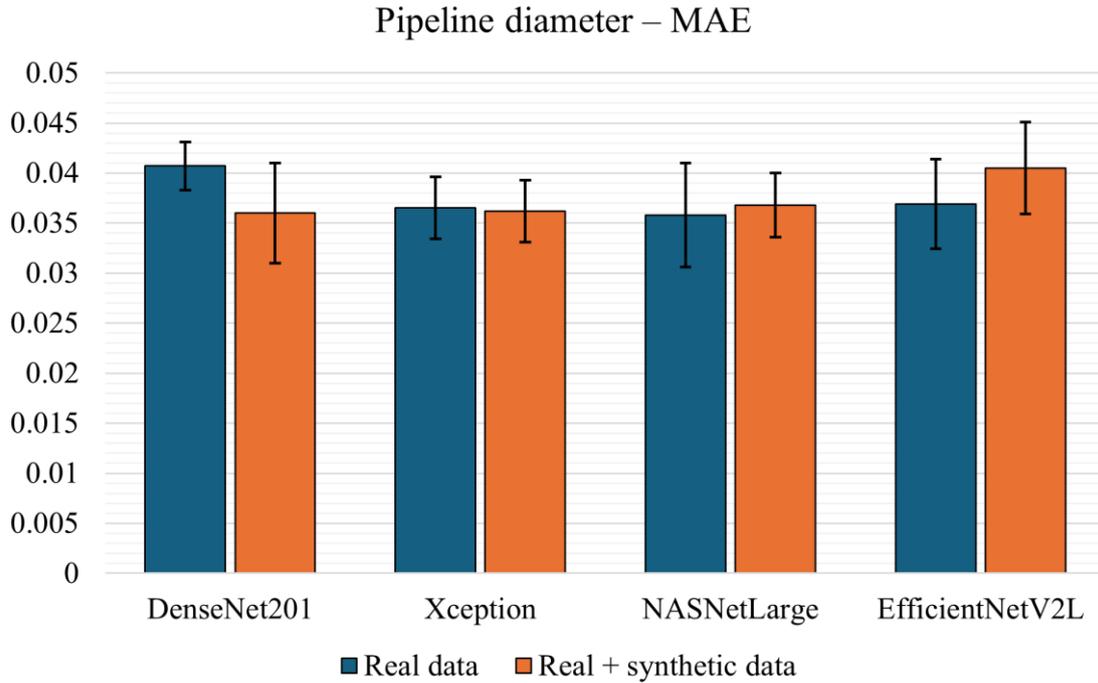
Figure 6.9 Performance comparison of different DL model architectures for pipeline diameter estimation (expressed as MAE) across real-world and combined (real + synthetic) B-scan data.

The quantitative results for intersection angle estimation across different DL architectures on real-world and combined (real + synthetic) B-scan data, expressed as R^2, MAE, and RMSE, are presented in Figures 6.10, 6.11, and B.4, respectively. The comprehensive numerical details are given in Tables B.7 and B.8. Assessing the intersection angle introduced predictive challenges similar to the pipeline diameter estimation task, which are reflected by moderate performances. Relying on real-world data only during the training phase, Xception showed the most robust feature extraction capabilities, with the highest initial R^2 of 0.6755 ± 0.0310, along with the lowest MAE (0.0577 ± 0.0056) and RMSE (0.0769 ± 0.0053). EfficientNetV2L achieved the second-highest accuracy with an R^2 of 0.6550 ± 0.0727, whereas DenseNet201 could not exceed an R^2 of 0.5914 ± 0.0279 on the non-augmented dataset. The introduction of synthetic data yielded architecture-specific responses that differed from those observed in TWTT, burial depth, and pipeline diameter estimation. Across all evaluated model architectures, a consistent improvement in mean R^2 is observed. Xception successfully leveraged combined input data, achieving an overall peak R^2 of 0.7182 ± 0.0250 (~4.3% improvement) and minimizing MAE and RMSE to 0.0495 ± 0.0062 and 0.0717 ± 0.0051, respectively. DenseNet201 also achieved a substantial gain, improving R^2 score by ~5% to reach 0.6409 ± 0.0808.

Figure 6.10 Performance comparison of different DL model architectures for intersection angle estimation (expressed as R2) across real-world and combined (real + synthetic) B-scan data.

Figure 6.11 Performance comparison of different DL model architectures for intersection angle estimation (expressed as MAE) across real-world and combined (real + synthetic) B-scan data.

While the average accuracy increased, this trend masked an underlying statistical instability. Although NASNetLarge and EfficientNetV2L benefited from synthetic data inclusion in the training process, as seen from improved performance measures, their standard deviations also increased significantly. Such an amplified variance directly implies that the synthetic data, while raising the performance threshold, fundamentally weakened the stability and generalization of those specific model architectures.

Analyzing the four descriptive underground infrastructure parameters revealed a clear hierarchy of estimation difficulty. Estimation of TWTT and burial depth yielded highly accurate and stable results across all baseline models, consistently surpassing R^2 scores of 0.92 and 0.94, respectively. Such consistency in performance is expected, given that TWTT is a directly observable parameter in B-scan data, characterized by the mostly clear apex of the hyperbola-shaped reflection. Nevertheless, the precise apex localization remains critical, regardless of whether the hyperbola-shaped reflection is produced by metallic, plastic, or concrete pipelines. Hence, the utilization of DL models leads to a more robust and efficient estimation of TWTT parameter. Since burial depth is closely related to TWTT, as seen from Pearson correlation matrix in Figure 2.15, and (2.1) and (2.2), also yielded robust performance. However, direct determination of burial depth requires prior estimation or measurement of the soil's EM wave velocity. Precise interpretation of hyperbola-shaped reflections is thus critical, as the reflection geometry inherently encodes the velocity information needed for burial depth estimation. According to the presented results, it is evident that the models effectively extracted and analyzed such information directly from the B-scan data. In contrast, estimating pipeline diameter and intersection angle presents a highly complex regression problem, as evidenced by the R^2 values ranging from 0.53 to 0.72. Estimating pipeline diameter is exceptionally challenging since the subtle hyperbolic width is often obscured by soil heterogeneity, noise, and partial reflections. The pipeline diameter requires contextual interpretation of the hyperbola-shaped curvature rather than a direct visual feature, explaining the reduced performance measures compared to TWTT and burial depth. Similarly, the intersection angle between the GPR survey line and pipeline is indirectly encoded in the asymmetry and elongation of hyperbola-shaped reflections, which are often subtle and obscured by noise or incomplete reflection profiles; this inherently challenging extraction process explains both the lower R^2 values and the higher fold-to-fold variability. To summarize, parameters directly encoded in hyperbola-shaped reflections (TWTT and burial depth) are reliably estimated, while indirectly inferred parameters (pipeline diameter and angle) require more sophisticated spatial reasoning.

Furthermore, soil heterogeneity scatters and distorts hyperbola-shaped reflections, which significantly complicates the extraction of parameters like pipeline diameter and angle as evidenced by performances of these parameters.

In every data configuration, Xception proved to be the most capable at feature extraction in underground infrastructure characterization tasks. Its underlying architecture appears well-suited for isolating the distinct spatial frequencies of GPR hyperbola-shaped reflections from complex background clutter. On the other hand, DenseNet201 showed a different behavioral trend. While initially underperforming on the real-world dataset, it benefited the most from the synthetically augmented feature space of the combined data. The aforementioned implies that the dense connectivity patterns within DenseNet architecture require a larger volume of varied data to properly adapt for regression tasks in GPR domain. Furthermore, the introduction of synthetic data into the training set resulted in non-uniform effects regarding generalization. It can be concluded that including synthetic data together with real-world data in training set effectively bridges the domain gap for DenseNet201, and Xception. However, larger and heavier architectures such as NASNetLarge and EfficientNetV2L require more complex domain adaptation to fully leverage the synthetic data without introducing noise. The achieved results validate Xception as the optimal backbone architecture when trained on combined dataset, around which the system for parameter estimation of underground infrastructure will be developed.

6.2.3 Performance of the Proposed Cross-Attention Spatio-Spectral Fusion Network

The experimental results validating the proposed hybrid DL architecture for underground infrastructure descriptive parameter estimation are presented in this subsection. Building on the previously selected optimal spatial backbone (Xception), the advanced stage aims to improve regression performance. This can be achieved through the inclusion of localized spectral features computed using various TFDs. Since hyperbola-shaped reflections found within B-scans encode both spatial geometry and distinct localized frequency characteristics, fusing these complementary domains can enhance the estimation performances of the four descriptive parameters: TWTT, burial depth, pipeline diameter, and intersection angle. In order to generate the corresponding TF representations, six different TFDs: WVD, PWVD, SPWVD, BJD, ZAMD, and CWD are applied to cropped B-scan segments. The

proposed dual-branch spatio-spectral architecture is specifically designed to capture complex relationships between the physical and frequency domains. Within the spatial feature extraction branch, an Xception model is utilized to process the cropped B-scans. At the same time, the spectral feature extraction branch processes the localized TFDs via a custom 3D CNN integrated with BiLSTM layers. These feature representations are afterwards fused via a multi-head cross-attention mechanism, which allows the model to dynamically map the most relevant frequency components to their corresponding spatial structures. The training procedure for hybrid architecture is performed through a progressive fine-tuning strategy to ensure optimal feature convergence. The spectral feature extraction branch is randomly initialized and trained utilizing the same hyperparameters established for the baseline models, while the weights of the ImageNet pretrained spatial branch (Xception) remain frozen. All training phases utilize an adaptive learning rate scheduler combined with early stopping mechanisms. Following the convergence of the spectral branch, its weights are thereafter frozen, and the spatial feature extraction branch is unfrozen to allow for domain-specific adaptation to the GPR data. In the final fine-tune phase, the entire hybrid model is unfrozen and trained end-to-end utilizing a reduced learning rate (10^{-5}). To guide the model optimization in the final training phase, penalty weights of 1.0 for TWTT and burial depth, 3.0 for pipeline diameter, and 5.0 for the intersection angle are applied within the custom weighted MSE loss function. This way, the model's representational capacity is directed toward the higher complexity geometrical estimations. In order to ensure the statistical reliability and robustness of the evaluation, all performance metrics are derived from a 5-fold cross-validation.

The quantitative results for TWTT estimation across the baseline spatial, standalone TFD, and proposed hybrid models, expressed as R^2, MAE, and RMSE, are presented in Figures 6.12, 6.13, and C.1, respectively. The comprehensive numerical details are given in Tables C.1 and C.2. The Xception baseline achieved an initial R^2 of 0.9689 ± 0.0060. However, incorporating the WVD into the proposed hybrid architecture yielded the highest overall performances for TWTT estimation, raising the R^2 to 0.9775 ± 0.0014 while simultaneously minimizing the MAE to 0.7470 ± 0.0176 and the RMSE to 0.9628 ± 0.0262. It is important to note that relying entirely on the spectral feature extraction branch (without spatial features) resulted in slightly lower performance than the baseline (e.g., an R^2 of 0.9671 ± 0.0034 for WVD standalone model). The results confirm that neither domain alone provides a complete representation. Fusing the spatial and spectral features creates a highly synergistic model.
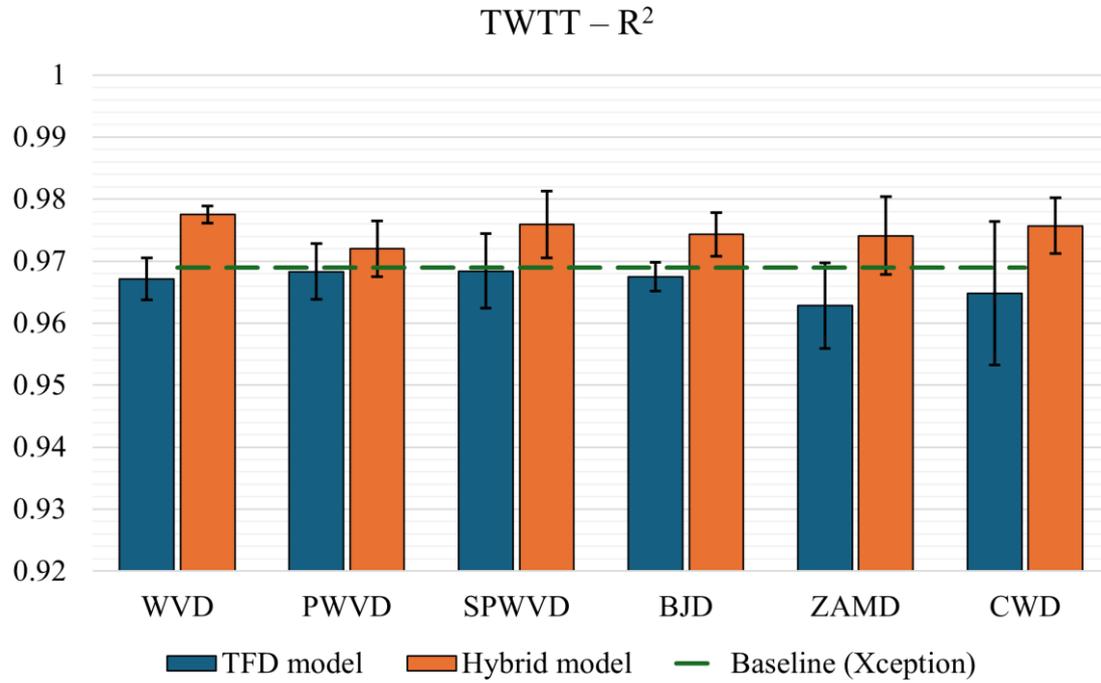
Figure 6.12 Performance comparison of TFD, hybrid, and baseline models for TWTT estimation (expressed as R^2) across different TFDs.
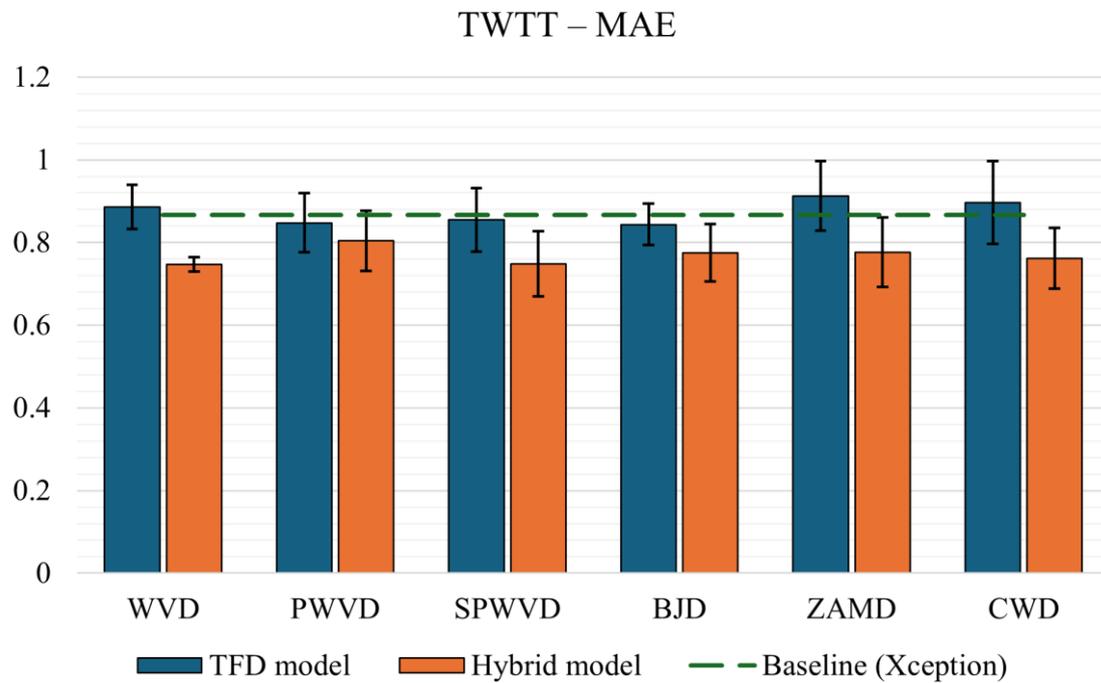
Figure 6.13 Performance comparison of TFD, hybrid, and baseline models for TWTT estimation (expressed as MAE) across different TFDs.

