

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**NOVEL DEEP LEARNING BASED
APPROACHES FOR CLUTTER REMOVAL
AND TARGET DETECTION IN GPR**



Ph.D. THESIS

Yavuz Emre KAYACAN

Department of Electronics and Telecommunication Engineering

Telecommunication Engineering Programme

FEBRUARY 2026

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**NOVEL DEEP LEARNING BASED
APPROACHES FOR CLUTTER REMOVAL
AND TARGET DETECTION IN GPR**

Ph.D. THESIS

**Yavuz Emre KAYACAN
(504192311)**

Department of Electronics and Telecommunication Engineering

Telecommunication Engineering Programme

Thesis Advisor: Prof. Dr. Işın ERER

FEBRUARY 2026

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**YNR'DE KARGAŞA GİDERME VE HEDEF ALGILAMA İÇİN
YENİ DERİN ÖĞRENME TABANLI YAKLAŞIMLAR**

DOKTORA TEZİ

**Yavuz Emre KAYACAN
(504192311)**

Elektronik ve Haberleşme Mühendisliği Anabilim Dalı

Telekomünikasyon Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Işın ERER

ŞUBAT 2026

Yavuz Emre KAYACAN, a Ph.D. student of ITU Graduate School student ID 504192311 successfully defended the thesis entitled “NOVEL DEEP LEARNING BASED APPROACHES FOR CLUTTER REMOVAL AND TARGET DETECTION IN GPR”, which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Prof. Dr. Işın ERER**
Istanbul Technical University

Jury Members : **Prof. Dr. Zümray ÖLMEZ**
Istanbul Technical University

Prof. Dr. Ender Mete EKŞİOĞLU
Istanbul Technical University

Asst. Prof. Dr. İpek ŞEN
Istanbul Bilgi University

Asst. Prof. Dr. Serdar KARGIN
Istanbul Arel University

Date of Submission : **02 January 2026**

Date of Defense **: 02 February 2026**





To my family,



FOREWORD

This doctoral thesis, which I pursued with great effort, patience, and dedication throughout my studies, represents one of the most important and instructive turning points in my academic life. This work, aimed at developing new deep learning-based approaches for processing Ground Penetrating Radar (GPR) data, went beyond being only a technical research project and played an important role in shaping my scientific thinking.

I would like to express my sincere gratitude to my doctoral advisor, Prof. Dr. Işın ERER, whose expertise, vision, and guidance were invaluable through this process. Her constructive feedback and unwavering support, both academically and personally, greatly enriched my experience. It was an honor to be her student.

I thank the thesis committee and defense jury members, Prof. Dr. Zümray ÖLMEZ and Asst. Prof. Dr. İpek ŞEN, for their time, constructive feedback, and valuable suggestions, which significantly improved the quality of this thesis.

I am especially thankful to Asst. Prof. Dr. Serdar KARGIN, as well as my professors and colleagues at Istanbul Technical University and Beykent University, for generously sharing their knowledge and experience, and for their consistent support throughout my academic studies.

I am deeply grateful to my wife, Emine SIRMALI KAYACAN, whose patience, love, and unwavering support have been my greatest motivation throughout this journey.

Finally, I thank my mother, father, and brother for their unwavering support, love, and belief in me. Their encouragement has been my greatest strength in completing this work. I hope this work will shed light on the field of GPR and deep learning for researchers and contribute to the scientific community.

FEBRUARY 2026

Yavuz Emre KAYACAN

TABLE OF CONTENTS

	Page
FOREWORD	ix
TABLE OF CONTENTS	xi
ABBREVIATIONS	xv
LIST OF TABLES	xvii
LIST OF FIGURES	xxi
SUMMARY	xxiii
ÖZET	xxv
1. INTRODUCTION	1
1.1 Problem Definition and Thesis Objective	1
1.2 Proposed Methods	2
1.3 Literature Review	4
2. BACKGROUND ON GROUND PENETRATING RADAR	11
2.1 Overview of the GPR System	11
2.2 Dataset Collection	16
2.3 Performance Evaluation Metrics	17
3. PROPOSED METHODS FOR CLUTTER REMOVAL IN GPR DATA ...	23
3.1 Proposed Unsupervised Model	23
3.1.1 Autoencoder-guided low-rank approximation (AE-LRA)	23
3.1.1.1 Experimental results on hybrid data	26
3.1.1.2 Experimental results on real-world data	27
3.2 Proposed Supervised Models	29
3.2.1 Declutter vision transformer (dc-vit)	29
3.2.1.1 Patching, patch embedding, and position embedding	32
3.2.1.2 Dense transformer block	32
3.2.1.3 Reconstruction block	33
3.2.2 Declutter vision mamba (dc-mamba)	34
3.2.2.1 Dynamic patch embedding	37
3.2.2.2 Tensor-train linear layer	37
3.2.2.3 Efficient dense block	38
3.2.2.4 EVSS block and group	39
3.2.2.5 Inverted residual block and group	40
3.2.2.6 Lightweight reconstruction block	41
3.2.3 Loss function	42
3.2.4 Experimental results of supervised models	43
3.2.4.1 Selection of model parameters	43
3.2.4.2 Experimental results on hybrid data	44
3.2.4.3 Experimental results on real-world data	44
4. CLUTTER REMOVAL AND TARGET DETECTION IN GPR DATA USING THE PROPOSED JOINT MODELS (DEC-YOLO)& RAFDEC-YOLO)	49
4.1 Proposed Joint Models	49
4.1.1 Model I: dec-yolo	49
4.1.2 Model II: rafdec-yolo	50
4.1.2.1 Shared backbone: multi-scale feature extraction	51
4.1.2.2 Clutter removal branch: decoder architecture	52
4.1.2.3 Attention-based up-sampling block	53
4.1.2.4 Feature integration: residual adapter fusion	54
4.1.3 Loss function	55
4.2 Experimental Results of Joint Models	56
4.2.1 Selection of model parameters	57
4.2.2 Results on hybrid test data	57

4.2.3 Results on real test data.....	62
5. CONCLUSION	69
REFERENCES	71
CURRICULUM VITAE	77







ABBREVIATIONS

AbUsB	: Attention-Based Up-Sampling Block
AE	: Autoencoder
CAE	: Convolutional Autoencoder
CAM	: Channel Attention Module
cGAN	: Conditional Generative Adversarial Network
CNN	: Convolutional Neural Network
CR-Net	: Clutter Removal Network
DCAE	: Deep Convolutional Autoencoder
DC-ViT	: Declutter Vision Transformer
DRMF	: Direct Robust Matrix Factorization
DWT	: Discrete Wavelet Transform
ESRTMDet	: Enhanced Super-Resolution and Target-Detection
FDTD	: Finite-Difference Time-Domain
FGSR	: Factor Group-Sparse Regularization
FLOPs	: Floating Point Operations
GAN	: Generative Adversarial Network
GoDec	: Go Decomposition
GPR	: Ground Penetrating Radar
LeFF	: Locally-enhanced Feed-Forward
LRSD	: Low-Rank Sparse Decomposition
mAP	: Mean Average Precision
MCA	: Morphological Component Analysis
MLP	: Multilayer Perceptron
NMF	: Non-Negative Matrix Factorization
PCA	: Principal Component Analysis
RAE	: Robust Autoencoder
RAF	: Residual Adapter Fusion
R-CNN	: Region-based Convolutional Neural Network
ReLU	: Rectified Linear Unit
RNMF	: Robust Non-Negative Matrix Factorization
ROSL	: Robust Orthonormal Subspace Learning
RPCA	: Robust Principal Component Analysis
RPN	: Region Proposal Network
SAM	: Spatial Attention Module
SAR	: Synthetic Aperture Radar
SNR	: Signal-to-Noise Ratio
SOD	: Small Object Detection
SSM	: State Space Models
SVD	: Singular Value Decomposition
ViT	: Vision Transformer
YOLO	: You Only Look Once



LIST OF TABLES

	<u>Page</u>
Table 2.1: Summary of experimental scenarios used in laboratory measurements.	18
Table 3.1: PSNR and MS-SSIM Results for Hybrid Dataset.	26
Table 3.2: AE-LRA result for different λ values.	27
Table 3.3: Performance Comparison of Evaluated Models.	44
Table 3.4: Comparative Analysis of SCR Improvement (dB) for Different Scenarios.	45
Table 4.1: Target detection comparison results using the hybrid dataset (Higher value is better).	60
Table 4.2: Model complexity metrics (FLOPs, Parameters) (Lower value is better ($\downarrow$)). Best results are in bold.	61
Table 4.3: Target detection comparison results using the laboratory measurements (Higher value is better).	66



LIST OF FIGURES

	<u>Page</u>
Figure 2.1: A-scan signal acquired from the GPR system.	13
Figure 2.2: B-scan image reconstructed from data acquired by the GPR system.	13
Figure 2.3: C-scan image reconstructed from data acquired by the GPR system.	14
Figure 2.4: Laboratory setup and equipment used for GPR measurements.	17
Figure 2.5: Experimental scenarios used in laboratory measurements: a) Scenario 1: Mobile phone and water-filled PET bottle. b) Scenario 2: Mobile phone and cylindrical rebar. c) Scenario 3: Targets buried under an asphalt layer to simulate complex surface clutter.	18
Figure 3.1: Architecture of an autoencoder.	25
Figure 3.2: Comparison results for a sample GPR image from the hybrid dataset. (a) Raw data, (b) Ground-truth, clutter removal results obtained by (c) SVD, (d) PCA, (e) NMF, (f) RPCA, (g) RNMF, (h) RAE, and (i) AE-LRA.	27
Figure 3.3: Comparison results for real data for a copper rod at a depth of 10 cm in sand (a) Raw data, clutter removal results obtained by (b) SVD, (c) PCA, (d) NMF, (e) RPCA, (f) RNMF, (g) RAE, and (h) proposed AE methods.	28
Figure 3.4: Comparison results for real data for a plastic bottle at a depth of 6 cm in sand (a) Raw data, clutter removal results obtained by (b) SVD, (c) PCA, (d) NMF, (e) RPCA, (f) RNMF, (g) RAE, and (h) AE.	28
Figure 3.5: Architecture of the DC-ViT. Patching & Patch Embedding: Focuses on local features and captures key information. Position Embedding: Adds location details. Dense Transformers Block: Distinguishes targets from clutter. Reconstruction Block: Refines the target image.	31
Figure 3.6: Proposed DC-MAMBA architecture. (a) The overall network structure illustrates the "Global to Local" hierarchy, where input data is processed by a Convolutional Patch Embedding module, followed by a series of Efficient Dense Blocks (comprising EVSS and Inverted Residual Groups), and finally a Reconstruction Block. (b) Detailed schematic of the Effective Visual State Space (EVSS) Block integrating the Global Branch (TensorSS2D) for long-range dependency modeling and the Local Branch (CNN) for fine detail extraction. (c) Structure of the Inverted Residual Block providing local feature enhancement using an expand-filter-compress strategy. (d) The SE Block is used to perform dynamic channel-based feature calibration within encoder modules.	36

Figure 3.7: Scenario 1 (S1). The sub-figures correspond to: (a) Raw Data, (b) CR-Net, (c) Attention U-Net, (d) DC-ViT, (e) DC-MAMBA V1, and (f) DC-MAMBA V2. The corresponding SCR signal plots are presented below each B-scan image.	46
Figure 3.8: Scenario 2 (S2). The sub-figures correspond to: (a) Raw Data, (b) CR-Net, (c) Attention U-Net, (d) DC-ViT, (e) DC-MAMBA V1, and (f) DC-MAMBA V2. The corresponding SCR signal plots are presented below each B-scan image.	47
Figure 3.9: Scenario 3 (S3). The sub-figures correspond to: (a) Raw Data, (b) CR-Net, (c) Attention U-Net, (d) DC-ViT, (e) DC-MAMBA V1, and (f) DC-MAMBA V2. The corresponding SCR signal plots are presented below each B-scan image.	48
Figure 4.1: Workflow of the proposed DeC-YOLO. A raw GPR image is first fed into the Clutter Removal Model, which produces a processed image. This processed image is then processed by the Target Detection Module (e.g., YOLO) to identify and localize buried targets. Both models are trained end-to-end within a joint architecture, enabling parallel optimization of the two tasks to enhance accuracy.....	51
Figure 4.2: The overall architecture of the proposed RAFDeC-YOLO framework. The model features a streamlined design where the Clutter Removal Branch directly leverages backbone features, and the Detection Branch is enhanced via Residual Adapter Fusion modules.....	52
Figure 4.3: Detailed architectures of the proposed sub-modules: a) The Attention-Based Up-Sampling Block (AbUsB) is utilized in the decoder branch. It mitigates noise propagation by filtering the skip-connection features (x_{skip}) with an Attention Gate before concatenating them with the upsampled features (x_{up}) from the preceding layer. b) Internal schematic of the Attention Gate mechanism. This module computes a spatial attention map α by processing the gating signal (x_{dec}) and input features (x_{skip}), suppressing clutter-related activations through element-wise multiplication. c) The Residual Adapter Fusion module is integrated into the detection neck. Instead of complex gating, this streamlined mechanism aligns the channel dimensions of the decoder features ($F_{decoder}$) using a 1×1 convolution and Batch Normalization, then fuses them with the raw neck features (F_{neck}) via element-wise addition to preserve gradient flow and feature magnitude.	53

- Figure 4.4:** Qualitative detection results on Hybrid Dataset (Sample 1) (Moderate Clutter). (a) Raw Input, (b) Ground Truth. **Sequential Models with:** (c) UNet, (d) CR-Net, (e) Att-UNet, (f) DC-ViT. **Proposed Model I** DeC-YOLO with (g) UNet, (h) CR-Net, (i) Att-UNet, (j) DC-ViT structure as clutter removal module. **SOTA Detectors:** (k) SuperYOLO, (l) ESRTMDet, (m) SOD. **Proposed Model II** (n) RAFDeC-YOLO12n, (o) RAFDeC-YOLO12s, (p) RAFDeC-YOLO12m. The proposed methods (n-p) exhibit the highest confidence and localization precision. 58
- Figure 4.5:** Qualitative detection results on Hybrid Dataset (Sample 2) (Heavy Clutter / Weak Targets). (a) Raw Input, (b) Ground Truth. **Sequential Models with:** (c) UNet, (d) CR-Net, (e) Att-UNet, (f) DC-ViT. **Proposed Model I** DeC-YOLO with (g) UNet, (h) CR-Net, (i) Att-UNet, (j) DC-ViT structure as clutter removal module. **SOTA Detectors:** (k) SuperYOLO, (l) ESRTMDet, (m) SOD. **Proposed Model II** (n) RAFDeC-YOLO12n, (o) RAFDeC-YOLO12s, (p) RAFDeC-YOLO12m. While many baselines fail to detect the weak target on the right, RAFDeC-YOLO12m (p) successfully localizes it with high confidence. 59
- Figure 4.6:** Detection results on Real Scenario 1 (Mobile Phone & PET Bottle). (a) Raw Input. **Sequential Models with:** (b) UNet, (c) CR-Net, (d) Att-UNet, (e) DC-ViT. **Proposed Model I** with (f) UNet, (g) CR-Net, (h) Att-UNet, (i) DC-ViT as clutter removal module. **SOTA:** (j) SuperYOLO, (k) ESRTMDet, (l) SOD. **Proposed Model II:** (m) RAFDeC-YOLO12n, (n) RAFDeC-YOLO12s, (o) RAFDeC-YOLO12m. Note that the proposed models (m-o) successfully detect the weak dielectric PET bottle. 63
- Figure 4.7:** Detection results on Real Scenario 2 (Mobile Phone & Rebar). (a) Raw Input. **Sequential Models with:** (b) UNet, (c) CR-Net, (d) Att-UNet, (e) DC-ViT. **Proposed Model I** with (f) UNet, (g) CR-Net, (h) Att-UNet, (i) DC-ViT as clutter removal module. **SOTA:** (j) SuperYOLO, (k) ESRTMDet, (l) SOD. **Proposed Model II:** (m) RAFDeC-YOLO12n, (n) RAFDeC-YOLO12s, (o) RAFDeC-YOLO12m. The proposed method effectively separates the strong pipe reflection from surface clutter. 64
- Figure 4.8:** Detection results on Real Scenario 3 (Asphalt Surface). (a) Raw Input. **Sequential Models with:** (b) UNet, (c) CR-Net, (d) Att-UNet, (e) DC-ViT. **Proposed Model I** with (f) UNet, (g) CR-Net, (h) Att-UNet, (i) DC-ViT as clutter removal module. **SOTA:** (j) SuperYOLO, (k) ESRTMDet, (l) SOD. **Proposed Model II:** (m) RAFDeC-YOLO12n, (n) RAFDeC-YOLO12s, (o) RAFDeC-YOLO12m. Despite heavy clutter caused by the asphalt layer, RAFDeC-YOLO12m (o) successfully identifies the buried targets..... 65



NOVEL DEEP LEARNING BASED APPROACHES FOR CLUTTER REMOVAL AND TARGET DETECTION IN GPR

SUMMARY

Ground Penetrating Radar (GPR) is a non-destructive geophysical tool that uses electromagnetic waves to create detailed images of shallow underground structures and buried objects.

The basic operating principle of this system is based on the interaction of waves emitted by the transmitter antenna with materials in the subsurface that have different dielectric properties (permittivity) and conductivity. Signals returning from different layers or objects are recorded by the receiver antenna and processed in the time-depth plane. Particularly in B-scan images obtained by moving the antenna along a line, point targets (e.g., pipes, mines, or archaeological remains) appear as characteristic hyperbolic (inverted U) signatures resulting from the change in distance as the antenna approaches and moves away from the target.

However, the raw GPR data obtained in practice is quite different from ideal theoretical models. Direct interaction between antennas (crosstalk), strong surface reflections at the air-soil boundary, scattering caused by inhomogeneous ground structure, and multiple reflections caused by the signal echoing between layers create intense noise components referred to in the literature as “clutter.” These noises generally mask weak signals returning from the target, reducing image quality and making target detection a serious signal processing and engineering problem.

To solve this problem, statistical subspace methods such as Singular Value Decomposition (SVD) and Principal Component Analysis (PCA) have been used in the literature for many years. However, these linear methods are insufficient for separating the target signal from noise in complex and variable ground conditions and often lead to the loss of target data. Moreover, although advanced techniques such as Low-Rank Sparse Decomposition (LRSD) offer improvements, they suffer from a critical dependency on the manual tuning of regularization parameters. Although Convolutional Neural Networks (CNNs) have become popular with the advent of deep learning, the “limited receptive field” problem of these architectures prevents the capture of long-range (global) context spread across the entire image, such as hyperbolic reflections in GPR radargrams. Furthermore, most existing studies treat the processes of clutter removal and target detection as separate, sequential blocks. In this approach, even the slightest data corruption during the cleaning phase can cause the target to be completely missed during the detection phase.

This thesis contributes four original and innovative architectures to the literature to overcome the aforementioned methodological and structural limitations. First, an Autoencoder Guided Low-Rank Approximation (AE-LRA) approach is introduced to

address the parameter dependency of classical LRSD methods by automatically learning the low-rank clutter distribution from the data. Second, DC-ViT (Densely Connected Vision Transformer), based on the Vision Transformer architecture, was developed to overcome the local focus limitation of CNNs. This model uses the “Self-Attention” mechanism to model the relationship between information at any point in the image and all other points, capturing the global context. At the same time, thanks to the densely connected connections and “Locally-enhanced Feed-Forward” (LeFF) blocks integrated into the structure, the preservation of local details, which Transformers are generally weak at, is ensured.

Third, the DC-Mamba architecture is presented as a solution to the quadratic computational load problem that Transformer models create with high-resolution data. This model adapts the “State Space Models” (Mamba) architecture, which can model long-range dependencies in data with linear computational complexity, to the GPR domain. In particular, the Tensor-Train (TT) decomposition technology used in the model’s V2 variant radically reduces the model’s number of parameters and memory usage, enabling it to operate with high efficiency on embedded systems with limited processing power (FPGA, mobile processors, etc.).

The fourth and most comprehensive solution, RAFDEC-YOLO, is an end-to-end “Joint Learning” architecture that combines clutter removal and target detection tasks in a single network. The “Residual Adapter Fusion” (RAF) mechanism developed in this model transfers noise-free feature maps obtained from cleaning layers to the detection neck without disrupting the gradient flow of the main detection network. Thus, the network learns not only to visually improve the image but also to highlight features that will maximize detection success.

The performance of the developed models was validated using extensive synthetic hybrid datasets and sandbox experiments conducted under challenging laboratory conditions. Comparative experiments indicated that the proposed AE-LRA method achieved greater Signal-to-Clutter Ratio (SCR) improvement than RPCA and NMF methods, without the need for manual parameter adjustment. The DC-Mamba architecture also delivered superior image quality and processing efficiency, producing the cleanest B-scan images, especially in conditions with intense signal attenuation, such as beneath asphalt, and high surface complexity. On the target detection side, RAFDEC-YOLO detected weak targets with high reliability even in data where sequential methods completely failed, and targets were lost in noise. Furthermore, thanks to the integrated training strategy, it was shown that the model’s bounding boxes encircle targets much more tightly and accurately.

This study shows that joint models considering the global context in GPR are efficient and can share information between tasks. As a result, they deliver results that are much better, more reliable, and more practical than traditional methods.

YNR'DE KARGAŞA GİDERME VE HEDEF ALGILAMA İÇİN YENİ DERİN ÖĞRENME TABANLI YAKLAŞIMLAR

ÖZET

Yere Nüfuzlu Radar (YNR), elektromanyetik dalga teorisine dayanan ve sığ yeraltı yapılarının görüntülenmesinde kullanılan ileri teknoloji bir jeofizik yöntemdir. Sisteme ait temel çalışma prensibi, verici anten tarafından gönderilen yüksek frekanslı (10 MHz – 3 GHz) elektromanyetik dalgaların, yeraltındaki farklı dielektrik özelliklere sahip materyallerle, yani toprak katmanları, boşluklar veya gömülü nesnelerle etkileşime girmesi ve bir kısmının yüzeye geri yansması esasına dayanır. Geri dönen sinyalin alıcı anten tarafından kaydedilmesiyle, hedefin derinliği ve konumu hakkında bilgi edinilir. Bu süreçte sinyalin hızı, ortamın dielektrik sabitine (ϵ_r) doğrudan bağlıdır; örneğin suyun yüksek dielektrik sabiti sinyali yavaşlatırken, kuru kumda sinyal daha hızlı ilerler. YNR verileri genellikle üç farklı formatta görselleştirilir: Tek bir noktadan alınan derinlik-genlik bilgisi olan A-tarama, antenin bir hat boyunca hareketiyle elde edilen ve yeraltının kesitsel görüntüsünü sunan B-tarama ve paralel B-taramaların birleştirilmesiyle oluşturulan hacimsel C-tarama. Bu tezde odaklanılan B-tarama görüntülerinde, nokta hedefler (örneğin borular veya mayınlar), antenin hedefe yaklaşıp uzaklaşması sırasında oluşan mesafe değişiminden dolayı karakteristik bir hiperbol, yani ters U şeklinde görülür. Ancak, ideal bir YNR görüntüsü elde etmek pratikte oldukça zordur çünkü alınan ham sinyal, sadece hedeften gelen yansımayı değil, aynı zamanda Kargaşa olarak adlandırılan yoğun gürültü bileşenlerini de içerir. Literatürde bu kargaşa sinyalleri; verici ve alıcı anten arasındaki doğrudan etkileşim ile hava-toprak sınırındaki güçlü yansımaların oluşturduğu ve hedefi maskeleyen yatay bantlar olan rırken zaman kargaşa sinyalleri, toprak içindeki taşlar ve kökler gibi heterojen yapıların neden olduğu Saçılma ve sinyalin katmanlar arasında gidip gelmesiyle oluşan Çoklu Yansımalar olmak üzere üç ana kategoride incelenir. Bu kargaşa sinyalleri enerjisi genellikle hedef sinyalinden çok daha yüksek olduğu için, zayıf hedef yansımalarının tespit edilmesi ve yorumlanması ciddi bir mühendislik problemi haline gelmektedir.

YNR veri işleme literatürü, basit sinyal işleme tekniklerinden karmaşık yapay zeka modellerine doğru evrilmiştir. İlk dönemlerde kullanılan Ortalama Çıkarma gibi yöntemler, durağan yatay bantları kaldırmada etkili olsa da, eğimli veya düzensiz yüzeylerde yetersiz kalmıştır. İstatistiksel yöntemlerin yükselişiyle birlikte, Tekil Değer Ayırımı (TDA) ve Temel Bileşen Analizi (TBA) gibi alt uzay yöntemleri popüler hale gelmiştir. Bu yaklaşımlar, veriyi en baskın bileşenlere ve daha zayıf bileşenlere ayırarak temizlemeyi amaçlasa da, bu lineer yöntemler karmaşık ve lineer olmayan toprak yapılarında hedef sinyalinin bir kısmını gürültü kabul ederek silme riski taşır. Daha ileri seviye olan Gürbüz Temel Bileşen Analizi (GTBA) ve Düşük Sıralı Seyrek Ayırıştırma (DSSA) yöntemleri, kargaşa sinyalleri matrisin "düşük sıralı", hedefi ise "seyrek" bileşeni olarak modelleyerek başarıyı artırmıştır. Buna rağmen,

bu yöntemlerin performansı matematiksel parametrelerin manuel olarak çok hassas ayarlanmasına bağlıdır ve işlem yükleri yüksektir. Bu noktada, derin öğrenme temelli Otokodlayıcı yaklaşımları, kargaşanın düşük dereceli yapısını parametre ayarına gerek duymadan öğrenerek umut verici bir alternatif sunmaktadır. Derin öğrenme devrimiyle birlikte, Konvolüsyonel Otomatik Kodlayıcılar (KOK) ve Çekişmeli Üretken Ağ (ÇÜA) mimarileri YNR alanına uyarlanmıştır. Bu modeller, veri dağılımını öğrenerek daha başarılı temizleme yapsalar da, temel bir yapısal kısıtla karşılaşmışlardır: CNN'lerin doğası gereği sahip olduğu "yerel alıcı alan", piksel komşuluklarına odaklanır; oysa YNR görüntülerindeki hiperbolik yansımalar ve katmanlı yapılar, görüntünün tamamına yayılan uzun menzilli (global) uzamsal bağımlılıklar içerir. Standart CNN'ler bu global bağlamı modellemede yetersiz kalmaktadır. Ayrıca, mevcut literatürdeki çalışmalar genellikle kargaşa giderme ve hedef tespiti birbirinden kopuk iki ayrı süreç olarak ele alır. Bu ardışık yaklaşımda, temizleme algoritması tespit aşımının ihtiyaçlarından habersizdir ve temizleme sırasında oluşan küçük bir bozulma, tespit aşımında hedefin tamamen kaçırılmasına yol açabilir.

