UNIVERSITY OF THESSALY

SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Imaging from antenna array data with machine learning techniques

Diploma Thesis

Barbounakis Sokratis

Supervisor: Argyriou Antonios

July 2025

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

UNIVERSITY OF THESSALY

SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Imaging from antenna array data with machine learning techniques

Diploma Thesis

Barbounakis Sokratis

Supervisor: Argyriou Antonios

July 2025

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Απεικόνιση από δεδομένα συστοιχιών κεραιών με τεχνικές μηχανικής μάθησης

Διπλωματική Εργασία

Μπαρμπουνάκης Σωκράτης

Επιβλέπων: Αργυρίου Αντώνιος

Ιούλιος 2025

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Approved by the Examination Committee:

Supervisor **Argyriou Antonios**
Professor, Department of Electrical and Computer Engineering, University of Thessaly

Member **Korakis Athanasios**
Professor, Department of Electrical and Computer Engineering, University of Thessaly

Member **Flegkas Parisis**
Assistant Professor, Department of Electrical and Computer Engineering, University of Thessaly

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Acknowledgements

I would like to express my sincere gratitude to my family for their unwavering support, encouragement and love throughout my academic journey.

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

DISCLAIMER ON ACADEMIC ETHICS AND INTELLECTUAL PROPERTY RIGHTS

«Being fully aware of the implications of copyright laws, I expressly state that this diploma thesis, as well as the electronic files and source codes developed or modified in the course of this thesis, are solely the product of my personal work and do not infringe any rights of intellectual property, personality and personal data of third parties, do not contain work / contributions of third parties for which the permission of the authors / beneficiaries is required and are not a product of partial or complete plagiarism, while the sources used are limited to the bibliographic references only and meet the rules of scientific citing. The points where I have used ideas, text, files and / or sources of other authors are clearly mentioned in the text with the appropriate citation and the relevant complete reference is included in the bibliographic references section. I also declare that the results of the work have not been used to obtain another degree. I fully, individually and personally undertake all legal and administrative consequences that may arise in the event that it is proven, in the course of time, that this thesis or part of it does not belong to me because it is a product of plagiarism».

The declarant

Barbounakis Sokratis

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Diploma Thesis

Imaging from antenna array data with machine learning techniques

Barbounakis Sokratis

Abstract

This study focuses on enhancing the quality of spatial spectral maps used for estimating the Direction of Arrival (DoA) of signals, through the development of a deep neural network, specifically a denoising autoencoder. The process begins with the incoming signal impinging on a Uniform Rectangular Array (URA) of antennas, where multiple elements capture the wavefront. A beamforming algorithm based on the two-dimensional Discrete Fourier Transform (2D-DFT) is then applied to generate a spatial spectral map, which represent the spatial distribution of the signal's power. The peak on these map indicates the estimated direction of arrival.

In low Signal-to-Noise Ratio (SNR) environments, these spectral maps often suffer from severe distortions that obscure the signal's structure, leading to inaccurate DoA estimations. To address this issue, the proposed denoising autoencoder enhances the quality of the spectral maps by restoring the clarity and reliability of the underlying signal features. This denoising stage significantly improves peak detection accuracy and consequently, the reliability of DoA estimation.

To evaluate the effectiveness of the proposed approach, a dataset of spatial spectral maps was generated and used to train the denoising autoencoder. The model's performance was assessed primarily based on its ability to accurately detect peaks within the spectral maps. Experimental results demonstrate high accuracy, with peak detection success rates consistently around 90% for SNR values ranging from -20 dB to 20 dB. However, at extremely low SNR levels (approximately -23 dB), the increased noise results in false or shifted peaks, reducing detection accuracy to approximately 61%. This value represents the lowest SNR level used in the study, as our goal was to identify the operational limit of the method— the point below which performance begins to degrade significantly.

Keywords:

Phased Arrays, Uniform Rectangular Array, Direction of Arrival Estimation, Deep Learning,

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Denoising Autoencoder

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Διπλωματική Εργασία

Απεικόνιση από δεδομένα συστοιχιών κεραιών με τεχνικές μηχανικής μάθησης

Μπαρμπουνάκης Σωκράτης

Περίληψη

Η παρούσα μελέτη επικεντρώνεται στη βελτίωση της ποιότητας των χωρικών φασματικών χαρτών που χρησιμοποιούνται για την εκτίμηση της κατεύθυνσης άφιξης (Direction of Arrival – DoA) σημάτων, μέσω της ανάπτυξης ενός βαθιού νευρωνικού δικτύου και συγκεκριμένα ενός αυτοκωδικοποιητή αποθορυβοποίησης. Η διαδικασία ξεκινά με την πρόσπτωση του εισερχόμενου σήματος σε μία ομοιόμορφη ορθογώνια διάταξη (Uniform Rectangular Array – URA) κεραιών, όπου τα επιμέρους στοιχεία καταγράφουν το προσπίπτον κύμα. Στη συνέχεια, εφαρμόζεται ένας αλγόριθμος κατευθυντικής διαμόρφωσης βασισμένος στον δισδιάστατο Διακριτό Μετασχηματισμό Fourier (2D-DFT), ο οποίος παράγει έναν χωρικό φασματικό χάρτη που απεικονίζει την κατανομή της ισχύος του σήματος στον χώρο. Η κορυφή αυτού του χάρτη υποδεικνύει την εκτιμώμενη κατεύθυνση άφιξης του σήματος.

Σε περιβάλλοντα με χαμηλό λόγο σήματος προς Θόρυβο (SNR), οι φασματικοί χάρτες παρουσιάζουν συχνά έντονες αλλοιώσεις που αποκρύπτουν τη δομή του σήματος, οδηγώντας σε ανακριβείς εκτιμήσεις της κατεύθυνσης άφιξης. Για την αντιμετώπιση αυτού του προβλήματος, ο προτεινόμενος αυτοκωδικοποιητής αποθορυβοποίησης βελτιώνει την ποιότητα των φασματικών χαρτών αποκαθιστώντας τη διαύγεια και αξιοπιστία των υποκείμενων χαρακτηριστικών του σήματος. Το στάδιο αυτό συμβάλλει σημαντικά στη βελτίωση της ακρίβειας εντοπισμού των κορυφών και κατ’ επέκταση στην αξιοπιστία της εκτίμησης DoA.

Για την αξιολόγηση της αποτελεσματικότητας της προτεινόμενης προσέγγισης, δημιουργήθηκε ένα σύνολο δεδομένων από χωρικούς φασματικούς χάρτες και χρησιμοποιήθηκε για την εκπαίδευση του αποθορυβοποιητικού αυτοκωδικοποιητή. Η απόδοση του μοντέλου αξιολογήθηκε κυρίως με βάση την ικανότητά του να εντοπίζει με ακρίβεια τις κορυφές στους φασματικούς χάρτες. Τα πειραματικά αποτελέσματα δείχνουν υψηλά ποσοστά επιτυχίας, με ακρίβεια εντοπισμού κορυφής να κυμαίνεται σταθερά γύρω στο 90% για τιμές SNR από -20 dB έως 20 dB. Ωστόσο, σε εξαιρετικά χαμηλές τιμές SNR (περίπου -23 dB), η αυξημένη παρουσία θορύβου οδηγεί σε ψευδείς ή μετατοπισμένες κορυφές, μειώνοντας την ακρίβεια

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

εντοπισμού σε περίπου 61%. Η τιμή αυτή αποτελεί το χαμηλότερο επίπεδο SNR που χρησιμοποιήθηκε στη μελέτη, καθώς στόχος μας ήταν να εντοπίσουμε το λειτουργικό όριο της μεθόδου, δηλαδή το σημείο κάτω από το οποίο η απόδοση αρχίζει να υποβαθμίζεται σημαντικά.

Λέξεις-κλειδιά:

Φασικές Συστοιχίες Κεραιών, Ομοιόμορφη Ορθογώνια Διάταξη Κεραιών, Εκτίμηση Κατεύθυνσης Άφιξης, Βαθιά Μάθηση, Αυτοκωδικοποιητής Αποθορυβοποίησης

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Table of contents

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

List of figures

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

List of tables

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Abbreviations

Radar	Radio detection and ranging
Sonar	Sound navigation and ranging
DoA	Direction of Arrival
URA	Uniform Rectangular Array
DFT	Discrete Fourier Transform
SNR	Signal-to-Noise Ratio
ULA	Uniform Linear Array
AWGN	Additive White Gaussian Noise
IFFT	Inverse Fast Fourier Transform
DL	Deep Learning
ML	Machine Learning
AI	Artificial Intelligence
etc.	et cetera (and so on)
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
MUSIC	Multiple Signal Classification
ESPRIT	Estimation of Signal Parameters via Rotational Invariance Techniques
MSE	Mean Square Error
DAE	Denoising Autoencoder
NaN	Not a Number
PSNR	Peak Signal-to-Noise Ratio
SSIM	Structural Similarity Index Measure
PDSR	Peak Detection Success Rate
PDE	Peak Distance Error
RMSE	Root Mean Square Error

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

vs	versus
dB	decibel
e.g.	exempli gratia (for example)

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Chapter 1

Introduction

In recent years, the demand for accurate and efficient spatial signal processing techniques has grown rapidly, driven by applications in radar, wireless communications, sonar, and autonomous systems. One of the most critical tasks in this domain is the estimation of the direction from which signals arrive at an antenna array—commonly referred to as Direction-of-Arrival (DoA) estimation. Accurate DoA estimation enables systems to localize sources, track targets and enhance communication by directing signal reception towards desired directions while suppressing interference. Despite the long history of research in the area, DoA estimation remains a challenging problem, especially under adverse conditions such as high noise levels, complex signal environments or limited computational resources. Building on this background, this thesis directs it's attention to this particular problem.

1.1 Thesis Subject

This thesis addresses the fundamental problem of Direction-of-Arrival estimation using Uniform Rectangular Array (URA) configurations, which enable two-dimensional spatial sampling for accurate determination of both azimuth and elevation angles. While traditional methods such as DFT-based beamforming—utilized in this work—are used for DoA estimation, their effectiveness may be compromised under challenging conditions, particularly in low signal-to-noise ratio (SNR) environments. To overcome these limitations, this study investigates the application of a denoising autoencoder designed to enhance the spatial spectrum produced by the above method, focusing on noise suppression and highlighting the spectral features that reveal the directions of arrival. Although the model is tested across a

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

wide range of SNR levels, from very high to very low, special emphasis is placed on its performance in low SNR scenarios, which constitute the core challenge addressed in this thesis.

1.1.1 Contributions

This thesis makes the following key contributions:

1. Integration of deep learning techniques within classical signal processing frameworks, illustrating the potential of data-driven models to enhance traditional methods of DoA estimation.

2. Development and implementation of a denoising autoencoder aimed at enhancing the spatial spectrum obtained from DFT-based beamforming, by reducing noise and reinforcing the spectral features that lead to signal directions.

1.2 Thesis Organization

This thesis is structured into six chapters. Chapter 1 introduces the subject of the thesis and outlines its main contributions. Chapter 2 provides the necessary background on phased array antennas and describes the steps involved for Direction-of-Arrival (DoA) estimation based on Uniform Rectangular Array (URA) configurations. Chapter 3 offers an overview of deep learning concepts, highlighting the motivation for applying deep learning to DoA estimation, with a particular focus on autoencoders and denoising autoencoders. Chapter 4 presents the experimental framework, including the generation of the dataset, as well as the design and training of the self-supervised denoising autoencoder. This chapter also describes the evaluation metrics used for performance assessment. Chapter 5 discusses the experimental results and finally, Chapter 6 concludes this thesis with a summary of the most important findings and recommendations for future research directions.

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Chapter 2

Phased Arrays and Direction-of-Arrival Estimation

2.1 Applications of DoA Estimation

DoA estimation and imaging have several practical applications in wireless communication, secure wireless communication [1], radar [2], acoustics, and radio astronomy [3]. With phased arrays, for example, we can use images that we generate so that we classify jammers or RFI sources [4]. In wireless communication, we can carefully direct the direction that we send our wireless information signal [1]. In radio astronomy, we can find the spatial direction of exo-galactic signal by searching with the beamforming algorithm [3]. The fundamental component in all these applications is that they require a phased array antenna system, which we will describe next. DoA estimation and imaging can be accomplished with methods like the MUSIC [5] or MVDR [3] beamforming. Here we will explore DFT-based beamforming to generate these images.