The hybrid model utilizing the SPWVD achieved a competitive R^2 of 0.9759 ± 0.0054, while the CWD configuration closely followed with an R^2 of 0.9757 ± 0.0045. Despite high performances of hybrid model approach, the standalone TFD models for SPWVD and CWD achieved lower accuracies (R^2 of 0.9684 ± 0.0060 and 0.9648 ± 0.0116, respectively). According to the results it can be concluded that localized spectral features lack the spatial context necessary for independent precision. Therefore, the proposed cross-attention fusion mechanism bridges the domain gap in this case.

The quantitative results for burial depth estimation across the baseline spatial, standalone TFD, and proposed hybrid models, expressed as R^2, MAE, and RMSE, are presented in Figures 6.14, 6.15, and C.2, respectively. The comprehensive numerical details are given in Tables C.3 and C.4. Evaluating burial depth estimation revealed behavioral trends consistent with the TWTT analysis. The direct physical correlation between the two parameters explains the outcome. The spatial baseline model (Xception) established a robust starting point with an R^2 of 0.9551 ± 0.0077. From the results, it can be seen that the introduction of TFD features further stabilized and improved the estimations, enabling the spatio-spectral hybrid model to outperform both standalone domains across all evaluated configurations.
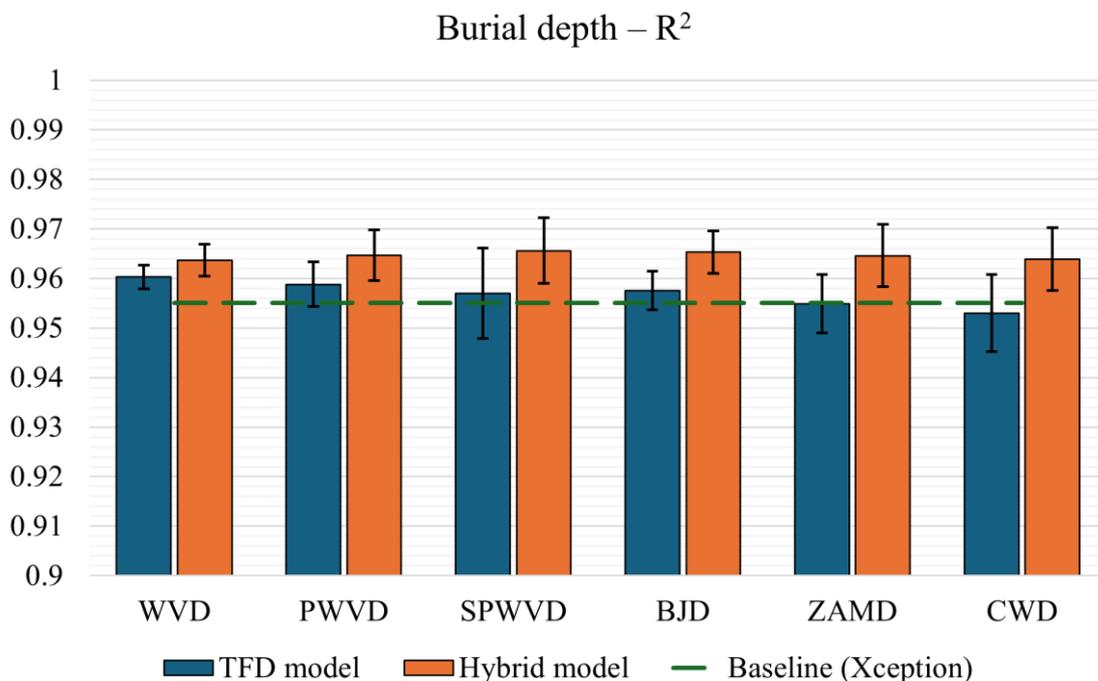
Figure 6.14 Performance comparison of TFD, hybrid, and baseline models for burial depth estimation (expressed as R^2) across different TFDs.
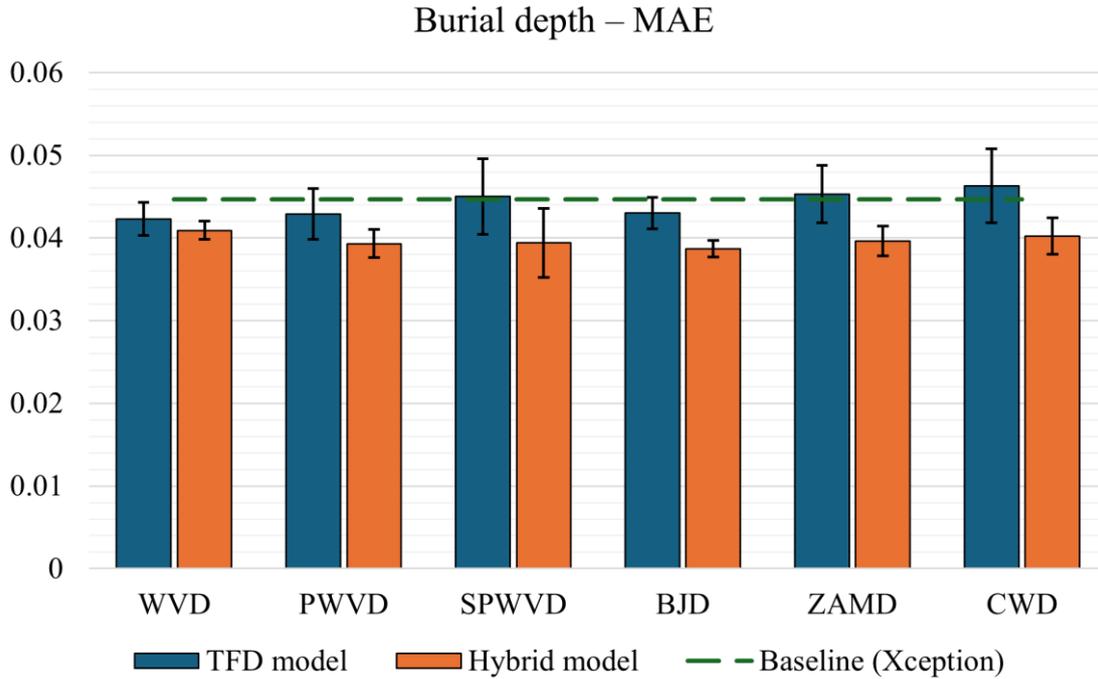
Figure 6.15 Performance comparison of TFD, hybrid, and baseline models for burial depth estimation (expressed as MAE) across different TFDs.

For burial depth estimation, the hybrid model utilizing the SPWVD achieved the best overall performance. The configuration achieved an R^2 of 0.9656 ± 0.0066, an MAE of 0.0394 ± 0.0042, and an RMSE of 0.0553 ± 0.0060. Furthermore, the BJD and PWVD configurations also showed high regression capabilities, achieving an R^2 score of 0.9653 ± 0.0043 and 0.9647 ± 0.0051, respectively. However, a standard deviation analysis reveals that the BJD is the more robust configuration. Although the SPWVD achieved the highest R^2, the BJD hybrid model minimized the overall MAE to 0.0387 ± 0.0010 alongside significantly lower standard deviations.

The quantitative results for pipeline diameter estimation across the baseline spatial, standalone TFD, and proposed hybrid models, expressed as R^2, MAE, and RMSE, are presented in Figures 6.16, 6.17, and C.3, respectively. The comprehensive numerical details are given in Tables C.5 and C.6. In the case of pipeline diameter, the spatial baseline (Xception) model revealed limitations by achieving an R^2 of 0.5910 ± 0.0430, confirming the high complexity of the parameter. However, the integration of spectral features produced significant performance improvements for this highly complex parameter. Outperforming the spatial baseline by ~10%, the WVD hybrid model achieved the highest overall R^2 score of 0.6863 ± 0.0325.
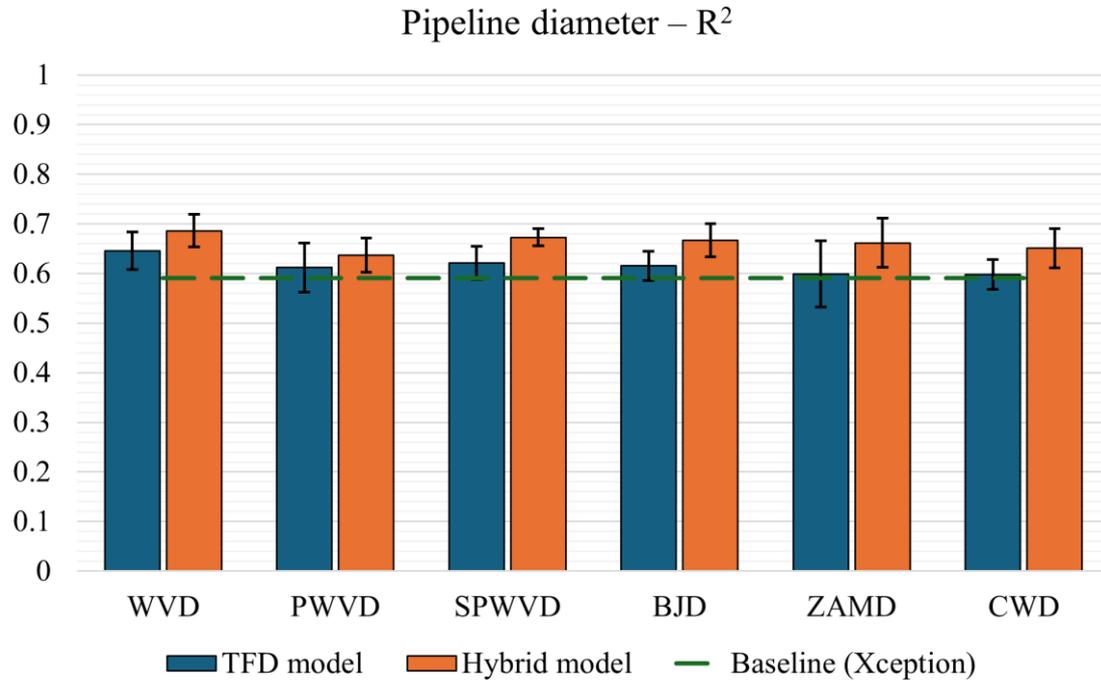
Figure 6.16 Performance comparison of TFD, hybrid, and baseline models for pipeline diameter estimation (expressed as R^2) across different TFDs.
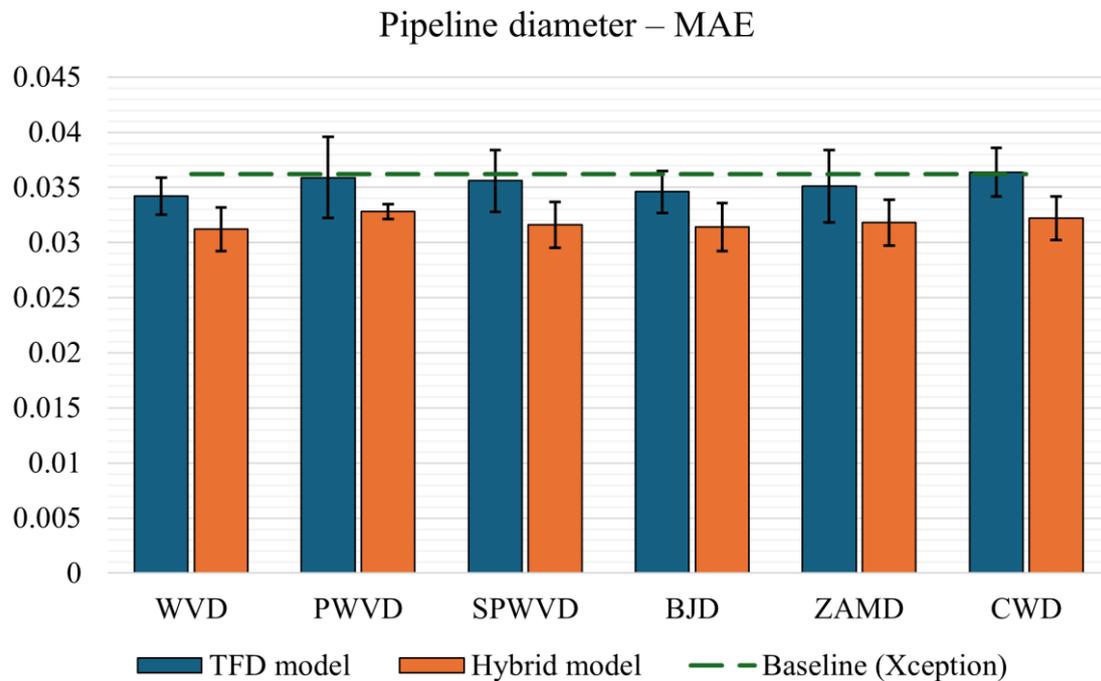
Figure 6.17 Performance comparison of TFD, hybrid, and baseline models for pipeline diameter estimation (expressed as MAE) across different TFDs.

Furthermore, the same configuration also achieved the lowest MAE (0.0312 ± 0.0020) and RMSE (0.0456 ± 0.0051). The SPWVD hybrid model closely followed with an R^2 of 0.6726 ± 0.0174. Experimental results revealed that relying entirely on the spectral domain proved more effective that the standalone spatial approach, with TFD models outperforming the Xception baseline by achieving R^2 values ranging from 0.5983 for CWD to 0.6454 for WVD. The behavioral shift from the TWTT and burial depth trends implies that localized frequency variations contain more descriptive information regarding pipeline dimensions than the raw geometric shape of the reflection alone.