Bu tez çalışması, belirtilen kısıtları aşmak için dört temel yenilikçi yaklaşım önermektedir. İlk aşamada, klasik matris ayrıştırma yöntemlerinin parametre bağımlılığını ortadan kaldırmak amacıyla Otokodlayıcı Güdümlü Düşük Dereceli Yaklaşım (AE-LRA) yöntemi sunulmuştur. Bu yöntem, verideki baskın kargaşa bileşenini bir otokodlayıcı ağı ile öğrenerek ve parametre ayarı gerektirmeyen gürbüz bir temizleme sağlar. İlk olarak, CNN'lerin global bağlam eksikliğini gidermek için Vision Transformer (ViT) mimarisini temel alan DC-ViT modeli geliştirilmiştir. DC-ViT, görüntüyü küçük yamalara bölerek ve "Self-Attention" mekanizmasını kullanarak, görüntüdeki her bir noktanın diğer tüm noktalarla ilişkisini hesaplar; böylece kargaşa sinyalleri ile hedef hiperbolleri arasındaki global yapısal farklar öğrenilir. Modelde, standart ViT'ten farklı olarak katmanlar arasında yoğun bağlantılar kurularak bilgi akışı güçlendirilmiş ve Transformer'ların zayıf olduğu yerel detayları yakalamak için LeFF blokları entegre edilmiştir.

İkinci olarak, Transformer modellerinin yüksek işlem yükü sorununa çözüm olarak, Mamba (State Space Models) mimarisi DC-Mamba adıyla ilk kez YNR alanına uyarlanmıştır. Mamba mimarisi, veriyi bir dizi olarak işleyerek uzun menzilli bağımlılıkları lineer hesaplama karmaşıklığıyla modelleyebilir, bu da yüksek çözünürlüklü YNR verilerinin çok daha hızlı işlenmesini sağlar. Model, global özellikleri öğrenen EVSS blokları ile yerel detayları işleyen Ters Residual bloklarının birleşiminden oluşan hiyerarşik bir yapıya sahiptir. Ayrıca, tezin en çarpıcı teknik katkılarından biri olan Tensor-Train (TT) teknolojisinin modele entegre edildiği V2 varyantı, performanstan ödün vermeden parametre sayısını ve bellek kullanımını radikal bir şekilde düşürerek gömülü sistemlere uygun hale getirilmiştir. Özellikle V2 varyantı, DC-ViT mimarisine kıyasla parametre sayısında %74 ve işlem yükünde %71 oranında sağladığı düşüşle, donanım kısıtı olan cihazlarda yüksek performans vaat etmektedir.

Üçüncü ve en kapsamlı model olan RAFDEC-YOLO ise kargaşa giderme ve hedef tespiti görevlerini tek bir ağda birleştirir. Model, YOLOv12 tabanlı bir omurga, kargaşa temizleyen bir kod çözücü ve tespit yapan bir kafa yapısından oluşur. Bu mimaride kullanılan "Residual Adapter Fusion (RAF)" mekanizması, kargaşa giderme dalından elde edilen temizlenmiş özellik haritalarını ana tespit ağının boyun bölgesine eklemeli

bir şekilde aktarır. Ayrıca, temizleme dalında kullanılan dikkat kapılı yukarı örnekleme blokları gürültüyü filtrelerken, ortak kayıp fonksiyonu sayesinde ağ uçtan uca eğitilerek temizleme modülünün tespit başarısını artıracak özellikleri belirginleştirmesi sağlanır.

Geliştirilen modellerin başarısı, öncelikle sentetik ve gerçek verilerin birleşiminden oluşan hibrit veri seti üzerinde, ardından zorlu laboratuvar koşullarında gerçekleştirilen kum havuzu deneyleriyle doğrulanmıştır. Yapılan karşılaştırmalı testlerde, önerilen AE-LRA yönteminin, parametre ayarı gerektiren NMF ve RPCA gibi klasik yöntemlere kıyasla daha yüksek sinyal-kargaşa oranı (SCR) artışı sağladığı ve hedef hiperbollerini daha net ortaya çıkardığı gözlemlenmiştir. Hibrit veri seti ile yapılan testlerde, önerilen DC-Mamba mimarisi (özellikle V1 varyantı), hem görüntü yeniden oluşturma kalitesi hem de gürültü bastırma yeteneği açısından geleneksel CNN modellerini ve Transformer tabanlı DC-ViT mimarisini geride bırakmıştır. Ayrıca, DC-Mamba'nın Tensor-Train teknolojisi ile optimize edilen V2 varyantı, DC-ViT'e kıyasla çok daha düşük bir hesaplama yükü ve parametre sayısına sahip olmasına rağmen, görsel kaliteyi koruyarak kargaşa sinyallerini temizlemede üstün bir başarı sergilemiş ve modelin verimliliğini hibrit veriler ve laboratuvarda yapılan ölçümler üzerinde kanıtlamıştır.

Kargaşa giderme ve hedef tespitini tek bir çatı altında birleştiren Joint (Bütünleşik) modellerin değerlendirilmesinde, önerilen RAFDEC-YOLO mimarisi rakiplerine kıyasla belirgin bir üstünlük sağlamıştır. Bu üstünlük, ilk olarak hibrit veri seti üzerinde yapılan kapsamlı testlerle doğrulanmış; ardışık eğitim stratejisi ile eğitilen temel modellere (UNet, CR-Net vb.) kıyasla, bütünleşik eğitim yaklaşımının tespit doğruluğunu (mAP) ve hassasiyeti (Precision) önemli ölçüde artırdığı sayısal olarak kanıtlanmıştır. Özellikle ardışık yöntemlerin başarısız olduğu ve hedeflerin kaybolduğu yoğun gürültülü gerçek saha verilerinde, bu bütünleşik yapı zayıf hedefleri dahi tespit edebilmiştir. Asfalt altı gibi karmaşık senaryolarda RAFDEC-YOLO, yüzey karmaşasını filtreleyerek derin hedefleri yüksek güvenilirlikle yakalamıştır. Ayrıca, sadece hedefin varlığını bulmak değil yerini tam olarak belirlemek de kritik olduğundan, bütünleşik eğitim stratejisi sayesinde modelin ürettiği tespit kutularının hedefleri çok daha sıkı ve doğru bir şekilde çevrelediği, rakiplerin aksine hatalı konumlanmış kutular üretmediği görsel analizlerle doğrulanmıştır. Sonuç olarak, bu tez çalışması GPR teknolojisinde yapay zeka kullanımını bir adım öteye taşıyarak; global bağlamı anlayan, hesaplama açısından verimli ve görevler arası bilgi paylaşımı yapabilen mimarilerin, özellikle karmaşık ve gürültülü saha koşullarında geleneksel yöntemlere kıyasla çok daha üstün ve güvenilir sonuçlar verdiğini kanıtlamıştır. Yapılan kapsamlı testler sonucunda, DC-Mamba V2 modeli metalik hedeflerde %50'nin üzerinde SCR artışı sağlarken, RAFDEC-YOLO mimarisi mAP@0.5:0.95 metriğinde mevcut literatürdeki en iyi modellere kıyasla %29'a varan performans farkı oluşturmuştur.

Gelecek çalışmalar kapsamında, bu tezde sunulan derin öğrenme yaklaşımlarının 3-boyutlu (3-B) YNR verilerine (C-tarama) uyarlanması hedeflenmektedir. Ayrıca, modellerin genelleştirme yeteneğini artırmak için farklı toprak tiplerine adaptasyon süreçlerinin iyileştirilmesi ve geliştirilen hafif mimarilerin gömülü sistemler üzerinde donanım hızlandırılmalı uygulamalarının gerçekleştirilerek gerçek zamanlı uygulamalara entegrasyonu planlanmaktadır.



1. INTRODUCTION

The non-destructive analysis of underground structures and objects is a strategic issue in today's engineering and fundamental science research. Among the technologies used in this context, Ground Penetrating Radar (GPR) stands out for its high-resolution imaging capability, application speed, and flexibility across various surfaces. GPR technologies started to develop after their first use as an alternative to seismic methods in lunar surface research in the 1970s. Today, they are used in many areas, such as uncovering archaeological remains, checking the structural integrity of important infrastructure like bridges, tunnels, and nuclear power plants, detecting anti-personnel and anti-tank mines in military operations, and analyzing geophysical layers. [1], [2]. GPR signals are created when electromagnetic waves sent underground reflect off layers with different dielectric properties. However, the uneven soil structure, direct reflections from the ground surface, and scatterings from objects that are not the target can cause a lot of noise, called "clutter." This clutter often hides weak signals from the target, making images less clear and detection harder. Because of this, it is important to use an effective "clutter removal" method to get useful information from GPR data.

1.1 Problem Definition and Thesis Objective

GPR technology is important in many fields because it can inspect materials without causing damage. Still, differences in soil structure, ground reflections, and signals from other objects can mask the target's unique hyperbolic reflections, making detection challenging. Researchers have used statistical matrix decomposition methods like Singular Value Decomposition (SVD) [3], [4], and Principal Component Analysis (PCA) [5] to address this issue. Because these methods are sensitive to noise, more robust techniques have been developed. For example, Robust Principal Component Analysis (RPCA) [6] and Robust Non-Negative Matrix Factorization (RNMF) [7] take advantage of the fact that pixel intensities are not negative. Godec [8], based

on the principle of random matrix-decomposition, has made significant progress in the literature. However, despite improvements in clutter-removal performance, these quantitative models have failed to deliver the desired results due to substantial data loss when separating the target signal from the clutter component in complex, nonlinear, and highly variable ground structures, and their performance is overly dependent on the selected parameters.

Although deep neural network architectures such as Convolutional Autoencoders (CAE) [9], [10], [11] and Generative Adversarial Networks (GAN) [12], [13], [14] have been introduced in recent years to tackle these difficulties, it has been seen that long-range spatial dependencies in radargrams cannot be adequately modeled due to the local processing constraints of convolutional layers.

The main objective of this thesis is to provide end-to-end, high-performance deep learning solutions that maximize the detectability of underground targets by eliminating complex clutter structures in GPR images. In pursuit of this general goal, this research intends to introduce an unsupervised Autoencoder Guided Low-Rank Approximation (AE-LRA) [15] approach to overcome the parameter dependency of classical decomposition methods, Transformer-based original models, such as DC-ViT [16], which can capture the global context to overcome the limited receiver area problem of classic CNN architectures, and State-Space Models, such as DC-Mamba, which enable highly efficient modeling in the time-depth plane, to the GPR literature. Within the scope of this thesis, we proposed a method that presents a state-of-the-art clutter-removal and target-detection system in an end-to-end framework. With this method, models that were previously trained separately on different networks are now combined under a single umbrella. This has both shortened the model training time and made the target detection system more robust by utilizing information from feature maps obtained during the clutter removal phase. Our proposed joint method also offers a significant advantage for real-time systems that will be used in the future.

1.2 Proposed Methods

The first fundamental approach proposed in this thesis is Vision Transformer (ViT)-based architectures that can model the global context and long-range spatial

dependencies in GPR radargrams. The DC-ViT (Declutter Vision Transformer) model, developed to address the limited receptive field of traditional convolutional neural networks (CNNs), enhances feature transmission and improves training consistency by adding dense connections to the standard ViT architecture. In this architecture, the relationships between image patches are learned through the “self-attention” mechanism, thereby more competently minimizing the characteristic spread of clutter components across the radargram. To overcome the data-starvation problem experienced by Transformer models, especially on small-scale GPR datasets, convolutional layers incorporated into the model maximize restoration quality by retaining local features. The second original approach is the DC-Mamba method based on State Space Models (SSM), which has attracted significant interest in the deep learning world as of 2024 and is presented as an effective alternative to Transformer architectures. The Mamba architecture, adapted for the first time to the GPR clutter removal problem in this thesis, can capture long-range dependencies in the data with linear computational complexity. This eliminates the quadratic computational expense linked with traditional Transformer models, especially when processing high-resolution, large-scale GPR data. The DC-Mamba model treats reflections in the radargram as a sequence, precisely modeling and cleaning noise originating from complex ground structures in the time-depth plane. In addition to these supervised architectures, we introduce an unsupervised method called Autoencoder Guided Low-Rank Approximation (AE-LRA) to overcome the parameter dependency found in traditional Low-Rank Sparse Decomposition (LRSD) techniques. Classical methods such as RPCA and RNMF require manual tuning of regularization parameters. In contrast, our method uses an autoencoder to automatically learn a low-rank clutter distribution, providing a robust, parameter-free way to remove clutter.

In the final stage of this thesis, a Joint Deep Learning Architecture is proposed that combines clutter removal and object detection processes within a single framework. These processes, which are generally designed as two independent sequential stages in the literature, have been combined in this thesis into a standard network architecture that can be trained end-to-end. Thanks to this approach, the denoising layers are optimized not only to improve visual quality but also to produce feature maps that maximize

the success of the target detection algorithm (YOLO/R-CNN). This proposed joint architecture will be a milestone for methods that achieve low error rates and high-speed target detection in future real-time field applications.

1.3 Literature Review

The clutter in GPR data is typically modeled as a high-energy, stationary (or semi-stationary) component caused by direct reflections between the antenna and the surface. The oldest and computationally simplest method proposed to solve this problem is the Average Subtraction process. One of the primitive methods used to eliminate clutter in GPR data is to apply statistical methods that partition the data into subspaces. In this approach, the GPR data is separated into subcomponents: target, clutter, and noise. Step frequency has compared methods such as averaging, likelihood ratio testing, and system identification for GPR data and found that, despite the computational ease of simple techniques, these methods are insufficient under complex subsurface conditions. In contrast, time-frequency methods such as Wavelet Transform provided a higher Signal-to-Noise Ratio (SNR) [3]. Over time, subspace methods based on the statistical properties of the data have become more popular. Singular Value Decomposition (SVD) and Principal Component Analysis (PCA) have been widely used to separate GPR data into “clutter” and “target” subspaces. Another study, noting that the mean subtraction method leaves residual noise, combined this method with an eigenvalue-based approach to make the target signal more distinct [4]. In subsurface mapping applications, when comparing PCA, moving median filters, and average removal methods, PCA was reported to perform best at reducing clutter [17]. Highlighting the risk of leakage, where PCA-based methods may mistakenly remove the target signal as noise, [18] proposed applying PCA only to target regions using a non-homogeneous detector (NHD) and developed the NHD-PCA and NHD-BSE (Blind Source Extraction) algorithms. Emphasizing the inadequacy of standard SVD in detecting weak targets, [19] proposed an optimized SVD approach that processes the actual measurement together with a non-target reference measurement. Despite performance issues in subspace models, Low-Rank Sparse Decomposition (LRSD) methods, which have been widely used in image processing in subsequent years and applied to many problems, are also used for

clutter removal. In GPR data, the low-rank component is treated as clutter, while the sparse component is treated as the target. The Robust Principal Component Analysis (RPCA) method [20], which is part of this approach, is used to reduce noise in GPR data [6]. The RPCA method employs the nuclear norm within a convex optimization framework that necessitates high-dimensional principal component analysis. To address the associated computational demands, several alternative approaches have been introduced. For example, the GoDec method [21] utilizes bilateral random projections, which are computationally less intensive than singular value decomposition. GoDec has demonstrated effectiveness in reducing noise in ground-penetrating radar (GPR) data [22], [8]. The Robust Orthonormal Subspace Learning (ROSL) method replaces the nuclear norm in RPCA with group sparsity coefficients under an orthonormal subspace to model the low-rank component [23], and has also been applied to GPR noise removal [24]. Robust Non-Negative Matrix Factorization (RNMF) [25] simultaneously decomposes data into low-rank and sparse components. The RNMF approach for GPR denoising is both faster and more effective than RPCA [7]. Additionally, the LRSD problem has been addressed using a Robust Autoencoder (RAE) as an alternative to RPCA for clutter removal in GPR data [26]. RPCA and related techniques typically rely on convex relaxations, such as the nuclear norm, which may result in suboptimal solutions. To overcome this limitation, the Direct Robust Matrix Factorization (DRMF) method was developed to perform direct matrix factorization without nuclear-norm relaxation [27]. Furthermore, the Factor Group-Sparse Regularization (FGSR) method was introduced to address the parameter sensitivity observed in RPCA and RNMF by employing a nonconvex approach [28]. This thesis proposes a method that eliminates the need for a λ parameter and generates a low-rank component representing the clutter within an Autoencoder (AE) [15]. RPCA and similar methods are generally based on convex relaxations (such as the nuclear norm), which can lead to suboptimal results. To solve this problem, they developed the Direct Robust Matrix Factorization (DRMF) method, which performs direct matrix factorization without nuclear-norm relaxation [27]. The Factor Group-Sparse Regularization (FGSR) method, proposed in another study, aims to address the parameter-sensitivity issues of RPCA and RNMF by adopting a nonconvex approach [28]. In our thesis work, we proposed a method that does

not require a λ parameter and generates a low-rank component representing the clutter component in an Autoencoder (AE) [15]. The target image obtained by removing the component derived from this method from the raw image outperformed results obtained using Non-Negative Matrix Factorization (NMF) and other LRSD methods, such as Robust Principal Component Analysis (RPCA), Robust NMF (RNMF), and Robust Autoencoder (RAE). Despite these developments, the performance of LRSD-based methods largely depends on the correct selection of regularization parameters, such as λ , and on the assumption that the chaos is low-rank. The weakening of these assumptions in complex and heterogeneous soils has led researchers to Deep Learning methods that can learn the data distribution. The first successful applications of deep learning to GPR noise reduction are autoencoder architectures that employ unsupervised or semi-supervised learning strategies. Among the pioneering works in this field are the Convolutional Autoencoder (CAE) and its deeper version, the Deep CAE (DCAE) structures [9]. In this work, raw GPR data were compressed into a lower-dimensional feature space (latent space) using successive convolutional and pooling layers. Then, clean images were reconstructed using decoder layers. It is reported that the network trained with simulation data performs better than methods such as MCA and RPCA, especially in complex scenarios, but shows comparable performance to classical methods on real data (especially in dry soil conditions). To enhance the feature extraction capabilities of standard CNN architectures, the U-Net architecture and its derivatives, which have proven successful in biomedical imaging, have been adapted to the GPR [10]. By integrating Attention Mechanisms into the standard U-Net architecture, the Attention U-Net model was proposed [11]. This approach includes a “Spatial Attention Module” (SAM), which allows the network to focus only on regions relevant to the target, and a “Channel Attention Module” (CAM), which highlights essential feature channels. Although attention mechanisms help the network suppress irrelevant background clutter, they cannot fully address the limited local receptive field, a fundamental constraint of CNNs. The success of deep learning models largely depends on the amount of labeled data. However, in the GPR field, it is nearly impossible to obtain both “cluttered” and “clean” (ground-truth) real data for the same region. To address this data scarcity, methods based on Generative

Adversarial Networks (GANs) have been developed. In a study called Declutter-GAN, an “image-to-image translation” approach that converts cluttered GPR images into clean images using the conditional GAN (cGAN) structure is presented [29]. The most significant contribution of this work is the use of a hybrid training set, created by adding real (non-target) clutter data to the simulation data, to enhance the network’s generalization. Another study addressing small datasets is the Wavelet-GAN model [13]. This method separates the GPR image into frequency sub-bands using the Discrete Wavelet Transform (DWT), processing the low-frequency components with a CNN and the high-frequency details with a GAN. Thus, even when trained with a limited amount of real data, the network can quickly adapt to new target types. A more recent approach, ULCR-Net, uses Diffusion Models to increase data diversity and develops a Contrastive Learning-based GAN architecture to learn clutter patterns [30]. This enables chaos prediction without requiring large-scale matched datasets. The aforementioned CNN- and GAN-based methods are highly effective at extracting local features. However, GPR B-scans contain long-range (global) dependencies, such as hyperbolic reflections and layered soil structures that span the entire image. The local convolution operations of CNNs are insufficient for modeling this global context. To overcome this limitation, the DC-ViT (Declutter Vision Transformer) architecture, based on Vision Transformers [31], has been proposed [16]. DC-ViT divides the image into patches and uses a Self-Attention mechanism to model the relationship between information at any point in the image and all other points. Furthermore, using convolutional layers (LeFF) [32] instead of standard MLPs in the Transformer encoders has also improved the ability to capture local details, which is weak in ViTs. This work is an essential milestone in the transition from CNN dominance to Transformer models in GPR clutter removal and constitutes one of the main motivations of this thesis.

The main purpose of interpreting GPR data is to accurately determine the locations of subsurface anomalies or buried objects. Recently, two-stage and single-stage detectors have been widely used in visual object detection. Convolutional Neural Networks (R-CNN), the forerunner of two-stage detectors, have achieved a significant increase in accuracy compared to standard approaches by applying convolutional feature extraction on region proposals based on selective search [33]. In the evolution of this

approach, the developed Faster R-CNN further improved computational efficiency and detection performance by end-to-end learning of the region proposal process through the integration of the Region Proposal Network (RPN) [34]. In contrast, single-stage detectors offer a speed-focused paradigm. The You Only Look Once (YOLO) architecture, introduced in 2016, treats object detection as a single regression problem and enables real-time processing with grid-based predictions on the image. This eliminates the region proposal cost of two-stage methods, providing superior performance, especially in applications where latency is critical [35]. In the current literature, subsequent iterations of YOLO (e.g., YOLOv8, YOLOv10, and beyond) and the rise of transformer-based hybrid models continue to improve the accuracy-speed tradeoff. The application of these general-purpose detectors to the GPR clutter removal problem has been achieved through various strategies. For example, the Faster R-CNN architecture was directly applied to GPR B-scans for object detection, yielding promising results [36]. Detecting subsurface targets from GPR B-scan images is a challenging problem due to signal clutter and low signal-to-noise ratio (SNR). To overcome these difficulties, methods that combine clutter removal and object detection have been proposed in the literature. However, the fundamental methodological issue in the existing literature is that the clutter removal and target detection processes are treated as separate, sequential (cascade) blocks. In this “discrete” approach, detection networks are either fed with raw data or the clutter-removal process is applied as a preprocessing step independent of the detection network. A study showed that synthetic data cleaned using the RNMF method and then trained with Faster R-CNN, EfficientDet, and YOLOv5 networks increased the detection success of the cleaned data [37]. However, in this sequential structure, distortions or data losses introduced by the clutter removal algorithm (e.g., RNMF or simple filters) in the target signal irreversibly degrade the performance of the subsequent detection network. Since the detection network is unaware of the clutter removal process (clutter-unaware), it cannot compensate for the features lost during the cleaning stage. Another study proposed a two-stage GAN architecture for clutter removal and dielectric constant extraction [14]. In a study using a similar GAN architecture, an unsupervised CycleGAN was employed to suppress clutter, followed by a separate target-localization algorithm [38]. In a study

proposing a framework called SpaDyn [39], anomaly detection is first performed using Faster-RCNN, followed by dynamic feature extraction applied to these regions. In the Dual Cascade model, which performs both clutter removal and target detection, the first stage (Cascade-1) performs sequential cleaning by directly filtering wave clutter. In contrast, the second network (Cascade-2) detects the target’s characteristic hyperbola, enhanced by GPR spatial information and attention mechanisms [40]. Although sequential methods offer a modular structure, running two separate models incurs computational overhead, especially for real-time GPR applications. More importantly, since the enhancement algorithm in the first stage operates independently of the detection task, there is a risk of losing features that could be critical for target detection. To overcome the computational load and information-loss limitations of sequential models, Joint frameworks have been developed that combine object detection and image reconstruction into a single model. These approaches typically share a common backbone for detection (high-level) and reconstruction (low-level) tasks. The models are optimized using a composite loss function that combines detection loss. Many studies use the restoration task as an auxiliary branch to train the backbone of the main detection network. For example, DSNet [41] and TogetherNet [42] share the RetinaNet and YOLOX backbones, respectively, with a de-hazing subnetwork. SuperYOLO [43] and ESRTMDet [44], developed for remote sensing (RSI), use super-resolution (SR) as an auxiliary task. Typically active only during training, these auxiliary branches enable the shared layers to learn more robust features and do not increase computational load during inference. For SAR images, [45] combines noise suppression and segmentation as auxiliary tasks. The integrated model approach has also begun to find its place in the GPR field. For example, DA-YOLO [46] uses a GAN-based domain adaptation (DA) module and pseudo-labeling strategy to reduce the domain gap between FDTD simulations and real data. Following the modern “Joint Learning” trend in the literature, this thesis presents an innovative GPR analysis framework that optimizes clutter removal and target detection in a single architecture using a common loss function.