2.2 Overview of Phased Array Antennas

Phased array antenna systems are an advanced technology that enables precise control over the directions in which signals are transmitted and received. Unlike traditional antennas that emit signals equally in all directions, phased arrays concentrate the signal energy into specific directions by forming what is known as a beam—a focused transmission of wave energy. A key advantage of these systems is their ability to steer the beam electronically,

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

without requiring any physical movement of the antennas as referred in [6]. This is achieved by adjusting the phase of the signal at each individual antenna element. When these adjustments are properly synchronized across the array, the signals from all elements combine to form a sharp, directional beam that can be quickly and accurately pointed toward different angles. This ability, known as electronic beam steering, allows phased arrays to scan wide areas efficiently, a feature referred to as sector coverage. In addition, phased arrays can shape their beam patterns to suppress signal emissions in undesired directions, reducing interference. This is done by designing the array to produce patterns with low sidelobes—weaker, unintended beams that could otherwise degrade system performance [7]. To support these capabilities, modern phased arrays are built using sophisticated hardware and advanced signal processing algorithms. This integration of technology offers the flexibility, speed, and precision needed to meet the complex demands of today's communication, radar and sensing systems.

2.2.1 URA and ULA

Phased array antenna systems can be categorized based on their geometric configurations, which significantly influence their radiation characteristics and application suitability. Two of the most common array geometries are the Uniform Linear Array (ULA) and the Uniform Rectangular Array (URA). The ULA consists of antenna elements arranged in a single straight line with uniform spacing, making it one of the simplest and most widely studied configurations. ULA's are especially useful for one-dimensional beam steering and are extensively employed in applications where scanning is needed in a single plane. On the other hand, the URA extends this concept into two dimensions by arranging elements in a uniformly spaced rectangular grid, enabling two-dimensional beam steering both in azimuth and elevation planes. URA's provide enhanced spatial resolution and flexibility, making them suitable for more complex scenarios such as two-dimensional (target localization, wide-area surveillance and multi-beam communication systems).

In this thesis, the Uniform Rectangular Array (URA) configuration is employed due to its superior capability for two-dimensional beam steering which enables 2D Direction of Arrival estimation.

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

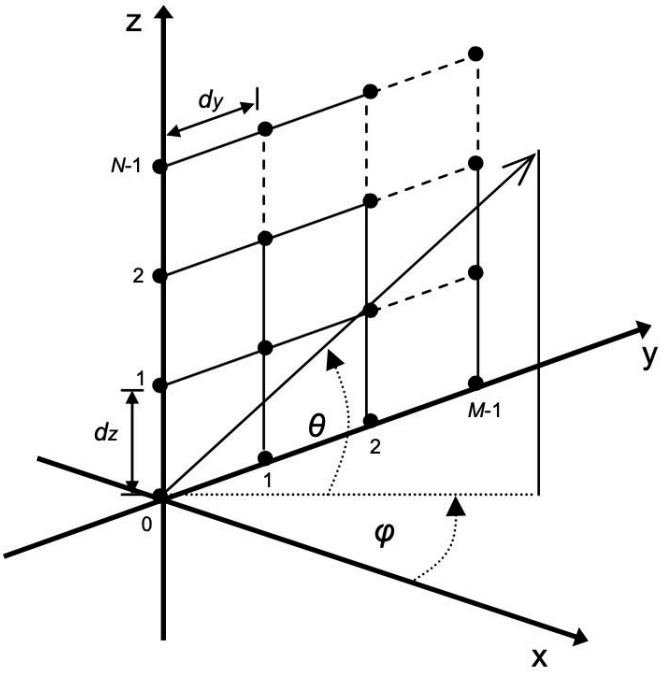

Figure 2.1: Uniform Rectangular Array (URA)
[8, 4].

2.3 Direction of Arrival Estimation based on URA Configurations

The process of DoA estimation begins with the reception of a signal by a Uniform Rectangular Array. As the signal arrives from a specific direction in space, it reaches each antenna element with a different phase shift. These phase shifts — caused by the geometry of the array and the direction of arrival — are encapsulated in what is known as the steering vector.

$$a_{\text{src}} = \exp\left(-j \cdot \pi \cdot f_c \cdot a_{\text{src1}}\right) \tag{2.1}$$

where

- a_{src1}: Beamforming of the signal using the Uniform Rectangular Array antenna positions and the angles of arrival (azimuth θ and elevation ϕ).
- f_c: The carrier frequency of the signal.

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

At this point, the received signal at the array can be modeled as:

$$\text{signal} = a_{\text{src}} \cdot x \tag{2.2}$$

where x is the transmitted signal.

This model captures the theoretical relationship, where the transmitted signal x is modulated by the steering vector asrc to produce the received signal at each sensor. In practice, since x is a sequence of signal samples over time we replicate asrc along the time dimension and perform an element-wise multiplication with x.

We make the received signal more realistic by adding Additive White Gaussian Noise (AWGN) so we have:

$$y = \text{signal} + \text{noise} \tag{2.3}$$

Now, we calculate the covariance matrix of the received signal y, which quantifies the statistical relationship between the signals received at different sensors across all time samples.

It is calculated as:

$$C_{ij} = \frac{1}{N-1} \sum_{k=1}^{N} \left(y_{2D,k,i} - \bar{y}_i\right) \left(y_{2D,k,j} - \bar{y}_j\right) \tag{2.4}$$

where:

- N is the total number of time samples,
- $y_{2D,k,i}$ and $y_{2D,k,j}$ denote the received signals at the i-th and j-th sensors during the k-th time sample,
- $\bar{y}_i$ and $\bar{y}_j$ are the average values of these signals computed over all time samples,
- C_{ij} represents the covariance between the signals received at these two sensors.

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

After computing the covariance matrix, we extract the first row, which contains the covariance values between the reference sensor ($i = 1$) and all other sensors in the array. This row is then reshaped into a 2D matrix of dimensions equal to the URA, allowing us to map these correlation values to the actual physical geometry of the array.

Once the covariance values have been reshaped to match the Uniform Rectangular Array geometry, we apply a 2D inverse Discrete Fourier Transform (2D IFFT) to the resulting matrix. This transformation converts the reshaped covariance data into the spatial frequency domain, producing a two-dimensional grid that represents the distribution of signal power across different spatial frequencies. Each element of this grid, known as a "bin," corresponds to a specific spatial frequency range, which can be directly related to a range of potential angles of arrival. The more bins there are, the smaller the spatial frequency range covered by each bin becomes. This means that potential DoAs are divided into finer segments, reducing the likelihood that multiple DoAs will fall into the same bin. As a result, the estimation derived from each azimuth and elevation bin becomes more accurate, allowing for finer discrimination between closely spaced DoA's [4, 3]. Last but not least, within the spatial frequency maps the highest power levels form clear peaks, marking the estimated positions in bins of the signal sources.

Comment:

The description of the above procedure is based on material presented in [4].

Moving forward, the transformation from spatial frequency bins to the estimated angles of arrival is performed using the following formulas, which calculate the angles in radians:

$$\phi = \sin^{-1}\left(\frac{2 \cdot l_{\text{max}}}{N_{\text{FFT}}} \cdot \frac{\lambda}{2d}\right) \tag{2.5}$$

$$\theta = \sin^{-1}\left(\frac{2 \cdot k_{\text{max}}}{N_{\text{FFT}}} \cdot \frac{\lambda}{2d \cdot \cos(\phi)}\right) \tag{2.6}$$

Where:

- k_{max} and l_{max} are the indices of the peak bin along the elevation and azimuth axes respectively,
- N_{FFT} is the transform size,

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

- λ is the wavelength,
- d is the sensor spacing,
- ϕ is the elevation angle,
- θ is the azimuth angle.

The calculated angles are then converted from radians to degrees for easier interpretation.

2.3.1 Visualized example of DoA estimation using the actual configurations employed in this thesis

Parameters:

- The number of sensors is 1024, arranged in a 32 by 32 array.
- The sensor spacing is 0.5 meters which corresponds to half the wavelength given a carrier frequency of 1 kHz and a propagation speed of 1000 m/s.
- The discrete Fourier transform is computed using 32 points ($nFFT = 32$).
- The total number of time samples is 300, with the pulse starting at sample 3 and having a width of 4 samples.

Following the procedure described in Section 2.2 and for a specific SNR value of 3 dB with an azimuth angle of 50 degrees and an elevation angle of 10 degrees, we obtain the following spectral map.

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

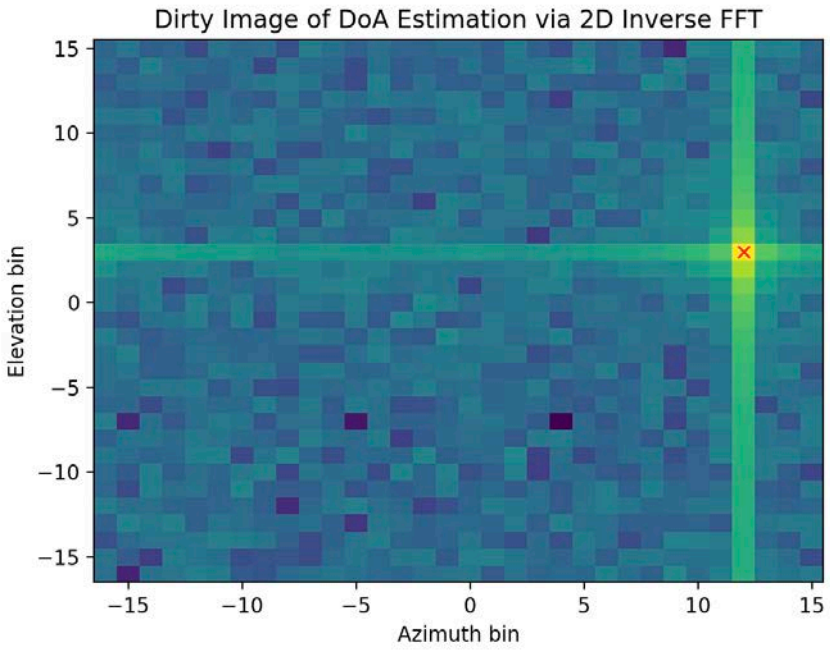

Figure 2.2: Spatial spectrum map with clear localization of the target bin corresponding to the signal's direction of arrival.

By applying the previously outlined formulas to convert the spectral bins into angular values, we obtain estimated angles of 49.78° in azimuth and 10.81° in elevation. These values closely match the true direction of arrival, indicating high estimation accuracy.

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Chapter 3

Deep Learning

3.1 Introduction to Deep Learning

Deep Learning (DL) is a subfield of Machine Learning (ML), which itself belongs to the broader domain of Artificial Intelligence (AI). While conventional machine learning algorithms rely heavily on feature engineering — a process that demands domain expertise to manually design and extract meaningful features from raw data — deep learning introduces a fundamentally different approach. By employing deep neural networks composed of multiple hidden layers, DL models are capable of automatically learning hierarchical feature representations directly from raw input data. This capability enables deep learning particularly effective in complex pattern recognition tasks such as image processing, object detection, video analysis, etc. Compared to traditional ML methods, DL has demonstrated significantly improved performance, especially in data-rich environments. Among the most widely used deep learning architectures are Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). CNNs are especially suitable for processing spatial data like images. Their main architectural design—where each neuron is locally connected to only a subset of neurons from the previous layer—enables the network to extract local spatial features efficiently, while also reducing the number of trainable parameters via weight sharing mechanisms. This structure supports robust feature extraction while minimizing computational cost. On the other hand, RNNs are designed to handle sequential data, such as speech signals or time series, where the temporal order of information is essential. These networks maintain a memory of past inputs, allowing them to model dependencies across time, which is crucial for tasks such as natural language processing and sequence prediction. [9]

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

3.2 Motivation for Deep Learning in DoA Estimation

Direction-of-Arrival (DoA) estimation is a fundamental task in array signal processing, with applications in radar, communications, and acoustics. Traditional DoA estimation techniques, including super-resolution methods such as MUSIC and ESPRIT, as well as DFT-based beamforming algorithms, typically rely on certain conditions to achieve reliable performance. These conditions include high signal-to-noise ratio, accurate calibration of the sensor array and a sufficient number of observations. However, in real-world environments where noise is high, measurements are limited and models are imperfect, the performance of these classical methods often degrades. This has motivated the exploration of data-driven techniques, particularly deep learning. Deep neural networks—and especially denoising autoencoders—are well suited for extracting meaningful features from noisy inputs and can help improve DoA estimation accuracy under challenging conditions. For example in [10], the authors demonstrated that deep learning models can significantly reduce estimation errors in low-SNR scenarios highlighting their ability to improve traditional methods.