The quantitative results for intersection angle estimation across the baseline spatial, standalone TFD, and proposed hybrid models, expressed as R^2, MAE, and RMSE, are presented in Figures 6.18, 6.19, and C.4, respectively. The comprehensive numerical details are given in Tables C.7 and C.8. Assessing the intersection angle introduced predictive challenges similar to the pipeline diameter estimation. However, the proposed hybrid fusion approach proved highly beneficial. The spatial Xception baseline yielded an R^2 of 0.7182 ± 0.0250. Across all evaluated TFDs, the cross-attention hybrid models consistently surpassed the spatial baseline threshold.
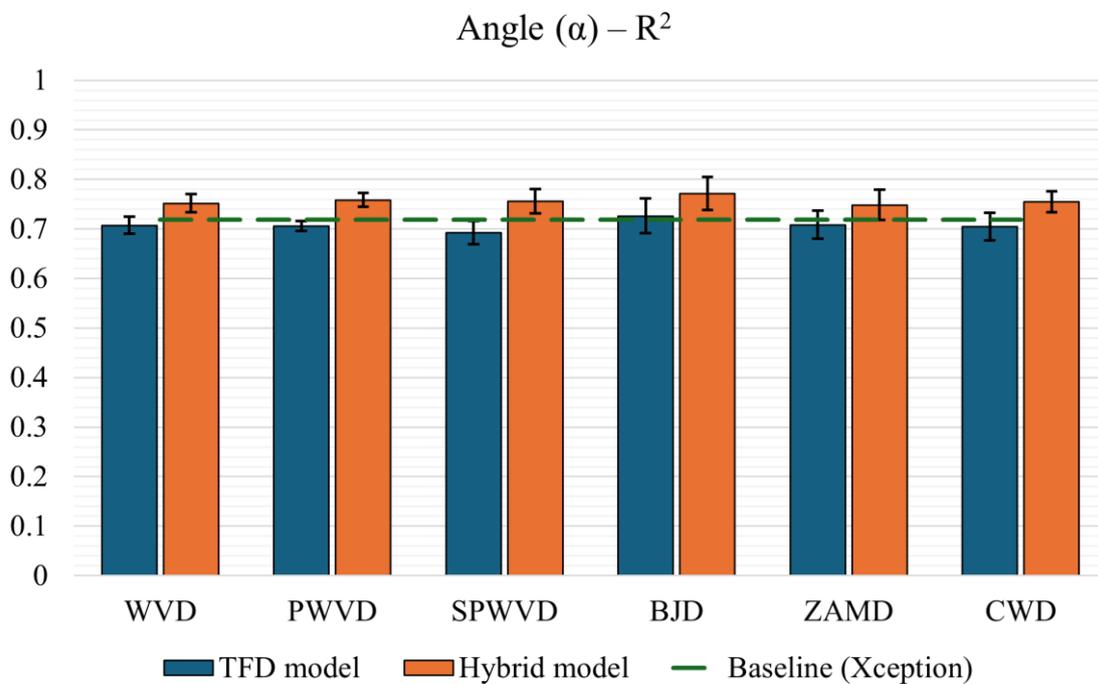
Figure 6.18 Performance comparison of TFD, hybrid, and baseline models for intersection angle estimation (expressed as R^2) across different TFDs.
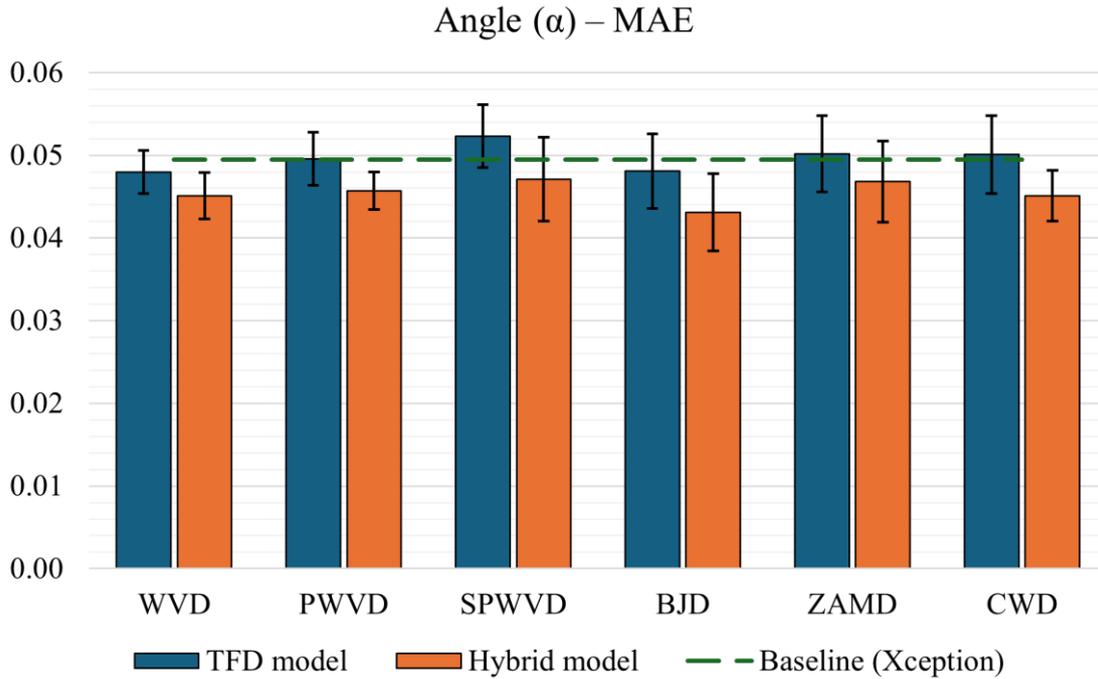
Figure 6.19 Performance comparison of TFD, hybrid, and baseline models for intersection angle estimation (expressed as MAE) across different TFDs.

According to the experimental results, the optimal configuration for intersection angle estimation utilized the BJD hybrid model, reaching a peak R^2 of 0.7713 ± 0.0336, which represents ~5% improvement over the baseline model. The BJD hybrid configuration simultaneously minimized the error, yielding an MAE of 0.0431 ± 0.0047 and an RMSE of 0.0645 ± 0.0065. As the second-best configuration, the PWVD hybrid model achieved an R^2 of 0.7585 ± 0.0139. Importantly, the standalone BJD spectral model marginally outperformed the spatial baseline by ~0.8%.

The comprehensive evaluation across the four descriptive parameters of underground infrastructure supports the superior representational capacity of the proposed spatio-spectral fusion architecture. Relying entirely on spatial data, however, revealed limitations when estimating complex geometrical characteristics such as pipeline diameter, and intersection angle. Such characteristics are encoded within localized frequency variations, which gives the spectral representations an analytical advantage. The results also reflect the aforementioned, as the standalone TFD models surpassed the spatial baseline in estimating both pipeline diameter and intersection angle. Despite the isolated strengths of each domain, the cross-attention hybrid model achieved the highest regression performance across every of four estimation parameters. The fusion approach maps interference-suppressed spectral features to the corresponding

spatial structures, effectively combining the most descriptive elements of both data sources. The specific performance optima for individual distributions (WVD for TWTT and pipeline diameter, and BJD for burial depth and intersection angle) prove that spectral feature selection should correspond to the physical complexity of the subsurface object.

The different optimal performances arise directly from how each TFD mathematical kernel interacts with the physical scattering phenomena of the GPR signals. TWTT and pipeline diameter estimations rely on the highly localized reflection apex of the hyperbola-shaped signature. For this reason, the corresponding estimations can benefit from the exact temporal positioning and maximum energy concentration preserved by the unsmoothed WVD. Estimating burial depth, however, introduces signal attenuation and frequency-dependent dispersion caused by the complex surrounding medium. By applying independent temporal and spectral smoothing, the SPWVD suppresses environmental clutter and stabilizes the depth profile. Thus, a necessary trade-off in absolute resolution is accepted in order to improve burial depth estimations. Estimation of the intersection angle introduces different geometric complexities, as non-perpendicular pipeline crossings elongate the standard hyperbola-shaped reflections into broader, often asymmetrical profiles. Such distortion, combined with the interference of scattered waves, masks the individual auto-terms required for accurate structural reconstruction. By suppressing distant cross-terms, the BJD mathematical formulation unmasks the required auto-terms, enabling the hybrid model to decouple the overlapping interference patterns without sacrificing the object geometry. According to the experimental results in this subsection, the proposed hybrid model proves to be a highly adaptable framework for precise underground infrastructure mapping by successfully integrating the spatial and spectral data domains.

6.2.4 Experimental Validation of the End-to-End Framework on Field GPR Data

The complete end-to-end framework is experimentally validated to establish a proof-of-concept on real-world, field-acquired GPR data. The survey samples are acquired utilizing a GSSI UtilityScan system, capturing physical underground pipelines situated within the typical subsurface clutter. Figure 6.20 illustrates the performance of the initial detection and localization stage. As observed, the YOLOv11x model successfully localizes the ground truth hyperbola-shaped reflections and generates precise bounding boxes.
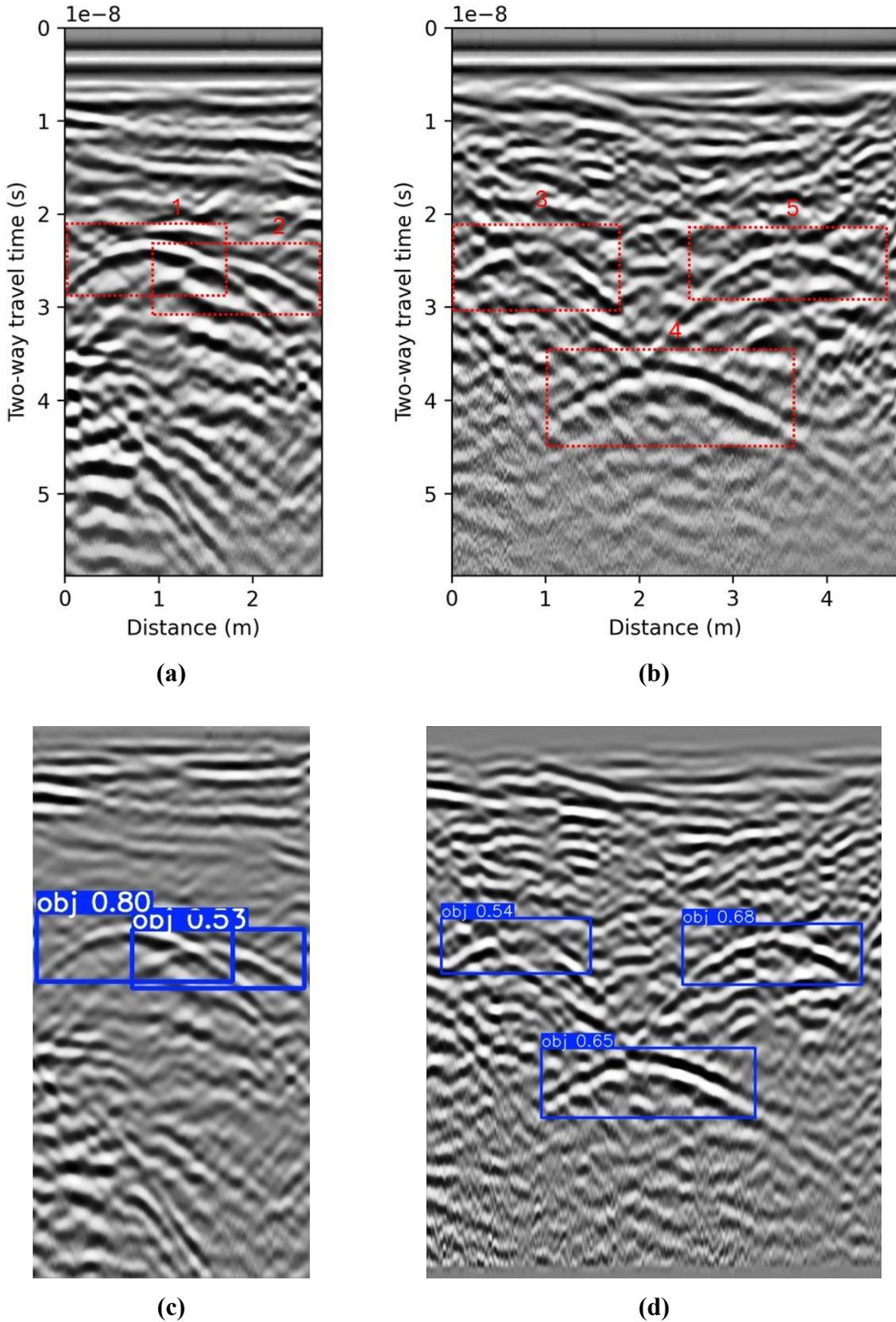
Figure 6.20 YOLOv11x detection performance on real-world GPR data acquired using a GSSI UtilityScan. The manually annotated ground-truth hyperbola-shaped reflections from two distinct locations are shown in (a) and (b), with the corresponding YOLOv11x predicted bounding boxes in (c) and (d).

However, field data inherently contain non-target hyperbola-shaped reflections generated by various subsurface features such as stones, tree roots, rebar, construction debris, air-filled voids, water-filled cavities, or other anomalies with sufficient dielectric contrast relative to the surrounding medium. An example of such a reflection is visible between boxes 3 and 5 in Figure 6.20 (b). The detection algorithm successfully excluded this feature, confirming that it is specifically trained to isolate pipeline-related hyperbola-shaped reflections and discard irrelevant clutter. Above-ground objects, such as fences or poles, can also produce hyperbolic diffraction signatures when the antenna is sufficiently close to receive lateral reflections. While the presence of aforementioned complex artifacts can reduce the overall detection performance through potential false positives, the proposed two-stage framework systematically overcomes this challenge. If the estimated parameters result in values outside the expected physical distribution for the pipeline in question, the reflection can be discarded as a false positive.

Following the successful localization, the cropped spatial and spectral representations are processed by the CASS-Fusion network to extract physical characteristics of the detected pipelines. Table 6.3 presents a quantitative comparison between the ground truth values and the estimated descriptive parameters for the reflections isolated in the preceding figure. The CASS-Fusion framework employs a parameter-specific spectral extraction strategy, utilizing the WVD to estimate the TWTT and pipeline diameter, and the BJD to estimate the burial depth and intersection angle. The dual-branch architecture proved highly reliable for positional parameter estimation, which resulted in sub-decimeter accuracy for the burial depths of reflections 1 through 4, with a minor 0.18 m underestimation observed for reflection 5.

Table 6.3 Quantitative comparison of the ground truth and estimated descriptive parameters of underground infrastructure in terms of TWTT, burial depth, pipeline diameter, and intersection angle, for the reflections shown in Figure 6.20.

Reflection	TWTT (ns)		Burial depth (m)		Pipeline diameter (m)		Angle $- |\alpha|$ (°)	
	True	Est.	True	Est.	True	Est.	True	Est.
1	22.32	23.46	1.10	1.11	0.30	0.23	51	49
2	23.93	24.27	1.18	1.15	0.25	0.26	51	46
3	22.03	21.37	0.95	0.96	0.11	0.14	55	73
4	36.02	34.44	1.75	1.67	0.21	0.25	45	54
5	23.07	21.53	1.15	0.97	0.10	0.12	53	56

Furthermore, the absolute error did not exceed 0.07 m in the case of pipeline diameter estimations. Regarding the intersection angle, estimations for reflections 1, 2, 4, and 5 remained within a narrow error margin of 2˚ to 9˚, whereas reflection 3 resulted in a more significant deviation. In conclusion, the precise alignment of the estimated descriptive parameters with the physical ground truth validates the integration of the proposed dual-branch architecture, confirming the complete two-stage framework as a reliable tool for automated utility mapping.

However, several limitations of the proposed framework should be acknowledged. First, the GPR data is acquired exclusively across multiple microlocations within the city of Rijeka, Croatia, which limits the geological diversity. Second, a single GSSI UtilityScan system equipped with 350 MHz antenna is utilized during data acquisition, restricting the generalizability of the results to other GPR devices and antenna configurations. Lastly, the range resolution of a 350 MHz antenna may be insufficient to directly resolve the geometry of small subsurface objects, such as cables or narrow-diameter pipelines, which effectively behave as point scatterers rather than spatially resolved targets.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

7.1 Conclusions

This thesis presented a two-stage DL-based system for the underground infrastructure descriptive parameters estimation from GPR data. The proposed framework integrates an object detection algorithm in the first stage with a proposed hybrid architecture that fuses spatial and spectral feature representations in the second stage. The experimental evaluation of the first stage showed that YOLOv11x achieves more robust and consistent detection and localization of hyperbola-shaped reflections compared to the transformer-based RT-DETR-X. The inclusion of synthetic data in the training set proved beneficial for the generalization of both object detection architectures. Moreover, evaluating the performances of the hyperbola fitting approach revealed critical limitations; the methodology lacked sufficient visual information and spatial reasoning required to reliably distinguish between changes in burial depth, pipeline diameter, and intersection angle. These limitations motivated the development of the primary contribution of the research, the CASS-Fusion network, which bridges the domain gap by integrating spatial B-scan features, extracted via an Xception backbone, with localized spectral representations derived from Cohen's class TFDs. The Xception backbone is selected as the optimal spatial feature extractor based on the DL baseline experiments, where it consistently outperformed DenseNet201, NASNetLarge, and EfficientNetV2L across all four descriptive parameters estimation. Additionally, the inclusion of synthetic data in the training set proved

beneficial for Xception, which further improved the regression performance. The hybrid model, on the other hand, outperformed both the standalone spatial baseline and standalone TFD models across all four descriptive parameters estimation. The most significant improvements are observed for pipeline diameter, where the WVD hybrid model improved R^2 by ~10% over the Xception baseline. By analyzing the experimental results, it can be concluded that the choice of the optimal TFD depends directly on the physical properties and complexity of the target parameter. Accordingly, the WVD achieved best results for TWTT and pipeline diameter estimation, while the BJD proved optimal for burial depth and intersection angle estimation. Finally, to provide proof-of-concept under real-world operating conditions, the complete end-to-end framework is experimentally validated using field data acquired with a GSSI UtilityScan system.