2. BACKGROUND ON GROUND PENETRATING RADAR

2.1 Overview of the GPR System

Ground Penetrating Radar (GPR) is a non-destructive geophysical method that uses high-frequency electromagnetic (EM) waves to obtain high-resolution images of shallow subsurface layers and buried objects. While it is fundamentally similar to the seismic reflection method, it uses EM waves in the microwave band (typically 10 MHz - 3 GHz) instead of acoustic waves.

GPR has a wide range of applications, from archaeology to civil engineering, mine detection to glacier research. The system's fundamental success relies on its ability to detect changes in the dielectric properties (electrical conductivity and dielectric constant) of the subsurface environment.

When the transmitted wave encounters an interface or object with different dielectric properties, part of the energy is reflected to the surface and the rest continues deeper into the subsurface.

The operating principle of a GPR system involves transmitting short electromagnetic pulses into the subsurface using a transmitter antenna and recording the reflected signals with a receiver antenna.

The process can be described in the following steps:

- **Signal Transmission:** The transmitter antenna sends a pulse into the ground at time $t = 0$.
- **Propagation and Reflection:** When the signal encounters a target with a different dielectric constant (ϵ_r), it scatters due to the change in impedance.

- **Signal Reception:** The reflected wave returns to the receiver antenna at the surface. The depth (d) of the target is found by using the round-trip time (t) of the signal and the speed (v) of the wave in the material, as shown below:

$$d = \frac{v \cdot t}{2} \quad (2.1)$$

Here, the velocity v is approximately as follows in terms of the speed of light in a vacuum ($c \approx 3 \times 10^8$ m/s) and the relative dielectric constant of the medium (ϵ_r) (assuming magnetic permeability is 1):

$$v = \frac{c}{\sqrt{\epsilon_r}} \quad (2.2)$$

This relationship points out the vital role of the medium's dielectric properties in depth estimation. For example, because water's dielectric constant ($\epsilon_r \approx 81$) is much higher than that of dry sand ($\epsilon_r \approx 3 - 6$), the wave velocity decreases on wet ground, and the signal attenuates more quickly.

GPR data visualization is performed in three basic formats, depending on the antenna's movement model and the data-acquisition strategy used.

The most fundamental of these is the A-Scan, which is a one-dimensional signal trace recorded while the antenna is stationary; here, the vertical axis represents time or depth, and the horizontal axis represents reflection amplitude.

However, the A-Scan format does not provide spatial information regarding the target's geometry. The B-Scan is generated by moving the antenna along a line and arranging sequential A-Scans side by side, resulting in a two-dimensional cross-sectional image of the subsurface.

In the B-scan format, which is also the main focus of this thesis, the horizontal axis represents the antenna's position on the surface, while the vertical axis represents the signal's round-trip time; point targets such as pipes or mines are displayed in a characteristic hyperbolic (inverted U) shape depending on the antenna's approach and retreat movement.

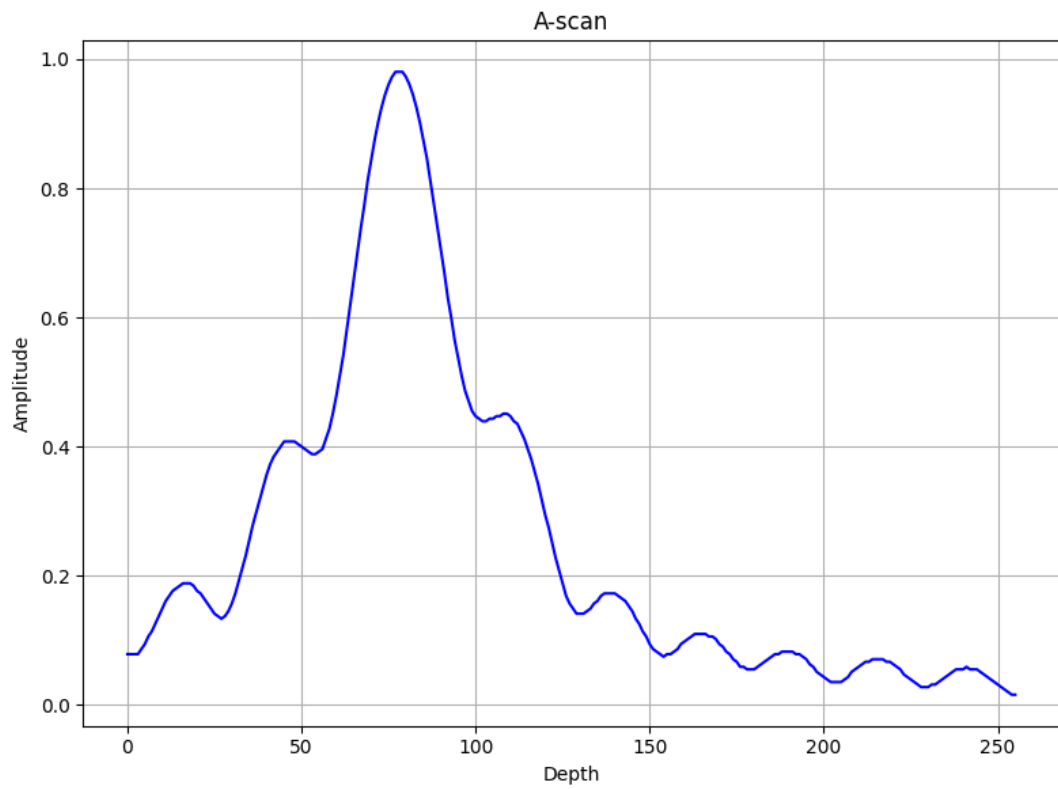


Figure 2.1: A-scan signal acquired from the GPR system.

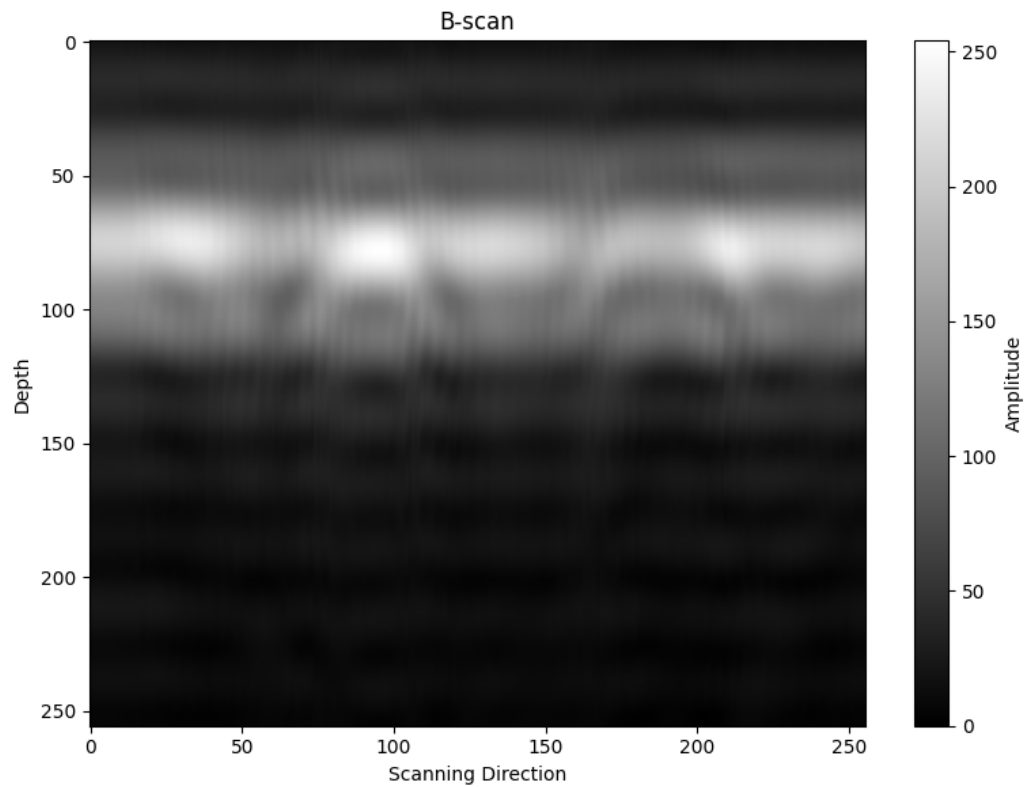


Figure 2.2: B-scan image reconstructed from data acquired by the GPR system.

The third format, C-Scan, is a three-dimensional volumetric image created by combining multiple B-scan lines that are parallel to each other in a specific area and is generally used to map horizontal cross-sections (time slices or width) at specific depth levels.

C-Scan

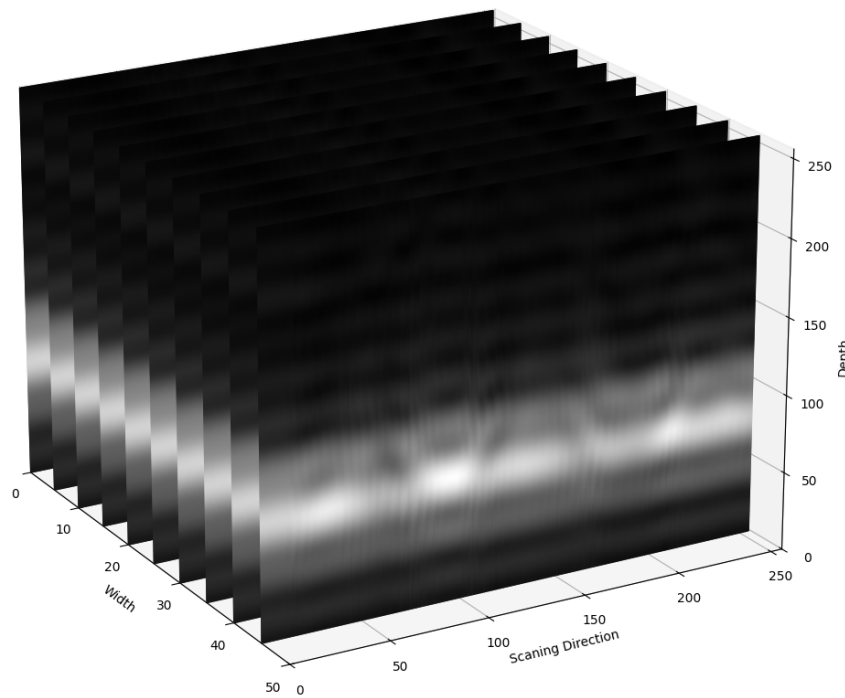


Figure 2.3: C-scan image reconstructed from data acquired by the GPR system.

Ideally, a GPR measurement would only capture reflections from the target, such as a mine or a buried pipe. In practice, though, GPR signals also include many unwanted components caused by the complex subsurface environment and hardware limitations. These unwanted signals, called “clutter” in the literature, are coherent signals created by the interaction between the electromagnetic wave and the medium, which makes them different from random noise.

GPR clutter can be grouped into three main categories based on where it comes from:

- **Inter-Antenna Interaction (Crosstalk) and Surface Reflection:** Some of the wave sent from the transmitter antenna goes straight to the receiver antenna through the air, without entering the ground. Also, because air and soil have very different dielectric

properties, much of the transmitted energy is reflected back from the surface. These signals, called “early-time clutter,” usually show up as horizontal, high-amplitude bands in B-scan images.

- **Scattering Clutter:** Heterogeneous structures underground (rocks, tree roots, soil voids) and local variations in dielectric properties cause the wave to scatter randomly. This creates a “speckled” noise texture in the image and distorts the target’s hyperbolic shape.
- **Multiple Reflections (Ringing/Multiples):** These occur when the signal bounces back and forth between the surface and the antenna, or between different underground layers.

Mathematically, the raw GPR signal $x(t)$ recorded by the receiver antenna at time t can be described as the sum of these components, following the superposition principle:

$$x(t) = s_{target}(t) + c_{clutter}(t) + n_{noise}(t) \quad (2.3)$$

Here:

- $s_{target}(t)$: The target signal, which is the signal reflected from the buried object and is to be detected. It is usually very low in amplitude due to propagation loss.
- $n_{noise}(t)$: Summed white Gaussian noise (AWGN), usually caused by receiver electronics and environmental electromagnetic interference. It has been omitted in our studies.
- $c_{clutter}(t)$: The clutter component, which includes the surface reflections and scatterings defined above.

The main problem with this model is that the energy of the $c_{clutter}(t)$ component is usually much higher than that of the $s_{target}(t)$ component. For shallow targets, strong surface reflections can completely hide the target signal. Also, because clutter signals have a frequency spectrum similar to the target signal, simple frequency filters cannot

separate them. This challenge limits the effectiveness of target-detection algorithms and points out the need for advanced clutter-removal methods using Deep Learning (DC-ViT and DC-MAMBA), as proposed in this thesis.

2.2 Dataset Collection

In this thesis study, the hybrid CLT-GPR dataset developed by [10] was used for training deep learning models. This dataset consists of three different categories aimed at maximizing data diversity. The first component of the dataset, the Synthetic Sub-Dataset, comprises 1,920 B-scan images obtained by simulating six different soil types and four surface conditions using the gprMax software and the FDTD method with a 1.5 GHz GSSI antenna model. The second component, the Sand Sub-Dataset, consists of 6,000 hybrid data points created by mathematically combining real field clutter collected with a multi-polarimetric system operating in the 1.3–6.0 GHz range with synthetic target signatures. Finally, the Concrete Sub-Dataset, representing urban infrastructure, contains 4,000 hybrid images based on real concrete clutter collected using monostatic and bistatic configurations of a single-polarization GPR system.

This comprehensive dataset, which integrates synthetic targets with real-world conditions, enables the developed models to generalize effectively to both controlled simulation environments and complex real-world scenarios.

The model was also tested in real-world conditions, not only with synthetic data. Field measurements took place in a 2×1 m test pool at the Radar Signal Processing Laboratory of Istanbul Technical University. Measurements were collected using 1000 data points in the 600 MHz-1300MHz frequency band. Three scenarios were designed to test detection performance, including distinguishing between metal and dielectric materials and detecting targets at different depths. Asphalt layers and other environmental factors were added to introduce clutter. Table 2.1 provides more details and objectives for these experiments.

This combined strategy, which combines controlled synthetic datasets using experimental field measurements, establishes a strong basis for evaluating the proposed

models. The physical experimental setup is shown in Figure 2.4, and schematic diagrams of the test scenarios are presented in Figure 2.5.

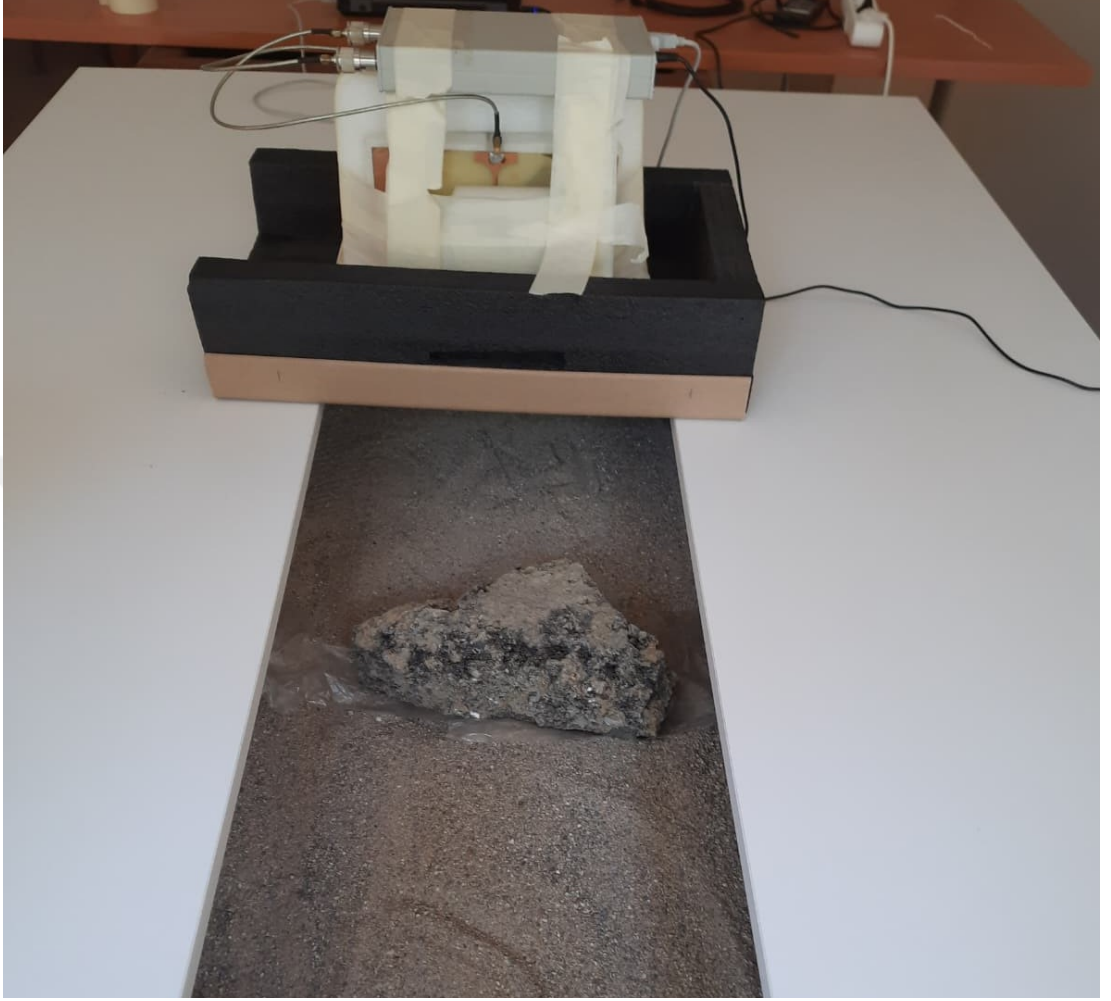


Figure 2.4: Laboratory setup and equipment used for GPR measurements.

The data were preprocessed before being used in the model. Each radargram was resized to 256×256 pixels, and pixel values were scaled to the $[0, 1]$ range. For supervised target detection, axis-aligned bounding boxes were manually drawn around hyperbolic reflections in the validation set to create ground-truth labels.

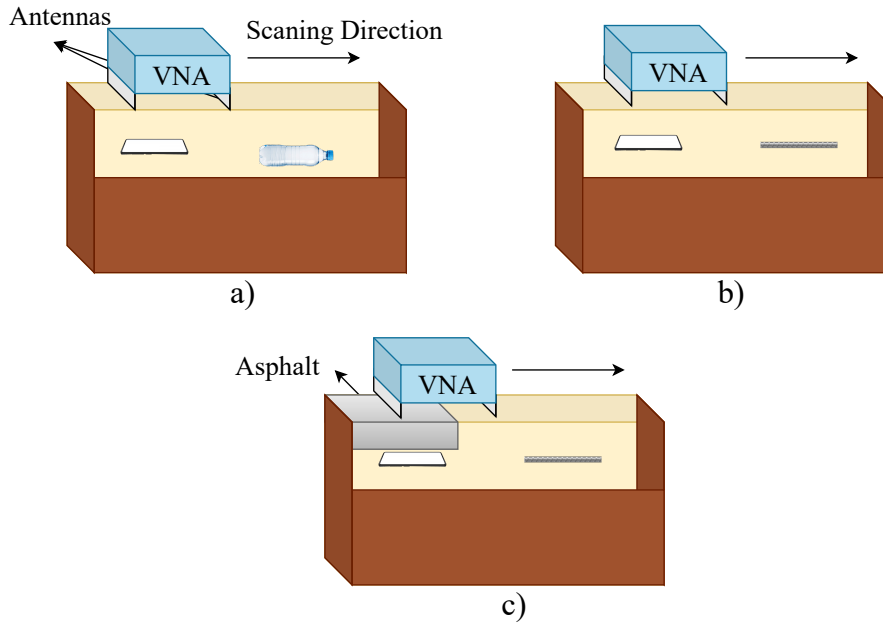
2.3 Performance Evaluation Metrics

To evaluate clutter-removal performance, we use the Peak Signal-to-Noise Ratio (PSNR) on the hybrid dataset, for which the ground truth is available. PSNR is defined as:

$$PSNR = 10 \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \quad (2.4)$$

Table 2.1: Summary of experimental scenarios used in laboratory measurements.

Scenario	Focus	Target Configuration	Objective
Scenario 1	Dielectric Contrast	Mobile Phone (10 cm depth) and Water-filled PET Bottle (5.5 cm depth), spaced 60 cm apart.	Evaluate detection capability for weak dielectric targets with low reflectivity contrast.
Scenario 2	Material Variety	Mobile Phone (10 cm depth) and Cylindrical Rebar (6 cm depth), spaced 60 cm apart.	Assess shape robustness and detection performance across different material types (metal vs. composite).
Scenario 3	Complex Environment	Same target configuration as Scenario 2, but with the sand surface covered by an asphalt layer.	Simulate challenging urban conditions with significant surface clutter and signal attenuation.

**Figure 2.5:** Experimental scenarios used in laboratory measurements: **a)** Scenario 1: Mobile phone and water-filled PET bottle. **b)** Scenario 2: Mobile phone and cylindrical rebar. **c)** Scenario 3: Targets buried under an asphalt layer to simulate complex surface clutter.

where MAX_I is the maximum possible pixel value of the image, and MSE is the Mean Squared Error between the original and processed images.

We also employ the Multi-Scale Structural Similarity Index (MS-SSIM) to assess the preservation of structural information. MS-SSIM calculates the structural similarity between two images across multiple scales, providing a more robust measure of image quality than single-scale SSIM:

$$\text{MS-SSIM}(x, y) = [l_M(x, y)]^{\alpha_M} \cdot \prod_{j=1}^M [c_j(x, y)]^{\beta_j} [s_j(x, y)]^{\gamma_j} \quad (2.5)$$

where l_M represents luminance comparison at the coarsest scale M , and c_j and s_j denote the contrast and structure comparisons at scale j , respectively. The exponents $\alpha_M, \beta_j, \gamma_j$ are weights for each component.

For real-world experiments lacking ground truth, we adopt the Improvement Factor (IF) based on the Signal-to-Clutter Ratio (SCR).

To evaluate with SCR, the GPR data $x \in R^{H \times W \times 1}$ is normalized in the $[-1, 1]$ range, and then the target energy function (TEF) is calculated for each A-Scan sequence that forms the GPR image. The calculations required to find TEF and SCR are shown in (2.6) and (2.7), respectively.

$$\text{TEF}(j) = \sum_{i=1}^H |x(i, j)|^2 \quad (2.6)$$

$$\text{SCR} = \frac{\max(\text{TEF}_{DTR})}{\max(\text{TEF}_{FAR})} \quad (2.7)$$

To distinguish signals from the target and clutter, the target and its near surroundings are defined as the detection region (DTR), and the remaining areas are defined as the false alarm region (FAR). Appropriate intervals for the target and clutter were selected from the calculated TEF array, and TEF_{DTR} and TEF_{FAR} . The power of the signal and clutter is estimated as the maximum of TEF_{DTR} and TEF_{FAR} and used to calculate the SCR.

The IF quantifies the enhancement in decibels:

$$IF = 20 \log_{10} \left(\left| \frac{\text{SCR}_{pred}}{\text{SCR}_{raw}} \right| \right) \quad (2.8)$$

where a higher IF indicates superior clutter suppression.

To quantitatively evaluate the target detection performance of the proposed ensemble models, the standard metrics accepted in the literature, namely Precision, Recall, and Mean Average Precision (mAP), were used.

Precision is the proportion of the model's predicted positives that are actually correct, i.e., the reliability of the predictions. Recall, on the other hand, measures how well the model can detect the actual targets in the dataset; in other words, the scope of its detection capability. These two metrics are mathematically expressed by the following formulas:

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN} \quad (2.9)$$

These equations use the Intersection over Union (IoU) criterion. True Positives (TP) are predicted boxes that match ground truth objects with an IoU above a set threshold. False Positives (FP) are detections with IoU below the threshold or background areas wrongly marked as objects. False Negatives (FN) are ground-truth objects the model misses or predicts with little overlap. IoU measures how much the predicted box (B_p) overlaps with the ground truth box (B_{gt}) as shown below:

$$\text{IoU} = \frac{\text{Area}(B_p \cap B_{gt})}{\text{Area}(B_p \cup B_{gt})} \quad (2.10)$$

To evaluate overall model performance, we use the Average Precision (AP) metric. AP is the area under the Precision-Recall (PR) curve and is calculated as the integral of precision over all recall levels:

$$AP = \int_0^1 p(r) dr \quad (2.11)$$

Here, $p(r)$ is the precision at a specific recall level r . In this study, we used two mAP versions to analyze model performance. The first, mAP@0.5, measures detection at a fixed IoU threshold of 0.5. The second, mAP@0.5:0.95, gives a stricter evaluation by averaging mAP values at IoU thresholds from 0.5 to 0.95 in steps of 0.05:

$$\text{mAP}@0.5 : 0.95 = \frac{1}{10} \sum_{i=0}^9 \text{mAP}_{\text{IoU}=0.5+0.05 \times i} \quad (2.12)$$

Especially in GPR data, the correct positioning of hyperbolas is very important. Therefore, it is important to use the “mAP@0.5:0.95” metric, which penalizes incorrectly placed bounding boxes more heavily. In addition, the operational efficiency and computational costs of the models were compared in terms of floating-point operations per second (FLOP) and total number of parameters (M).