3.3 Autoencoders

Autoencoders were first introduced in 1986 by Geoffrey Hinton and are a class of deep learning architectures designed to learn efficient data representations. Their primary objective is to encode input data into a compressed form and subsequently reconstruct the original input as accurately as possible. An autoencoder consists of two main components: the encoder and the decoder. The encoder processes the input data and maps it to a lower-dimensional latent space representation. This latent space captures the most essential and informative features of the input, effectively summarizing the underlying structure and patterns within the data in a more compact and abstract form. The decoder then takes this latent representation and attempts to reconstruct the original input from it. It is important to note, that the overall architecture is typically symmetric, with the decoder mirroring the encoder structure, however this is not a strict rule and can vary depending on the specific application. During training, the autoencoder involves minimizing the difference between the input and the reconstructed output, typically through backpropagation. To quantify reconstruction quality during training, a loss function is employed—most commonly the Mean Squared Error (MSE) or cross-entropy loss. MSE is used when the output is continuous-valued, such as in image reconstruction tasks

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

where pixel intensities are real-valued within a range.The MSE loss is mathematically defined as:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2 \tag{3.1}$$

where x_i is the ground truth value (e.g., pixel intensity), $\hat{x}_i$ is the reconstructed value and n is the total number of samples (e.g., number of pixels).

In contrast, cross-entropy loss is more appropriate when the output represents categorical data or class probabilities, such as in classification tasks. For the purposes of this thesis, where the autoencoder is employed to reconstruct continuous data (images), MSE serves as the suitable loss function. [11]

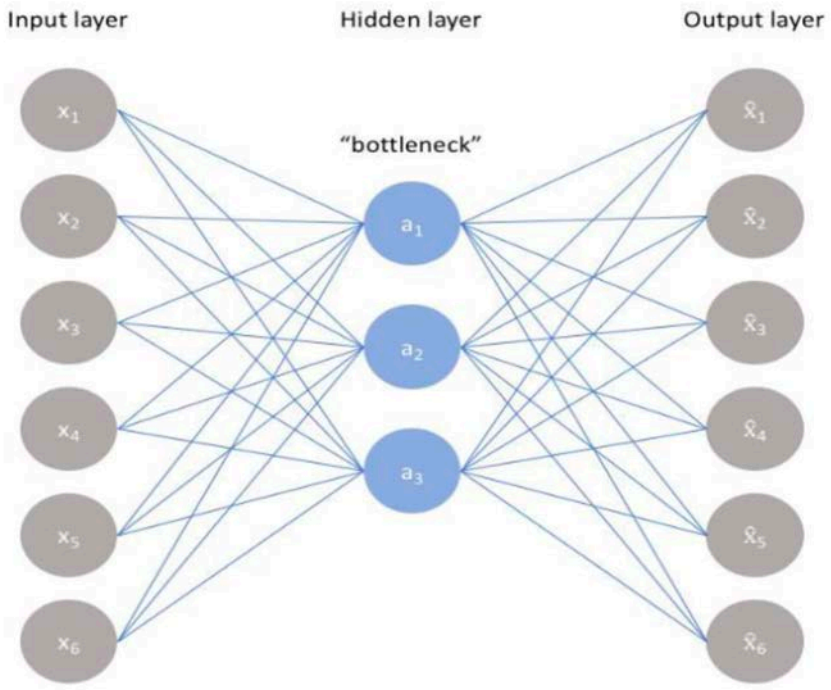

Figure 3.1: Illustration of an autoencoder model architecture

[12]

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

3.3.1 Denoising Autoencoders

Denoising Autoencoders (DAEs) are a specialized variant of autoencoders designed specifically to address the problem of noise in data, particularly in images. In real-world scenarios, images often become corrupted by noise introduced through various sources such as camera sensor defects, transmission errors, or environmental factors. Noise appears as random variations in color or brightness, degrading the quality of images and potentially obscuring important details. The primary goal of DAEs is to learn how to remove such noise and recover the clean, original image from its corrupted version. During training, DAEs receive noisy images as input and are trained to reconstruct their clean counterparts, which are used as the ground truth targets for the loss function. This forces the model to focus on the underlying structure of the image rather than the noise itself. If the learning process is successful, the model gradually learns to distinguish noise from relevant features. As a result, the encoder captures meaningful and noise-robust representations of the data, which the decoder then uses to generate a cleaner version of the input. This ability to extract and preserve important visual information makes DAEs particularly effective in image restoration tasks, where maintaining detail and structure is critical. [13]

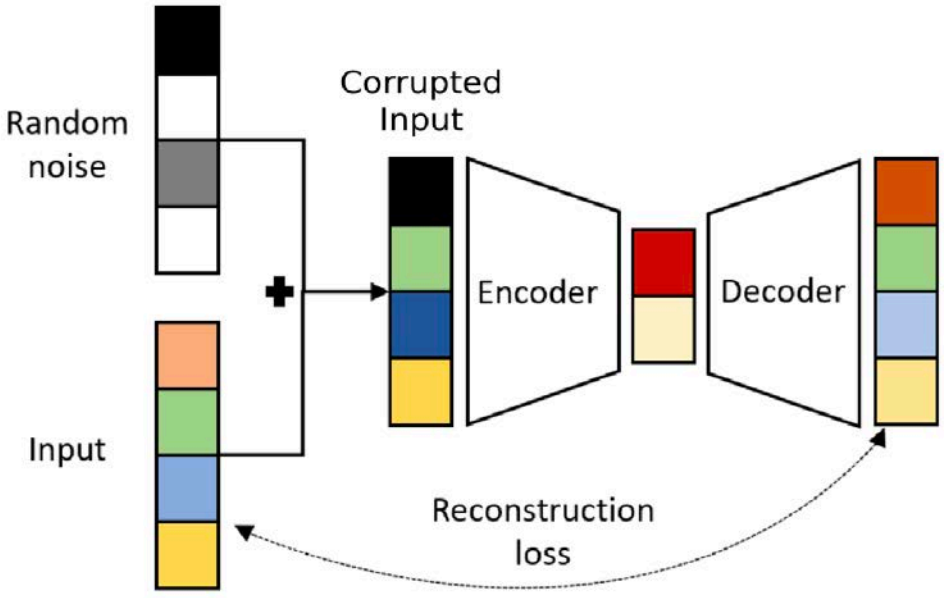

Figure 3.2: Denoising Autoencoder

[14]

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Chapter 4

Experimental Framework for the self-supervised Denoising Autoencoder

4.1 Dataset

In Section 2.2, we described the process of generating beamformed 2D spatial power maps for DoA estimation based on URA configurations using a 32×32 antenna grid. Given a known steering vector and an incoming plane wave from a certain azimuth and elevation angle, we applied a DFT–based beamforming technique to localize the signal source within the spatial domain. The resulting output was a 2D image where the signal energy peak at the spatial bins corresponded to the true source location. By locating this peak, we identified the elevation and azimuth bins in which the target was found. Subsequently, through appropriate conversion formulas, these bin indices was translated into angle estimates, providing the final DoA estimation.

Building on this foundation, we constructed a dataset consisting of 2D beamformed spatial power maps with a total of 2.187 clean beamformed images, each with a resolution of 128×128 pixels. These images correspond to all possible combinations of azimuth and elevation angles in the range of −70° to +70°, with a step size of 10° in both directions. The 10° step size was intentionally chosen to reduce the likelihood of different angle combinations mapping to the same spatial bin. Using a finer step, such as 5°, would increase the chance of multiple angles mapping to the same bin, thereby reducing the diversity of peak locations in the dataset. This diversity is important for effectively training the autoencoder to generalize across varying peak positions within the spatial maps.

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

For each clean beamformed image, additive white Gaussian noise was applied at multiple Signal-to-Noise Ratio (SNR) levels: −23 dB, −20 dB, −15 dB, −10 dB, −5 dB, 0 dB, 5 dB, 10 dB, 15 dB and 20 dB. This approach generated a set of noisy counterparts for each clean image, thereby creating a paired noisy-clean dataset of 128×128 spatial power maps. These image pairs were used to train a self-supervised denoising autoencoder — a model that learns to recover clean spatial structures from synthetically corrupted inputs without external labels. By reconstructing the clean image from noisy inputs, the autoencoder improves the robustness of DoA estimation, especially under low-SNR conditions, by preserving crucial spatial features such as the signal peak.

Comment 1:

It is important to note that some combinations of azimuth and elevation bins result in invalid angle estimates, producing NaN (Not a Number) values during the conversion from spatial bins to physical angles. Based on the angle calculation formulas presented in Section 2.2, this occurs due to the domain restrictions of the arcsine function. Specifically, when the arguments inside the arcsine functions fall outside the interval [−1, 1], the result is undefined (NaN). This can happen especially when the cosine of the elevation angle is close to zero, causing the azimuth calculation to become invalid. Therefore, to ensure meaningful and physically valid training data, only those angle combinations are included in the dataset for which the formulas yield valid (non-NaN) elevation and azimuth estimates. Training the autoencoder on bins that correspond to undefined angle estimates would not be beneficial, as these bins do not represent physically realizable source directions.

Comment 2:

The decision to restrict azimuth and elevation angles to the range of −70° to +70° is motivated by both practical and theoretical considerations related to DFT-based beamforming on a URA. Although, in theory, angle estimation can extend up to ±90°, the performance of the DFT begins to degrade as we approach these extreme angles. The spatial frequency representation becomes increasingly distorted due to the periodic nature of the DFT and the spatial sampling of the array. This can lead to aliasing and the appearance of mirror peaks in the spatial power map, causing confusion in the source localization. The steering vectors also become less distinguishable near ±90°, which further reduces the accuracy of DoA es-

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

timation. Therefore, to ensure that the training data reflects scenarios where the estimation process is physically valid and reliable, we set a practical threshold of ±70°. Including directions beyond this range would result in images that correspond to ambiguous or misleading localization outputs — much like the NaN-producing cases — and would not help the autoencoder learn to denoise meaningful signal structures.

Dataset Visualization:

Examples of image pairs from the dataset are shown below, illustrating the correspondence between clean and noisy spatial power maps. These pairs highlight how the noise affects the structure of the spatial energy distribution and demonstrate the kinds of input-output relationships the denoising autoencoder is trained to model. Note that only a subset of the SNR values used in the dataset is shown here for visualization purposes.