In summary, the research confirms both hypotheses. High performance in the detection and localization of underground infrastructure objects is achievable through DL-based object detection, with YOLOv11x demonstrating robust and stable performance. Furthermore, the simultaneous estimation of multiple descriptive parameters, such as TWTT, burial depth, pipeline diameter, and the intersection angle, is achievable through the proposed hybrid architecture that integrates spatial and spectral features. The scientific contributions of the thesis, namely the GIS-based GPR data labeling protocol, the DL detection and localization model, and the CASS-Fusion hybrid estimation model, collectively form a complete and experimentally validated system for automated underground infrastructure characterization from GPR data.

7.2 Future work

Despite the promising results achieved by the proposed framework, several directions for future work are identified that could further extend its capabilities. First, the dataset should be expanded by incorporating GPR surveys from diverse geographic locations with varying soil compositions, moisture levels, and geological conditions. This expansion should specifically target underrepresented parameter ranges, particularly intersection angles below 50˚, in order to improve the generalization capability of the framework. In parallel, the framework's applicability should be investigated across multiple GPR systems with different antenna frequencies and spatial sampling characteristics, since the current evaluation is limited to a

single GSSI UtilityScan radar. Currently, the CASS-Fusion network utilizes a parameter-specific, fixed TFD selection strategy; the optimal TFD for each descriptive parameter is selected independently based on its cross-validation performance. An extension would be to develop an adaptive TFD fusion mechanism that dynamically weights contributions from multiple distributions within a single forward pass. Finally, integrating the parameter estimates produced by the framework across multiple B-scans acquired in a mesh survey pattern could enable 3D reconstruction of the subsurface infrastructure. Such a capability would allow an automated generation of georeferenced underground infrastructure maps directly from GPR survey data.

BIBLIOGRAPHY

[1]   Afrasiabi, A., Faramarzi, A., Chapman, D., & Keshavarzi, A. (2025). Optimising Ground Penetrating Radar data interpretation: A hybrid approach with AI-assisted Kalman Filter and Wavelet Transform for detecting and locating buried utilities. *Journal of Applied Geophysics*, *232*, 105567.

[2]   Alkhammash, E. H. (2025). Multi-classification using YOLOv11 and hybrid YOLO11n-MobileNet models: A fire classes case study. *Fire*, *8*(1), 17.

[3]   Al-Nuaimy, W., Huang, Y., Nakhkash, M., Fang, M. T. C., Nguyen, V. T., & Eriksen, A. (2000). Automatic detection of buried utilities and solid objects with GPR using neural networks and pattern recognition. *Journal of applied Geophysics*, *43*(2-4), 157-165.

[4]   Annan, A. P. (2009). Electromagnetic principles of ground penetrating radar. *Ground penetrating radar theory and applications*, 1-40.

[5]   Bai, X., Yang, Y., Wei, S., Chen, G., Li, H., Li, Y., ... & Cui, H. (2023). A comprehensive review of conventional and deep learning approaches for ground-penetrating radar detection of raw data. *Applied Sciences*, *13*(13), 7992.

[6]   Baker, G. S., & Jol, H. M. (Eds.). (2007). *Stratigraphic analyses using GPR* (Vol. 432). Boulder, CO, USA: Geological Society of America.

[7]   Baker, G. S., Jordan, T. E., & Pardy, J. (2007). An introduction to ground penetrating radar (GPR).

[8]   Barkataki, N., Tiru, B., & Sarma, U. (2023). Size estimation of underground targets from GPR frequency spectra: A deep learning approach. *Journal of Applied Geophysics*, *213*, 105009.

[9]   Benedetto, F., & Tosti, F. (2013). GPR spectral analysis for clay content evaluation by the frequency shift method. *Journal of Applied Geophysics*, *97*, 89-96.

[10]   Berenger, J. P. (1994). A perfectly matched layer for the absorption of electromagnetic waves. *Journal of computational physics*, *114*(2), 185-200.

[11] Besaw, L. E., & Stimac, P. J. (2015, May). Deep convolutional neural networks for classifying GPR B-scans. In *Detection and sensing of mines, explosive objects, and obscured targets XX* (Vol. 9454, pp. 385-394). SPIE.

[12] Birkenfeld, S. (2010, September). Automatic detection of reflexion hyperbolas in GPR data with neural networks. In *2010 World Automation Congress* (pp. 1-6). IEEE.

[13] Boashash, B. (1988). Note on the use of the Wigner distribution for time-frequency signal analysis. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *36*(9), 1518-1521.

[14] Boashash, B. (Ed.). (2016). *Time-frequency signal analysis and processing: A comprehensive reference (2nd ed.)*. Academic press.

[15] Borgli, R. J., Stensland, H. K., Riegler, M. A., & Halvorsen, P. (2019, May). Automatic hyperparameter optimization for transfer learning on medical image datasets using bayesian optimization. In *2019 13th international symposium on medical information and communication technology (ISMICT)* (pp. 1-6). IEEE.

[16] Bracewell, R., & Kahn, P. B. (1966). The Fourier transform and its applications. *American Journal of Physics*, *34*(8), 712-712.

[17] Candelieri, A., Perego, R., & Archetti, F. (2018). Bayesian optimization of pump operations in water distribution systems. *Journal of Global Optimization*, *71*(1), 213-235.

[18] Chandramouli, A., Song, H., Liu, M., Damai, A., Narman, H. S., & Alzarrad, A. (2025, October). Deep Learning Approaches for Railroad Infrastructure Monitoring: Comparing YOLO and Vision Transformers for Defect Detection. In *2025 IEEE 16th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)* (pp. 0205-0211). IEEE.

[19] Cheng, Q., Cui, F., Chen, B., Dong, G., Wang, R., Zhang, G., & Fu, J. (2024). Attenuation of non-stationary random noise in ground penetrating radar data based on time-varying filtering. *Measurement*, *236*, 115169.

[20] Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *Peerj computer science*, *7*, e623.

[21] Choi, H. I., & Williams, W. J. (1989). Improved time-frequency representation of multicomponent signals using exponential kernels. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *37*(6), 862-871.

[22] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251-1258).

[23] Cohen, L. (1995). Time-frequency analysis. *Englewood Cliffs*.

[24] Cohen, L. (2002). Time-frequency distributions-a review. *Proceedings of the IEEE*, *77*(7), 941-981.

[25] Conyers, L. B. (2023). *Ground-penetrating radar for archaeology*. Rowman & Littlefield.

[26] Cordero, E., De Gosson, M., & Nicola, F. (2018). On the reduction of the interferences in the Born–Jordan distribution. *Applied and Computational Harmonic Analysis*, *44*(2), 230-245.

[27] Daniels, D. J. (1996). Surface-penetrating radar. *Electronics & Communication Engineering Journal*, *8*(4), 165-182.

[28] Daniels, D. J. (Ed.). (2004). *Ground penetrating radar* (Vol. 1). Iet.

[29] Davis, J. L., & Annan, A. P. (1989). Ground-penetrating radar for high-resolution mapping of soil and rock stratigraphy 1. *Geophysical prospecting*, *37*(5), 531-551.

[30] De Freitas, N., & Wang, Z. (2013). Bayesian optimization in high dimensions via random embeddings.

[31] Dou, Q., Wei, L., Magee, D. R., & Cohn, A. G. (2016). Real-time hyperbola recognition and fitting in GPR data. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(1), 51-62.

[32] Drozdova, I., & Petrov, A. (2018). World practice and Russian experience of housing and utilities sector digitization. In *SHS Web of Conferences* (Vol. 44, p. 00031). EDP Sciences.

[33] Economou, N. (2016). Time-varying band-pass filtering GPR data by self-inverse filtering. *Near Surface Geophysics*, *14*(2), 207-217.

[34] Eriksson, D., Pearce, M., Gardner, J., Turner, R. D., & Poloczek, M. (2019). Scalable global optimization via local Bayesian optimization. *Advances in neural information processing systems*, *32*.

[35] Feng, J., Yang, L., Wang, H., Song, Y., & Xiao, J. (2020, May). Gpr-based subsurface object detection and reconstruction using random motion and depthnet. In *2020 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 7035-7041). IEEE.

[36] Feurer, M., Klein, A., Eggensperger, K., Springenberg, J., Blum, M., & Hutter, F. (2015). Efficient and robust automated machine learning. *Advances in neural information processing systems*, *28*.

[37] Flandrin, P. (1984, March). Some features of time-frequency representations of multicomponent signals. In *ICASSP'84. IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 9, pp. 266-269). IEEE.

[38] Flandrin, P., & Escudié, B. (1984). An interpretation of the pseudo-Wigner-Ville distribution. *Signal Processing*, *6*(1), 27-36.

[39] Geophysical Survey Systems. UtilityScan Ground Penetrating Radar System, accessed: 14.09.2025. URL: *https://www.geophysical.com/products/utilityscan*

[40] Giannopoulos, A. (2005). Modelling ground penetrating radar by GprMax. *Construction and building materials*, *19*(10), 755-762.

[41] Hansen, L. H., van Son, R., Weiser, A., & Kjems, E. (2021). Addressing the elephant in the underground: An argument for the integration of heterogeneous data sources for reconciliation of subsurface utility data. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (pp. 43-48).

[42] He, L. H., Zhou, Y. Z., Liu, L., Cao, W., & Ma, J. H. (2025). Research on object detection and recognition in remote sensing images based on YOLOv11. *Scientific Reports*, *15*(1), 14032.

[43] He, W., & Lai, W. W. L. (2024). Unified optimization-based analysis of GPR hyperbolic fitting models. *Tunnelling and Underground Space Technology*, *146*, 105633.

[44] He, W., & Lai, W. W. L. (2024). Angle-corrected GPR hyperbolic fitting models for improved parameter estimation. *Tunnelling and Underground Space Technology*, *147*, 105741.

[45] He, W., Zhang, Y., Xu, T., An, T., Liang, Y., & Zhang, B. (2025, February). Object detection for medical image analysis: Insights from the RT-DETR model. In *Proceedings of the 2025 International Conference on Artificial Intelligence and Computational Intelligence* (pp. 415-420).

[46] Hlawatsch, F., & Auger, F. (Eds.). (2013). *Time-frequency analysis*. John Wiley & Sons.

[47] Hlawatsch, F., & Boudreaux-Bartels, G. F. (2002). Linear and quadratic time-frequency signal representations. *IEEE signal processing magazine*, *9*(2), 21-67.

[48]   Hlawatsch, F., & Flandrin, P. (1997). The interference structure of the Wigner distribution and related time-frequency signal representations. In *The Wigner Distribution-Theory and Applications in Signal Processing* (pp. 59-133). Elsevier.

[49]   Hlawatsch, F., Manickam, T. G., Urbanke, R. L., & Jones, W. (1995). Smoothed pseudo-Wigner distribution, Choi-Williams distribution, and cone-kernel representation: Ambiguity-domain analysis and experimental comparison. *Signal Processing*, *43*(2), 149-168.

[50]   Ho, K. C., Gader, P. D., Wilson, J. N., Lee, W., & Glenn, T. C. (2005, June). Landmine detection using frequency domain features from GPR measurements and their fusion with time domain features. In *Detection and Remediation Technologies for Mines and Minelike Targets X* (Vol. 5794, pp. 1141-1150). SPIE.

[51]   Hou, F., Rui, X., Fan, X., & Zhang, H. (2022). Review of GPR activities in civil infrastructures: Data analysis and applications. *Remote Sensing*, *14*(23), 5972.

[52]   Hou, L., Zhang, Q., & Du, Y. (2024). Width estimation of hidden cracks in tunnel lining based on time-frequency analysis of GPR data and back propagation neural network optimized by genetic algorithm. *Automation in Construction*, *162*, 105394.

[53]   Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

[54]   Huang, G., Liu, Z., Pleiss, G., Van Der Maaten, L., & Weinberger, K. Q. (2019). Convolutional networks with dense connectivity. *IEEE transactions on pattern analysis and machine intelligence*, *44*(12), 8704-8716.

[55]   Iwai, S., & Mizutani, T. (2024). Millimeter accuracy depth estimation in concrete using smartphone-equipped handheld GPR and hyperbolic curve fitting with antenna distance consideration. *Heliyon*, *10*(19).

[56]   Jazayeri, S., Klotzsche, A., & Kruse, S. (2018). Improving estimates of buried pipe diameter and infilling material from ground-penetrating radar profiles with full-waveform inversion. *Geophysics*, *83*(4), H27-H41.

[57]   Jeong, J., & Williams, W. J. (1992). Kernel design for reduced interference distributions. *IEEE Transactions on Signal Processing*, *40*(2), 402-412.

[58]   Jiang, G., Zhou, X., Li, J., & Chen, H. (2019). A cable-mapping algorithm based on ground-penetrating radar. *IEEE Geoscience and Remote Sensing Letters*, *16*(10), 1630-1634.

[59] Jiang, Y., Dai, Z., Chen, Y., & Chu, Z. (2025, July). An Improved Multi-task Model for Instance Segmentation and Pose Estimation Based on YOLOv11. In *International Conference on Intelligent Computing* (pp. 218-229). Singapore: Springer Nature Singapore.

[60] Jin, A., Chen, C., Yang, B., Zou, Q., Wang, Z., Yan, Z., ... & Zhou, J. (2024). Gprformer: Detection and parametric reconstruction of hyperbolas in gpr b-scan images with transformers. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1-13.

[61] Jol, H. M. (Ed.). (2008). *Ground penetrating radar theory and applications*. Elsevier.

[62] Jones, D. R., Schonlau, M., & Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, *13*(4), 455-492.

[63] Jun, E. L. T., Tham, M. L., & Kwan, B. H. (2024, June). A Comparative Analysis of RT-DETR and YOLOv8 for Urban Zone Aerial Object Detection. In *2024 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)* (pp. 340-345). IEEE.

[64] Khan, Z. Y., & Niu, Z. (2021). CNN with depthwise separable convolutions and combined kernels for rating prediction. *Expert Systems with Applications*, *170*, 114528.

[65] Khanam, R., & Hussain, M. (2024). Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725*.

[66] Kodituwakku, S., Kennedy, R. A., & Abhayapala, T. D. (2010). Radial function based kernel design for time-frequency distributions. *IEEE Transactions on Signal Processing*, *58*(6), 3395-3400.

[67] Kong, Y., Shang, X., & Jia, S. (2024). Drone-DETR: Efficient small object detection for remote sensing image using enhanced RT-DETR model. *Sensors*, *24*(17), 5496.

[68] Krause, A., Perciavalle, P., Johnson, K., Owens, B., Frodl, D., Sarni, W., & Foundry, W. (2018). The digitization of water. *In: Boston*.