3. PROPOSED METHODS FOR CLUTTER REMOVAL IN GPR DATA

3.1 Proposed Unsupervised Model

3.1.1 Autoencoder-guided low-rank approximation (AE-LRA)

Autoencoders are an unsupervised learning method, and since the representation is learned automatically from the data, the learned representation is specific to that data.

To implement autoencoders, three functions are required: an encoding function, a decoding function, and a loss function to calculate the loss between inputs and outputs.

The encoder processes raw input data ($\mathbf{x}$) in the input layer. Subsequent hidden layers gradually reduce the size of the data by applying activation functions such as ReLU or sigmoid. Each layer contributes to the development of a compressed representation by extracting specific features and patterns.

The bottleneck layer or hidden space in the encoder forms the core of the autoencoder. This layer has a much lower dimensionality than the input data and serves as the compressed encoding or hidden representation ($\mathbf{z}$). The hidden space stores the most critical information needed to reconstruct the original data. The hidden representation ($\mathbf{z}$) is expressed as follows.

$$\mathbf{z} = f_{\theta_e}(\mathbf{x}) = \phi_e(\mathbf{W}_e \mathbf{x} + \mathbf{b}_e) \quad (3.1)$$

Where f_{θ_e} is the encoder function, ϕ_e is an activation function such as a sigmoid function or a rectified linear unit (ReLU), $\mathbf{W}_e$ is a weight matrix, and $\mathbf{b}_e$ is a bias vector. Here, θ_e represents the learnable parameters of the encoder, consisting of $\mathbf{W}_e$ and $\mathbf{b}_e$.

The decoder starts with the encoded representation ($\mathbf{z}$) from the latent space. Hidden layers progressively increase the data's dimensionality to recover information lost during compression. The activation functions in these layers help to learn the mapping

back to the original input space. The final layer produces the reconstructed output data ($\hat{x}$). The reconstruction result can be expressed as follows.

$$\hat{x} = g_{\theta_d}(z) = \phi_d(\mathbf{W}_d \mathbf{z} + \mathbf{b}_d) \quad (3.2)$$

where g_{θ_d} , ϕ_d , $\mathbf{W}_d$, and $\mathbf{b}_d$ are decoder function, activation function, weight matrix, and bias vector, respectively.

To implement autoencoders, three functions are required: an encoding function, a decoding function, and a loss function to calculate the loss between inputs and outputs.

The encoder and decoder functions are parametric functions, such as Artificial Neural Networks (ANNs), whose derivatives can be taken according to the distance function used in the loss function. Thus, the encoder and decoder functions can be optimized using Stochastic Gradient Descent (SGD) to minimize the reconstruction loss.

The autoencoder diagram is shown in Figure 3.1. The dimensions of the input to the encoder block and the reconstructed data output from the decoder block are the same.

Autoencoder neurons are connected through random weights (w) and biases (b) to avoid getting stuck in local minima during training. Random initialization helps break symmetry and allows for exploring different solutions.

In this study, to obtain an image containing the target component from GPR data, we propose a method that reconstructs the image containing clutter components using an autoencoder and extracts it from the raw GPR image.

The proposed method leverages the power of autoencoders, which are artificial neural networks trained to reconstruct their inputs. The encoder takes raw GPR data as input. This data is typically represented as a collection of samples, denoted by $\mathbf{x}^{M \times N}$. The encoder function $f_{\theta_e}(\mathbf{x})$ decreases the size of the data and generates a compressed representation ($\mathbf{z}$) from Eq. 3.1. The decoder function, $g_{\theta_d}(\mathbf{x})$, utilizes this compressed representation for reconstruction. The objective function used to reconstruct the low-rank component is given below

$$\min_{\theta_e, \theta_d} \|\mathbf{x} - g_{\theta_d}(f_{\theta_e}(\mathbf{x}))\|_F^2 \quad (3.3)$$

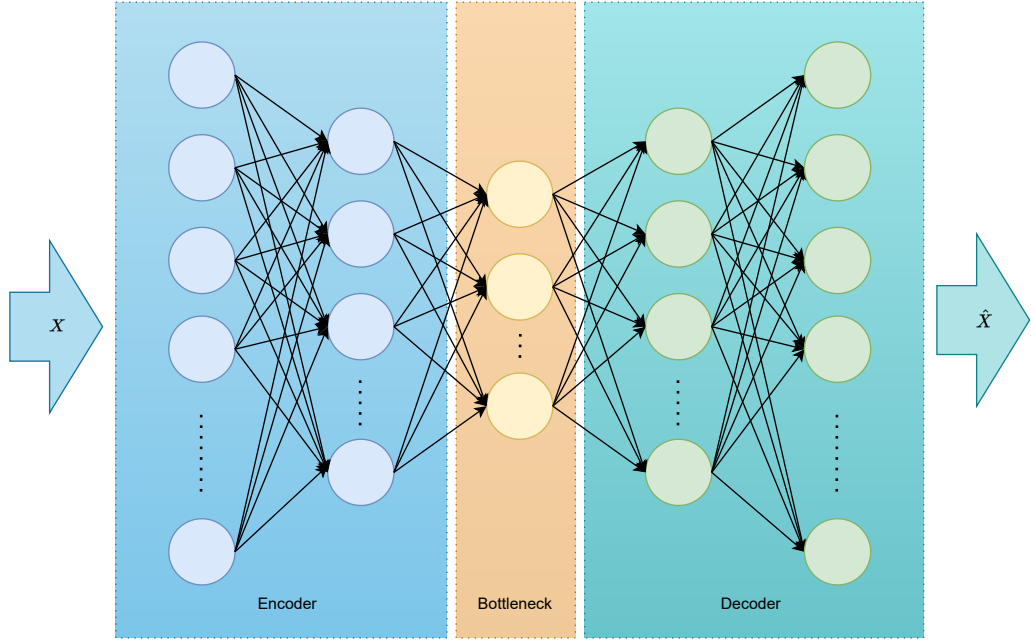


Figure 3.1: Architecture of an autoencoder.

The Frobenius norm is widely used in the low-rank approximation. The loss function to measure reconstruction error can be defined as

$$L(\mathbf{x}, \hat{\mathbf{x}}) = \|\mathbf{x} - \hat{\mathbf{x}}\|_F^2, \quad \hat{\mathbf{x}} = g_{\theta_d}(f_{\theta_e}(\mathbf{x})) \quad (3.4)$$

Briefly, the optimal values of the weights that minimize the objective function (θ_e, θ_d) are obtained by training the autoencoder using gradient descent. The objective function allows the network to be trained to learn low-rank components as outputs, and thus, the network does not require any training data. The target output is obtained by subtracting data containing low-rank components from the raw data. The proposed autoencoder-based GPR decluttering method is summarized in Algorithm 1. After training the autoencoder to reconstruct the clutter component ($\hat{x}$), the reconstructed clutter component is subtracted from the raw GPR data (x) to obtain the target component (x_t). This expression is as follows:

$$\mathbf{x}_t = \mathbf{x} - \hat{\mathbf{x}}. \quad (3.5)$$

Algorithm 1 Proposed Autoencoder-Based Clutter Removal Method

```
1: Input: GPR data  $\mathbf{x}$ 
2: Parameters: Encoder function  $f_{\theta}(x)$ , Decoder function  $g_{\theta}(z)$ 
3: Loss function:  $L(\mathbf{x}, \hat{\mathbf{x}}) = \|\mathbf{x} - g_{\theta_d}(f_{\theta_e}(\mathbf{x}))\|_F^2$ 
4: Initialize encoder weights  $\theta_e$  and decoder weights  $\theta_d$ 
5: Set learning rate  $\alpha$ , number of epochs  $N$ 
6: for epoch = 1 to  $N$  do
7:   Compute encoder output:  $z = f_{\theta_e}(\mathbf{x})$ 
8:   Compute decoder output:  $\hat{\mathbf{x}} = g_{\theta_d}(z)$ 
9:   Compute loss:  $L(\mathbf{x}, \hat{\mathbf{x}})$ 
10:  Update encoder weights:  $\theta_e \leftarrow \theta_e - \alpha \frac{\partial L}{\partial \theta_e}$ 
11:  Update decoder weights:  $\theta_d \leftarrow \theta_d - \alpha \frac{\partial L}{\partial \theta_d}$ 
12: end for
13: Output:  $\mathbf{x}_t = \mathbf{x} - \hat{\mathbf{x}}$ 
```

In the study, we used hidden-layer dimensions (128, 64, 32, 64, 128) for the encoder and decoder, set the learning rate to 0.001, and used Stochastic Gradient Descent to update the weights.

3.1.1.1 Experimental results on hybrid data

We used the test set from the hybrid dataset [10] to compare our proposed method with SVD, PCA, NMF, RPCA, RNMF, and RAE [26]. PSNR and MS-SSIM results are given in Table 3.1.

Table 3.1: PSNR and MS-SSIM Results for Hybrid Dataset.

Methods	SVD [5]	PCA [5]	NMF [47]	RPCA [6]	RNMF [7]	RAE [26]	AE-LRA [15]
Mean PSNR (dB)	24.48	24.41	24.49	27.72	28.60	28.18	28.94
MS-SSIM (Mean)	0.841	0.833	0.841	0.875	0.889	0.868	0.892

Table 3.1 and Figure 3.2 show that our proposed method outperforms SVD, PCA, NMF, RNMF, and RPCA. The recommended λ is given by;

$$\lambda = 1/\sqrt{\max(H, W)} \quad (3.6)$$

in [6], and we found $\lambda = 0.0625$ for the data sizes in our dataset. For this value, our method achieved higher PSNR and MS-SSIM. The results obtained using different λ values in RAE are shown in Table 3.2. The Table 3.2 shows that the λ value generated by the empirical formula in RAE does not consistently yield the optimal outcomes. Our

proposed method does not need λ parameter and soft-thresholding to calculate sparse components as in RAE. Instead, it computes only the low-rank component and recovers the target image by subtracting the raw image from it.

Table 3.2: AE-LRA result for different λ values.

λ	0.05	0.0625	0.001	0.005	0.0001
Mean PSNR (db)	26.73	28.18	29.48	29.46	28.59
Mean MS-SSIM (db)	0.8568	0.8684	0.8934	0.8923	0.8748

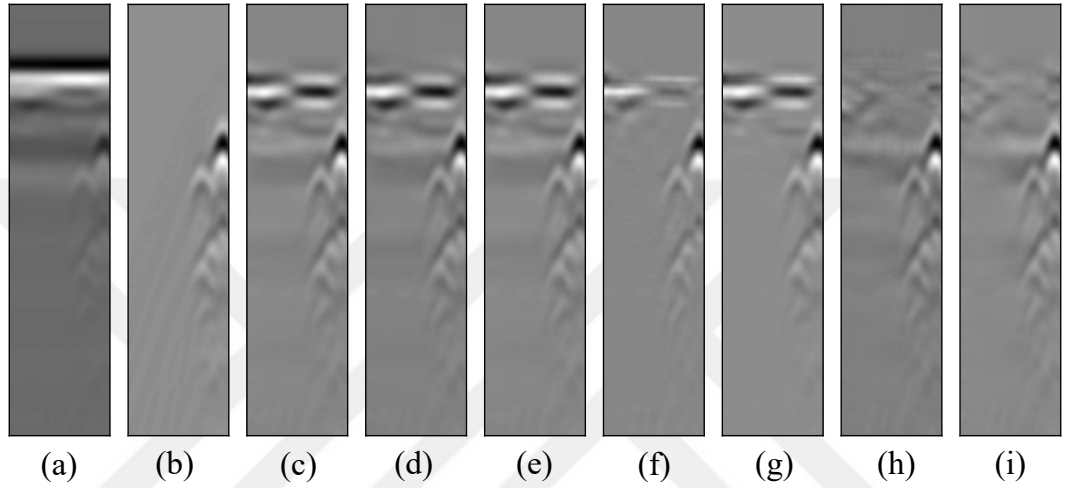


Figure 3.2: Comparison results for a sample GPR image from the hybrid dataset. (a) Raw data, (b) Ground-truth, clutter removal results obtained by (c) SVD, (d) PCA, (e) NMF, (f) RPCA, (g) RNMF, (h) RAE, and (i) AE-LRA.

3.1.1.2 Experimental results on real-world data

Figures 3.3 and 3.4 show the results obtained for real data. In Figure 3.3, clutter-removal results for a copper rod buried in sand at a depth of 10 cm are shown. Although the target is recovered in the SVD, PCA, RPCA, and RNMF results, residual clutter remains on the right side of the target. This clutter is absent in RAE and present in AE-LRA. NMF fails to remove the clutter.

Clutter-removal results for plastic buried in sand at a depth of 6 cm are shown in Figure 3.4. Although all methods fail to completely remove the clutter due to the complexity of the scenario, the distinction between clutter and target components is more evident in the results of RNMF, RAE, and the proposed method AE.

3.2 Proposed Supervised Models

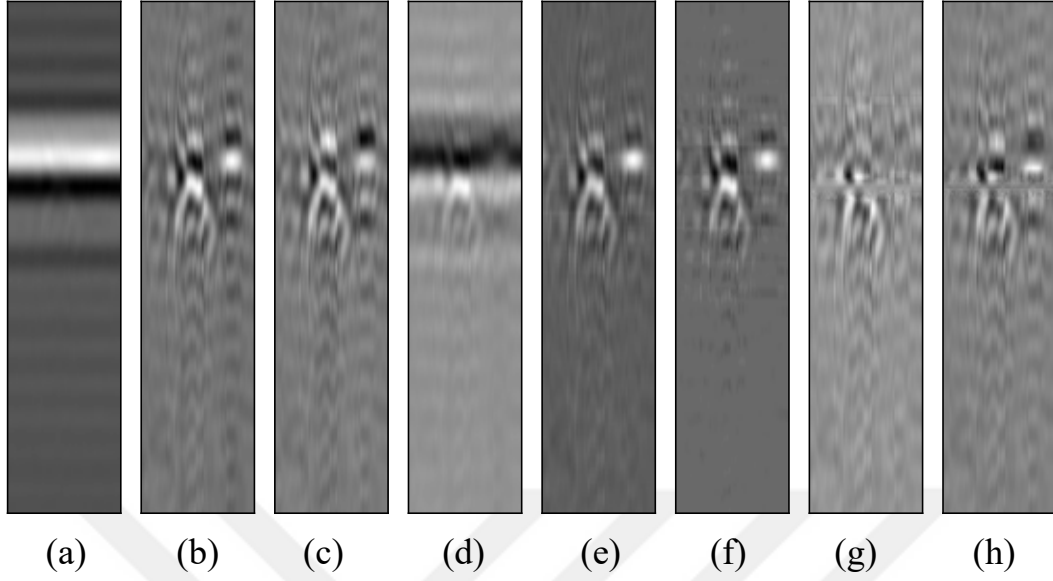


Figure 3.3: Comparison results for real data for a copper rod at a depth of 10 cm in sand (a) Raw data, clutter removal results obtained by (b) SVD, (c) PCA, (d) NMF, (e) RPCA, (f) RNMF, (g) RAE, and (h) proposed AE methods.

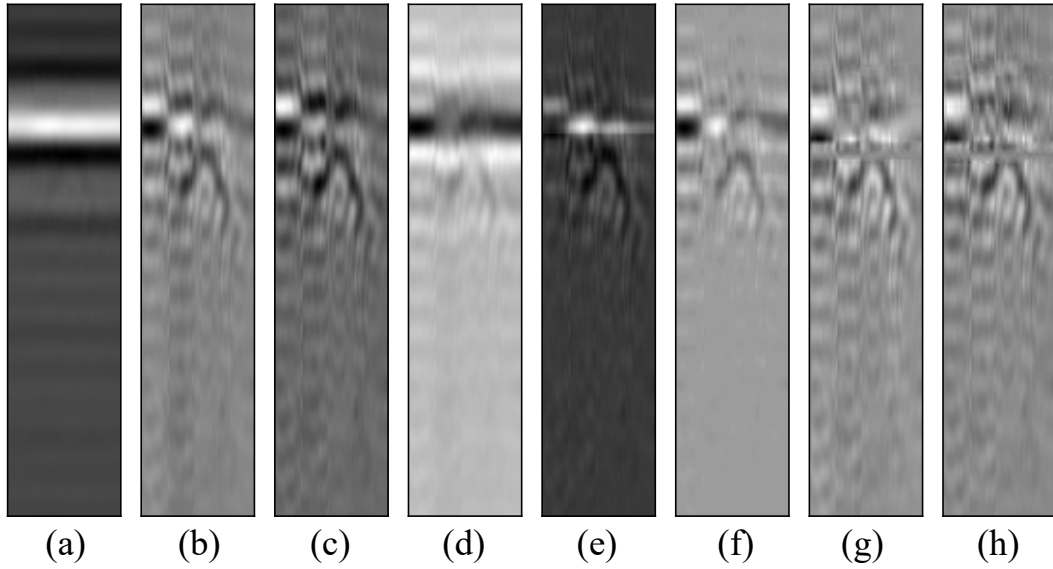


Figure 3.4: Comparison results for real data for a plastic bottle at a depth of 6 cm in sand (a) Raw data, clutter removal results obtained by (b) SVD, (c) PCA, (d) NMF, (e) RPCA, (f) RNMF, (g) RAE, and (h) AE.

3.2.1 Declutter vision transformer (dc-vit)

Traditional subspace-based clutter removal techniques, such as SVD and PCA, which were widely used before the era of deep learning, aim to project raw GPR data onto clutter and target subspaces. [5]. However, these methods may not perform well when target components are split into multiple subimages. RPCA, on the other hand, aims to optimize the low-rank and sparse matrix representation problem to separate noise and target responses [6]. Nevertheless, it can struggle when target responses exhibit broad, flat features and requires careful hyperparameter selection. Robust Non-Negative Matrix Factorization (RNMF) optimizes a matrix representation problem to separate low-rank complexity components from sparse target components [7]. However, like RPCA, RNMF's performance also depends on parameter selection, making it somewhat sensitive to specific features of the underground scenario. Sparse representation-based techniques such as morphological component analysis (MCA) [48] and dictionary learning (DL) effectively separate noise and target components. However, dictionary selection significantly affects their performance and complexity. More versatile noise removal methods are needed to overcome these challenges in subsurface environments. In recent years, deep learning-based approaches have attracted attention for GPR noise removal and have demonstrated their potential through the accuracy results they have achieved. In [9], a convolutional autoencoder is proposed for deconvolution, and the study demonstrates how parameters such as convolutional filter size, number of filter channels, and network depth affect network performance and superiority over traditional methods. [10] proposes an encoder-decoder architecture built with the U-NET architecture. In the skip connection section, Radial Dense Blocks are used between the encoder and decoder instead of a direct connection to preserve target-response features and reduce unwanted noise. Although Convolutional Neural Networks (CNNs) are successful in capturing local features in images, their limited filter sizes are not sufficiently successful in capturing long-range dependencies between pixels. Furthermore, the static filter weights in CNNs also pose a challenge. Visual Transformers (ViTs) overcome these limitations by using the Self-Attention mechanism, which enables them to capture long-range dependencies and achieve notable success in

various tasks [31]. The SA mechanism, which calculates the response at a given pixel as the weighted sum of all other positions, is a more powerful and dynamic alternative to CNN-based models for capturing long-range pixel dependencies. Multi-head self-attention (MSA) incorporates multiple instances of parallel self-attention and increases the model's capacity to capture various patterns and dependencies in the data, making it preferable to SA. Due to their advantages, Vision Transformers are widely used in many image processing tasks such as denoising, image restoration, and classification [49], [50], [51], [52].

This study proposes Declutter Vision Transformers (DC-ViT) to remove clutter in 2D GPR images, also known as B-Scan. ViTs successfully capture long-range dependencies through the self-attention mechanism. Still, since they are insufficient to capture local information, the Convolutional block, which is better at capturing local contexts, was used in DC-ViT instead of the MLP block. Since DC-ViT is a deep network, dense connections are used to reduce losses and interruptions in information flow between layers.

GPR data can be expressed as follows:

$$x = x_c + x_t \quad (3.7)$$

where $x \in R^{H \times W}$, $x_c \in R^{H \times W}$, and $x_t \in R^{H \times W}$ denote GPR data, clutter component, and the target component, respectively. H and W are the depth index and the number of antenna locations. The proposed method for obtaining a target image containing a hyperbolic target component, after removing clutter, from the given raw GPR data, x , is explained in three parts. Firstly, raw images are transformed into a sequence of informative elements. These elements are then processed through a novel Dense Transformers Block. Finally, the processed features are used to reconstruct the image.

3.2.1.1 Patching, patch embedding, and position embedding

In the initial part of the model, the input GPR image $x \in R^{H \times W \times 1}$ is divided into patches of size using (3.8), where $P(\cdot)$ represents the patching operation.

$$x_P = P(x), x_P \in R^{\frac{H \times W}{p^2} \times p^2} \quad (3.8)$$

These patches then undergo the Patch and Positional Embedding processes. Patch embeddings flatten the patches, and they are then passed through a fully connected layer. In this step, as seen in (3.9), feature embedding consisting of N D-length vectors is obtained where $N = \frac{H \times W}{p^2}$, and $H_{PatchEmb}(\cdot)$ is the patch embedding process.

$$x_{PE} = H_{PatchEmb}(x_P), x_{PE} \in R^{N \times D} \quad (3.9)$$

Positional encoding is essential for enabling the model to comprehend the order and spatial relationships inherent in the input sequence. This encoding provides crucial information regarding the positions and sequential arrangement of elements within the input embeddings. Without this framework, models would struggle to process temporal or spatial hierarchies effectively, thereby hindering their overall performance on tasks that rely on a clear understanding of sequence dynamics. Standard learnable 1D positional embeddings, similar to those in [31], have been used in the study. After the patch embedding and position embedding processes, the required feature embeddings to feed the Dense Transformer Network are obtained as follows:

$$F_0 = x_{PE} + H_{PosEmb}(x_0) \quad (3.10)$$

Here, F_0 represents the vector sequence that feeds the Dense Transformer Network, and $H_{PosEmb}(\cdot)$ is the Position Embedding process.

3.2.1.2 Dense transformer block

The Dense Transformer Block consists of dense connections within Transformer Encoder Groups, which are sequential transformer encoder connections [53]. Instead

of MLP used in traditional Transformer encoders, Locally-enhanced Feed-Forward Network (LeFF) was used in the study. Since LeFF more effectively captures local context in image data [54], it more effectively distinguishes local clutter from real targets. The LeFF used includes 3x3 Depth Convolution, and ReLU was used as an activation function. Following this, Batch Normalization was used to normalize the activations of the convolutional layers, reducing internal covariance shift and stabilizing training for faster convergence.

Direct connections are established from any layer to all subsequent layers in Dense Transformer Networks. Thus, the i -th layer receives the feature maps of all preceding layers, and it can be represented as follows:

$$f_i = H_i(\text{Con}[f_0, f_1, \dots, f_{i-1}]) \quad (3.11)$$

Here, f is the i -th feature map, and $H(\cdot)$ is a composite function that involves operations such as Transformer Encoder Groups and Linear Projection (LP). Linear Projection can help in simplifying the representation of clutter and target features. This can make it easier for the network to separate clutter from meaningful signals. Connections between Transformers are established using the Concatenation operator represented as $\text{Con}[\cdot]$ in (3.11). Linear Projection is used to reduce the dimension after the Concatenation operation.

As a result, the deep features are obtained as below,

$$F_D = H_{DTN}(F_0) \quad (3.12)$$

here, $H_{DTN}(F_0)$ represents the proposed Dense Transformer Network.

3.2.1.3 Reconstruction block

The obtained deep features are fed into a reconstruction block to obtain the target image, and the result is expressed as follows:

$$x_T = H_{rec}(F_D) \quad (3.13)$$

where $H_{rec}(F_D)$ represents the reconstruction process. In the reconstruction block, reshaping is performed first, converting the $N \times D$ sequence into the $\frac{H}{p} \times \frac{W}{p} \times D$ feature

map, and then it is passed through a 1×1 convolutional layer. During this process, the depth dimension is increased by a factor of 4 through UpSampling to enrich the deep feature. To enhance resolution, the depth-to-space layer is used, resulting in a feature map of size $H \times W \times \frac{4D}{p^2}$. Finally, using a 1×1 convolution block, the number of channels is reduced to one and the target image $x_t \in R^{H \times W \times 1}$ is obtained. This reconstruction step further helps isolate and refine the target GPR image while suppressing clutter. Techniques like depth-to-space operations and convolutional layers enhance this process.

3.2.2 Declutter vision mamba (dc-mamba)

Current research shows a transition away from conventional signal processing to deep learning (DL) methods, particularly Convolutional Neural Networks (CNNs) [9]. Architectures such as CR-Net [10] and Attention U-Net [11] have produced effective results in noise suppression. However, the limited receptive fields inherent to convolutional operations frequently have difficulty capturing the long-range dependencies required to distinguish target hyperbolas from broad noise. To overcome these constraints, Vision Transformer (ViT) architectures have been adapted to the GPR domain [16]. In our previous work, DC-ViT [16], we proposed a transformer-driven approach that uses self-attention to model global context. However, the complexity of self-attention scales quadratically with image resolution, creating a severe bottleneck for high-resolution processing.

More recently, State Space Models (SSMs) like Mamba [55] have emerged, maintaining linear processing time complexity while achieving the modeling power of Transformers. Based on this, Vision Mamba [56] and VMamba [57] adapted selective scanning to 2D data. In particular, Efficient VMamba [58] uses atrous selective scanning to improve feature extraction.