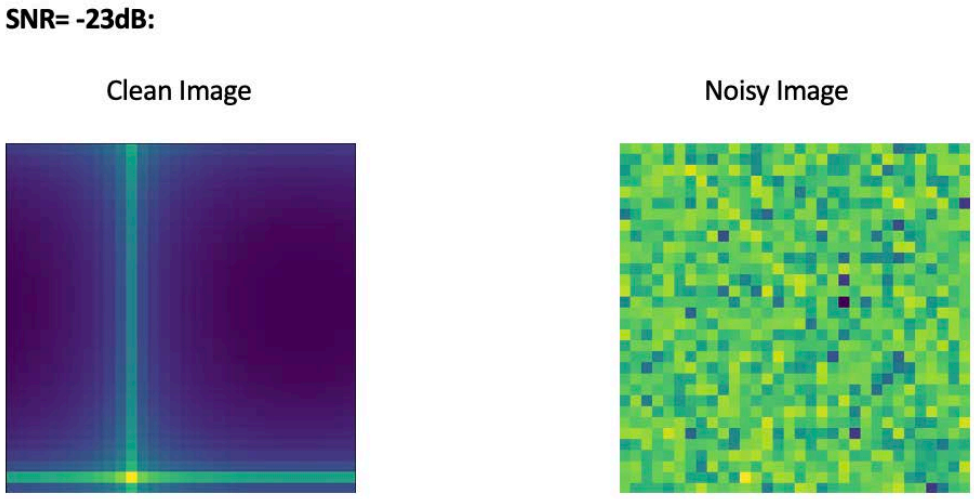

Figure 4.1: Example of image pairs for -23 dB

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

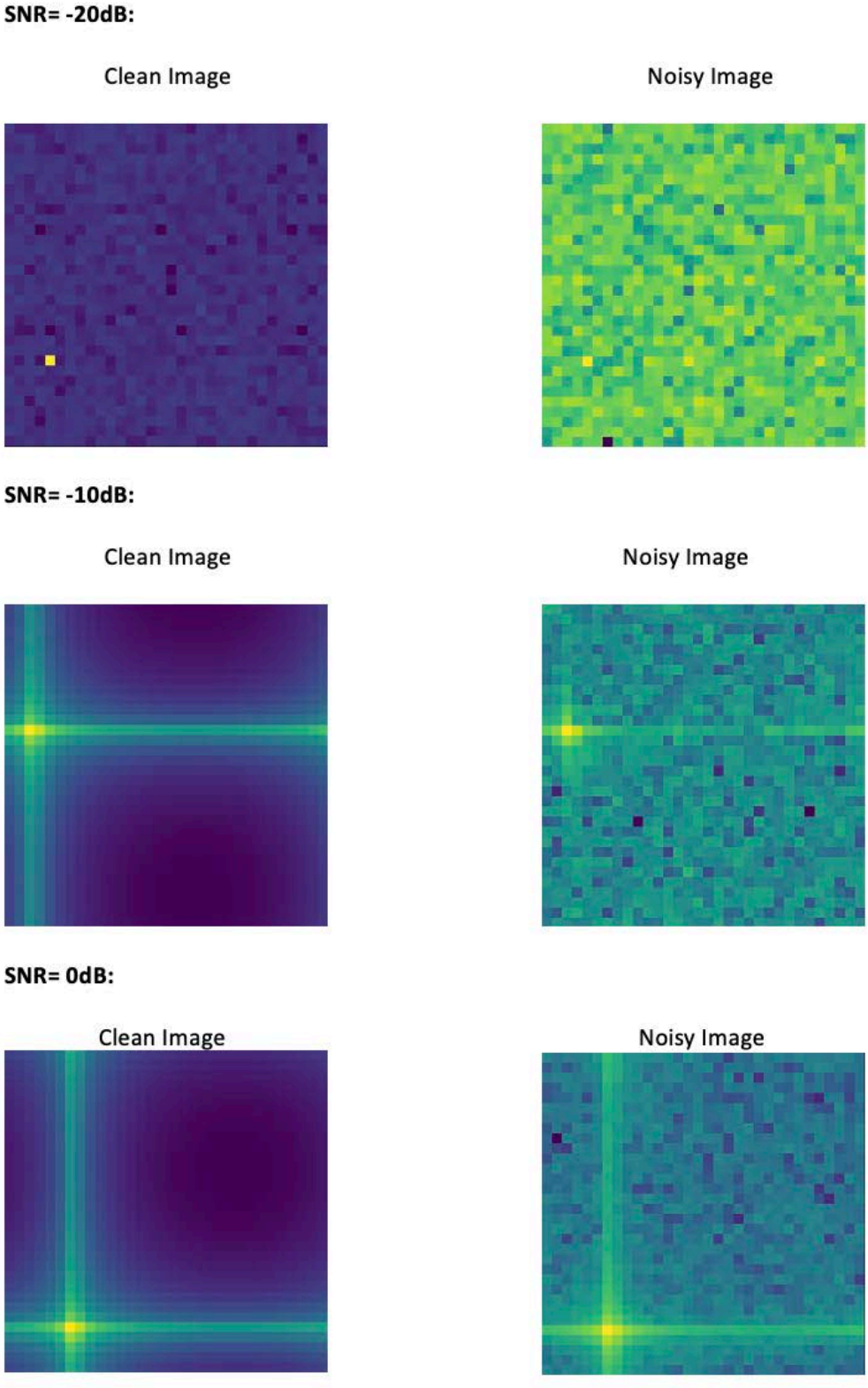

Figure 4.2: Examples of image pairs for -20, -10 and 0 dB

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

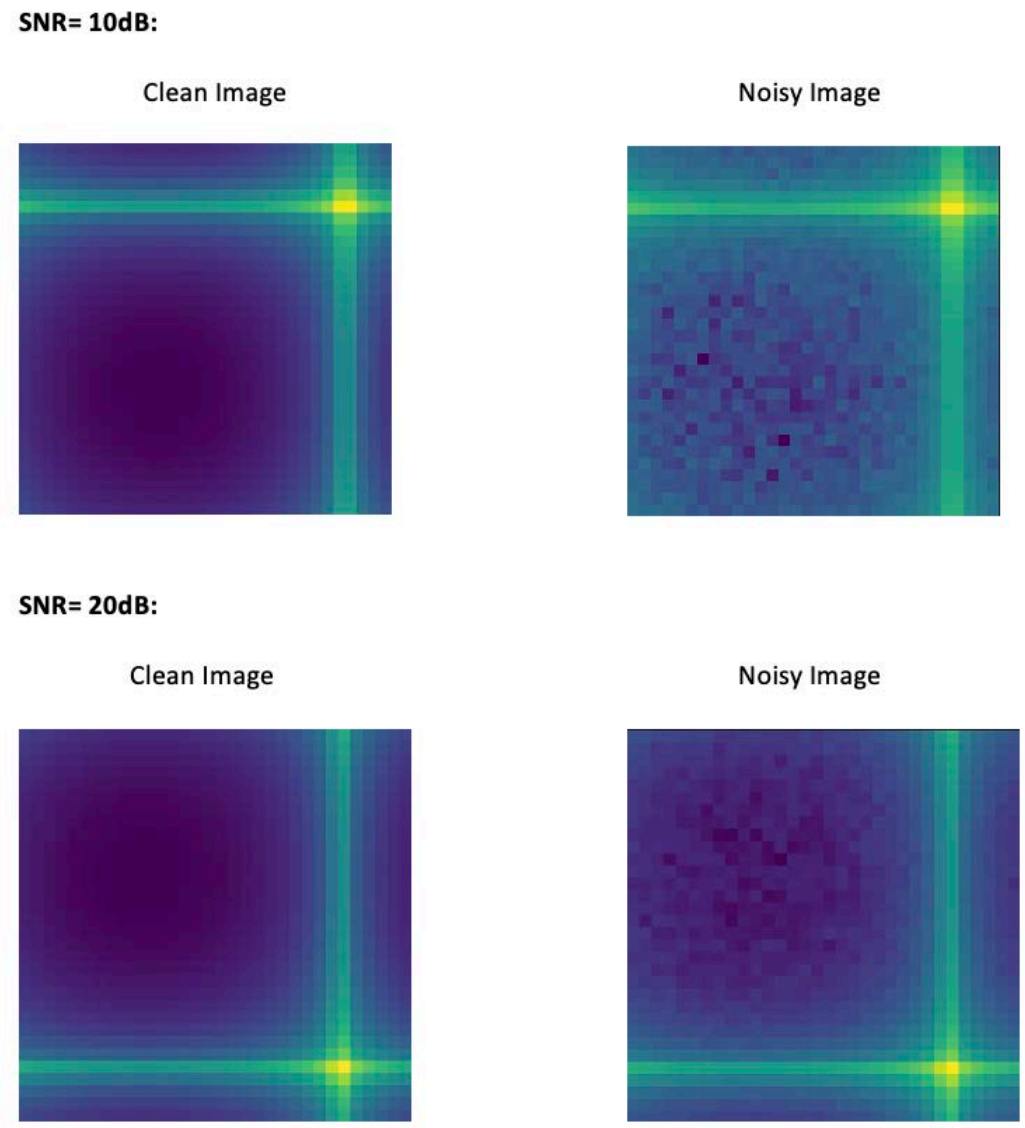

Figure 4.3: Examples of image pairs for 10 and 20 dB

4.2 Data Split

The dataset was split into three parts: 70% for training, 10% for validation and 20% for testing. The training set provides the data from which the model learns, while the validation set helps to monitor its performance and guide parameter tuning throughout the training process. Finally, the test set is kept separate to provide a fair evaluation of how well the model works on new, unseen data.

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

The table below shows a detailed breakdown of the data split based on SNR values.

SNR (dB)	Train	Validation	Test
-23	148	15	38
-20	164	14	40
-15	175	14	32
-10	146	18	57
-5	163	12	46
0	163	20	38
5	162	22	37
10	155	22	44
15	152	15	54
20	146	23	52

Table 4.1: (Train-Validation-Test) Data Split per SNR

4.3 Denoising Autoencoder Architecture

As we have previously mentioned, the denoising autoencoder is designed to reconstruct clean spatial power maps from noisy input data by effectively reducing noise while preserving critical signal structures. The following architecture, adopts a classic encoder–decoder framework, enhanced with a self-attention module at the bottleneck. This addition enables the network to better model global spatial relationships, which is especially beneficial in distinguishing signal features from background noise.

Encoder

The encoder is composed of a sequence of convolutional layers that systematically reduce the spatial dimensions of the input by half at each layer. Simultaneously, the number of feature maps — each representing different learned aspects of the input [15]— increases progressively. This compression process transforms the input into a compact yet information-rich representation that captures essential spatial characteristics. Each convolutional layer applies multiple learnable kernels that scan the input to detect localized spatial patterns — for instance, signal peaks or directionally structured regions — which are particularly important

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

for the DoA estimation task. Increasing the number of feature maps allows the network to express a more diverse and detailed set of spatial features.

Self-Attention Module

At the bottleneck — the point of maximum compression, where the spatial resolution is lowest and the feature depth is highest — a self-attention mechanism is applied. Unlike conventional convolutional layers that operate on spatially local neighborhoods, self-attention enables the model to capture global dependencies by directly relating every spatial position in the feature map to every other. This mechanism allows the network to dynamically assess the importance of each region in the input, regardless of its spatial distance. Technically, the self-attention mechanism operates by first applying three separate learned linear transformations to the input feature map, resulting in three distinct representations commonly referred to as queries (Q), keys (K), and values (V). Each of these representations serves a specific role within the attention computation: the queries act as vectors searching for relevant information, the keys act as reference vectors which the queries compare to in order to determine relevance, and the values hold the actual information that will be combined and aggregated according to the attention scores derived from the query-key comparisons. To compute the attention, each query is compared with all keys using the dot product — a mathematical operation where corresponding components of two vectors are multiplied and summed, producing a single score that indicates their similarity. These raw scores are then passed through a softmax function, which converts them into normalized attention weights — that is, positive values that sum to one. This step ensures that the model focuses proportionally on the most relevant parts of the input. The resulting attention weights are used to compute a weighted sum of the values, producing an output vector for each query. This allows the model to highlight important patterns — such as structured signals or spatial peaks — while suppressing irrelevant or noisy data, even when important features are located far apart. Such capability is especially useful in tasks like DoA estimation, where spatially scattered but semantically related information must be captured. Finally, the self-attention output — a refined set of feature maps enriched with global context — is combined element-wise with the original input via a learnable scaling parameter, commonly referred to as gamma. This parameter controls how much influence the attention output should have relative to the original features. Gamma is initialized to zero and is updated during training via backpropagation, just like other learnable parameters in

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

the network. By adjusting gamma over the course of training, the model can learn to balance between preserving the original features and integrating the new, globally-informed context provided by the attention mechanism. At the end, this refined representation is passed to the decoder for reconstruction. [16]

Decoder

The decoder mirrors the encoder's structure in reverse. It progressively restores the spatial resolution of the feature maps by doubling their dimensions at each stage using transposed convolutional layers. This gradual expansion transforms the compressed representation back into a full-sized image that approximates the original, noise-free spatial power map. The final layer uses a sigmoid activation function to constrain output values between 0 and 1, ensuring consistency with the normalized range of the input data.

Comment:

To reduce the size of the images at each stage, we apply convolutional layers with a kernel size of 3×3, stride of 2, and padding of 1. According to the formula in [17],

$$\text{Output size} = \left\lfloor \frac{\text{Input size} - \text{kernel size} + 2 \times \text{padding}}{\text{stride}} \right\rfloor + 1 \tag{4.1}$$

this configuration reduces the dimensions of the images to half their size at every layer.