[69] Lai, W. W. L., Kind, T., Sham, J. F. C., & Wiggenhauser, H. (2016). Correction of GPR wave velocity at different oblique angles between traverses and alignment of line objects in a common offset antenna setting. *Ndt & E International*, *82*, 36-43.

[70] Lau, P. K. W., Cheung, B. W. Y., Lai, W. W. L., & Sham, J. F. C. (2021). Characterizing pipe leakage with a combination of GPR wave velocity algorithms. *Tunnelling and Underground Space Technology*, *109*, 103740.

[71] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436-444.

[72] Lei, W., Hou, F., Xi, J., Tan, Q., Xu, M., Jiang, X., ... & Gu, Q. (2019). Automatic hyperbola detection and fitting in GPR B-scan image. *Automation in Construction*, *106*, 102839.

[73] Li, Y., Zhao, Z., Luo, Y., & Qiu, Z. (2020). Real-time pattern-recognition of GPR images with YOLO v3 implemented by tensorflow. *Sensors*, *20*(22), 6476.

[74] Li, Y., Zhang, W., Lv, S., Yu, J., Ge, D., Guo, J., & Li, L. (2025). YOLOv11-CAFM model in ground penetrating radar image for pavement distress detection and optimization study. *Construction and Building Materials*, *485*, 141907.

[75] Liang, H., Yao, D., & Shen, H. (2022). Pseudo Wigner-Ville distribution for 3D white light scanning interferometric measurement. *Optics Express*, *30*(22), 40540-40556.

[76] Liu, C., Han, P., & Kang, J. (2025). Integrating short-time linear canonical transform and joint space-time-frequency analysis for advanced representation of subsurface information in ground penetrating radar. *Scientific Reports*, *15*(1), 23315.

[77] Liu, H., Lin, C., Cui, J., Fan, L., Xie, X., & Spencer, B. F. (2020). Detection and localization of rebar in concrete by deep learning using ground penetrating radar. *Automation in construction*, *118*, 103279.

[78] Liu, H., Yue, Y., Liu, C., Spencer Jr, B. F., & Cui, J. (2023). Automatic recognition and localization of underground pipelines in GPR B-scans using a deep learning model. *Tunnelling and Underground Space Technology*, *134*, 104861.

[79] Liu, H., Chen, J., Zhang, X., Dai, D., Cui, J., & Spencer, B. F. (2024). Collaborative imaging of subsurface cavities using ground-pipeline penetrating radar. *IEEE Geoscience and Remote Sensing Letters*, *21*, 1-5.

[80] Liu, Z., Wu, W., Gu, X., Li, S., Wang, L., & Zhang, T. (2021). Application of combining YOLO models and 3D GPR images in road detection and maintenance. *Remote Sensing*, *13*(6), 1081.

[81] Lu, G., Zhao, W., Forte, E., Tian, G., Li, Y., & Pipan, M. (2020). Multi-frequency and multi-attribute GPR data fusion based on 2-D wavelet transform. *Measurement*, *166*, 108243.

[82] Lujan-Moreno, G. A., Howard, P. R., Rojas, O. G., & Montgomery, D. C. (2018). Design of experiments and response surface methodology to tune machine learning hyperparameters, with a random forest case-study. *Expert Systems with Applications*, *109*, 195-205.

[83] Molyneaux, T. C. K., Millard, S. G., Bungey, J. H., & Zhou, J. Q. (1995). Radar assessment of structural concrete using neural networks. *NDT & E International*, *28*(5), 281-288.

[84] Namdari, H., Moradikia, M., Petkie, D. T., Askari, R., & Zekavat, S. (2023, September). Comprehensive GPR signal analysis via descriptive statistics and machine learning. In *2023 IEEE International Conference on Wireless for Space and Extreme Environments (WiSEE)* (pp. 127-132). IEEE.

[85] Orini, M., Bailón, R., Mainardi, L. T., Laguna, P., & Flandrin, P. (2011). Characterization of dynamic interactions between cardiovascular signals by time-frequency coherence. *IEEE transactions on biomedical engineering*, *59*(3), 663-673.

[86] Padilla, R., Netto, S. L., & Da Silva, E. A. (2020, July). A survey on performance metrics for object-detection algorithms. In *2020 international conference on systems, signals and image processing (IWSSIP)* (pp. 237-242). IEEE.

[87] Palermo, S. A., Maiolo, M., Brusco, A. C., Turco, M., Pirouz, B., Greco, E., ... & Piro, P. (2022). Smart technologies for water resource management: An overview. *Sensors*, *22*(16), 6225.

[88] Pak, C. (2024). *Monitoring of soil movement due to deep excavation works using Time Lapse Ground Penetrating Radar (TLGPR)* (Doctoral dissertation, RMIT University).

[89] Park, S., Kim, J., Jeon, K., Kim, J., & Park, S. (2021). Improvement of gpr-based rebar diameter estimation using yolo-v3. *Remote Sensing*, *13*(10), 2011.

[90] Pasolli, E., Melgani, F., & Donelli, M. (2009). Automatic analysis of GPR images: A pattern-recognition approach. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(7), 2206-2217.

[91] Peplinski, N. R., Ulaby, F. T., & Dobson, M. C. (1995). Dielectric properties of soils in the 0.3-1.3-GHz range. *IEEE transactions on Geoscience and Remote sensing*, *33*(3), 803-807.

[92] Pham, H., Guan, M., Zoph, B., Le, Q., & Dean, J. (2018, July). Efficient neural architecture search via parameters sharing. In *International conference on machine learning* (pp. 4095-4104). PMLR.

[93] Qi, Y., Wang, X., Yang, X., Sun, T., Razzaq, I., Yang, L., ... & Rasool, G. (2024). Nonexcavation Localization Method for Buried PE Pipes Based on Elastic Wave Reflection Imaging and the BPA Method. *IEEE Sensors Journal*, *24*(11), 17987-17998.

[94] Rasol, M., Pais, J. C., Pérez-Gracia, V., Solla, M., Fernandes, F. M., Fontul, S., ... & Assadollahi, H. (2022). GPR monitoring for road transport infrastructure: A systematic review and machine learning insights. *Construction and Building Materials*, *324*, 126686.

[95] Rasol, M., Elseicy, A., Solla, M., Celaya, M., & Schmidt, F. (2024). Role of intelligent data analysis to enhance GPR data interoperability: road transports. In *Interpretable Machine Learning for the Analysis, Design, Assessment, and Informed Decision Making for Civil Infrastructure* (pp. 159-184). Woodhead Publishing.

[96] Ristic, A. V., Petrovacki, D., & Govedarica, M. (2009). A new method to simultaneously estimate the radius of a cylindrical object and the wave propagation velocity from GPR data. *Computers & Geosciences*, *35*(8), 1620-1630.

[97] Rohani, A., Taki, M., & Abdollahpour, M. (2018). A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I). *Renewable Energy*, *115*, 411-422.

[98] Savelyev, T. G., Van Kempen, L., Sahli, H., Sachs, J., & Sato, M. (2006). Investigation of time–frequency features for GPR landmine discrimination. *IEEE transactions on geoscience and remote sensing*, *45*(1), 118-129.

[99] Shahrabi, M. A., & Hashemi, H. (2021). Analysis of GPR hyperbola targets using image processing techniques. *J Seism Explor*, *6*, 561-575.

[100] Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & De Freitas, N. (2015). Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, *104*(1), 148-175.

[101] Sham, J. F., & Lai, W. W. (2016). Development of a new algorithm for accurate estimation of GPR's wave propagation velocity by common-offset survey method. *NDT & E International*, *83*, 104-113.

[102] Shi, X., Liu, L., Bao, X., Pan, B., & Hussain, S. (2025). Dynamic gated fusion network with hierarchical multi-scale attention for hyperspectral image classification. *Scientific Reports*, *15*(1), 44289.

[103] Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, *25*.

[104] Snoek, J., Rippel, O., Swersky, K., Kiros, R., Satish, N., Sundaram, N., ... & Adams, R. (2015, June). Scalable bayesian optimization using deep neural networks. In *International conference on machine learning* (pp. 2171-2180). PMLR.

[105] Solimene, R., D'Alterio, A., & Soldovieri, F. (2011). Determining Fresnel reflection coefficients in 3D half-space geometry by GPR multistatic data. *Near Surface Geophysics*, *9*(3), 265-276.

[106] Srinivas, N., Krause, A., Kakade, S. M., & Seeger, M. (2009). Gaussian process optimization in the bandit setting: No regret and experimental design. *arXiv preprint arXiv:0912.3995*.

[107] Sunil, C. K., & Jaidhar, C. D. (2021). Cardamom plant disease detection approach using EfficientNetV2. *Ieee Access*, *10*, 789-804.

[108] Swersky, K., Snoek, J., & Adams, R. P. (2013). Multi-task bayesian optimization. *Advances in neural information processing systems*, *26*.

[109] Štifanić, D., Musulin, J., Miočević, A., Baressi Šegota, S., Šubić, R., & Car, Z. (2020). Impact of COVID-19 on Forecasting Stock Prices: An Integration of Stationary Wavelet Transform and Bidirectional Long Short-Term Memory. *Complexity*, *2020*(1), 1846926.

[110] Štifanić, D., Štifanić, J., Anđelić, N., & Car, Z. (2025). An Automated System for Underground Pipeline Parameter Estimation from GPR Recordings. *Remote Sensing*, *17*(20), 3493.

[111] Taflove, A., Hagness, S. C., & Piket-May, M. (2005). Computational electromagnetics: the finite-difference time-domain method. *The Electrical Engineering Handbook*, *3*(629-670), 15.

[112] Tan, M., & Le, Q. (2021, July). Efficientnetv2: Smaller models and faster training. In *International conference on machine learning* (pp. 10096-10106). PMLR.

[113] Victoria, A. H., & Maragatham, G. (2021). Automatic tuning of hyperparameters using Bayesian optimization. *Evolving Systems*, *12*(1), 217-223.

[114] Ville, J. (1948). Theorie et application dela notion de signal analysis. *Câbles et transmissions*, *2*(1), 61-74.

[115] Wagner, S. (2022). *Design of a 0.3-12 GHz Ground-Penetrating Radar Underground Imaging System* (Doctoral dissertation, University of California, Davis).

[116] Warren, C., Giannopoulos, A., & Giannakis, I. (2016). gprMax: Open source software to simulate electromagnetic wave propagation for Ground Penetrating Radar. *Computer Physics Communications*, *209*, 163-170.

[117] Wei, X., Yin, L., Zhang, L., & Wu, F. (2024). DV-DETR: improved UAV aerial small target detection algorithm based on RT-DETR. *Sensors*, *24*(22), 7376.

[118] Wen, Z., Lin, W., Wang, T., & Xu, G. (2023). Distract your attention: Multi-head cross attention network for facial expression recognition. *Biomimetics*, *8*(2), 199.

[119] Williams, C. K., & Rasmussen, C. E. (2006). *Gaussian processes for machine learning* (Vol. 2, No. 3, p. 4). Cambridge, MA: MIT press.

[120] Wirtz, B. W., Müller, W. M., & Schmidt, F. W. (2021). Digital public services in smart cities–An empirical analysis of lead user preferences. *Public Organization Review*, *21*(2), 299-315.

[121] Wong, T. T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern recognition*, *48*(9), 2839-2846.

[122] Wu, J., Chen, X. Y., Zhang, H., Xiong, L. D., Lei, H., & Deng, S. H. (2019). Hyperparameter optimization for machine learning models based on Bayesian optimization. *Journal of Electronic Science and Technology*, *17*(1), 26-40.

[123] Xia, T. (2017). GPR Technologies for Underground Sensing. *Underground Sensing: Monitoring and Hazard Detection for Environment and Infrastructure*, 115.

[124] Xie, F., Wu, C. G. W., Lai, W. W. L., & Sham, J. F. C. (2018). Correction of multi-frequency GPR wave velocity with distorted hyperbolic reflections from GPR surveys of underground utilities. *Tunnelling and Underground Space Technology*, *76*, 76-91.

[125] Xu, Q., Wang, Z., & Li, X. (2023). Depth and radius joint estimation for underground pipeline using GPR. *IEEE Sensors Journal*, *23*(13), 14628-14639.

[126] Yamaguchi, T., Mizutani, T., Meguro, K., & Hirano, T. (2022). Detecting subsurface voids from GPR images by 3-D convolutional neural network using 2-D finite difference time domain method. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 3061-3073.

[127] Yamaguchi, T., Mizutani, T., & Nagayama, T. (2020). Mapping subsurface utility pipes by 3-D convolutional neural network and Kirchhoff migration using GPR images. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(8), 6525-6536.

[128] Yang, B., Yang, Z., Wang, X., Mu, B., Xu, J., & Li, H. (2025). YOLOv11-XRBS: Enhanced Identification of Small and Low-Detail Explosives in X-Ray Backscatter Images. *Sensors*, *25*(19), 6130.

[129] Yelf, R. (2004, June). Where is true time zero?. In *Proceedings of the Tenth International Conference on Grounds Penetrating Radar, 2004. GPR 2004.* (Vol. 1, pp. 279-282). IEEE.

[130] Zadhoush, H., & Giannopoulos, A. (2022). Optimizing GPR time-zero adjustment and two-way travel time wavelet measurements using a realistic three-dimensional numerical model. *Near Surface Geophysics*, *20*(2), 208-226.

[131] Zhang, J., Liu, Q., & Nath, B. (2004, August). Landmine feature extraction and classification of GPR data based on SVM method. In *International Symposium on Neural Networks* (pp. 636-641). Berlin, Heidelberg: Springer Berlin Heidelberg.

[132] Zhang, X., Han, L., Robinson, M., & Gallagher, A. (2021). A GANs-based deep learning framework for automatic subsurface object recognition from ground penetrating radar data. *IEEE Access*, *9*, 39009-39018.

[133] Zhao, Y., Atlas, L. E., & Marks, R. J. (1990). The use of cone-shaped kernels for generalized time-frequency representations of nonstationary signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *38*(7), 1084-1091.

[134] Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., ... & Chen, J. (2024). Detrs beat yolos on real-time object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 16965-16974).

[135] Zhu, C., & Ye, H. (2023). A modular method for gpr hyperbolic feature detection and quantitative parameter inversion of underground pipelines. *Remote Sensing*, *15*(8), 2114.

[136] Zhu, M., & Kong, E. (2024). Multi-scale fusion uncrewed aerial vehicle detection based on RT-DETR. *Electronics*, *13*(8), 1489.

[137] Zong, Z., Chen, C., Mi, X., Sun, W., Song, Y., Li, J., ... & Yang, B. (2019). A deep learning approach for urban underground objects detection from vehicle-borne ground penetrating radar data in real-time. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*, 293-299.

[138] Zoph, B., Vasudevan, V., Shlens, J., & Le, Q. V. (2018). Learning transferable architectures for scalable image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8697-8710).