In this study, we propose DC-MAMBA, a new architecture built upon the Effective Visual State Space (EVSS) blocks of Efficient VMamba [58]. We developed two variations: DC-MAMBA V1, which uses standard fully connected (FC) layers, and DC-MAMBA V2, which employs Tensor-Train Linear (TT-Linear) layers [59] in linear projections and dense connections. This structural improvement in V2 considerably reduces the parameter load compared to V1 while retaining the capability to represent

global dependencies. Furthermore, a resource-efficient Lightweight Reconstruction Block has been designed to convert deep features into high-resolution GPR images. Unlike the single-step expansion in DC-ViT, this block gradually enlarges the image using PixelShuffle and depth-wise separable convolutions. This technique lowers the total GFLOPS cost while preventing checkerboard artifacts.

In this study, a hybrid architecture is proposed to eliminate clutter in GPR B-scan images, featuring two distinct variants: DC-MAMBA V1 and DC-MAMBA V2. This architecture overcomes the scalability limitations of previous DC-ViT work by leveraging the linear complexity ($O(N)$) of State Space Models in both variants and the parameter efficiency of the Tensor-Train (TT) decomposition, specifically in DC-MAMBA V2. In the structure where the GPR signal model is defined by the superposition $x = x_c + x_t$, where x is the raw data, x_c is the clutter, and x_t is the target, the target estimation ($\hat{x}_t$) follows an integrated pipeline consisting of three fundamental stages (Figure 3.6a). First, the input data is transformed into the feature space via a convolutional patch embedding. In the encoding stage, Mamba blocks, which capture global context, and CNN blocks, which process local details, are used sequentially, with TT-Linear-supported dense connections that improve parameter efficiency. In the final stage, dense features are processed using a Pixel Shuffle-based Reconstruction mechanism to obtain a high-resolution target image.

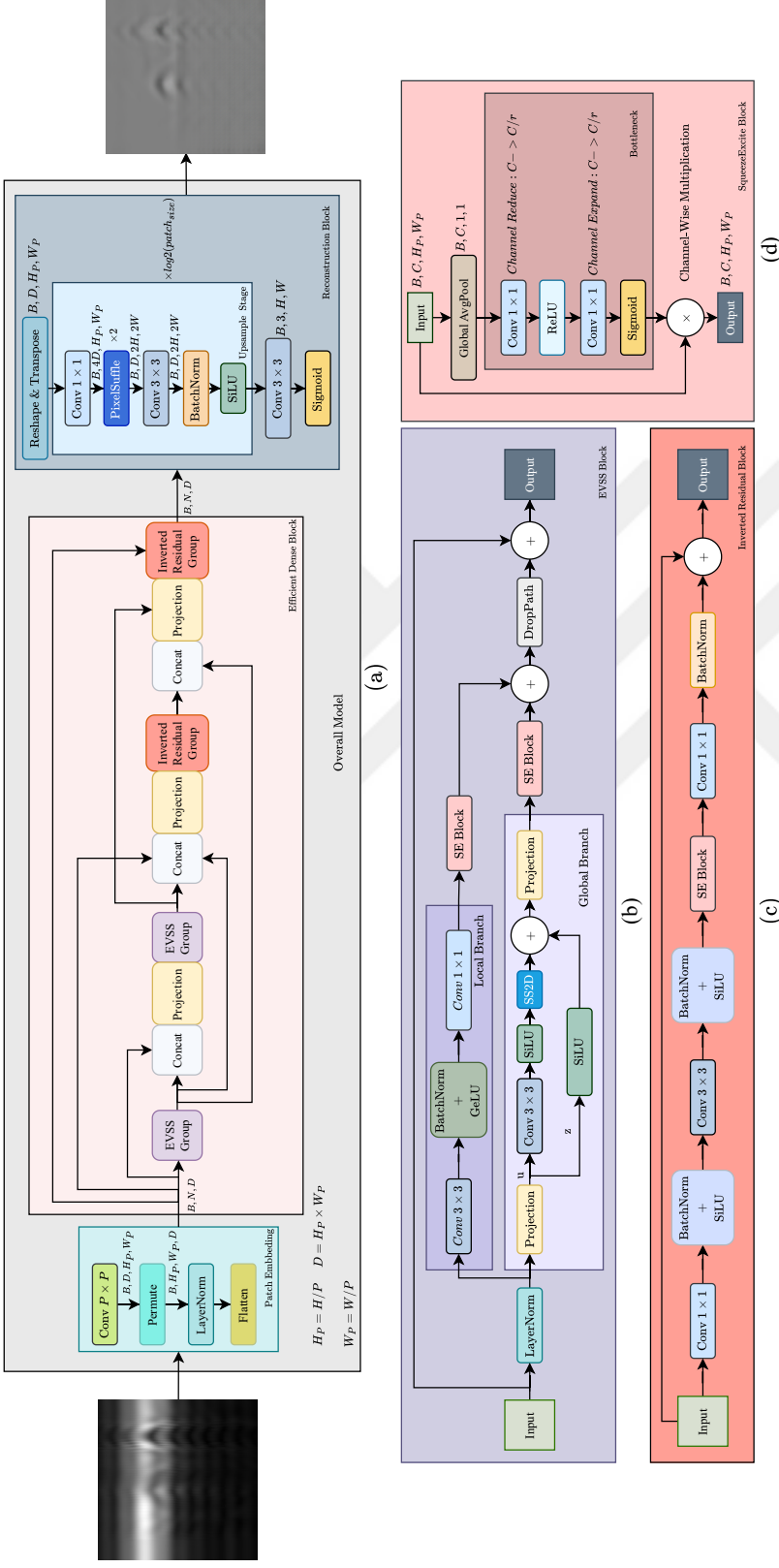


Figure 3.6: Proposed DC-MAMBA architecture. (a) The overall network structure illustrates the "Global to Local" hierarchy, where input data is processed by a Convolutional Patch Embedding module, followed by a series of Efficient Dense Blocks (comprising EVSS and Inverted Residual Groups), and finally a Reconstruction Block. (b) Detailed schematic of the Effective Visual State Space (EVSS) Block integrating the Global Branch (TensorSS2D) for long-range dependency modeling and the Local Branch (CNN) for fine detail extraction. (c) Structure of the Inverted Residual Block providing local feature enhancement using an expand-filter-compress strategy. (d) The SE Block is used to perform dynamic channel-based feature calibration within encoder modules.

3.2.2.1 Dynamic patch embedding

At the input stage, the raw GPR data of size $x \in \mathbb{R}^{H \times W \times C_{in}}$ is transformed into the feature space using a convolutional operation that preserves local spatial continuity, unlike standard approaches where non-overlapping patches are flattened. Using a 2D convolutional layer with a $P \times P$ kernel size and a P stride, the integrity of the fine hyperbolic structures in the GPR B-scans is preserved. The resulting feature sequence x_p is expressed as follows:

$$x_p \in \mathbb{R}^{N \times D}, \quad N = \frac{H \cdot W}{P^2} \quad (3.14)$$

Here, N denotes the number of tokens, and D denotes the feature dimension. The normalization applied following this process ensures that the data is transferred to subsequent encoder blocks with a stable distribution.

3.2.2.2 Tensor-train linear layer

In traditional fully connected layers, the weight matrix $W \in \mathbb{R}^{M \times K}$ is of dimension $M \times K$, and the number of parameters increases exponentially depending on the size of the input and output dimensions. This situation creates a significant memory bottleneck, especially for embedded GPR systems intended for real-time applications. In this work, the Tensor-Train (TT) Decomposition method proposed by [59] is adopted to alleviate the computational load. In this approach, instead of physically storing the large W matrix, the input (M) and output (K) dimensions are virtually decomposed into d submodes (tensor dimensions) ($M = \prod_{k=1}^d m_k$ and $K = \prod_{k=1}^d n_k$). Thus, the weight matrix is represented as a series of interconnected core tensors ($G_k \in \mathbb{R}^{r_{k-1} \times m_k \times n_k \times r_k}$) rather than a single massive block. This factorization preserves the model's expressive power while achieving a logarithmic reduction in the number of parameters. When running the model, the input is reshaped and multiplied by the kernels to produce the output immediately, so there is no need to build the weight matrix. The output is then turned into a single line of values.

$$W(i_1, \dots, i_d; j_1, \dots, j_d) = G_1[i_1, j_1]G_2[i_2, j_2] \cdots G_d[i_d, j_d] \quad (3.15)$$

Here, $W(i_1, \dots, i_d; j_1, \dots, j_d)$ represents the element of the weight tensor W indexed by the output tuple $(i_1, \dots, i_d)$ and the input tuple $(j_1, \dots, j_d)$. The values $\{r_k\}_{k=0}^d$ are called the TT-rank ($r_0 = r_d = 1$). This method lowers the parameter complexity from $O(MK)$ to $O(\sum_{k=1}^d m_k n_k r_{k-1} r_k)$. As a result, it helps avoid memory bottlenecks when working with multi-dimensional GPR data and enables the use of deeper architectures.

3.2.2.3 Efficient dense block

The Efficient Dense Block, which determines the model's feature extraction capacity, is a hybrid module in which G computation groups are arranged in a hierarchical filter structure. In the architectural design, a dense connection strategy has been implemented to prevent weak target scatterings from being lost in deeper layers, enabling the direct transfer of features from previous layers to subsequent layers. In this structure, the l -th computation group processes the cumulative tensor (X_{cat}^l) formed by concatenating the block input X_0 and the outputs of all preceding groups along the channel axis:

$$X_{cat}^l = \text{Concat}([X_0, Y_0, Y_1, \dots, Y_{l-1}]) \quad (3.16)$$

The channel expansion resulting from the concatenation process can increase computational cost. To overcome this bottleneck, the high-dimensional X_{cat}^l tensor is projected onto the model's base dimension (D) using a standard fully-connected layer (as in DC-MAMBA V1) or a high-parameter-efficiency TT-Linear projection layer (in DC-MAMBA V2) (T_{proj}) to reduce it to the base dimension (D) of the model:

$$X_{in}^l = T_{proj}(X_{cat}^l) \quad (3.17)$$

The arrangement of groups within the block follows a “Global-to-Local” hierarchy appropriate for GPR signal characteristics. The EVSS Groups, located in the first half of the block ($l < G/2$), model long-range hyperbolic tails thanks to their broad receiver areas. The Inverted Residual Groups located in the second half ($l \geq G/2$) highlight

target-focused fine details through local convolutions. Thus, the tasks of global clutter suppression and local target enhancement are integrated within a single block.

3.2.2.4 EVSS block and group

The EVSS Group has been implemented in the first layers of the network ($l < G/2$) to address the challenge of separating spatially extensive clutter from target signals. The objective is to capture the extensive spatial relationships and typical complex patterns of clutter within the hyperbolic reflections in the GPR image. The group consists of Efficient Visual State Space (EVSS) blocks, as shown in Figure 3.6b. This architecture processes the normalized input ($Z_{norm} = \text{LayerNorm}(Z_{k-1})$) through parallel Global and Local branches and then combines their outputs to form a comprehensive feature map.

The Global Branch (TensorSS2D) uses the linear complexity ($O(N)$) of State Space Models to capture global context. First, the input goes through a projection layer and is split into a signal path (u) and a gate path (z). In the DC-MAMBA V1 version, this is done with a standard linear layer. In the V2 version, a more efficient TT-Linear layer is used:

$$[u, z] = \text{Split}(\text{TT-Linear}(Z_{norm})) \quad (3.18)$$

To preserve local neighborhood information before the state space mechanism, the signal branch is processed with a 3×3 depth convolution, followed by SiLU activation. Simultaneously, the gate branch is activated to regulate the information flow:

$$u' = \text{SiLU}(\text{Conv}_{dw}(u)), \quad z' = \text{SiLU}(z) \quad (3.19)$$

The processed signal u' is fed into the 2D Selective Scanning (SS2D) mechanism. To comprehensively capture spatial interactions throughout the entire B-scan image, the signal is scanned in four main directions.

To improve computational efficiency, an atrous sampling strategy is used. The module solves discrete state equations ($h_t = \bar{A}h_{t-1} + \bar{B}u'_t$) in each direction, which helps track the continuity of hyperbolic structures even when clutter is present.

The resulting global features are then modulated by the z' gate and reflected back to the feature dimension using a second TT-Linear layer:

$$X_{global} = \text{TT-Linear}_{out}(\text{SS2D}(u') \odot z') \quad (3.20)$$

The Local Branch operates concurrently with the Global Branch and is dedicated to extracting high-frequency components and fine textural details. This module applies a series of depthwise convolutions, batch normalization, GELU activation, and pointwise convolutions. This configuration is designed to highlight local statistical differences between background clutter and target responses:

$$X_{local} = \text{Conv}_{pw}(\text{GELU}(\text{BN}(\text{Conv}_{dw}(Z_{norm})))) \quad (3.21)$$

In the final stage, feature maps from both branches are adaptively weighted and combined. The Squeeze and Excitation (SE) blocks are used to balance the contributions of global hyperbolic forms and local features before adding the residual connection:

$$Z_k = Z_{k-1} + (\text{SE}(X_{global}) + \text{SE}(X_{local})) \quad (3.22)$$

3.2.2.5 Inverted residual block and group

Global context is crucial for clutter suppression, but precise target localization needs high-frequency detail. Thus, in the deeper network layers ($l \geq G/2$), the InvertedResidualGroup module focuses on enhancing general structures and defining target boundaries. This structure consists of M consecutive Inverted Residual blocks (Figure 3.6c) and applies the “Expansion-Filtering-Compression” strategy known from MobileNetV2 to preserve parameter efficiency. Intra-block operations begin by shifting the input to a higher channel dimension ($\gamma \cdot D$) using a 1×1 convolution to prevent

weak target signals from being lost and to increase representational capacity:

$$X_{exp} = \sigma(\text{BN}(\text{Conv}_{1 \times 1}(X_{in}))) \quad (3.23)$$

Here, γ is the expansion factor. In the expanded space, a 3×3 depthwise convolution is applied to resolve spatial neighborhoods without increasing computational load:

$$X_{dw} = \sigma(\text{BN}(\text{Conv}_{dw}(X_{exp}))) \quad (3.24)$$

Among the extracted local features, the *Squeeze – and – Excitation* (SE) mechanism (Figure 3.6d) is used to highlight critical information such as hyperbolic peaks:

$$X_{se} = X_{dw} \cdot \text{Sigmoid}(\text{MLP}(\text{AvgPool}(X_{dw}))) \quad (3.25)$$

The weighted features are then reduced back to the original channel dimension (D) in the final stage using a 1×1 linear convolution. The absence of an activation function in this step is crucial to prevent information loss:

$$Y_{block} = \text{BN}(\text{Conv}_{proj}(X_{se})) \quad (3.26)$$

When the input and output dimensions are compatible, the process is completed by adding a residual connection ($X_{out} = X_{in} + Y_{block}$). This approach leverages the discriminative power of CNN-based filtering to preserve target signals after clutter removal.

3.2.2.6 Lightweight reconstruction block

In the final stage of the proposed architecture, the denoised target image ($\hat{x}_t$) is recovered from the dense features (Z_{enc}) obtained from the encoder. In this process, a Pixel Shuffle-based progressive enlargement strategy is preferred to avoid the high computational cost and checkerboard artifacts that can cause false target perception, which are commonly caused by inverse convolution layers used in the literature. At each magnification stage, repeated until the original resolution is reached, the additional

channel information required for spatial expansion is first generated using a 1×1 convolution:

$$Z'_k = \text{Conv}_{1 \times 1}(Z_{k-1}) \quad (3.27)$$

Here, the number of channels is increased to four times the target output ($r = 2$). Subsequently, the resolution is increased by applying a parameterless tensor operation that rearranges the information in the depth axis to the spatial axis:

$$Z''_k = PS(Z'_k), \quad Z''_k \in \mathbb{R}^{C_{out} \times 2H_{in} \times 2W_{in}} \quad (3.28)$$

Any discontinuities that may arise after reordering are eliminated using deep convolutions, normalization, and activation layers with high parameter efficiency:

$$Z_k = \text{SiLU}(\text{BN}(\text{Conv}_{dw}(Z''_k))) \quad (3.29)$$

After all stages, a final convolution and Sigmoid activation compress the pixel values into the range $[0, 1]$ to produce the final prediction. This process is repeated until the original resolution is reached.

3.2.3 Loss function

Loss functions are used to measure the difference between the target images provided by the network and the reference target images (the actual value for the target component). The Mean Absolute Error (MAE), which calculates the average absolute error between the target and reference target images, is a precise and fast method for handling significant variations and is expressed as follows:

$$\text{MAE} = \frac{1}{H \times W} \sum |Y - \hat{Y}| \quad (3.30)$$

where H , W , Y , and $\hat{Y}$ represent the image height, image width, reference target image, and target image, respectively.

This method, commonly used in image reconstruction problems, was employed as the loss function in the clutter-removal networks in this study.

3.2.4 Experimental results of supervised models

This section evaluates the proposed DC-MAMBA architecture through quantitative analysis on a hybrid dataset and qualitative assessment of real-world GPR measurements. To ensure a fair comparison, the benchmark models CR-Net and Attention U-Net were configured as in their original studies. The encoder backbone used for DC-MAMBA comprises 2 EVSS and 2 Inverted Residual group blocks, whereas the transformer-based DC-ViT uses 4 Transformer group blocks. Three sequentially connected blocks are used in the group blocks. The patch size was set to 8×8 and the embedding dimension (D) to 512. For DC-MAMBA, the state dimension was fixed at 16 and the expansion ratio at 1 to keep a lean computational footprint. All models were trained for 100 epochs using the MSE loss function and optimized with the AdamW algorithm. Reconstruction quality and computational efficiency were first evaluated on the synthetic test set, and the models' generalization capabilities were assessed using the signal-to-clutter improvement metric on experimental data.

The hybrid GPR dataset provided by [10] was used for training and validation, comprising 2000 synthetic, clutter-free images, partitioned into 85% training and 15% validation. To evaluate the model's generalization capability under real-world conditions, we used field measurements as shown in Table 2.1 and Figure 2.5.

All models were trained for 100 epochs using the MSE loss function and optimized with the AdamW algorithm.

3.2.4.1 Selection of model parameters

To ensure fair comparison, benchmark models CR-Net and Attention U-Net were configured as per their original studies. The encoder backbone used for DC-MAMBA utilizes 2 EVSS and 2 Inverted Residual group blocks, respectively, while the transformer-based DC-ViT uses 4 Transformer group blocks. Three sequentially connected blocks are used in the group blocks. The patch size was set to 8×8 with an embedding dimension (D) of 512. For DC-MAMBA, the state dimension was fixed at 16 and the expansion ratio at 1 to keep a lean computational footprint. All models were trained for 100 epochs using the MSE loss function and optimized with the AdamW

algorithm. Reconstruction quality and computational efficiency were first evaluated on the synthetic test set, followed by the models' generalization capabilities assessed using the signal-to-clutter improvement metric on experimental data.

3.2.4.2 Experimental results on hybrid data

Table 3.3 compares the proposed DC-MAMBA models (V1 and V2) with CR-Net, Attention U-Net, and DC-ViT. DC-MAMBA V1 uses a standard fully connected layer and achieved the best result with an average PSNR of 53.61 dB. It outperformed both DC-ViT (53.43 dB) and CR-Net (51.30 dB). DC-MAMBA V2 uses the Tensor-Train (TT) Linear layer. This model is very efficient in terms of size. As seen in Table I, DC-MAMBA V2 has only 7.90 M parameters. This is about 74% smaller than DC-ViT and 37% smaller than CR-Net. Despite the reduced size, it still achieved a high PSNR of 52.86 dB. Also, both V1 and V2 models have much lower computational cost (12.57 GFLOPS) compared to DC-ViT (43.60 GFLOPS). This confirms that the Mamba-based design is computationally efficient.

Table 3.3: Performance Comparison of Evaluated Models.

Model	Avg. PSNR (dB)	PSNR Std. Dev. (dB)	Total GFLOPs	Total Params (M)
CR-Net [10]	51.2987	4.1779	33.0844	12.490
Attention UNet [11]	52.3591	2.8100	14.0180	8.812
DC-ViT [16]	53.4317	3.0830	43.5956	30.450
DC-MAMBA V1	53.6075	2.7962	12.5722	15.140
DC-MAMBA V2	52.8586	3.0332	12.5722	7.900

3.2.4.3 Experimental results on real-world data

The visual and numerical results for the S1, S2, and S3 scenarios are presented in Figures 3.7-3.9 and Table 3.4. In the S1 scenario, where weak dielectric targets were tested, CNN-based models (CR-Net and Attention U-Net) and the Transformer-based DC-ViT were insufficient to suppress background noise, achieving low IF scores of 0.31 dB, 0.27 dB, and 1.48 dB, respectively. In contrast, DC-MAMBA V2 clearly revealed the target hyperbola with a score of 9.22 dB. The SCR graphs below the B-scan images also show this quantitative advantage. In these graphs, the DC-MAMBA V2 trace has a

high-amplitude peak at the target location and reduces the surrounding clutter. Other models, however, show more clutter components.

In the S2 scenario, which included metal targets that posed challenges due to noise tails, CR-Net produced a negative result (-0.26 dB) while Attention U-Net was ineffective (0.74 dB). Although DC-ViT showed a notable success with 7.98 dB, DC-MAMBA V2 successfully isolated dominant reflections, reaching 12.20 dB and outperforming DC-ViT by approximately 53%. As illustrated in the corresponding SCR signal graphs, DC-MAMBA V2 significantly dampens the high-amplitude oscillation clutter tails characteristic of metal targets, providing a sharp contrast to the CR-Net and Attention U-Net traces, which fail to distinguish the target signal from the background interference.

Finally, in the S3 scenario, in which an asphalt layer was added only on top of the mobile phone target to simulate heterogeneous urban conditions, DC-MAMBA V1 showed superior resilience to signal attenuation. With an IF score of 16.11 dB, the model outperformed DC-ViT (14.75 dB) by approximately 9% and produced the cleanest B-scan images among all tested architectures. Furthermore, the SCR analysis presented in the figures confirms this resilience, showing that DC-MAMBA V1 maintains the highest signal integrity and amplitude retrieval in the target region despite the asphalt-induced attenuation.

Table 3.4: Comparative Analysis of SCR Improvement (dB) for Different Scenarios.

Scenario	CR-Net [10]	Attention UNet [11]	DC-ViT [16]	DC-MAMBA V1	DC-MAMBA V2
S1	0.31	0.27	1.48	5.60	9.22
S2	-0.26	0.74	7.98	2.78	12.20
S3	0.07	1.59	14.75	16.11	12.73

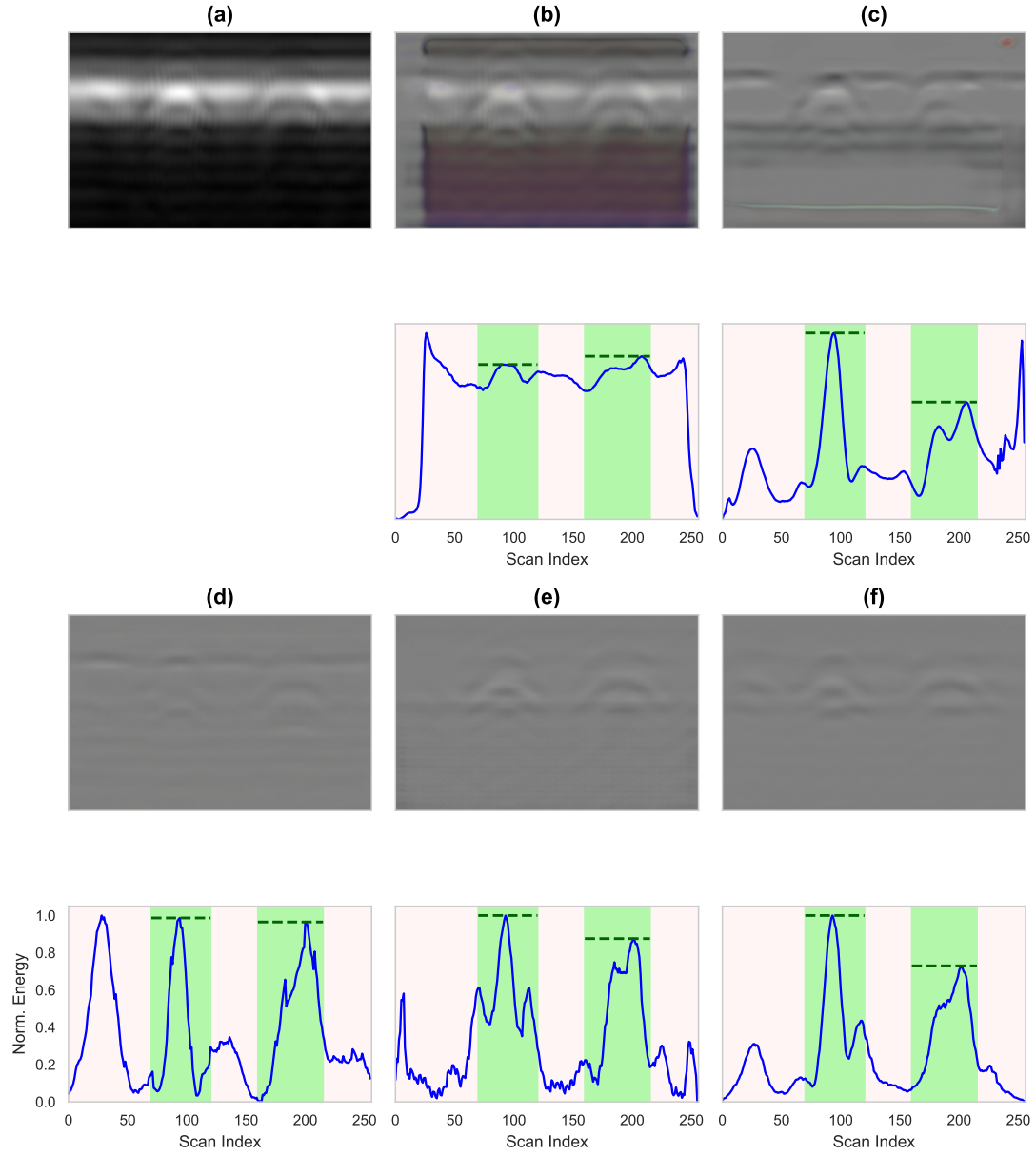


Figure 3.7: Scenario 1 (S1). The sub-figures correspond to: (a) Raw Data, (b) CR-Net, (c) Attention U-Net, (d) DC-ViT, (e) DC-MAMBA V1, and (f) DC-MAMBA V2. The corresponding SCR signal plots are presented below each B-scan image.