In the decoder, the transposed convolutional layers use the same kernel size, stride, and padding as in the encoder, but they also include output-padding = 1 to make sure the upsampled output reaches the correct size. This extra padding fixes the one-pixel difference caused by rounding down during the downsampling in the encoder, helping to accurately recover the original image dimensions. The calculation of the output size in transposed convolution operations, follows the formula which is presented below as described in [18].

$$\text{Output Size} = (\text{Input Size} - 1) \times \text{Stride} + \text{Kernel Size} - 2 \times \text{Padding} + \text{Output-Padding} \tag{4.2}$$

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Below is a table presenting for each stage of the autoencoder, the processing block type , the number of feature maps, as well as the input and output image dimensions.

Stage	Block Type	Feature Maps	Input Size	Output Size
Encoder 1	Conv2D	$1 \rightarrow 64$	128×128	64×64
Encoder 2	Conv2D	$64 \rightarrow 128$	64×64	32×32
Encoder 3	Conv2D	$128 \rightarrow 256$	32×32	16×16
Bottleneck	Self-Attention	256	16×16	16×16
Decoder 1	ConvTranspose2D	$256 \rightarrow 128$	16×16	32×32
Decoder 2	ConvTranspose2D	$128 \rightarrow 64$	32×32	64×64
Decoder 3	ConvTranspose2D	$64 \rightarrow 1$	64×64	128×128

Table 4.2: Summary of each autoencoder stage, specifying the processing block type, the number of feature maps, as well as the input and output image dimensions.

4.4 Training

The training process was carefully monitored using the validation loss to avoid overfitting. Early stopping was set up so that training would stop if the validation loss didn't improve after several consecutive epochs. However, that never happened — the validation loss kept going down smoothly, so the training continued until the last epoch. This suggests that the model was learning well without overfitting, as also shown in the plot below.

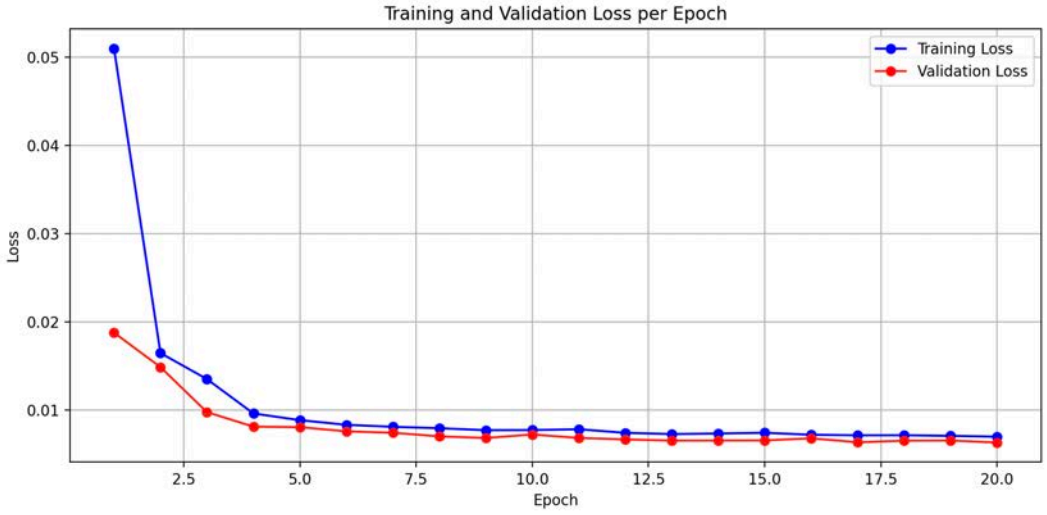

Figure 4.4: Training and Validation Loss per epoch

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

For completeness and transparency, the training parameters and their selected values are summarized below. These choices reflect the configuration under which the model achieved its final performance.

Training Parameter	**Value**
Optimizer	AdamW
Learning Rate	0.001
Weight Decay	0.0001
Loss Function	MSELoss
Batch Size	32
Number of Epochs	20
Early Stopping Patience	5 epochs

Table 4.3: Training Parameters and their assigned values.

Below are brief explanations of two of the parameters listed in Table 4.3, included for clarity:

Weight Decay: Weight decay is a regularization technique that adds a penalty to the loss function based on the size of the model's weights. By discouraging large weight values, it helps prevent overfitting, encouraging the model to learn simpler and more generalizable patterns. This results in better performance on unseen data.

Batch Size: Batch size refers to the fixed number of training samples processed together before the model updates its parameters (such as the weights). In each epoch, the model goes through all the training data divided into multiple batches of 32 samples each. This approach balances the accuracy of gradient estimates with computational efficiency.

Before concluding the training section, we provide a detailed explanation of the back-propagation technique, which is fundamental to training neural networks like convolutional denoising autoencoders. Understanding this process clarifies how the model learns by adjusting its parameters to minimize prediction errors. [19]

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

1. **Forward Propagation**

   The training process begins with forward propagation, during which the input data flows through the network layer by layer. Once the data has passed through all layers, the network produces an output.

2. **Error Calculation**

   This step involves evaluating how far this output deviates from the expected target value. The error is computed as the difference between the predicted output and the actual target. In networks with multiple output neurons, this error is calculated separately for each output neuron. Then, the total error is calculated. This stage provides a measure of the model's performance and determines the extent of adjustments needed in the subsequent step.

3. **Backpropagation**

   Once the error is calculated, the backpropagation phase begins. During this process, the error is propagated backward through the network, from the output layer toward the input layer. The goal is to evaluate how each learnable parameter—such as weights, biases, and other trainable variables—affects the overall error. To achieve this, the chain rule from calculus is applied to compute the gradient of the error with respect to each parameter. These gradients, known as partial derivatives, indicate how a small change in any given parameter would impact the total error.

4. **Parameter Update**

   This information guides the network in adjusting the parameters to the right direction and magnitude, ultimately reducing the prediction error.

5. **Iterative Optimization**

   All of the above stages are repeated for many training cycles (epochs). This iterative optimization continues until the error is sufficiently minimized, or a pre-defined stopping condition is met, such as a maximum number of epochs.

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

4.5 Evaluation Metrics

In this section, the evaluation metrics employed to measure the model's performance are described. Aligned with the primary objective of this thesis, the main metrics of interest are the Peak Detection Success Rate on the spectral maps and the Peak Distance Error. These metrics serve as the primary evaluation criteria because the core focus of this work is the precise localization of peaks corresponding to the target bins. For a more comprehensive assessment of the denoised outputs, additional metrics such as Loss, Peak Signal-to-Noise-Ratio (PSNR), and Structural Similarity Index (SSIM) are also employed to provide complementary insights into the model's performance.

4.5.1 Peak Detection Success Rate and Peak Distance Error

Peak Detection Success Rate:

Peak Detection Success Rate (PDSR) measures how often the model correctly identifies the location of the target in the denoised spectral map. Specifically, it checks whether the bin with the highest value—the peak—matches the corresponding bin in the ground truth map. The final score is expressed as a percentage of successful detections reflecting the model's ability to accurately locate the target despite the presence of noise.

Peak Distance Error:

In cases where the peak is not successfully detected, the Peak Distance Error (PDE) is computed as the Euclidean distance between the ground truth bin and the detected one. This distance represents the number of bins by which the predicted location deviates from the true target position. After calculating the PDE for each sample, with lower PDE values indicating more accurate localization, the values are averaged to produce a mean error.

4.5.2 Loss

As previously mentioned, the Mean Squared Error (MSE) loss is used during training to guide the model in minimizing the difference between the denoised output and the corresponding clean spectral map. In addition to its role as a training loss, MSE is also computed on the test set after training to evaluate the model's performance on previously unseen data.

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

4.5.3 PSNR

PSNR is a widely used metric for evaluating the quality of reconstructed images. Although it is directly related to the mean squared error, it provides a more interpretable and standardized way to assess reconstruction quality, which is why it is commonly used. It measures the logarithmic ratio between the square of the maximum possible pixel value and the MSE between the original and the reconstructed image, expressed in decibels. In this work, since the images are normalized within the range [0,1], the maximum pixel value is set to 1. The PSNR is calculated using the following formula:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{MaxValue^2}{\text{MSE}} \right) \tag{4.3}$$

This can equivalently be written in terms of the root mean squared error (RMSE) as:

$$\text{PSNR} = 20 \cdot \log_{10} \left(\frac{MaxValue}{\text{RMSE}} \right) \tag{4.4}$$

since $\text{RMSE} = \sqrt{\text{MSE}}$

In conclusion, higher PSNR values indicate better image reconstruction quality. [20]

4.5.4 SSIM

The Structural Similarity Index Measure (SSIM) is a metric developed to assess the similarity between two images by modeling how the human visual system perceives changes in visual information. Unlike conventional approaches that compare images on a pixel-by-pixel basis—simply measuring absolute differences—SSIM examines three critical components that influence human perception: luminance, contrast, and structural information. Luminance, refers to the overall brightness of the image which affects how humans detect variations in light intensity. Contrast, captures the variation in intensity between neighboring pixels reflecting how clearly details and edges stand out and finally, structural information relates to the spatial arrangement and relationship between pixels, emphasizing patterns and textures that form meaningful content in an image.

Based on these components, SSIM yields a similarity score between 0 and 1, where values closer to 1 indicate a higher degree of structural resemblance, with 1 representing a perfect match between the compared images. [20]

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Chapter 5

Results

In this section, the performance of the implemented self-supervised denoising autoencoder is presented and analyzed. The evaluation is based on the metrics described previously. All the following results refer to the test set, with the model's performance evaluated separately for each SNR level.

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

5.1 Evaluation Metrics Results

1. Peak Detection Success Rate vs SNR

The following table reports the Peak Detection Success Rate for each SNR value.

SNR (dB)	Peak Success Percentage
-23	61%
-20	88%
-15	91%
-10	95%
-5	89%
0	97%
5	92%
10	93%
15	93%
20	94%

Table 5.1: Peak Success Percentages vs SNR

At extremely low SNR levels, such as –23 dB, the peak detection success rate drops noticeably. This occurs because, at such low SNRs, noise can introduce fake peaks or cause the true peaks to shift from their original positions. In these cases, the model may be misled and attempt to denoise or enhance incorrect regions of the spectral map—focusing on false peaks instead of the true ones. As a result, the detected peaks may not align with the ground truth, reducing detection accuracy. Outside of this extreme scenario, the performance remains stable and high, with success rates consistently around 90% across the SNR range.

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

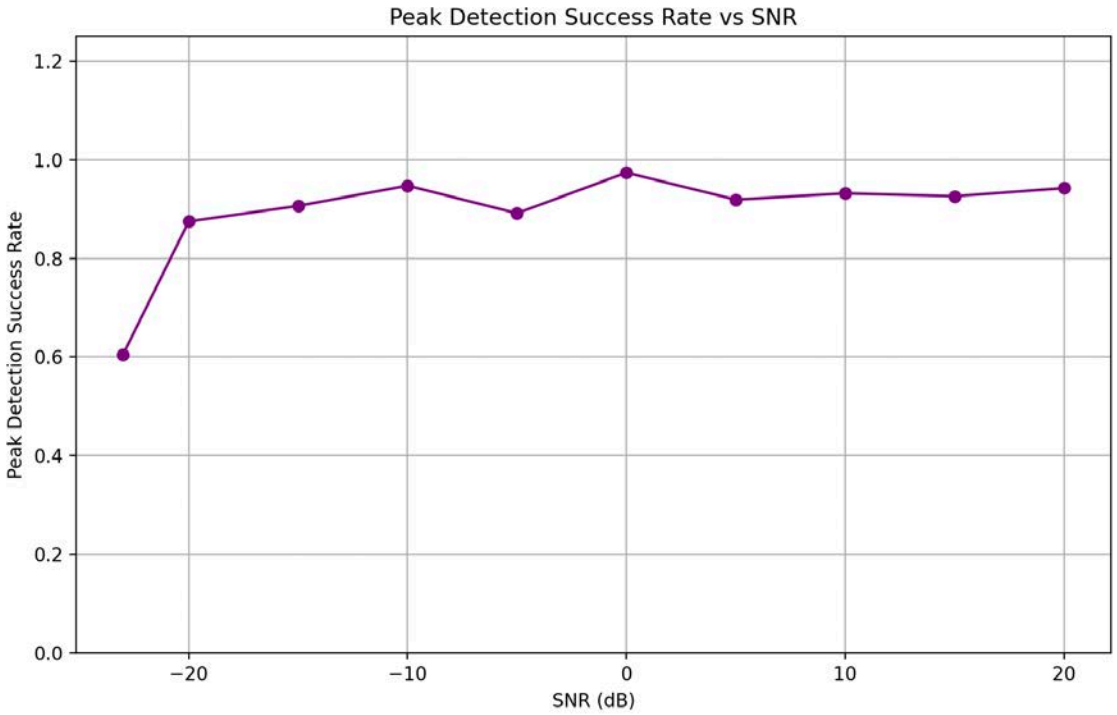

Figure 5.1: Peak Detection Success Rate vs SNR

2. Peak Distance Error

The following table reports the average Peak Distance Error in bins for each SNR value.