LIST OF ABBREVIATIONS

ADC	Analog-to-digital converter
AI	Artificial intelligence
AIFI	Attention-based intra-scale feature interaction
ANN	Artificial neural networks
AP	Average precision
AUC	Area under the curve
BiLSTM	Bidirectional long short-term memory
BJD	Born-Jordan distribution
BO	Bayesian optimization
C2PSA	Convolutional block with parallel spatial attention
C3k2	Cross stage partial block with kernel size 2
CASS-Fusion	Cross-attention spatio-spectral fusion network
CCFF	Cross-scale feature fusion
CNN	Convolutional neural network
CROPOS	Croatian positioning system
CSD	Cone-shaped distribution
CWD	Choi-Williams distribution
DL	Deep learning
EI	Expected improvement
EM	Electromagnetic
FC	Fully connected
FDTD	Finite-difference time-domain
FFT	Fast Fourier transform
FN	False negative
FP	False positives
GA	Genetic algorithm
GAN	Generative adversarial network

GAP2D	Global average pooling 2D
GIS	Geographic information system
GNSS	Global navigation satellite system
GP	Gaussian process
GPR	Ground penetrating radar
GPS	Global positioning system
GSSI	Geophysical Survey Systems Inc.
HCI	Human-computer interaction
IAF	Instantaneous autocorrelation function
ICT	Information and communication technology
IoU	Intersection over union
LSTM	Long short-term memory
MAE	Mean absolute error
MBConv	Mobile inverted bottleneck convolution
MSE	Mean squared error
mAP	Mean average precision
ML	Machine learning
NAS	Neural architecture search
NMS	Non-maximum suppression
PEC	Perfect electric conductor
PML	Perfectly matched layers
PWVD	Pseudo Wigner-Ville distribution
QGIS	Quantum geographic information system
ReLU	Rectified linear unit
RMSE	Root mean squared error
RT-DETR	Real-time detection transformer
Rx	Receiver antenna
SPPF	Spatial pyramid pooling-fast
SPWVD	Smoothed pseudo Wigner-Ville distribution
SSD	Single shot multibox detector
SVM	Support vector machines
TF	Time-frequency
TFD	Time-frequency distribution

TP	True positives
TWTT	Two-way travel time
Tx	Transmitter antenna
UCB	Upper confidence bound
WVD	Wigner-Ville distribution
YOLO	You only look once
ZAMD	Zhao-Atlas-Marks distribution

LIST OF SYMBOLS

Latin symbols:

A_{bbox}	Predicted bounding box
$A_{bbox} \cap B_{bbox}$	Intersection area of predicted and ground truth bounding boxes
$A_{bbox} \cup B_{bbox}$	Union area of predicted and ground truth bounding boxes
B	Magnetic field density
B_a	Half of the antenna separation
B_{bbox}	Ground truth bounding box
c	Velocity of the electromagnetic wave in free space
D	Electric displacement field
D_0	Burial depth measured from the ground surface to the top of the object
D_1	Distance from the antenna transmitter to the cylindrical object
D_2	Distance from the cylindrical object to the antenna receiver
E	Electric field
$EI(x)$	Expected improvement acquisition function
f	Frequency variable
$f(x)$	Objective function in Bayesian optimization
$f(x^+)$	Current best observed value of the objective function
$\mathcal{F}$	Fourier transform
$g(t)$	Time-smoothing window function
H	Magnetizing field
$h(\tau)$	Window function applied in the lag domain
J_c	Conduction current density
J_s	Impressed current density
k	Number of folds in cross-validation
k_{ucb}	Exploration-exploitation trade-off parameter
$k(x, x')$	Covariance kernel function

$K_s(t, \tau)$	Instantaneous autocorrelation function of the signal
l	Length-scale parameter of the kernel
$m(x)$	Mean function of the Gaussian process
n	Number of samples
N	Total number of discrete points in the isolated A-scan
q_v	Volume electric charge density
r	Radius of the cylindrical object
R^2	Coefficient of determination
$s(t)$	Time-domain signal
$s^*(t)$	Complex conjugate of s(t)
$S(f)$	Fourier transform of s(t)
SS_{res}	Residual sum of squares
SS_{tot}	Total sum of squares
t	Time
t_0	Vertical position expressed in terms of two-way travel time at the hyperbola apex
t_i	Vertical position expressed in terms of two-way travel time at measurement point i
$UCB(x)$	Upper confidence bound acquisition function
v	Propagation velocity of electromagnetic wave within the medium
x_0	Horizontal position of the antenna system at the hyperbola apex
xi	Horizontal position of the antenna system at measurement point i
x_{TFD}	Input TFD being normalized
x_{TFD_max}	Maximum value used in TFD normalization
x_{TFD_min}	Minimum value used in TFD normalization
x_{TFD_norm}	Normalized TFD
$\hat{y}_i$	Estimated value of sample i
y_i	Ground truth value of sample i
$\bar{y}$	Mean of all observed ground truth values
$\|x - x'\|^2$	Squared Euclidean distance

Greek symbols:

α	Intersection angle between the ground penetrating radar survey line and the cylindrical object
Δt	Temporal step size in the finite-difference time-domain simulation
Δx	Spatial step size in the x-direction in the finite-difference time-domain simulation
Δy	Spatial step size in the y-direction in the finite-difference time-domain simulation
Δz	Spatial step size in the z-direction in the finite-difference time-domain simulation
ε	Relative permittivity
μ	Relative permeability
$\mu_{ucb}(x)$	Mean of Gaussian process predictions at point x
∇	Nabla operator used in Maxwell's equations
σ	Kernel width parameter in the Choi-Williams distribution
$\sigma_{ucb}(x)$	Standard deviation of Gaussian process predictions at point x
τ	Lag variable

LIST OF FIGURES

LIST OF TABLES

APPENDICES

Appendix A

DETECTION AND LOCALIZATION RESULTS

Table A.1 Performances of the YOLOv11x model trained on real-world versus combined (real + synthetic) data for detection and localization of hyperbola-shaped reflections in B-scan data. Performance measures are expressed as mean ± standard deviation from 5-fold cross-validation.

YOLOv11x				
Dataset	**Precision**	**Recall**	**mAP@0.5**	**mAP@0.5~0.95**
Real	0.7789 ± 0.0238	0.7502 ± 0.0317	0.8354 ± 0.0182	0.5112 ± 0.0154
Real + synthetic	0.7732 ± 0.0400	0.7517 ± 0.0190	0.8388 ± 0.0193	0.5202 ± 0.0139

Table A.2 Performances of the RT-DETR-X model trained on real-world versus combined (real + synthetic) data for detection and localization of hyperbola-shaped reflections in B-scan data. Performance measures are expressed as mean ± standard deviation from 5-fold cross-validation.

RT-DETR-X				
Dataset	**Precision**	**Recall**	**mAP@0.5**	**mAP@0.5~0.95**
Real	0.7319 ± 0.0418	0.7335 ± 0.0262	0.7721 ± 0.0221	0.4725 ± 0.0142
Real + synthetic	0.7256 ± 0.0340	0.7443 ± 0.0328	0.7822 ± 0.0295	0.4837 ± 0.0271

Appendix B

DEEP LEARNING REGRESSION BASELINES

Figure B.1 Performance comparison of different DL model architectures for TWTT estimation (expressed as RMSE) across real-world and combined (real + synthetic) B-scan data.

Table B.1 Performances of DL feature extraction backbones (DenseNet201, Xception, NASNetLarge, and EfficientNetV2L) in TWTT estimation from hyperbola-shaped reflections, utilizing only real-world data for model training. Performance measures are expressed as R^2, MAE, and RMSE (mean ± standard deviation) derived from 5-fold cross-validation.

TWTT – Real data			
Backbone	$R^2 \pm \sigma$	**MAE ± σ**	**RMSE ± σ**
DenseNet201	0.9202 ± 0.0164	1.4668 ± 0.2340	1.8081 ± 0.2182
Xception	0.9644 ± 0.0048	0.9527 ± 0.0911	1.2114 ± 0.1002
NASNetLarge	0.9587 ± 0.0065	0.9990 ± 0.0565	1.3010 ± 0.0961
EfficientNetV2L	0.9209 ± 0.0479	1.4066 ± 0.5275	1.7335 ± 0.5442

Table B.2 Performances of DL feature extraction backbones (DenseNet201, Xception, NASNetLarge, and EfficientNetV2L) in TWTT estimation from hyperbola-shaped reflections, utilizing combined data for model training. Performance measures are expressed as R^2, MAE, and RMSE (mean ± standard deviation) derived from 5-fold cross-validation.

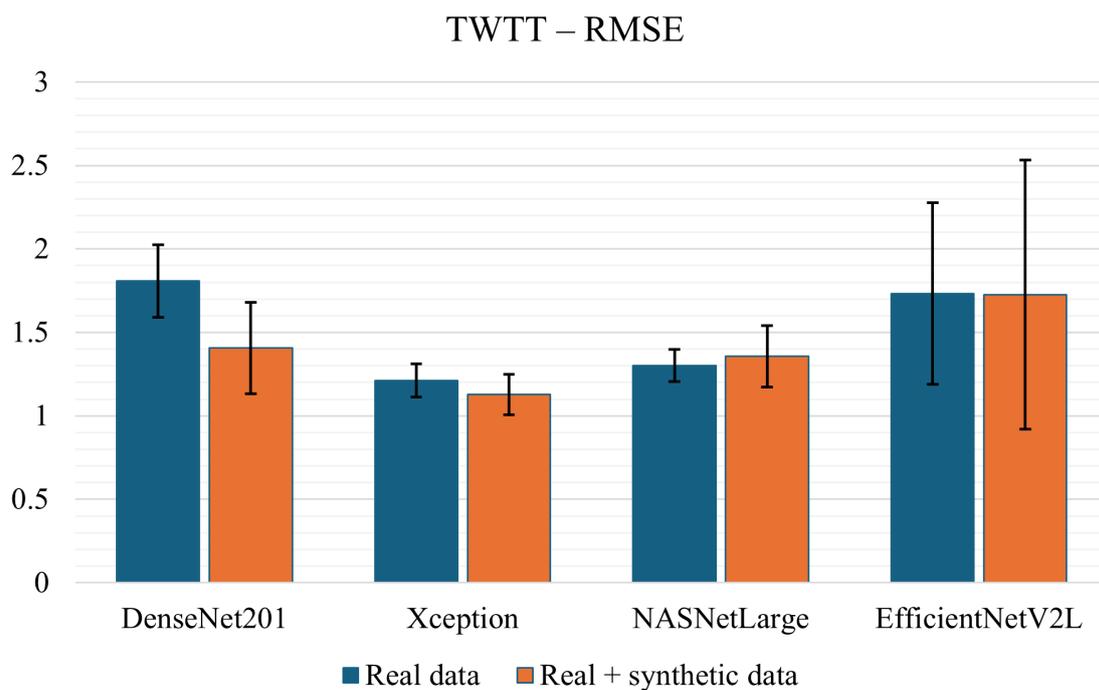| TWTT – Real + synthetic data			
Backbone	$R^2 \pm \sigma$	**MAE ± σ**	**RMSE ± σ**
DenseNet201	0.9505 ± 0.0180	1.1513 ± 0.2466	1.4055 ± 0.2746
Xception	0.9689 ± 0.0060	0.8670 ± 0.0937	1.1278 ± 0.1206
NASNetLarge	0.9550 ± 0.0108	1.0659 ± 0.1628	1.3565 ± 0.1830
EfficientNetV2L	0.9112 ± 0.0899	1.3656 ± 0.6857	1.7254 ± 0.8065

Burial depth – RMSE

Figure B.2 Performance comparison of different DL model architectures for burial depth estimation (expressed as RMSE) across real-world and combined (real + synthetic) B-scan data.

Table B.3 Performances of DL feature extraction backbones (DenseNet201, Xception, NASNetLarge, and EfficientNetV2L) in burial depth estimation from hyperbola-shaped reflections, utilizing only real-world data for model training. Performance measures are expressed as R^2, MAE, and RMSE (mean ± standard deviation) derived from 5-fold cross-validation.

Burial depth – Real data			
Backbone	**$R^2 \pm \sigma$**	**MAE $\pm \sigma$**	**RMSE $\pm \sigma$**
DenseNet201	0.9447 ± 0.0144	0.0525 ± 0.0072	0.0698 ± 0.0085
Xception	0.9544 ± 0.0089	0.0472 ± 0.0038	0.0636 ± 0.0066
NASNetLarge	0.9474 ± 0.0072	0.0507 ± 0.0039	0.0684 ± 0.0040
EfficientNetV2L	0.9440 ± 0.0147	0.0530 ± 0.0087	0.0701 ± 0.0106

Table B.4 Performances of DL feature extraction backbones (DenseNet201, Xception, NASNetLarge, and EfficientNetV2L) in burial depth estimation from hyperbola-shaped reflections, utilizing combined data for model training. Performance measures are expressed as R^2, MAE, and RMSE (mean ± standard deviation) derived from 5-fold cross-validation.

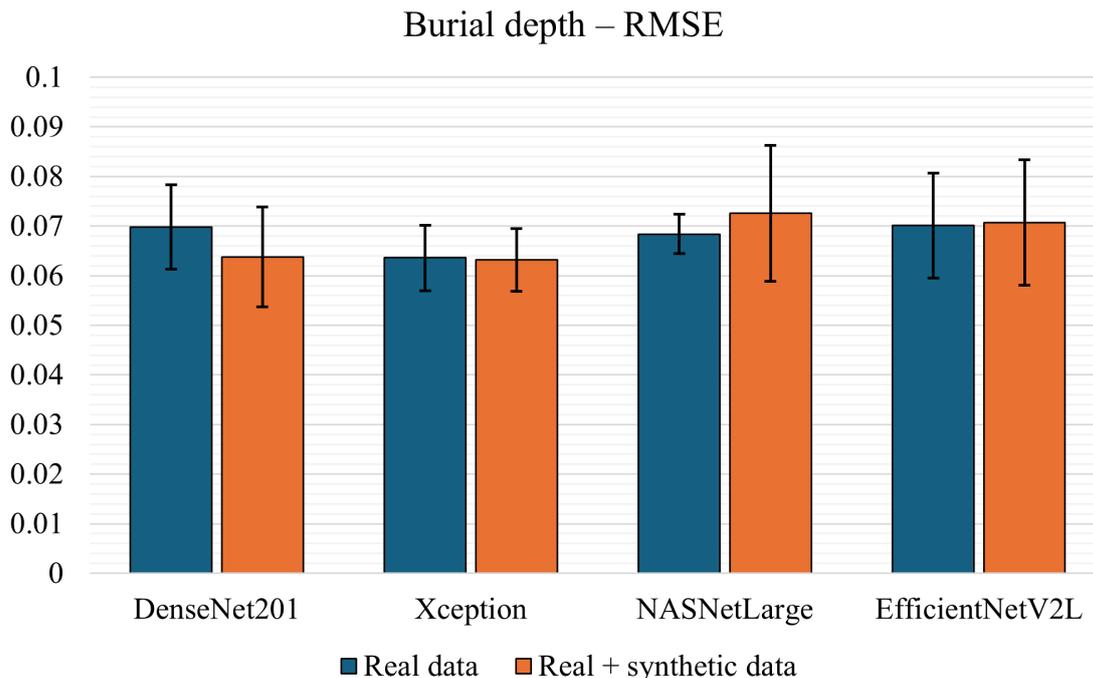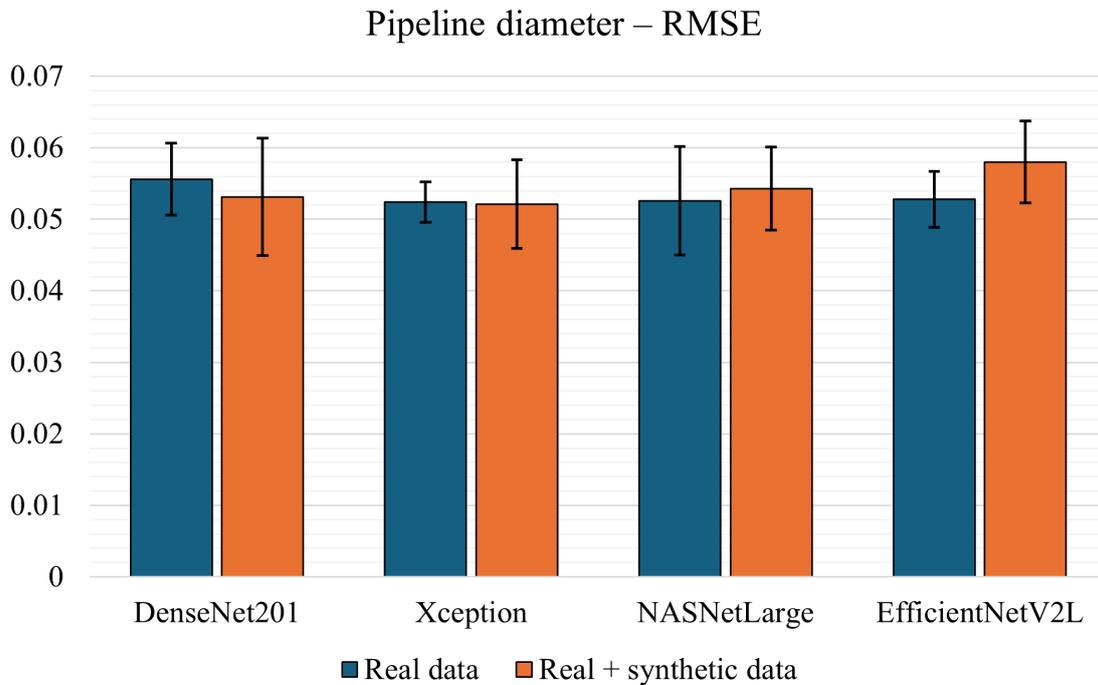| Burial depth – Real + synthetic data			
Backbone	**$R^2 \pm \sigma$**	**MAE ± σ**	**RMSE ± σ**
DenseNet201	0.9538 ± 0.0135	0.0464 ± 0.0066	0.0638 ± 0.0101
Xception	0.9551 ± 0.0077	0.0447 ± 0.0036	0.0632 ± 0.0063
NASNetLarge	0.9392 ± 0.0230	0.0535 ± 0.0090	0.0726 ± 0.0137
EfficientNetV2L	0.9423 ± 0.0213	0.0513 ± 0.0117	0.0707 ± 0.0126

Figure B.3 Performance comparison of different DL model architectures for pipeline diameter estimation (expressed as RMSE) across real-world and combined (real + synthetic) B-scan data.