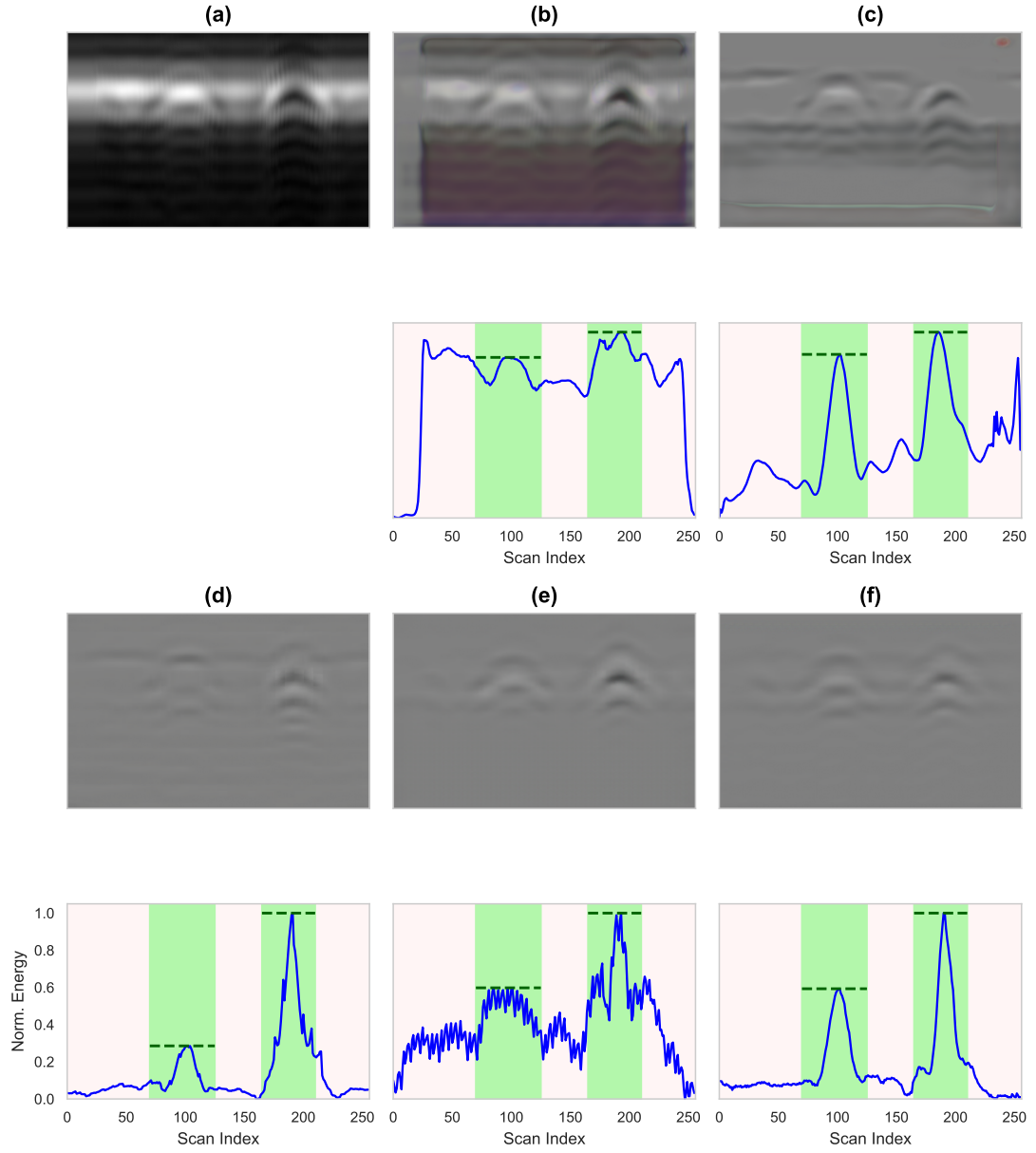


Figure 3.8: Scenario 2 (S2). The sub-figures correspond to: (a) Raw Data, (b) CR-Net, (c) Attention U-Net, (d) DC-ViT, (e) DC-MAMBA V1, and (f) DC-MAMBA V2. The corresponding SCR signal plots are presented below each B-scan image.

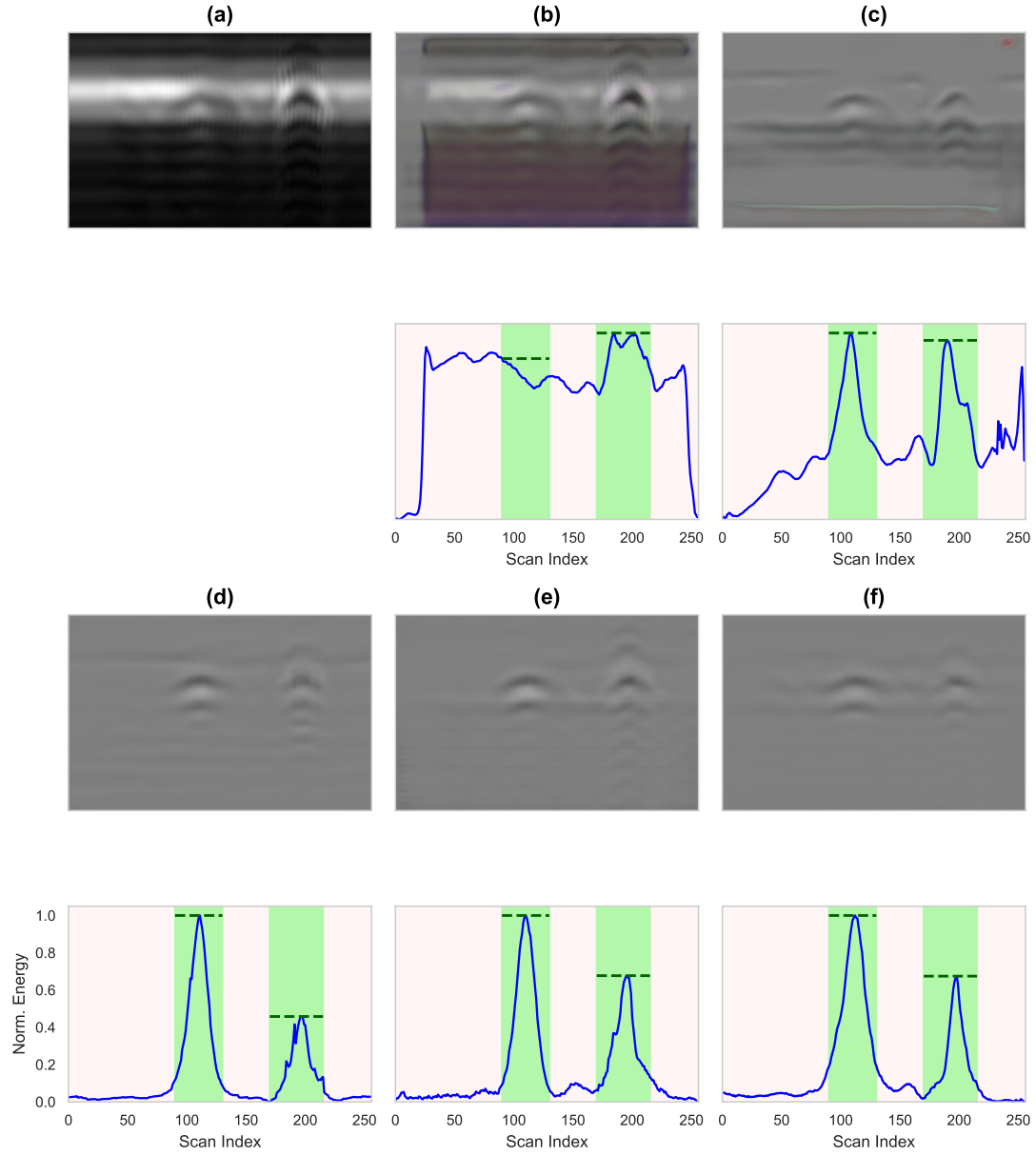


Figure 3.9: Scenario 3 (S3). The sub-figures correspond to: (a) Raw Data, (b) CR-Net, (c) Attention U-Net, (d) DC-ViT, (e) DC-MAMBA V1, and (f) DC-MAMBA V2. The corresponding SCR signal plots are presented below each B-scan image.

4. CLUTTER REMOVAL AND TARGET DETECTION IN GPR DATA USING THE PROPOSED JOINT MODELS (DEC-YOLO)& RAFDEC-YOLO)

4.1 Proposed Joint Models

GPR data used to detect subsurface targets is based on the reflection of electromagnetic waves sent into the ground due to dielectric contrast and conductivity between different media. The two-dimensional radargrams (B-scan images) obtained in this process are modeled as data matrices containing spatial information in the lateral (x -axis) and depth (y -axis) dimensions. A single B-scan image X can be mathematically expressed as follows [9,16]:

$$X = X_c + X_t, \quad (4.1)$$

where X_c represents the clutter components and X_t represents the signals from the targets. In practice, clutter (X_c) caused by direct antenna coupling, surface reflections, antenna resonance, and layers near heterogeneous surfaces often masks target signals (X_t) and reduces target detection accuracy. Therefore, clutter removal is crucial for effective target detection.

4.1.1 Model I: dec-yolo

The first architecture performs both clutter reduction and target detection in each epoch by jointly optimizing a clutter-reduction and a target-detection network with a single loss function. Thus, the effect of clutter reduction on target detection can be directly evaluated. The proposed hybrid architecture consists of two main modules, as shown in Figure 4.1. The first module includes a deep learning-based image enhancement network developed to produce a processed radargram output ($\hat{X}_t$) by removing the clutter components in the input radargrams.

$$\hat{X}_t = F_{CR}(X; \theta_{CR}), \quad (4.2)$$

where $F_{\text{CR}}(\cdot)$ represents the clutter removal process and θ_{CR} represents the corresponding network parameters. The second component includes an object detection network that directly detects targets in the processed radargrams. At this stage, a single-stage real-time object detection method such as YOLOv12 has been preferred. Thus, bounding boxes and class scores are produced from the processed radargram ($\hat{X}_t$):

$$Y = F_{\text{DET}}(\hat{X}_t; \theta_{\text{DET}}). \quad (4.3)$$

In the proposed joint approach, clutter removal and target detection are performed simultaneously and optimized using a joint loss function. The total loss function (L) is a weighted combination of the clutter removal loss (L_{CR}) and the target detection loss (L_{DET}):

$$L_{\text{total}} = \lambda_{\text{cr}} L_{\text{cr}} + \lambda_{\text{det}} L_{\text{det}} \quad (4.4)$$

Where λ_{cr} and λ_{det} are parameters that adjust the relative importance of the two tasks and are determined experimentally. Thanks to this structure, noise removal and target detection performance are directly coupled, and the weights of both models are updated simultaneously at each iteration. During the parameter update phase, each task is updated simultaneously using optimization algorithms appropriate to its nature (e.g., SGD, Adam, AdamW):

$$\theta_{\text{CR}} \leftarrow \theta_{\text{CR}} - \eta_{\text{CR}} \nabla_{\theta_{\text{CR}}} L, \quad (4.5)$$

$$\theta_{\text{DET}} \leftarrow \theta_{\text{DET}} - \eta_{\text{DET}} \nabla_{\theta_{\text{DET}}} L. \quad (4.6)$$

This architecture allows optimizing both networks' outputs within a single training cycle and enables direct interaction between the noise-suppression and target-accuracy processes. Unlike traditional sequential models, the proposed method achieves more consistent performance by simultaneously addressing inter-network information transfer and error backpropagation.

4.1.2 Model II: rafdec-yolo

In this study, as a second method, we propose RAFDeC-YOLO, a multi-task learning framework designed to simultaneously suppress clutter in GPR B-scan images and detect buried targets. As illustrated in Figure 4.2, RAFDeC-YOLO comprises a shared

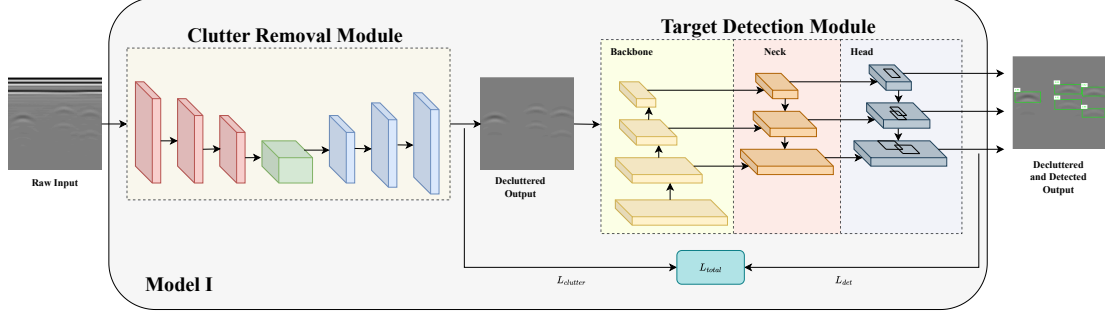


Figure 4.1: Workflow of the proposed DeC-YOLO. A raw GPR image is first fed into the Clutter Removal Model, which produces a processed image. This processed image is then processed by the Target Detection Module (e.g., YOLO) to identify and localize buried targets. Both models are trained end-to-end within a joint architecture, enabling parallel optimization of the two tasks to enhance accuracy.

backbone (encoder) for feature extraction, a Clutter Removal Branch (decoder) for image reconstruction, and a Detection Branch for target localization. These two parallel paths are interconnected by a streamlined ‘Neck-Fusion’ mechanism, Residual Adapter Fusion, which efficiently integrates cleaned features from the decoder into the detection neck.

4.1.2.1 Shared backbone: multi-scale feature extraction

The feature extraction module in the proposed architecture uses the YOLOv12 backbone to process the input GPR B-scan image $I \in \mathbb{R}^{H \times W \times C}$ in a hierarchical way. This backbone includes C3k2 blocks for efficient gradient flow, which help turn raw radar signals into useful features, and A2C2f (Attention-Based C2f) blocks for feature selection. As data moves through these layers, the network creates feature maps at different resolutions. These maps are collected using skip connections S_1, S_2, S_3 , and S_4 as shown in Figure

refjointOverall2, and the process ends with the output B_{out} at the deepest point. This pyramidal structure is meant to capture both the spectral and spatial features of the GPR data at the same time. The shallow layers near the input (S_1 and S_2) keep fine texture details and high-frequency elements like random noise, while also preserving high spatial resolution. This information is important for accurately finding the target’s boundaries. In deeper layers (S_4 and B_{out}), the network focuses on more abstract information because of the larger receptive field. These layers help model large-scale

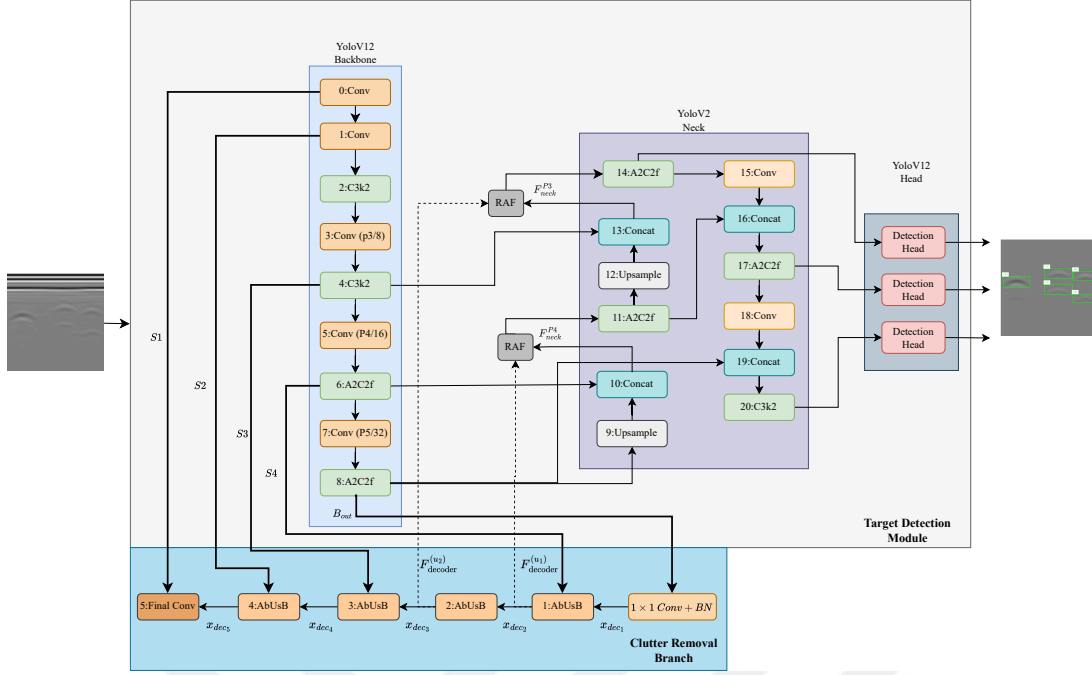


Figure 4.2: The overall architecture of the proposed RAFDeC-YOLO framework. The model features a streamlined design where the Clutter Removal Branch directly leverages backbone features, and the Detection Branch is enhanced via Residual Adapter Fusion modules.

clutter, such as soil layers, and identify the global hyperbolic patterns of buried targets. By combining information from all these levels, the backbone gives the model both the understanding needed to detect targets and the precision to locate them underground.

4.1.2.2 Clutter removal branch: decoder architecture

The clutter removal branch functions as a U-Net-style specialized decoder designed to reconstruct a noise-free GPR image from compressed feature representations generated by the encoder. The reconstruction process begins with the input of the deepest feature map (B_{out}) obtained from the backbone, which contains the most abstract semantic information about the target. To keep the process efficient and maintain important information, this high-level feature map goes through a 1×1 Convolution layer and then Batch Normalization. The feature maps are then gradually restored in space using a series of Attention-Based Up-Sampling Blocks. These blocks not only increase spatial resolution layer by layer, but also selectively highlight target-related features and suppress background noise through attention mechanisms. The process ends with a final convolutional layer that converts the deep features back into the source image space.

This produces a cleaned B-scan image in which clutter is removed and hyperbolic target signatures are preserved.

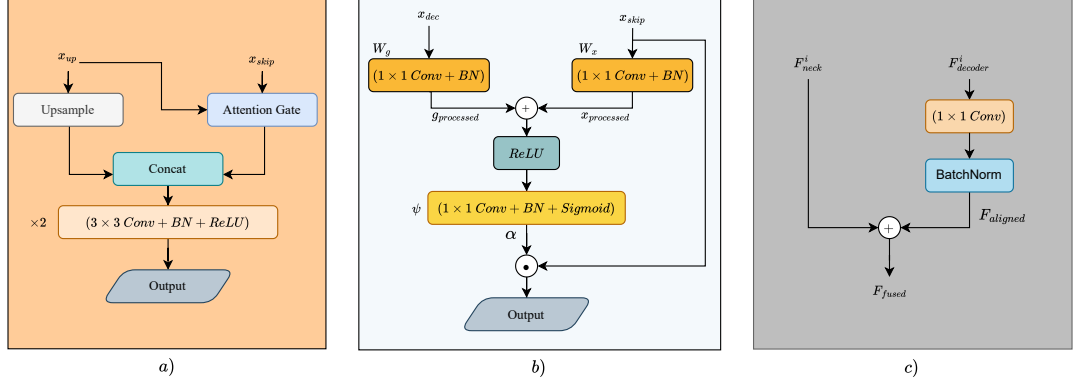


Figure 4.3: Detailed architectures of the proposed sub-modules: **a)** The Attention-Based Up-Sampling Block (AbUsB) is utilized in the decoder branch. It mitigates noise propagation by filtering the skip-connection features (x_{skip}) with an Attention Gate before concatenating them with the upsampled features (x_{up}) from the preceding layer. **b)** Internal schematic of the Attention Gate mechanism. This module computes a spatial attention map α by processing the gating signal (x_{dec}) and input features (x_{skip}), suppressing clutter-related activations through element-wise multiplication. **c)** The Residual Adapter Fusion module is integrated into the detection neck. Instead of complex gating, this streamlined mechanism aligns the channel dimensions of the decoder features ($F_{decoder}$) using a 1×1 convolution and Batch Normalization, then fuses them with the raw neck features (F_{neck}) via element-wise addition to preserve gradient flow and feature magnitude.

4.1.2.3 Attention-based up-sampling block

In each upsampling step within the decoder, the proposed Attention-Based Upsampling Block (AbUsB) is used (Figure 4.3-a). This block enhances standard skip connections by adding an “Attention Gate” mechanism, as shown in Figure 4.3(b). The Attention Gate takes two inputs: features from the deep decoder (x_{dec}) and skip connections from the encoder (x_{skip}). First, both inputs are aligned to a common channel size via 1×1 convolutions and Batch Normalization (BN); these are performed by the W_x and W_g processing blocks:

$$x_{processed} = \text{BN}(\text{Conv}_{1 \times 1}(x_{skip})) \quad (4.7)$$

$$g_{processed} = \text{BN}(\text{Conv}_{1 \times 1}(x_{dec})) \quad (4.8)$$

The aligned features are summed and passed through a ReLU activation. The resulting map is then transformed by the ψ block (1×1 Conv + BN + Sigmoid) to produce

spatial attention coefficients $\alpha \in [0, 1]$:

$$\alpha = \sigma(\psi(\text{ReLU}(x_{\text{processed}} + g_{\text{processed}}))) \quad (4.9)$$

These coefficients modulate the original skip connection via element-wise multiplication to obtain denoised features ($\hat{x}_{\text{skip}}$):

$$\hat{x}_{\text{skip}} = x_{\text{skip}} \odot \alpha \quad (4.10)$$

This process suppresses regions dominated by clutter" ($\alpha \approx 0$) while preserving hyperbolic signatures relevant to the target ($\alpha \approx 1$). In parallel with the attention mechanism, the feature map (x_{dec}) from the deep layer is upsampled to match the spatial resolution. Then, this upsampled map is concatenated with the filtered skip connection ($\hat{x}_{\text{skip}}$) in the channel dimension. In the final step, this combined feature set is passed through a block comprising two consecutive 3×3 Convolution, Batch Normalization (BN), and ReLU activation layers to refine features in the decoder path.

4.1.2.4 Feature integration: residual adapter fusion

To transfer the denoised information to the detection task without disrupting the gradient flow of the pre-trained backbone, we propose a robust Residual Adapter Fusion (RAF) mechanism as shown in Figure 4.3(c). Unlike complex gating mechanisms that may introduce optimization difficulties, this approach relies on the additive residual learning principle, treating the cleaned features as a supportive signal to the raw backbone features.

This mechanism is strategically integrated into the YOLOv12 Neck architecture at two critical scales where clutter suppression is most beneficial. At the intermediate Scale P4 (Layer 10), backbone features are fused with the u_1 feature map obtained from the First Attention-Based Up-Sampling Block. This fusion is important for capturing medium-sized hyperbolic signatures while minimizing noise. At Scale P3 (Layer 13), which processes high-resolution features for small, near-surface targets, the detection path is integrated with the u_2 output from the Second Attention-Based Up-Sampling Block. This integration preserves spatial details and reduces the surface clutter.

Let $F_{\text{neck}}^{(i)}$ represent the feature map from the Detection Branch (where $i \in \{10, 13\}$), and let $F_{\text{decoder}}^{(j)}$ represent the corresponding source feature map from the Clutter Removal

Branch ($j \in \{u_1, u_2\}$). The fusion process is executed in two sequential steps. First, a channel adaptation step is performed to align the dimensions; since the channel depth of the decoder outputs differs from the Neck features, $F_{\text{decoder}}^{(j)}$ passes through a 1×1 convolutional adapter followed by Batch Normalization (BN), resulting in the aligned feature map F_{aligned} as defined in Eq. (4.11):

$$F_{\text{aligned}} = \text{BN}(\text{Conv}_{1 \times 1}(F_{\text{decoder}}^{(j)})) \quad (4.11)$$

Subsequently, residual aggregation is applied, where the final fused feature map is computed via element-wise addition as shown in Eq. (4.12):

$$F_{\text{fused}} = F_{\text{neck}}^{(i)} + F_{\text{aligned}} \quad (4.12)$$

This additive fusion strategy functions as a residual connection, enabling the detection head to utilize the original texture details from the backbone (F_{neck}) while incorporating explicit "denoising guidance" from the decoder branch (F_{aligned}). This approach maintains training stability and prevents the loss of weak target signatures that could be over-smoothed by aggressive filtering.