SNR (dB)	Average Peak Distance Error (bins)
-23	9.2
-20	5.1
-15	1.0
-10	1.0
-5	1.0
0	1.0
5	1.0
10	1.0
15	1.0
20	1.0

Table 5.2: Average Peak Distance Error in bins vs SNR

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

As observed in the above table, at very low SNR values such as –23 dB, the peak distance error is significantly higher, averaging around 9.2 bins. This large displacement is caused by frequent peak shifts and the presence of false peaks introduced by noise, resulting in substantial errors in detected peak positions. At –20 dB, false peaks and peak shifts still occur but less frequently and to a smaller extent, reducing the average error to 5.1 bins. For higher SNR values, where the true peaks remain mostly stable, the average distance error consistently drops to approximately 1 bin, demonstrating a much better peak localization.

Example of Failed Peak Detection for SNR = 20dB

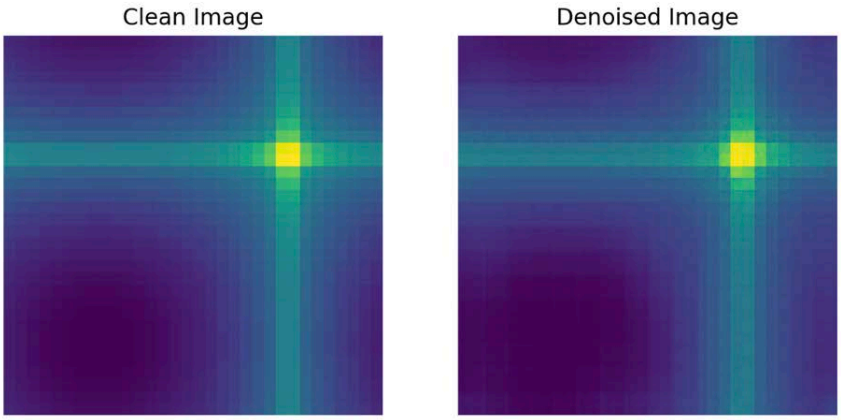

Figure 5.2: Example of Failed Peak Detection for SNR = 20dB

In the above figure, several bins around the true peak also appear with relatively high intensity. This can mislead the model during denoising, causing it to estimate the peak position slightly off, resulting in the observed 1-bin error in peak detection.

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

3. Average Loss vs SNR

The following table reports the average reconstruction loss for each SNR value.

SNR (dB)	Average Loss Value
-23	0.017163
-20	0.014696
-15	0.013952
-10	0.009428
-5	0.005726
0	0.003624
5	0.002369
10	0.001494
15	0.001097
20	0.001074

Table 5.3: Average Loss Values vs SNR

As observed in the table, the average loss decreases steadily as the SNR increases. This behavior is consistent with expectations, since higher SNR values correspond to less noise and allow the model to perform more accurate reconstructions. This trend is also clearly reflected in the following plot, which visually supports the numerical results.

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

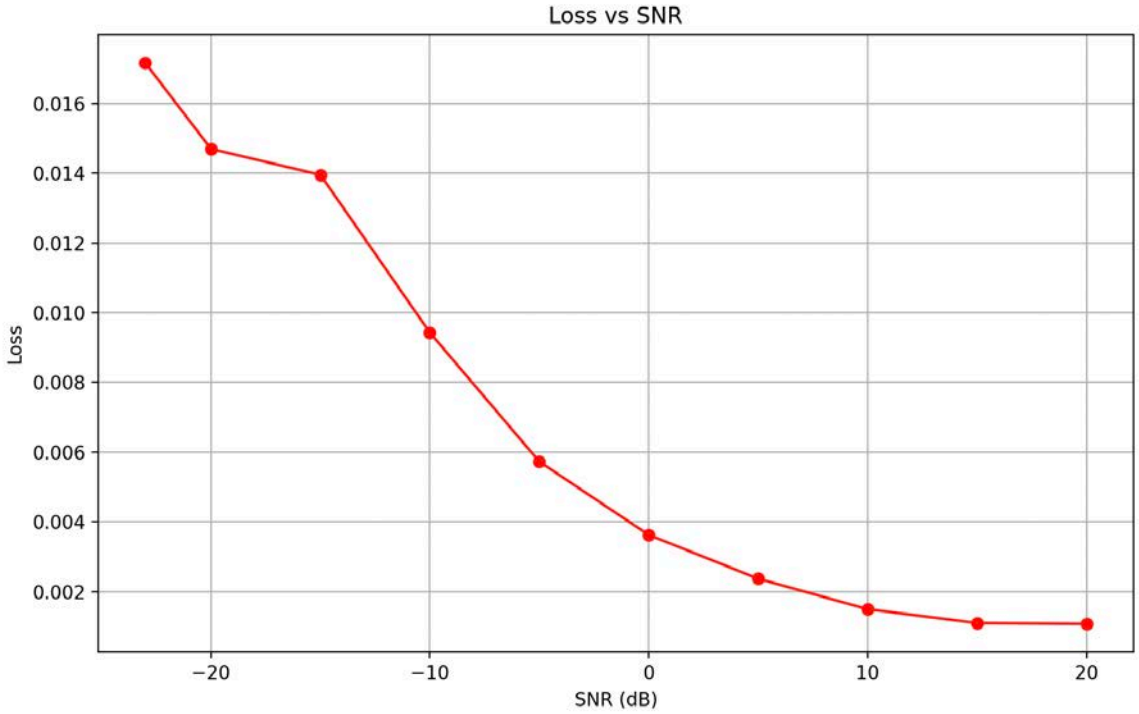

Figure 5.3: Average Loss vs SNR

4. Average PSNR vs SNR

The following table reports the average PSNR values for each SNR.

SNR (dB)	Average PSNR (dB)
-23	17.95
-20	18.67
-15	18.72
-10	20.49
-5	22.73
0	24.81
5	27.02
10	29.28
15	31.06
20	31.00

Table 5.4: Average PSNR Values vs SNR

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

The table indicates an overall increase in PSNR values as the SNR improves, reflecting the expected behavior since PSNR is inversely related to the mean squared error (MSE) used in the reconstruction loss. As the noise level decreases (higher SNR), the MSE reduces — as previously shown in Figure 5.3 — resulting in higher PSNR values. This positive trend is also clearly depicted in the figure below.

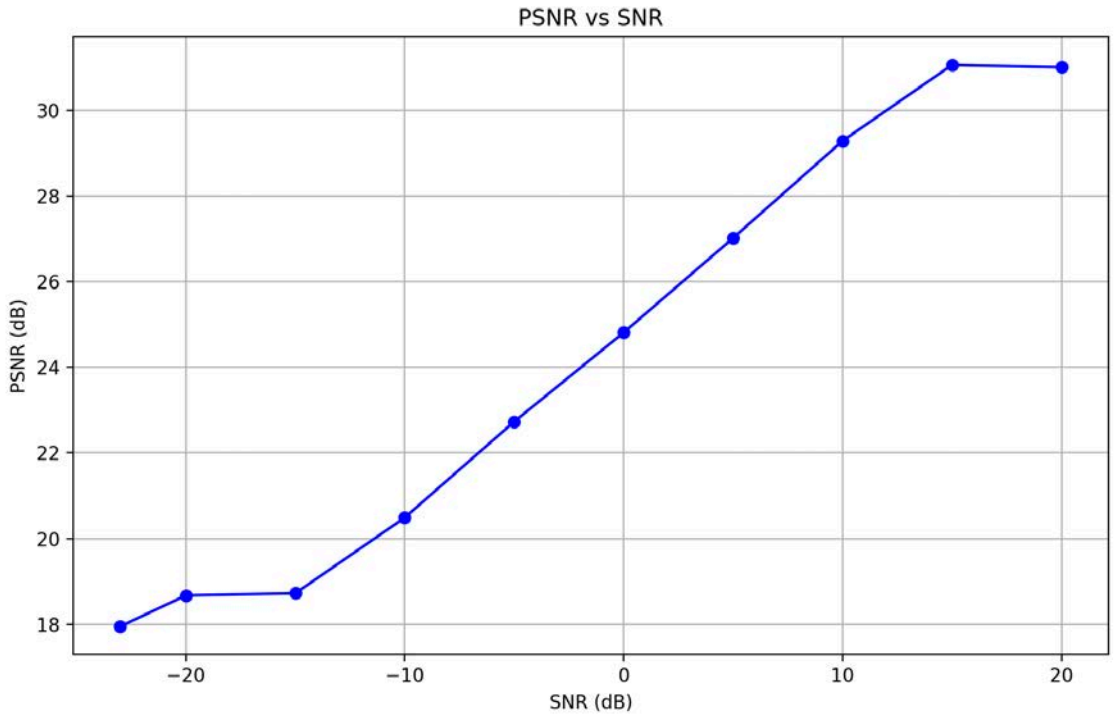

Figure 5.4: Average PSNR vs SNR

To provide a more detailed perspective, the following figures present the individual PSNR values obtained for all test samples, using SNR = –15 dB and SNR = 15 dB as indicative examples of low and high snr conditions. In each case, the average value is also plotted, offering insight into the distribution and consistency of the reconstruction quality at both low and high SNR levels.

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

SNR = -15dB:

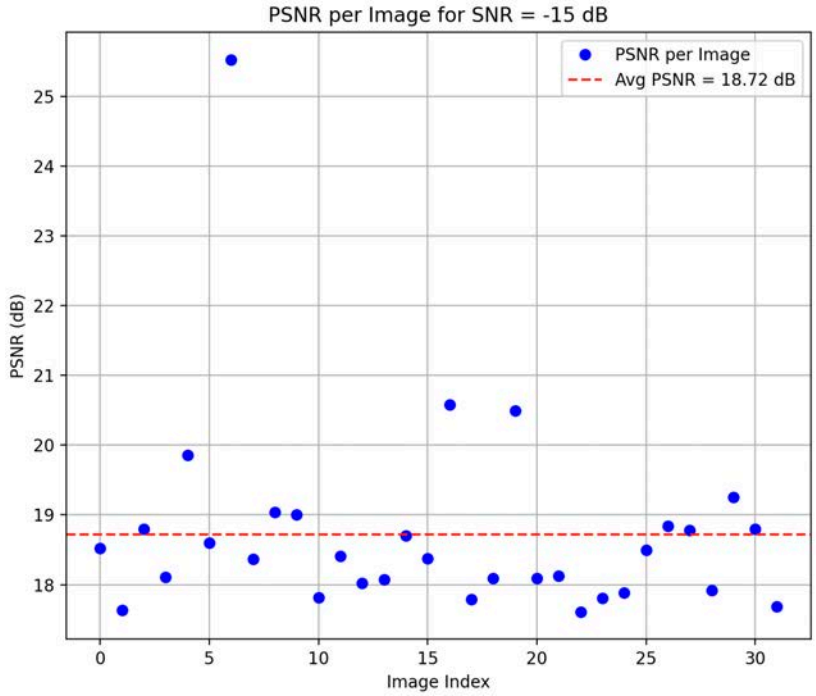

Figure 5.5: PSNR per Image for SNR = -15dB

SNR = 15dB:

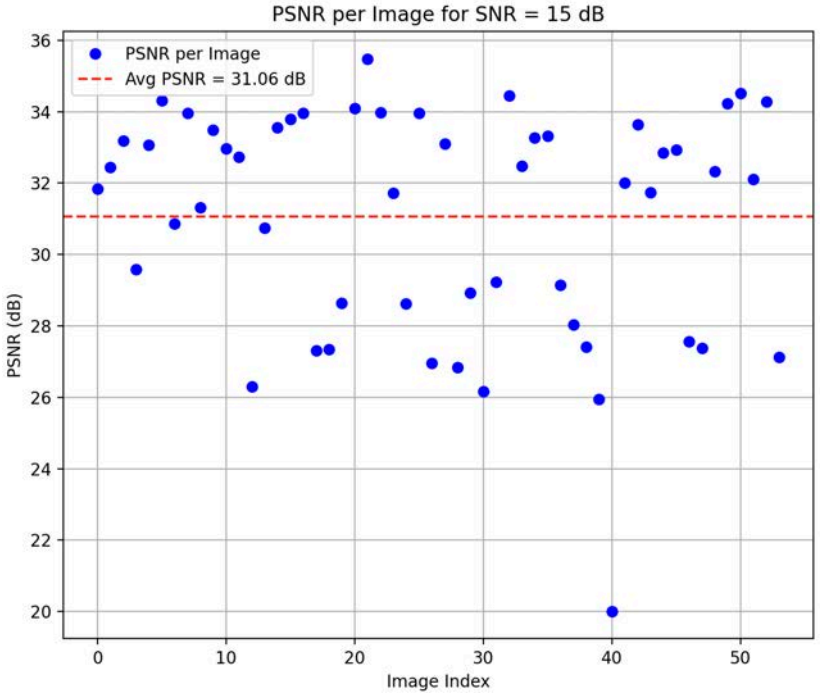

Figure 5.6: PSNR per Image for SNR = 15dB

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

5. Average SSIM vs SNR

The following table reports the average SSIM values for each SNR.