Table B.5 Performances of DL feature extraction backbones (DenseNet201, Xception, NASNetLarge, and EfficientNetV2L) in pipeline diameter estimation from hyperbola-shaped reflections, utilizing only real-world data for model training. Performance measures are expressed as R^2, MAE, and RMSE (mean ± standard deviation) derived from 5-fold cross-validation.

Pipeline diameter – Real data			
Backbone	$R^2 \pm \sigma$	**MAE ± σ**	**RMSE ± σ**
DenseNet201	0.5319 ± 0.0403	0.0407 ± 0.0024	0.0556 ± 0.0050
Xception	0.5789 ± 0.0716	0.0365 ± 0.0031	0.0524 ± 0.0028
NASNetLarge	0.5802 ± 0.0875	0.0358 ± 0.0052	0.0526 ± 0.0076
EfficientNetV2L	0.5753 ± 0.0527	0.0369 ± 0.0045	0.0528 ± 0.0039

Table B.6 Performances of DL feature extraction backbones (DenseNet201, Xception, NASNetLarge, and EfficientNetV2L) in pipeline diameter estimation from hyperbola-shaped reflections, utilizing combined data for model training. Performance measures are expressed as R^2, MAE, and RMSE (mean ± standard deviation) derived from 5-fold cross-validation.

Pipeline diameter – Real + synthetic data			
Backbone	$R^2 \pm \sigma$	**MAE ± σ**	**RMSE ± σ**
DenseNet201	0.5702 ± 0.0963	0.0360 ± 0.0050	0.0531 ± 0.0082
Xception	0.5910 ± 0.0430	0.0362 ± 0.0031	0.0521 ± 0.0062
NASNetLarge	0.5520 ± 0.0681	0.0368 ± 0.0032	0.0543 ± 0.0058
EfficientNetV2L	0.4843 ± 0.0899	0.0405 ± 0.0046	0.0580 ± 0.0057

Angle (α) – RMSE
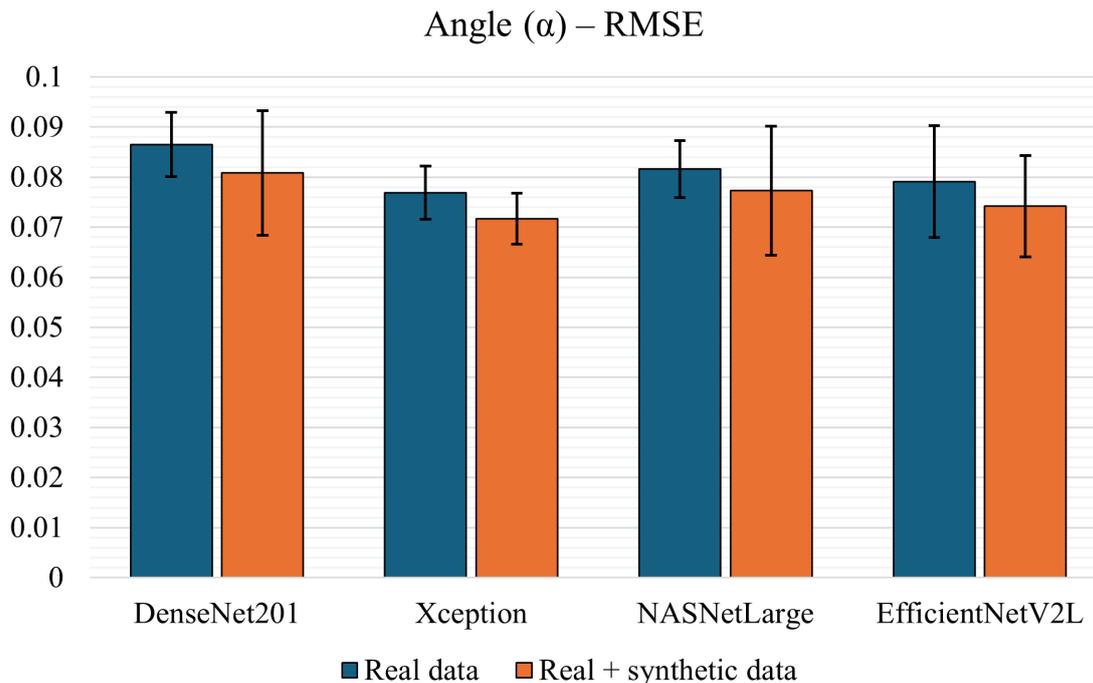
Figure B.4 Performance comparison of different DL model architectures for intersection angle estimation (expressed as RMSE) across real-world and combined (real + synthetic) B-scan data.

Table B.7 Performances of DL feature extraction backbones (DenseNet201, Xception, NASNetLarge, and EfficientNetV2L) in intersection angle estimation from hyperbola-shaped reflections, utilizing only real-world data for model training. Performance measures are expressed as R^2, MAE, and RMSE (mean ± standard deviation) derived from 5-fold cross-validation.

Angle ($α$) – Real data			
Backbone	**R^2 ± σ**	**MAE ± σ**	**RMSE ± σ**
DenseNet201	0.5914 ± 0.0279	0.0688 ± 0.0035	0.0865 ± 0.0064
Xception	0.6755 ± 0.0310	0.0577 ± 0.0056	0.0769 ± 0.0053
NASNetLarge	0.6356 ± 0.0261	0.0628 ± 0.0056	0.0816 ± 0.0057
EfficientNetV2L	0.6550 ± 0.0727	0.0620 ± 0.0101	0.0791 ± 0.0112

Table B.8 Performances of DL feature extraction backbones (DenseNet201, Xception, NASNetLarge, and EfficientNetV2L) in intersection angle estimation from hyperbola-shaped reflections, utilizing combined data for model training. Performance measures are expressed as R^2, MAE, and RMSE (mean ± standard deviation) derived from 5-fold cross-validation.

Angle (α) – Real + synthetic data			
Backbone	**$R^2 \pm \sigma$**	**MAE ± σ**	**RMSE ± σ**
DenseNet201	0.6409 ± 0.0808	0.0626 ± 0.0133	0.0808 ± 0.0124
Xception	0.7182 ± 0.0250	0.0495 ± 0.0062	0.0717 ± 0.0051
NASNetLarge	0.6678 ± 0.0976	0.0583 ± 0.0137	0.0773 ± 0.0129
EfficientNetV2L	0.6913 ± 0.0923	0.0536 ± 0.0089	0.0742 ± 0.0101

Appendix C
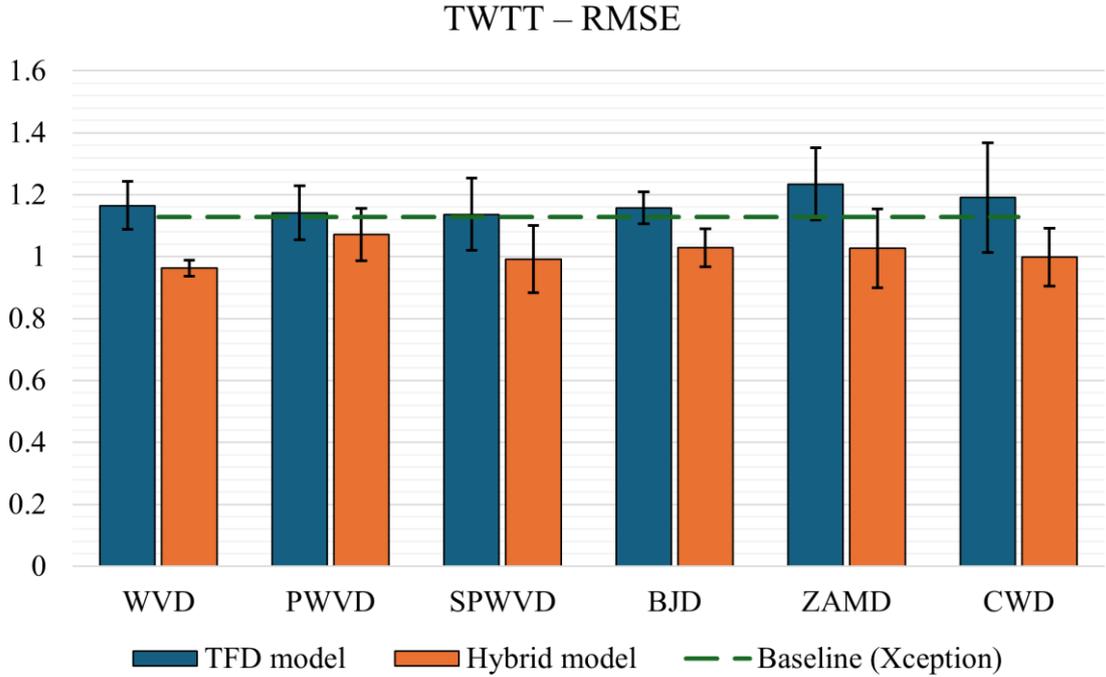
CASS-FUSION NETWORK RESULTS

TWTT – RMSE

Figure C.1 Performance comparison of TFD, hybrid, and baseline models for TWTT estimation (expressed as RMSE) across different TFDs.

Table C.1 Performance of the standalone TFD model for TWTT estimation from hyperbola-shaped reflections, utilizing various TFDs. Performance measures are expressed as R^2, MAE, and RMSE (mean ± standard deviation) derived from 5-fold cross-validation.

TWTT – TFD model			
TFD	**$R^2 \pm \sigma$**	**MAE $\pm \sigma$**	**RMSE $\pm \sigma$**
WVD	0.9671 ± 0.0034	0.8859 ± 0.0537	1.1651 ± 0.0774
PWVD	0.9683 ± 0.0045	0.8477 ± 0.0718	1.1413 ± 0.0868
SPWVD	0.9684 ± 0.0060	0.8550 ± 0.0768	1.1366 ± 0.1164
BJD	0.9675 ± 0.0023	0.8437 ± 0.0498	1.1576 ± 0.0521
ZAMD	0.9628 ± 0.0069	0.9124 ± 0.0844	1.2344 ± 0.1169
CWD	0.9648 ± 0.0116	0.8971 ± 0.1000	1.1907 ± 0.1766

Table C.2 Performance of the hybrid model for TWTT estimation from hyperbola-shaped reflections, utilizing various TFDs. Performance measures are expressed as R^2, MAE, and RMSE (mean ± standard deviation) derived from 5-fold cross-validation.

TWTT – hybrid model			
TFD	**$R^2 \pm \sigma$**	**MAE $\pm \sigma$**	**RMSE $\pm \sigma$**
WVD	0.9775 ± 0.0014	0.7470 ± 0.0176	0.9628 ± 0.0262
PWVD	0.9720 ± 0.0045	0.8039 ± 0.0731	1.0713 ± 0.0845
SPWVD	0.9759 ± 0.0054	0.7482 ± 0.0790	0.9920 ± 0.1080
BJD	0.9743 ± 0.0035	0.7749 ± 0.0696	1.0282 ± 0.0607
ZAMD	0.9741 ± 0.0063	0.7766 ± 0.0838	1.0265 ± 0.1271
CWD	0.9757 ± 0.0045	0.7615 ± 0.0734	0.9986 ± 0.0936

Figure C.2 Performance comparison of TFD, hybrid, and baseline models for burial depth estimation (expressed as RMSE) across different TFDs.

Table C.3 Performance of the standalone TFD model for burial depth estimation from hyperbola-shaped reflections, utilizing various TFDs. Performance measures are expressed as R^2, MAE, and RMSE (mean ± standard deviation) derived from 5-fold cross-validation.

Burial depth – TFD model			
TFD	**$R^2 \pm \sigma$**	**MAE ± σ**	**RMSE ± σ**
WVD	0.9603 ± 0.0024	0.0423 ± 0.0020	0.0596 ± 0.0026
PWVD	0.9588 ± 0.0045	0.0429 ± 0.0031	0.0607 ± 0.0039
SPWVD	0.9570 ± 0.0091	0.0450 ± 0.0046	0.0618 ± 0.0074
BJD	0.9576 ± 0.0039	0.0430 ± 0.0019	0.0616 ± 0.0037
ZAMD	0.9549 ± 0.0059	0.0453 ± 0.0035	0.0635 ± 0.0049
CWD	0.9530 ± 0.0078	0.0463 ± 0.0045	0.0647 ± 0.0059

Table C.4 Performance of the hybrid model for burial depth estimation from hyperbola-shaped reflections, utilizing various TFDs. Performance measures are expressed as R^2, MAE, and RMSE (mean ± standard deviation) derived from 5-fold cross-validation.