4.1.3 Loss function

The proposed models (models 1 and 2) are trained using an end-to-end approach. The total loss function (L_{total}) is a weighted sum of the detection loss (L_{det}) and the denoising loss ($L_{\text{decluster}}$) (Equation 4.13).

$$L_{\text{total}} = \lambda_{\text{det}} L_{\text{det}} + \lambda_{\text{decluster}} L_{\text{decluster}} \quad (4.13)$$

This multi-task training structure forces the shared backbone to learn more generalizable, robust features that improve both tasks. The denoising task serves as self-supervised regularization for the detection task. The network must learn not only the features associated with the targets, but also the foundational physics and structure of the entire GPR image (ground layers, clutter statistics, etc.). This integrated comprehension improves performance on both tasks compared with training them separately.

4.2 Experimental Results of Joint Models

In this section, we present the results of our proposed joint models (Model I), which integrate YOLO12n with clutter-removal models such as Unet, CR-Net, Attention Unet, and DC-ViT, and RAFDeC-YOLO (Model II). These models operate sequentially for noise reduction and target detection, and we demonstrate that they perform as well as, or better than, existing models such as SuperYOLO, ESRDTM-Det, and SOD. In the experimental studies, the training hyperparameters and configurations of the RAFDeC-YOLO variants proposed in the first section, as well as the parameters for clutter-removal network architectures such as UNet, CR-Net, Attention-UNet, and DC-ViT, are detailed. The hybrid dataset and real-world scenarios are described in the Dataset section. Subsequently, the detection performance of these models is compared under two training paradigms: the “Sequential” method, where clutter removal and detection are performed independently, and the “Joint” approach, which uses end-to-end multi-task learning. First, models trained on the Hybrid dataset were evaluated on a test set drawn from the same dataset.

The hybrid ground-penetrating radar (GPR) dataset from [10] was used for both training and validation. This dataset includes synthetic raw radargrams made with gprMax and clutter-free target images, covering a wide range of dielectric properties for concrete and soil. Out of 900 randomly chosen samples, 600 were used for training the network and 300 for validation.

The model was also tested in real-world conditions, not only with synthetic data. Table 2.1 and Figure 2.5 provides more details and objectives for these experiments. In this context, the practical applicability of the model has been verified through field experiments conducted on objects such as rebar and water-filled PET bottles with low dielectric properties, as well as plastic-metal composite objects (e.g., cell phones) buried in complex asphalt layers.

Quantitative assessments were performed using standard object detection metrics, including Precision, Recall, and mAP@0.5 and mAP@0.5:0.95, as well as model complexity metrics such as FLOPs and parameter count. In addition, visual analysis was performed on some of these test results. In the final stage of the study, comparisons were

made using real data obtained from laboratory measurements to test the generalizability of the proposed approach under challenging real-world conditions.

4.2.1 Selection of model parameters

In Model I, the first method we propose, the UNet architecture used 64, 128, 256, and 512 channels in the encoder/decoder layers, respectively, and 1,024 channels in the bottleneck layer. ReLU activation and BatchNorm layers were preferred in all convolutional blocks, and upsampling stages were performed using bilinear interpolation. The same parameters as in [10] were used for CR-Net. In the DC-ViT architecture, a 4×4 patch size and a 128-dimensional embedding vector were used, as in [16]. The parameters from [11] were also selected for Attention Unet. The YOLO12 nano model used in the target detection phase was run with depth-width scales of 0.33/0.25. Detector headers were placed on P3–P5 scales, and connection sets were automatically generated. The relative weights of the box, class, and Distribution Focus Loss terms in the loss function were set to 7.5, 0.5, and 1.5, respectively. In Model II (RAFDcC-YOLO), the standard YOLO12 model was used (n/s/m). In the Clutter Removal Branch, the channel counts were selected as 512, 256, 128, and 64. All models were trained for 100 epochs under AdamW optimization with a learning rate of 1×10^{-4} and a weight decay of 1×10^{-2} .

4.2.2 Results on hybrid test data

To validate the efficacy of the proposed frameworks and benchmark them against sequential models and state-of-the-art (SOTA) methods (e.g., SuperYOLO, ESRTMDet, and SOD), extensive experiments were conducted on the hybrid test dataset. The evaluation is twofold: a comparative analysis of detection performance under different training paradigms and an assessment of computational efficiency.

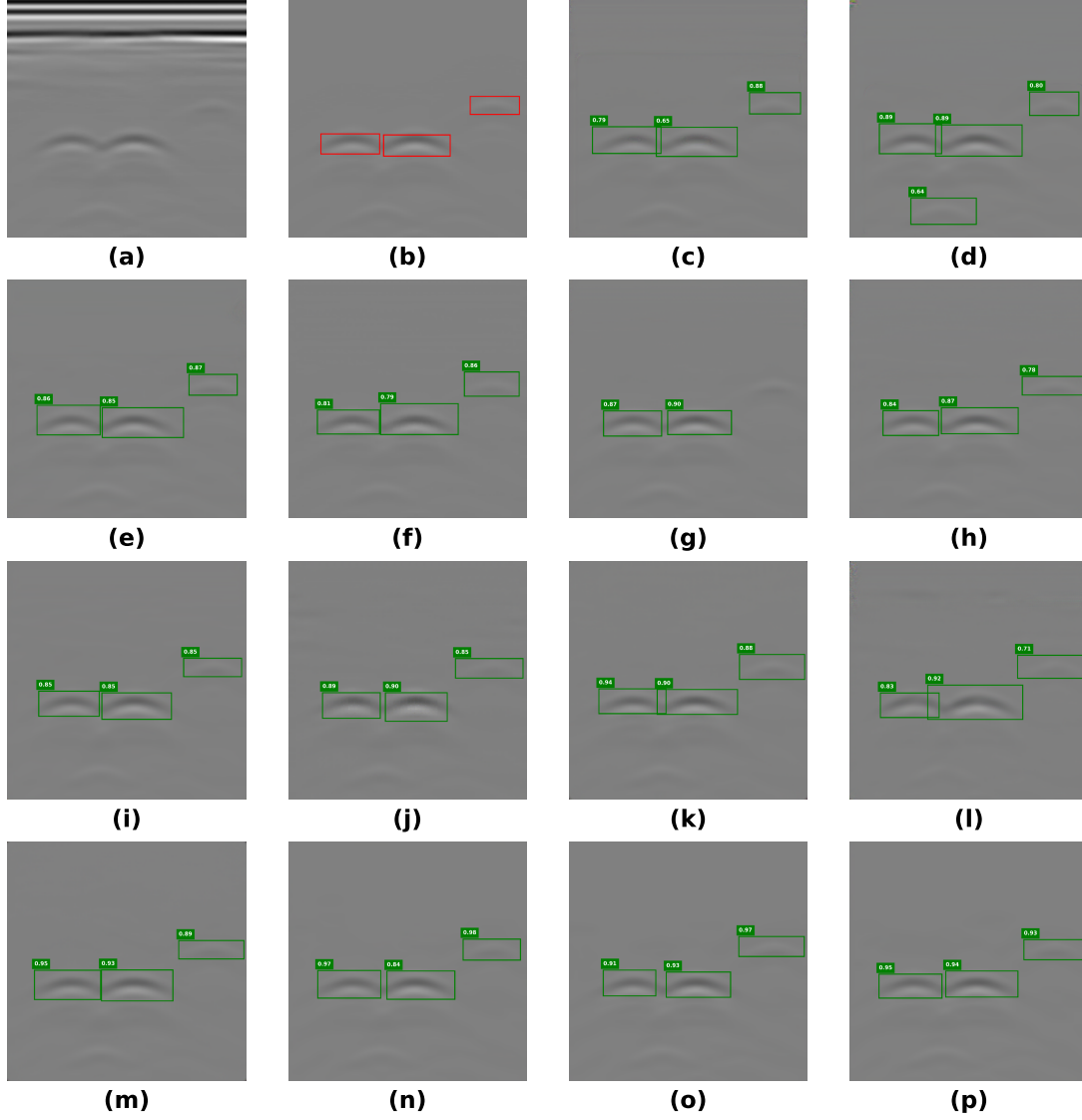


Figure 4.4: Qualitative detection results on Hybrid Dataset (Sample 1) (Moderate Clutter). (a) Raw Input, (b) Ground Truth. **Sequential Models with:** (c) UNet, (d) CR-Net, (e) Att-UNet, (f) DC-ViT. **Proposed Model I** DeC-YOLO with (g) UNet, (h) CR-Net, (i) Att-UNet, (j) DC-ViT structure as clutter removal module. **SOTA Detectors:** (k) SuperYOLO, (l) ESRTMDet, (m) SOD. **Proposed Model II** (n) RAFDeC-YOLO12n, (o) RAFDeC-YOLO12s, (p) RAFDeC-YOLO12m. The proposed methods (n-p) exhibit the highest confidence and localization precision.

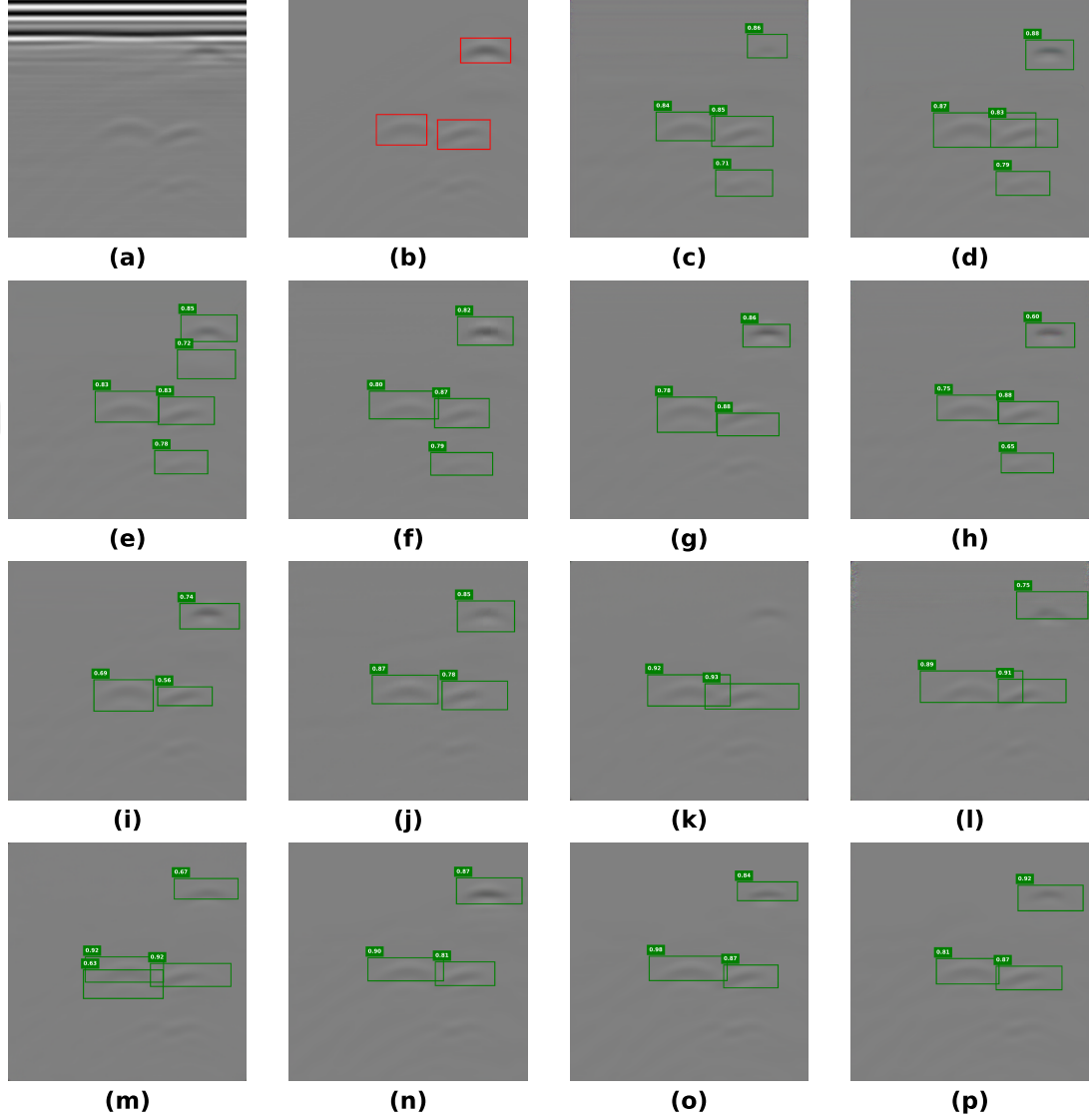


Figure 4.5: Qualitative detection results on Hybrid Dataset (Sample 2) (Heavy Clutter / Weak Targets). (a) Raw Input, (b) Ground Truth. **Sequential Models with:** (c) UNet, (d) CR-Net, (e) Att-UNet, (f) DC-ViT. **Proposed Model I** DeC-YOLO with (g) UNet, (h) CR-Net, (i) Att-UNet, (j) DC-ViT structure as clutter removal module. **SOTA Detectors:** (k) SuperYOLO, (l) ESRTMDet, (m) SOD. **Proposed Model II** (n) RAFDeC-YOLO12n, (o) RAFDeC-YOLO12s, (p) RAFDeC-YOLO12m. While many baselines fail to detect the weak target on the right, RAFDeC-YOLO12m (p) successfully localizes it with high confidence.

Table 4.1: Target detection comparison results using the hybrid dataset (Higher value is better).

Training Type	Model	Precision	Recall	mAP@0.5	mAP@0.5:0.95
Sequential	YOLO12n+UNet	0.7528	0.7622	0.6669	0.2502
	YOLO12n+CR-Net	0.7869	0.7966	0.7139	0.2590
	YOLO12n+Attention UNet	0.7932	0.8240	0.7269	0.2631
	YOLO12n+DC-ViT	0.7929	0.8019	0.7267	0.2681
Joint	DeC-YOLO with UNet (Model I)	0.7972	0.7648	0.7093	0.2621
	DeC-YOLO with CR-Net (Model I)	0.8162	0.8010	0.7607	0.3008
	DeC-YOLO with Attention UNet (Model I)	0.8393	0.8496	0.8049	0.3174
	DeC-YOLO with DC-ViT (Model I)	0.8623	0.8638	0.8299	0.3213
	SuperYOLO [43]	0.7872	0.7754	0.7022	0.2647
	ESRTMDet [44]	0.7340	0.7541	0.6528	0.2484
	SOD [60]	0.8114	0.7913	0.7150	0.2715
	RAFDeC-YOLO12n (Model II)	0.8209	0.7904	0.7384	0.2833
	RAFDeC-YOLO12s (Model II)	0.8453	0.8213	0.7815	0.3103
	RAFDeC-YOLO12m (Model II)	0.8533	0.8488	0.8141	0.3416

Table 4.2: Model complexity metrics (FLOPs, Parameters) (Lower value is better ($\downarrow$)). Best results are in bold.

Model	FLOPs (G) $\downarrow$	Parameters (M) $\downarrow$
DeC-YOLO with UNet	690.8	33.60
DeC-YOLO with CR-Net	421.2	15.05
DeC-YOLO with Attention-UNet	683.8	35.00
DeC-YOLO with DC-ViT	565.1	26.82
SuperYOLO [43]	68.1	4.93
ESRTMDet [44]	128.6	7.68
SOD [60]	68.8	5.19
RAFDDeC-YOLOn	38.7	5.56
RAFDDeC-YOLOs	55.1	12.59
RAFDDeC-YOLOm	103.4	23.78

Table 4.1 presents a quantitative comparison of RAFDeC-YOLO with recent clutter-removal and detection networks under "Sequential" and "Joint" training strategies. The results unequivocally demonstrate the superiority of the Joint Learning paradigm, as all baseline models (e.g., UNet, CR-Net) achieved significantly higher detection metrics when trained jointly rather than sequentially. This validates our hypothesis that end-to-end optimization enables the clutter-removal branch to learn more discriminative features for the detection task.

Among the evaluated models, the proposed RAFDeC-YOLO12m variant achieved state-of-the-art results across all key metrics. In terms of detection confidence, RAFDeC-YOLO12m exhibited a 5.2% increase in Precision and a 7.3% increase in Recall compared to the leading SOTA method, SOD. Regarding general detection accuracy (mAP@0.5), the proposed model surpassed SOD by approximately 13.9% and outperformed the domain-specific SuperYOLO architecture by nearly 16%. RAFDeC-YOLO12m showed a 25.8% improvement over SOD and more than 29% over SuperYOLO in the strict localization metric (mAP@0.5:0.95). Our model also outperformed SuperYOLO and SOD across all main metrics, demonstrating that the Residual Adapter Fusion mechanism works well.

Figure 4.4 and Figure 4.5 provide a visual comparison between our RAFDeC-YOLO models, standard detectors, and sequential training baselines.

The results from Sample 1, obtained from the hybrid dataset, are shown in Figure 4.4. Here, it can be seen that sequential models (c-f) and state-of-the-art detectors such as SuperYOLO and ESRTMDet generally locate targets correctly. Our proposed RAFDeC-YOLO12 demonstrated superior performance by accurately determining the full extent of hyperbolas without misclassification and with higher confidence scores (> 0.95). These results show that the Residual Adapter Fusion mechanism plays an important role in preserving the spatial information necessary for accurate localization. Figure 4.5 shows the test results performed in Sample 2. In this scenario, the baseline models (c-f) also exhibit significant issues, including high false-negative rates and an inability to reliably distinguish real targets from a noisy background. In contrast, the RAFDeC-YOLO variants (n-p) successfully detected these targets where other methods failed. This improved performance can be attributed to the integration of a specialized denoising branch that pre-processes the signal before feature fusion, enabling the detection head to identify real targets more consistently.

Although detection accuracy is essential, real-time computational performance is equally important for practical GPR applications in field settings. Table 4.2 details the computational requirements and highlights significant efficiency advantages. The RAFDeC-YOLOn (Nano) model achieves notable computational efficiency at 38.7 GFLOPs. Compared to existing methods, this model reduces computational overhead by approximately 43% relative to SuperYOLO (68.1 G) and by 70% compared to ESRTMDet (128.6 G), making it particularly suitable for deployment on edge devices with limited power resources. For scenarios prioritizing maximum detection accuracy over computational constraints, the RAFDeC-YOLOm (Medium) variant offers optimal detection performance at 103.4 GFLOPs, providing adaptable options for diverse operational requirements.

4.2.3 Results on real test data

The generalization capability and clutter robustness of the proposed RAFDeC-YOLO architecture were evaluated under real-world conditions through comprehensive laboratory tests. The experimental scenarios encompassed various material types, including metal and dielectric, multiple depths, and challenging surface clutter such as

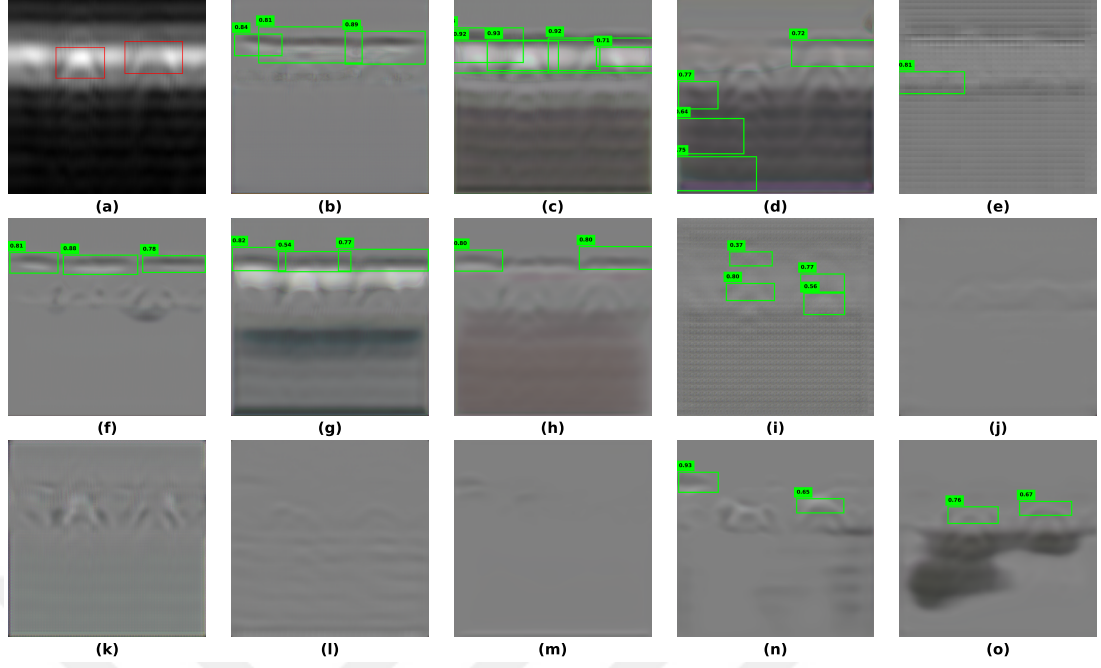


Figure 4.6: Detection results on Real Scenario 1 (Mobile Phone & PET Bottle). (a) Raw Input. **Sequential Models with:** (b) UNet, (c) CR-Net, (d) Att-UNet, (e) DC-ViT.

Proposed Model I with (f) UNet, (g) CR-Net, (h) Att-UNet, (i) DC-ViT as clutter removal module. **SOTA:** (j) SuperYOLO, (k) ESRTMDet, (l) SOD. **Proposed Model II:** (m) RAFDeC-YOLO12n, (n) RAFDeC-YOLO12s, (o) RAFDeC-YOLO12m. Note that the proposed models (m-o) successfully detect the weak dielectric PET bottle.

asphalt. The obtained numerical detection results are summarized in Table 4.3, while qualitative visual comparisons are presented in Figures 4.6, 4.7, and 4.8. Quantitative analysis reveals a dramatic performance gap between the proposed method and existing approaches. Table 4.3 shows that models with sequential training strategies (UNet, CR-Net, DC-ViT, etc.) perform relatively poorly on real data. In particular, the failure of CR-Net and DC-ViT models to achieve high scores in the sequential mode indicates that the disconnect between the clutter-removal and detection networks leads to information loss in complex real-world data. In contrast, when switching to the Joint training strategy proposed as Model I, significant performance improvements were observed across all models.

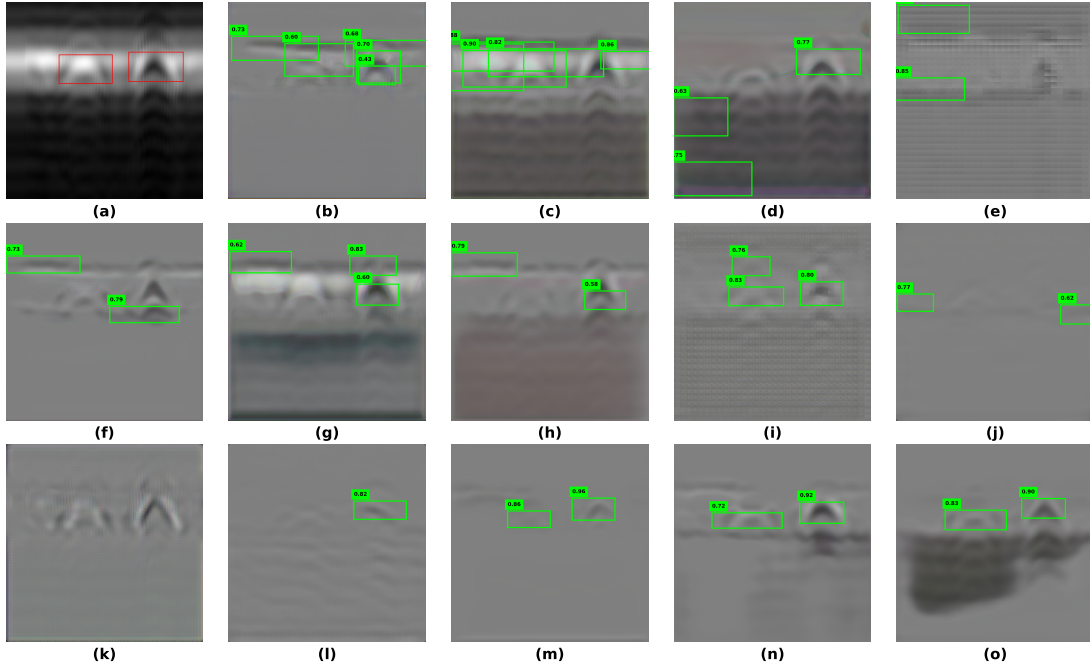


Figure 4.7: Detection results on Real Scenario 2 (Mobile Phone & Rebar). (a) Raw Input. **Sequential Models with:** (b) UNet, (c) CR-Net, (d) Att-UNet, (e) DC-ViT.

Proposed Model I with (f) UNet, (g) CR-Net, (h) Att-UNet, (i) DC-ViT as clutter removal module. **SOTA:** (j) SuperYOLO, (k) ESRTMDet, (l) SOD. **Proposed Model II:** (m) RAFDeC-YOLO12n, (n) RAFDeC-YOLO12s, (o) RAFDeC-YOLO12m. The proposed method effectively separates the strong pipe reflection from surface clutter.