SNR (dB)	Average SSIM
-23	0.47
-20	0.49
-15	0.50
-10	0.57
-5	0.64
0	0.70
5	0.75
10	0.77
15	0.83
20	0.80

Table 5.5: Average SSIM Values vs SNR

The table shows that SSIM values generally increase as the SNR improves, reflecting better preservation of luminance (brightness), contrast, and structural information in the reconstructed images. This indicates that at higher SNR levels, the model more effectively maintains the overall image quality and important details. This trend is clearly visible in the corresponding figure below.

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

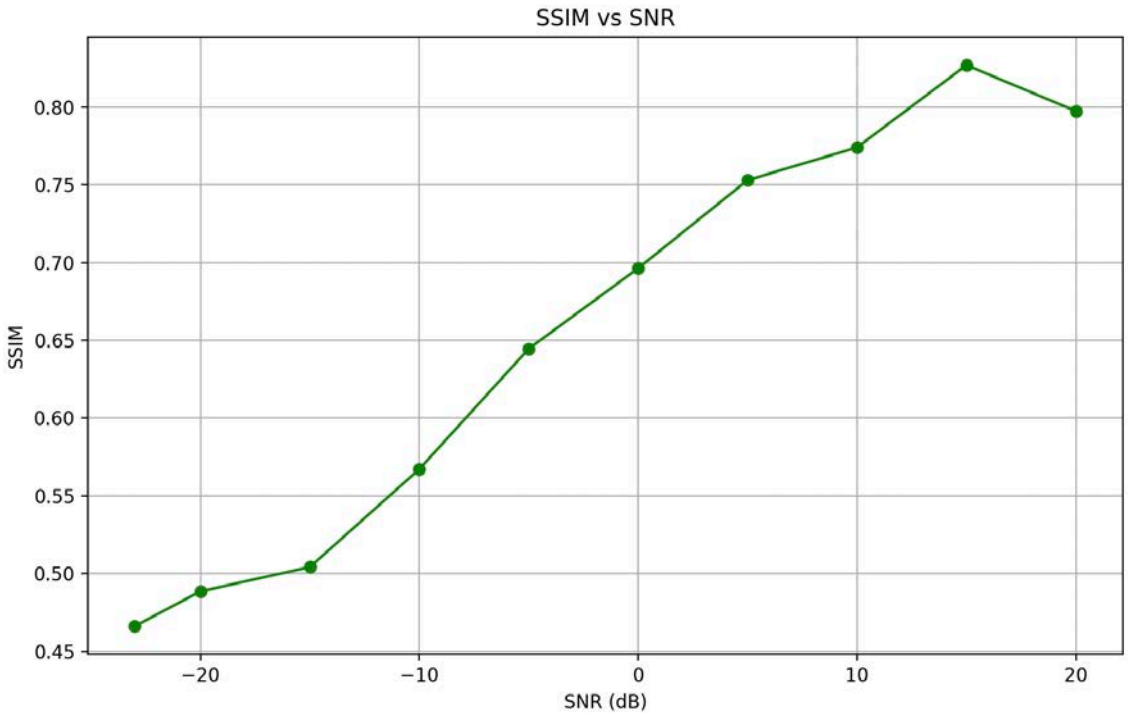

Figure 5.7: Average SSIM vs SNR

As we did for the PSNR evaluation metric, we present the individual SSIM values obtained for each test sample at the same SNR levels: –15 dB and 15 dB.

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

SNR = -15dB:

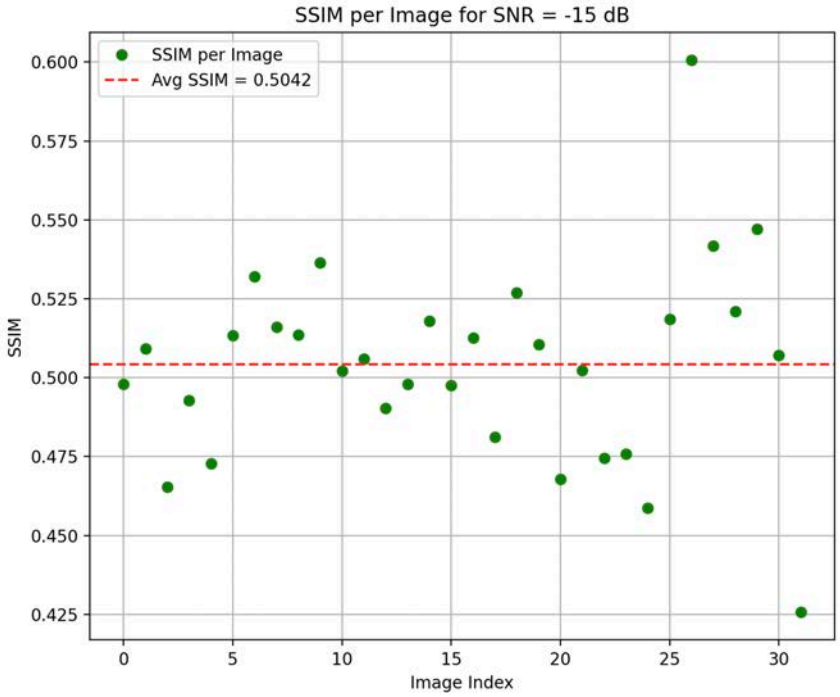

Figure 5.8: SSIM per Image for SNR = -15dB

SNR = 15dB:

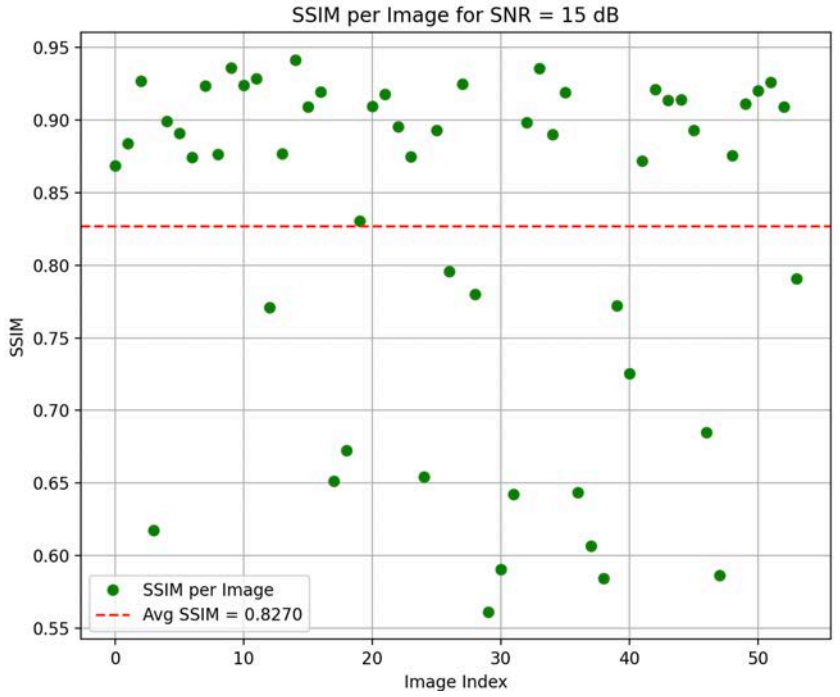

Figure 5.9: SSIM per Image for SNR = 15dB

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

5.2 Visual Evaluation of the Denoising Results

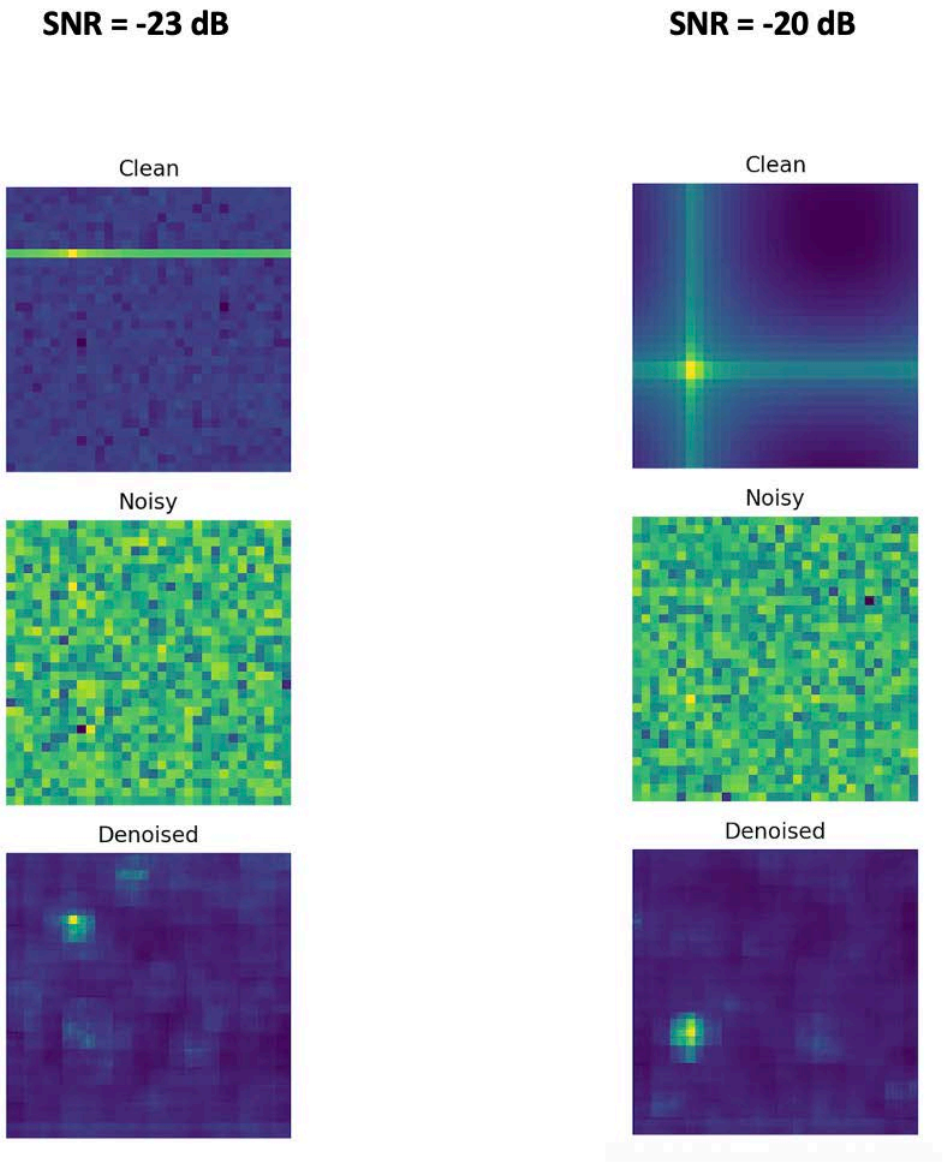

Figure 5.10: Examples of clean, noisy, and denoised images under the most challenging noise conditions in our experiments (-23 and -20 dB)