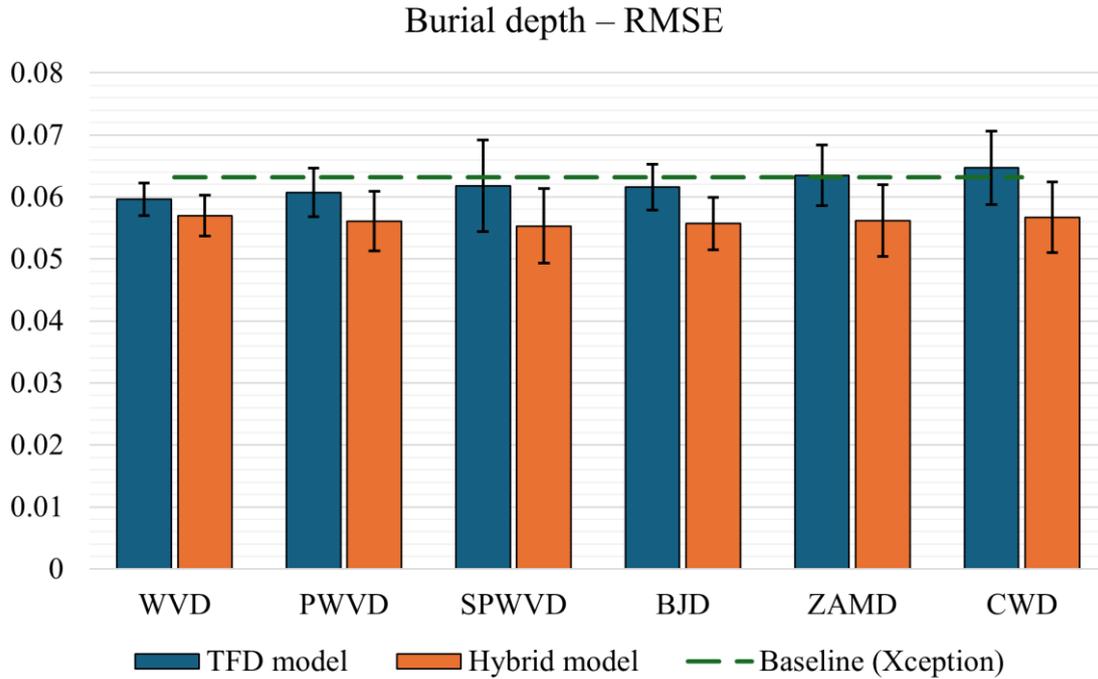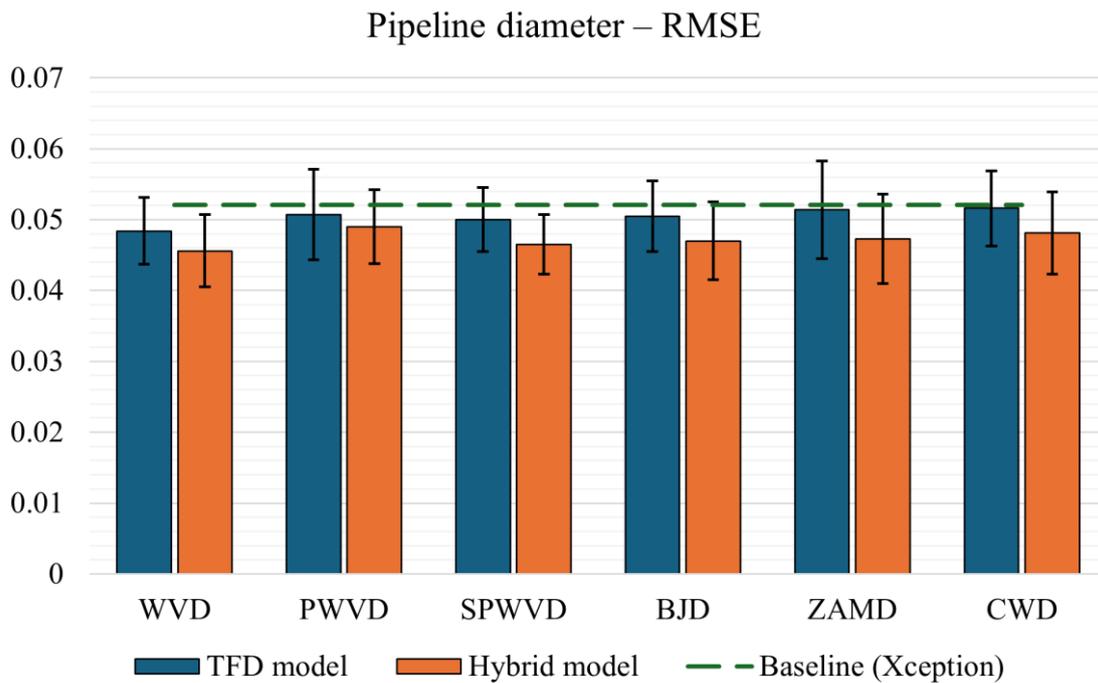| Burial depth – hybrid model			
TFD	**$R^2 \pm \sigma$**	**MAE $\pm \sigma$**	**RMSE $\pm \sigma$**
WVD	0.9637 ± 0.0032	0.0409 ± 0.0011	0.0570 ± 0.0033
PWVD	0.9647 ± 0.0051	0.0393 ± 0.0017	0.0561 ± 0.0048
SPWVD	0.9656 ± 0.0066	0.0394 ± 0.0042	0.0553 ± 0.0060
BJD	0.9653 ± 0.0043	0.0387 ± 0.0010	0.0557 ± 0.0042
ZAMD	0.9646 ± 0.0063	0.0396 ± 0.0018	0.0562 ± 0.0058
CWD	0.9639 ± 0.0063	0.0402 ± 0.0022	0.0567 ± 0.0057

Figure C.3 Performance comparison of TFD, hybrid, and baseline models for pipeline diameter estimation (expressed as RMSE) across different TFDs.

Table C.5 Performance of the standalone TFD model for pipeline diameter estimation from hyperbola-shaped reflections, utilizing various TFDs. Performance measures are expressed as R^2, MAE, and RMSE (mean ± standard deviation) derived from 5-fold cross-validation.

Pipeline diameter – TFD model			
TFD	**$R^2 \pm \sigma$**	**MAE $\pm \sigma$**	**RMSE $\pm \sigma$**
WVD	0.6454 ± 0.0377	0.0342 ± 0.0017	0.0484 ± 0.0047
PWVD	0.6120 ± 0.0495	0.0359 ± 0.0037	0.0507 ± 0.0064
SPWVD	0.6212 ± 0.0333	0.0356 ± 0.0028	0.0500 ± 0.0045
BJD	0.6155 ± 0.0296	0.0346 ± 0.0019	0.0505 ± 0.0050
ZAMD	0.5990 ± 0.0665	0.0351 ± 0.0033	0.0514 ± 0.0069
CWD	0.5983 ± 0.0301	0.0364 ± 0.0022	0.0516 ± 0.0053

Table C.6 Performance of the hybrid model for pipeline diameter estimation from hyperbola-shaped reflections, utilizing various TFDs. Performance measures are expressed as R^2, MAE, and RMSE (mean ± standard deviation) derived from 5-fold cross-validation.

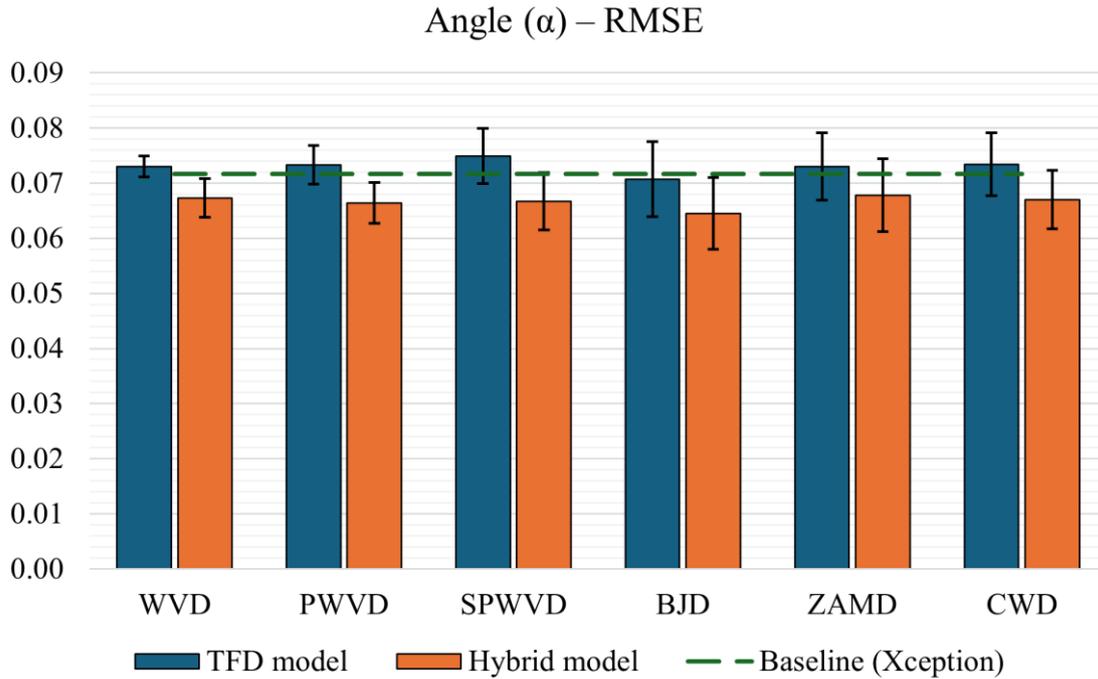| Pipeline diameter – hybrid model			
TFD	**$R^2 \pm \sigma$**	**MAE $\pm \sigma$**	**RMSE $\pm \sigma$**
WVD	0.6863 ± 0.0325	0.0312 ± 0.0020	0.0456 ± 0.0051
PWVD	0.6366 ± 0.0343	0.0328 ± 0.0007	0.0490 ± 0.0052
SPWVD	0.6726 ± 0.0174	0.0316 ± 0.0021	0.0465 ± 0.0042
BJD	0.6666 ± 0.0336	0.0314 ± 0.0022	0.0470 ± 0.0055
ZAMD	0.6615 ± 0.0497	0.0318 ± 0.0021	0.0473 ± 0.0063
CWD	0.6511 ± 0.0394	0.0322 ± 0.0020	0.0481 ± 0.0058

Figure C.4 Performance comparison of TFD, hybrid, and baseline models for intersection angle estimation (expressed as RMSE) across different TFDs.

Table C.7 Performance of the standalone TFD model for intersection angle estimation from hyperbola-shaped reflections, utilizing various TFDs. Performance measures are expressed as R^2, MAE, and RMSE (mean ± standard deviation) derived from 5-fold cross-validation.

Angle (α) – TFD model			
TFD	**$R^2 \pm \sigma$**	**MAE $\pm \sigma$**	**RMSE $\pm \sigma$**
WVD	0.7072 ± 0.0175	0.0480 ± 0.0026	0.0730 ± 0.0019
PWVD	0.7058 ± 0.0100	0.0496 ± 0.0032	0.0733 ± 0.0035
SPWVD	0.6924 ± 0.0231	0.0523 ± 0.0038	0.0749 ± 0.0050
BJD	0.7261 ± 0.0353	0.0481 ± 0.0045	0.0707 ± 0.0068
ZAMD	0.7084 ± 0.0283	0.0502 ± 0.0046	0.0730 ± 0.0061
CWD	0.7047 ± 0.0275	0.0501 ± 0.0047	0.0734 ± 0.0057

Table C.8 Performance of the hybrid model for intersection angle estimation from hyperbola-shaped reflections, utilizing various TFDs. Performance measures are expressed as R^2, MAE, and RMSE (mean ± standard deviation) derived from 5-fold cross-validation.

Angle (α) – hybrid model			
TFD	**$R^2 \pm \sigma$**	**MAE $\pm \sigma$**	**RMSE $\pm \sigma$**
WVD	0.7518 ± 0.0181	0.0451 ± 0.0028	0.0673 ± 0.0035
PWVD	0.7585 ± 0.0139	0.0457 ± 0.0023	0.0664 ± 0.0037
SPWVD	0.7563 ± 0.0244	0.0471 ± 0.0051	0.0667 ± 0.0052
BJD	0.7713 ± 0.0336	0.0431 ± 0.0047	0.0645 ± 0.0065
ZAMD	0.7483 ± 0.0307	0.0468 ± 0.0049	0.0678 ± 0.0066
CWD	0.7546 ± 0.0215	0.0451 ± 0.0031	0.0670 ± 0.0053

ACKNOWLEDGEMENTS

CURRICULUM VITAE

Daniel Štifanić, mag. ing. el., was born on May 29, 1994, in Pula, Croatia. He graduated from the Faculty of Engineering at the University of Rijeka in 2019, defending his master's thesis on artificial intelligence-based feature extraction methods for bladder cancer diagnosis. Following his graduation, he enrolled as a doctoral candidate in the field of technical sciences at the same institution. His doctoral research focuses on the automated analysis and processing of ground penetrating radar data utilizing advanced deep learning methodologies. He also actively participates in diverse scientific research related to expert systems and the application of artificial intelligence in smart cities, medicine, economics, and robotics. Furthermore, he has contributed to several high-impact projects, including the CEKOM SmartCity.4D intelligent infrastructure and the BrainClock study for early dementia detection. The results of his research have been published in several international peer-reviewed journals and presented at scientific conferences.

LIST OF PUBLICATIONS

Scientific papers published in journals

1. Štifanić, D., Štifanić, J., Anđelić, N., & Car, Z. (2025). An Automated System for Underground Pipeline Parameter Estimation from GPR Recordings. *Remote Sensing*, *17*(20), 3493.

2. Štifanić, J., Štifanić, D., Anđelić, N., & Car, Z. (2025). Explainable AI for oral cancer diagnosis: Multiclass classification of histopathology images and Grad-CAM visualization. *Biology*, *14*(8), 909.

3. Musulin, J., Štifanić, D., Zulijani, A., Ćabov, T., Dekanić, A., & Car, Z. (2021). An enhanced histopathology analysis: An ai-based system for multiclass grading of oral squamous cell carcinoma and segmenting of epithelial and stromal tissue. *Cancers*, *13*(8), 1784.

4. Štifanić, D., Musulin, J., Miočević, A., Baressi Šegota, S., Šubić, R., & Car, Z. (2020). Impact of COVID-19 on Forecasting Stock Prices: An Integration of Stationary Wavelet Transform and Bidirectional Long Short-Term Memory. *Complexity*, *2020*(1), 1846926.

5. Mrzljak, V., Štifanić, D., Meštrić, H., & Car, Z. (2022). Thermodynamic Analysis of CO2 Closed-Cycle Gas Turbine for Marine Applications at Various Pressure Ratios. *Tehnički vjesnik: znanstveno-stručni časopis tehničkih fakulteta Sveučilišta u Osijeku*, *29*(1), 51-59.

Conferences

1. Štifanić, D., Nadarević, T., Štifanić, J., Barković, I., & Car, Z. (2025, September). Semantic segmentation of chest computed tomography scans for quantifying pathologically altered lung parenchyma in patients with post-COVID syndrome. In *3rd International Conference on Chemo and BioInformatics (ICCBIKG)* (pp. 398-402).

2. Štifanić, D., Štifanić, J., & Car, Z. (2024, May). Impact of Cylindrical Object Diameter on Dielectric Constant in GPR Recordings. In *Serbian International Conference on Applied Artificial Intelligence* (pp. 158-165). Cham: Springer Nature Switzerland.

3. Štifanić, D., Štifanić, J., Šegota, S. B., Anđelić, N., & Car, Z. (2023, May). Material Classification of Underground Objects from GPR Recordings Using Deep Learning Approach. In *Serbian International Conference on Applied Artificial Intelligence* (pp. 149-158). Cham: Springer Nature Switzerland.

4. Musulin, J., Štifanić, D., Zulijani, A., Baressi Šegota, S., Anđelić, N., Lorencin, I., ... & Car, Z. (2022). Histopathological H&E-Stained Image Analysis Based on AI. In *1st Serbian International Conference on Applied Artificial Intelligence (SICAAI)*.

5. Musulin, J., Štifanić, D., Zulijani, A., Šegota, S. B., Lorencin, I., Anđelić, N., & Car, Z. (2021, October). Automated grading of oral squamous cell carcinoma into multiple classes using deep learning methods. In *2021 IEEE 21st international conference on bioinformatics and bioengineering (BIBE)* (pp. 1-6). IEEE.