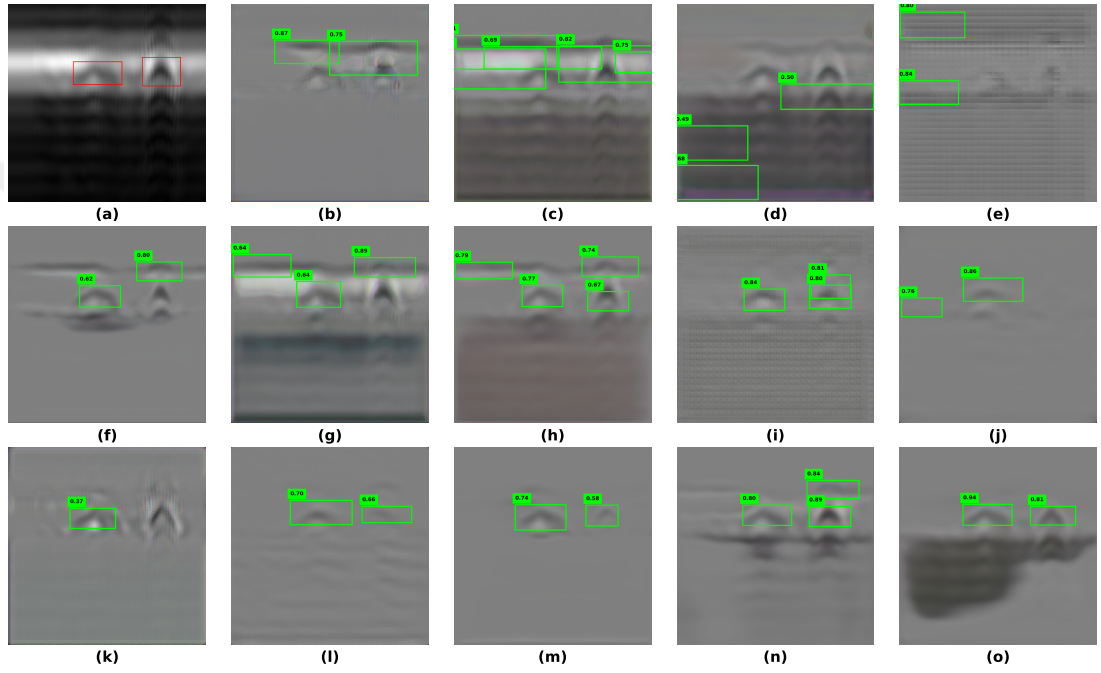


Figure 4.8: Detection results on Real Scenario 3 (Asphalt Surface). (a) Raw Input. **Sequential Models with:** (b) UNet, (c) CR-Net, (d) Att-UNet, (e) DC-ViT. **Proposed Model I** with (f) UNet, (g) CR-Net, (h) Att-UNet, (i) DC-ViT as clutter removal module. **SOTA:** (j) SuperYOLO, (k) ESRTMDet, (l) SOD. **Proposed Model II:** (m) RAFDeC-YOLO12n, (n) RAFDeC-YOLO12s, (o) RAFDeC-YOLO12m. Despite heavy clutter caused by the asphalt layer, RAFDeC-YOLO12m (o) successfully identifies the buried targets.

Table 4.3: Target detection comparison results using the laboratory measurements (Higher value is better).

Training Type	Model	Precision	Recall	mAP@0.5	mAP@0.5:0.95
Sequential	YOLO12n+UNet	0.1000	0.1667	0.0238	0.0079
	YOLO12n+CR-Net	0.0000	0.0000	0.0000	0.0000
	YOLO12n+Attention UNet	0.1538	0.3333	0.2333	0.0778
	YOLO12n+DC-ViT	0.0000	0.0000	0.0000	0.0000
Joint	DeC-YOLO with UNet (Model I)	0.1429	0.1667	0.0238	0.0079
	DeC-YOLO with CR-Net (Model I)	0.1111	0.1667	0.0333	0.0222
	DeC-YOLO with Attention UNet (Model I)	0.1250	0.1667	0.0333	0.0222
	DeC-YOLO with DC-ViT (Model I)	0.6000	1.0000	0.5000	0.1666
	SuperYOLO [43]	0.2000	0.1667	0.1667	0.0556
	ESRTMDet [44]	0.0000	0.0000	0.0000	0.0000
	SOD [60]	1.0000	0.5000	0.3333	0.1111
	RAFDeC-YOLO12n (Model II)	1.0000	0.6667	0.3222	0.1185
	RAFDeC-YOLO12s (Model II)	0.7143	0.8333	0.4167	0.1574
	RAFDeC-YOLO12m (Model II)	1.0000	1.0000	0.5583	0.2083

In the proposed Model 2, the RAFDeC-YOLO12m model achieved the highest accuracy and localization precision among all tested methods. In terms of localization precision ($mAP@0.5:0.95$), RAFDeC-YOLO12m achieved an impressive 87.5% improvement over its nearest competitor, SOD, demonstrating superior robustness against domain shifts. Similarly, for the $mAP@0.5$ metric, the proposed model outperformed SOD by approximately 67.5%. When analyzing individual detection metrics, although SOD achieved a high Precision score comparable to our model, its Recall was 50% lower than that of RAFDeC-YOLO12m. This indicates that, although the competitor missed half of the targets, the proposed RAFDeC-YOLO12m successfully identified all buried objects with no false negatives.

The visual results (Figures 4.6-4.8) corroborate these numerical findings. In the figures, red boxes indicate reference (ground truth) targets, while green boxes indicate model predictions. The detection of the water-filled PET bottle with weak reflection in Scenario 1, shown in Figure 4.6, was challenging for most models. While sequential models and ESRTMDet (k) completely missed this target, RAFDeC-YOLO12m (o) successfully detected the weak dielectric target and assigned it a high confidence score, thanks to its integrated clutter-removal capability. Scenario 2, featuring a rebar and a cell phone, is shown in Figure 4.7. In this scenario, while models such as DC-ViT (i) and SOD (l) successfully found the targets, the bounding boxes did not fully enclose them or were fragmented. The proposed RAFDeC-YOLO variations (m-o), especially the Medium model, visually supported the high $mAP@0.5:0.95$ score by producing tight and precise boxes that also included the hyperbolic arms of the targets. In the third scenario, shown in Figure 4.8, which is the most challenging test, the asphalt-under scenario, the raw data (a) contains intense clutter remains. Most sequential models (b-e) lost the targets or made incorrect detections in this clutter. RAFDeC-YOLO12m (o), however, successfully distinguished even deep targets by filtering out asphalt-induced clutter thanks to the Residual Adapter Fusion mechanism. In conclusion, both quantitative metrics and visual analyses demonstrate that the proposed RAFDeC-YOLO architecture offers more reliable, accurate, and clutter-resistant detection performance than existing methods, especially in complex, noisy real-world conditions.



5. CONCLUSION

The thesis has developed new Deep Learning designs to address the main problems of removing unwanted signals and detecting hidden objects in GPR data. By going beyond the limits of older signal processing methods and basic CNN-based approaches, the thesis has introduced a set of improved methods that significantly enhance underground imaging and target detection. The main achievements of the thesis are listed below, grouped by the specific design changes and their effects.

The first main achievement of the thesis is the development of an unsupervised AE-LRA. This approach was developed to address the problems of LRSD methods, which require extensive manual adjustments and often cannot handle various types of unwanted signals. By using an autoencoder, the AE-LRA method, which learns clutter patterns without training, eliminates the need for parameter adjustment. The results also show that the proposed method outperforms traditional matrix separation methods.

The thesis introduced DC-ViT, an unsupervised approach to overcome the limited field-of-view problem in traditional CNNs. While CNNs struggle to capture long-range dependencies, DC-ViT effectively captures broader context across the entire GPR image via its self-attention mechanism. Additionally, a LeFF was added to the model to prevent the loss of small details that can occur in Transformers. This combined design, which handles both large, complex patterns and small target features, provided much better image quality than standard deep learning models and preserved important target shapes.

Although transformer-based models achieve better results than CNNs, their high computational requirements can pose challenges for real-time applications. To overcome this, the thesis proposes a DC-MAMBA design based on State-Space Models for GPR. This new approach enables much more efficient processing of long-range connections in time-depth data using less computational power. DC-MAMBA V2 has significantly reduced the number of parameters and computational cost by using Tensor-Train

decomposition. This lighter design makes the model much smaller while also effectively eliminating complexity, making it a good choice for GPR systems with limited resources. DC-MAMBA models have proven their ability to detect difficult target signals that even other advanced models cannot find.

One of the most important contributions of this thesis is the proposal of the RAFDeC-YOLO architecture, a joint learning framework for clutter removal and target detection in GPR. Unlike traditional sequential methods that treat clutter removal and object detection as independent tasks, RAFDeC-YOLO optimizes both processes simultaneously within a single end-to-end trainable network. With the proposed Residual Adapter Fusion mechanism, features from the clutter-removal branch are transferred to the neck of the target detection network. This allows the target detection network to detect features in decluttered images, thereby improving target detection performance. This holistic approach has enabled the model to outperform even the most advanced sensors and sequential basis values, particularly in consistently detecting weak and deeply embedded targets. Furthermore, the developed “Nano” variant of this architecture offers a highly efficient solution providing real-time operation suitable for field applications.

In conclusion, the thesis has successfully bridged the gap between advanced deep learning technologies—such as Vision Transformers and State Space Models—and GPR signal processing. By addressing critical issues related to parameter dependency, global context modeling, and computational efficiency, the proposed methods have significantly advanced the state of the art. The developed joint frameworks not only enhance the visual quality of GPR images but also provide robust, real-time automated detection capabilities, making a substantial contribution to the practical use of GPR in modern engineering and geophysical applications.

Future research will focus on using these designs for C-Scan GPR data, improving their performance across different soil types, and testing faster versions of the lightweight models on specialized embedded systems to enable their use in commercial devices.

REFERENCES

- [1] **VanderGaast, B.W., McFee, J.E., Russell, K.L. and Faust, A.A.** (2015). Design and validation of inert homemade explosive simulants for ground penetrating radar, *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XX*, volume 9454, SPIE, pp.373–385.
- [2] **da Costa, A.M.B., Martins, D., Rodrigues, D., Fernandes, L., Moura, R. and Áurea Madureira-Carvalho** (2022). Ground Penetrating Radar for Buried Explosive Devices Detection: A Case Studies Review, *Australian Journal of Forensic Sciences*, 54(4), 559–578.
- [3] **Abujarad, F., Jostingmeier, A. and Omar, A.** (2004). Clutter removal for landmine using different signal processing techniques, *Proceedings of the Tenth International Conference on Grounds Penetrating Radar, 2004. GPR 2004.*, pp.697–700.
- [4] **Khan, U.S. and Al-Nuaimy, W.** (2010). Background removal from GPR data using Eigenvalues, *Proceedings of the XIII International Conference on Ground Penetrating Radar*, pp.1–5.
- [5] **Verma, P.K., Gaikwad, A.N., Singh, D. and Nigam, M.** (2009). Analysis of clutter reduction techniques for through wall imaging in UWB range, *Progress In Electromagnetics Research B*, 17, 29–48.
- [6] **Song, X., Xiang, D., Zhou, K. and Su, Y.** (2017). Improving RPCA-based clutter suppression in GPR detection of antipersonnel mines, *IEEE Geoscience and Remote Sensing Letters*, 14(8), 1338–1342.
- [7] **Kumlu, D. and Erer, I.** (2019). Improved clutter removal in GPR by robust nonnegative matrix factorization, *IEEE Geoscience and Remote Sensing Letters*, 17(6), 958–962.
- [8] **Kumlu, D. and Erer, I.** (2020). Ground penetrating radar clutter removal via randomized low rank and sparse decomposition for missing data case, *International Journal of Remote Sensing*, 41(19), 7680–7699.
- [9] **Temlioglu, E. and Erer, I.** (2021). A novel convolutional autoencoder-based clutter removal method for buried threat detection in ground-penetrating radar, *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–13.
- [10] **Sun, H.H., Cheng, W. and Fan, Z.** (2022). Learning to remove clutter in real-world GPR images using hybrid data, *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–14.

- [11] **Panda, S.L., Sahoo, U.K., Maiti, S. and Sasmal, P.** (2024). An Attention U-Net-Based Improved Clutter Suppression in GPR Images, *IEEE Transactions on Instrumentation and Measurement*, 73, 1–11.
- [12] **Ni, Z.K., Shi, C., Pan, J., Zheng, Z., Ye, S. and Fang, G.** (2022). Declutter-GAN: GPR B-scan data clutter removal using conditional generative adversarial nets, *IEEE Geoscience and Remote Sensing Letters*, 19, 1–5.
- [13] **Ge, J., Sun, H., Shao, W., Liu, D., Liu, H., Zhao, F., Tian, B. and Liu, S.** (2024). Wavelet-GAN: A GPR Noise and Clutter Removal Method Based on Small Real Datasets, *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1–14.
- [14] **Xu, T., Yuan, D., Sun, W., Yang, G. and Li, B.** (2023). Coupled-Learning GAN for Inversion of GPR Pipe Images, *IEEE Geoscience and Remote Sensing Letters*, 20, 1–5.
- [15] **Kayacan, Y.E. and Erer, I.** (2024). Autoencoder Guided Low-Rank Approximation Approach for Clutter Removal in GPR Images, *2024 47th International Conference on Telecommunications and Signal Processing (TSP)*, pp.332–335.
- [16] **Kayacan, Y.E. and Erer, I.** (2024). A Vision-Transformer-Based Approach to Clutter Removal in GPR: DC-ViT, *IEEE Geoscience and Remote Sensing Letters*, 21, 1–5.
- [17] **Tjora, S., Eide, E. and Lundheim, L.** (2004). Evaluation of methods for ground bounce removal in GPR utility mapping, *Proceedings of the Tenth International Conference on Grounds Penetrating Radar, 2004. GPR 2004.*, volume 1, pp.379–382.
- [18] **Liu, J., Wu, R., Li, T. and Zhang, B.** (2006). Novel Ground Bounce Removal Algorithms Based on Non-homogeneous Detector, *2006 CIE International Conference on Radar*, pp.1–5.
- [19] **Grzegorzcyk, T.M., Zhang, B. and Cornick, M.T.** (2013). Optimized SVD Approach for the Detection of Weak Subsurface Targets From Ground-Penetrating Radar Data, *IEEE Transactions on Geoscience and Remote Sensing*, 51(3), 1635–1642.
- [20] **Candès, E.J., Li, X., Ma, Y. and Wright, J.** (2011). Robust principal component analysis?, *J. ACM*, 58(3), <https://doi.org/10.1145/1970392.1970395>.
- [21] **Zhou, T. and Tao, D.** (2011). GoDec: randomized low-rank & sparse matrix decomposition in noisy case, *Proceedings of the 28th International Conference on International Conference on Machine Learning, ICML'11*, Omnipress, Madison, WI, USA, p.33–40.

- [22] **Song, X., Xiang, D., Zhou, K. and Su, Y.** (2019). Fast Prescreening for GPR Antipersonnel Mine Detection via Go Decomposition, *IEEE Geoscience and Remote Sensing Letters*, 16(1), 15–19.
- [23] **Shu, X., Porikli, F. and Ahuja, N.** (2014). Robust Orthonormal Subspace Learning: Efficient Recovery of Corrupted Low-Rank Matrices, *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp.3874–3881.
- [24] **Kumlu, D. and Erer, I.** (2020). GPR Clutter Reduction by Robust Orthonormal Subspace Learning, *IEEE Access*, 8, 74145–74156.
- [25] **Zhang, L., Chen, Z., Zheng, M. and He, X.** (2011). Robust non-negative matrix factorization, *Frontiers of Electrical and Electronic Engineering in China*, 6(2), 192–200, <https://doi.org/10.1007/s11460-011-0128-0>.
- [26] **Ni, Z.K., Ye, S., Shi, C., Li, C. and Fang, G.** (2022). Clutter Suppression in GPR B-Scan Images Using Robust Autoencoder, *IEEE Geoscience and Remote Sensing Letters*, 19, 1–5.
- [27] **Xiong, L., Chen, X. and Schneider, J.** (2011). Direct Robust Matrix Factorization for Anomaly Detection, *2011 IEEE 11th International Conference on Data Mining*, pp.844–853.
- [28] **Liu, L., Wu, Z., Xu, H., Wang, B. and Li, J.** (2022). GPR Clutter Removal Based on Factor Group-Sparse Regularization, *IEEE Geoscience and Remote Sensing Letters*, 19, 1–5.
- [29] **Ni, Z.K., Shi, C., Pan, J., Zheng, Z., Ye, S. and Fang, G.** (2022). Declutter-GAN: GPR B-Scan Data Clutter Removal Using Conditional Generative Adversarial Nets, *IEEE Geoscience and Remote Sensing Letters*, 19, 1–5.
- [30] **Dai, Q., Lee, Y.H., Sun, H.H., Qian, J., Yusof, M.L.M., Lee, D. and Yucel, A.C.** (2024). Learning From Clutter: An Unsupervised Learning-Based Clutter Removal Scheme for GPR B-Scans, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17, 19668–19681.
- [31] **Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. and Houshy, N.** (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, *CoRR*, abs/2010.11929, <https://arxiv.org/abs/2010.11929>, 2010.11929.
- [32] **Yuan, K., Guo, S., Liu, Z., Zhou, A., Yu, F. and Wu, W.** (2021). *Incorporating Convolution Designs into Visual Transformers*, <https://arxiv.org/abs/2103.11816>, 2103.11816.
- [33] **Girshick, R., Donahue, J., Darrell, T. and Malik, J.** (2014). *Rich feature hierarchies for accurate object detection and semantic segmentation*, <https://arxiv.org/abs/1311.2524>, 1311.2524.

- [34] **Ren, S., He, K., Girshick, R. and Sun, J.** (2016). *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*, <https://arxiv.org/abs/1506.01497>, 1506.01497.
- [35] **Redmon, J., Divvala, S., Girshick, R. and Farhadi, A.** (2016). *You Only Look Once: Unified, Real-Time Object Detection*, <https://arxiv.org/abs/1506.02640>, 1506.02640.
- [36] **Pham, M.T. and Lefèvre, S.** (2018). Buried Object Detection from B-Scan Ground Penetrating Radar Data Using Faster-RCNN, *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pp.6804–6807.
- [37] **Köprücü, F., Erer, I. and Kumlu, D.** (2021). Clutter Aware Deep Detection for Subsurface Radar Targets, *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pp.4868–4871.
- [38] **Hou, F., Fang, M., Xianghuan Luo, T., Fan, X. and Guo, Y.** (2024). Dual-Task GPR Method: Improved Generative Adversarial Clutter Suppression Network and Adaptive Target Localization Algorithm in GPR Image, *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1–13.
- [39] **Wei, C., Zhou, X., Liu, S. and Chen, H.** (2024). Enhanced Anomaly Detection in GPR Data by Combining Spatial and Dynamic Information, *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1–10.
- [40] **Wang, H., Liu, Q. and Zhou, L.** (2025). Underground target localization method for ground penetrating radar based on deep learning, *Measurement*, 253, 117647, <https://www.sciencedirect.com/science/article/pii/S0263224125010061>.
- [41] **Huang, S.C., Le, T.H. and Jaw, D.W.** (2021). DSNet: Joint Semantic Learning for Object Detection in Inclement Weather Conditions, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(8), 2623–2633.
- [42] **Wang, Y., Yan, X., Zhang, K., Gong, L., Xie, H., Wang, F.L. and Wei, M.** (2022). TogetherNet: Bridging Image Restoration and Object Detection Together via Dynamic Enhancement Learning, *Computer Graphics Forum*, 41(7), 465–476.
- [43] **Zhang, J., Lei, J., Xie, W., Fang, Z., Li, Y. and Du, Q.** (2023). SuperYOLO: Super Resolution Assisted Object Detection in Multimodal Remote Sensing Imagery, *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1–15.
- [44] **Liu, F., Chen, R., Zhang, J., Ding, S., Liu, H., Ma, S. and Xing, K.** (2023). ESRTMDet: An End-to-End Super-Resolution Enhanced Real-Time Rotated Object Detector for Degraded Aerial Images, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16, 4983–4998.

- [45] **Zhao, M., Zhang, X. and Kaup, A.** (2023). Multitask Learning for SAR Ship Detection With Gaussian-Mask Joint Segmentation, *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1–16.
- [46] **Wang, Y., Qin, H., Zhang, D., Wu, T. and Pan, S.** (2024). Domain Adaption YOLO Network to Enhance Target Detection in GPR Images, *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1–15.
- [47] **Kumlu, D. and Erer, I.** (2018). Clutter removal in GPR images using non-negative matrix factorization, *Journal of Electromagnetic Waves and Applications*, 32(16), 2055–2066.
- [48] **Temlioglu, E. and Erer, I.** (2016). Clutter removal in ground-penetrating radar images using morphological component analysis, *IEEE Geoscience and Remote Sensing Letters*, 13(12), 1802–1806.
- [49] **Strudel, R., Garcia, R., Laptev, I. and Schmid, C.** (2021). Segmenter: Transformer for Semantic Segmentation, *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp.7262–7272.
- [50] **Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S. and Yang, M.H.** (2022). Restormer: Efficient Transformer for High-Resolution Image Restoration, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.5728–5739.
- [51] **Thisanke, H., Deshan, C., Chamith, K., Seneviratne, S., Vidanaarachchi, R. and Herath, D.** (2023). Semantic segmentation using Vision Transformers: A survey, *Engineering Applications of Artificial Intelligence*, 126, 106669, <https://www.sciencedirect.com/science/article/pii/S0952197623008539>.
- [52] **Wang, Y., Deng, Y., Zheng, Y., Chattopadhyay, P. and Wang, L.** (2025). Vision Transformers for Image Classification: A Comparative Survey, *Technologies*, 13(1), <https://www.mdpi.com/2227-7080/13/1/32>.
- [53] **Yao, C., Jin, S., Liu, M. and Ban, X.** (2022). Dense residual Transformer for image denoising, *Electronics*, 11(3), 418.
- [54] **Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J. and Li, H.** (2022). Uformer: A general u-shaped transformer for image restoration, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp.17683–17693.
- [55] **Gu, A. and Dao, T.** (2023). Mamba: Linear-Time Sequence Modeling with Selective State Spaces, *arXiv preprint arXiv:2312.00752*.
- [56] **Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W. and Wang, X.** (2024). Vision Mamba: Efficient Visual Representation Learning with Bidirectional State Space Model, *R. Salakhutdinov, Z. Kolter, K. Heller, A. Weller,*

- N. Oliver, J. Scarlett and F. Berkenkamp, editors, Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, PMLR, pp.62429–62442, <https://proceedings.mlr.press/v235/zhu24f.html>.
- [57] **Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., Jiao, J. and Liu, Y.** (2024). VMamba: Visual State Space Model, *A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak and C. Zhang, editors, Advances in Neural Information Processing Systems*, volume 37, Curran Associates, Inc., pp.103031–103063, https://proceedings.neurips.cc/paper_files/paper/2024/file/baa2da9ae4bfed26520bb61d259a3653-Paper-Conference.pdf.
- [58] **Pei, X., Huang, T. and Xu, C.** (2025). EfficientVMamba: Atrous Selective Scan for Light Weight Visual Mamba, *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(6), 6443–6451, <https://ojs.aaai.org/index.php/AAAI/article/view/32690>.
- [59] **Novikov, A., Podoprikin, D., Osokin, A. and Vetrov, D.P.** (2015). Tensorizing Neural Networks, *C. Cortes, N. Lawrence, D. Lee, M. Sugiyama and R. Garnett, editors, Advances in Neural Information Processing Systems*, volume 28, Curran Associates, Inc., pp.442–450, https://proceedings.neurips.cc/paper_files/paper/2015/file/6855456e2fe46a9d49d3d3af4f57443d-Paper.pdf.
- [60] **Liu, J., Zhang, J., Ni, Y., Chi, W. and Qi, Z.** (2024). Small-Object Detection in Remote Sensing Images With Super-Resolution Perception, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17, 15721–15734.

CURRICULUM VITAE

Name SURNAME: Yavuz Emre KAYACAN

EDUCATION:

- **B.Sc.:** Kocaeli University, Faculty of Engineering, Electrical Engineering
- **M.Sc.:** Istanbul Technical University, Faculty of Electrical-Electronics, Electrical Engineering

PROFESSIONAL EXPERIENCE AND REWARDS:

- Research Assistant, Beykent University (Biomedical Engineering Dept.)
- Consultant, Hardware Development for Electromyogram Device, Magnus Electronics
- Reviewer, IEEE Journal of Biomedical and Health Informatics

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Kayacan, Y. E., Erer, I.** (2026). Efficient GPR Clutter Removal via State Space Modeling *IEEE Geoscience and Remote Sensing Letters*, (Research Article - In preparation)
- **Kayacan, Y. E., Erer, I., Paker, S.** (2026). Joint Deep Learning for Simultaneous Clutter Removal and Buried Object Detection in GPR. *IEEE Access*, (Research Article - Accepted)
- **Kayacan, Y. E., Erer, I.** (2024). A Vision-Transformer-Based Approach to Clutter Removal in GPR: DC-ViT. *IEEE Geoscience and Remote Sensing Letters*, 21, 1-5. (Research Article - Published)
- **Kayacan, Y. E., Erer, I.** (2024). Autoencoder Guided Low-Rank Approximation Approach for Clutter Removal in GPR Images. *2024 47th International Conference on Telecommunications and Signal Processing (TSP)*, 332-335. (Conference Paper - Published)

OTHER PUBLICATIONS, PRESENTATIONS AND PATENTS:

- **Kayacan, Y. E., Ozdogan, K., Colak, M. U., Aydin, O., Turan, A. E., Erer, I.** (2024). Combining Image Restoration with Deep Detectors for Improved Target Detection in GPR. *2024 47th International Conference on Telecommunications and Signal Processing (TSP)*. (Conference Paper - Published)
- **Kayacan, Y. E., Erer, I.** (2024). Enhancing Deep Learning Networks Performance By Using B-Spline Activation Functions for Clutter Removal in GPR. *2024 47th International Conference on Telecommunications and Signal Processing (TSP)*. (Conference Paper - Published)