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

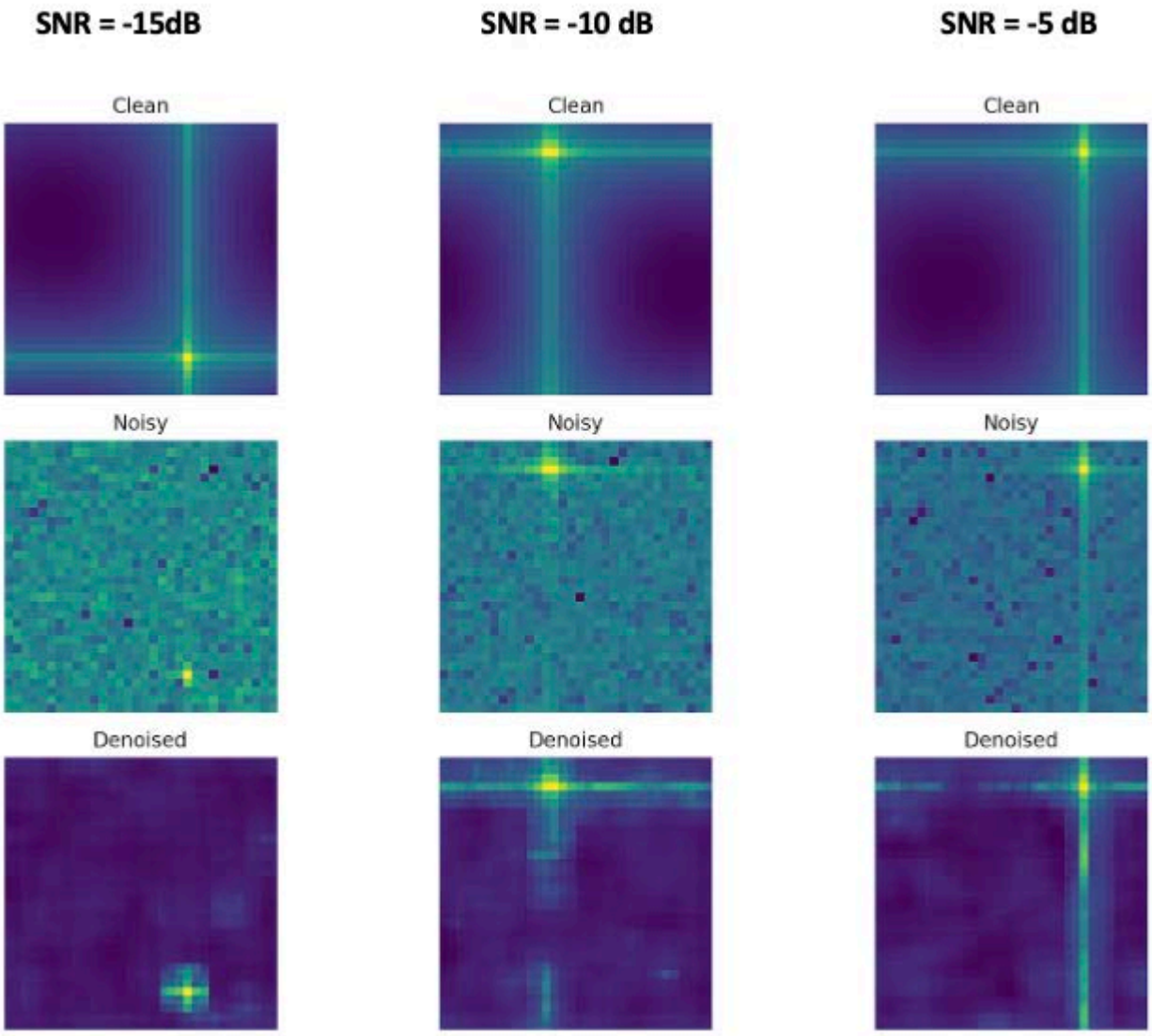

Figure 5.11: Examples of clean, noisy, and denoised images (-15, -10 and -5 dB)

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

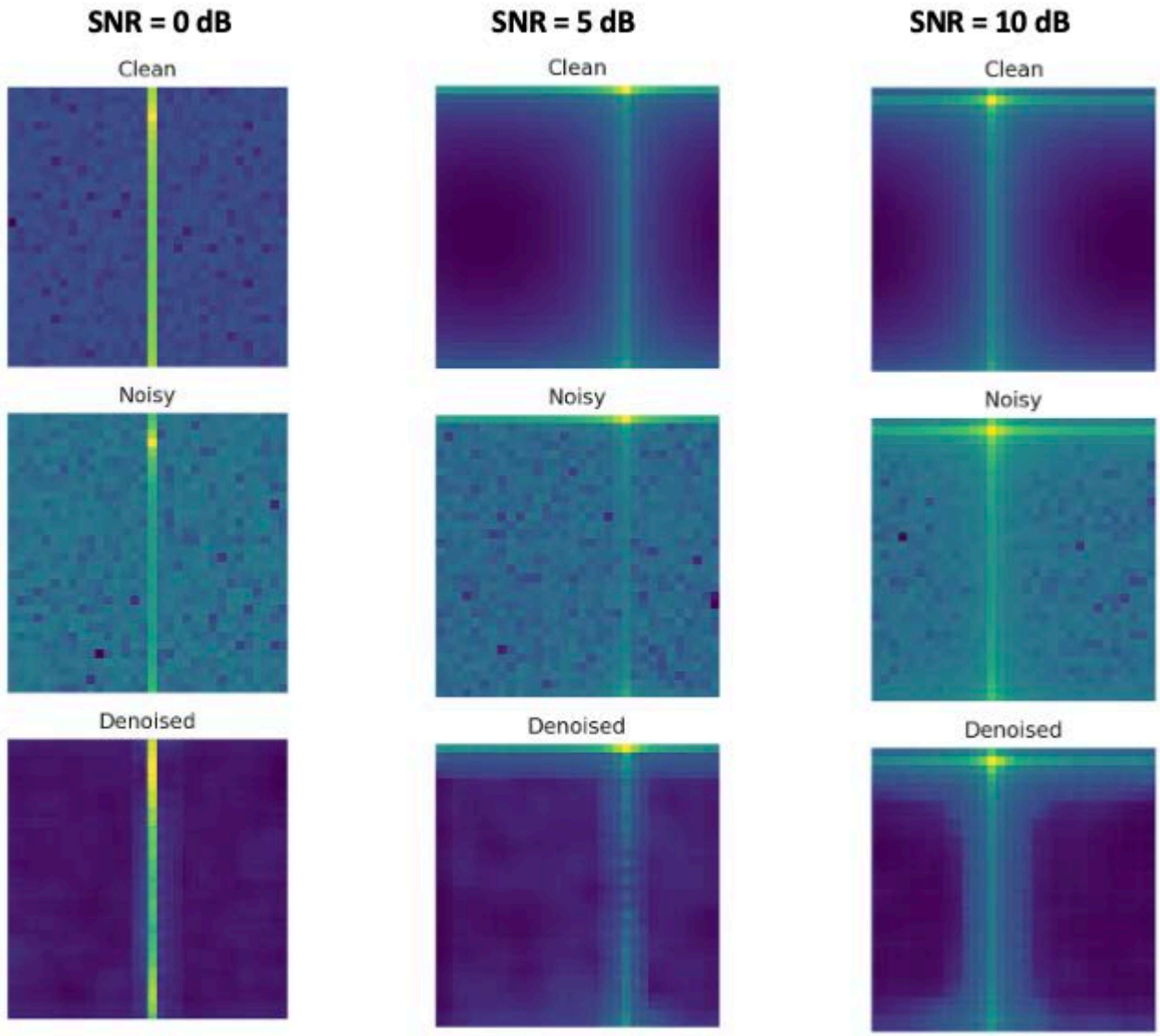

Figure 5.12: Examples of clean, noisy, and denoised images (0, 5 and 10 dB)

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

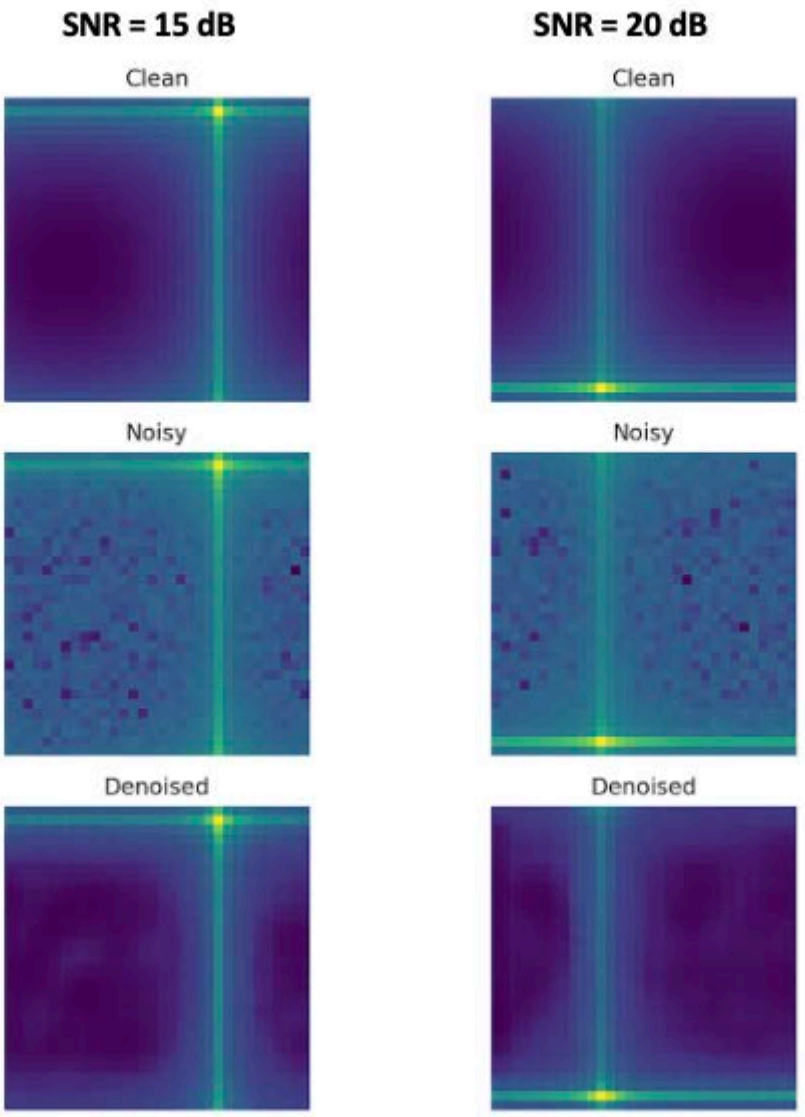

Figure 5.13: Examples of clean, noisy, and denoised images (15, 20 dB)

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Chapter 6

Conclusion

6.1 Summary and Key Findings

This thesis investigated the use of deep learning methods for Direction-of-Arrival estimation in uniform rectangular array configurations. The study began with an introduction to phased arrays and direction of arrival estimation. Next, the focus turned to Deep Learning with emphasis on autoencoders and their capacity to learn robust representations from noisy data. Building upon this theoretical background, we implemented and evaluated a self-supervised denoising autoencoder, trained on synthetically generated URA-based dataset. The model was specifically developed to denoise spatial spectrum maps affected by noise, enabling more accurate DoA estimation. The performance of the proposed method was assessed using a range of evaluation metrics, including Peak Detection Success Rate, Peak Distance Error, reconstruction loss, PSNR and SSIM.

The experimental results showed that the denoising autoencoder generally performs well in reducing noise and supporting accurate peak-based DoA estimation. The model achieved strong performance across most tested noise levels, with particularly good results in terms of Peak Detection Success Rate, which was one of the most important evaluation metrics used in this thesis. However, under extremely low SNR conditions—specifically around −23 dB—its performance began to degrade. At these noise levels, the DoA peak tended to shift or become less distinguishable, resulting in false peak detections. This suggested that the method struggled noticeably when the SNR was around −23dB, as expected due to peak ambiguity caused by those high noise levels. Nevertheless, it is precisely in these challenging scenarios that the denoising autoencoder proves most useful. While in higher SNR regimes

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

the conventional DFT-based method is generally sufficient for identifying the DoA peaks, the deep learning approach that was implemented here offers a clear advantage in low SNR conditions that make the target peak difficult to distinguish.

6.2 Future Extensions

While the proposed method showed promising results, several directions can be explored to further improve and extend this work. One potential extension involves training the model on more diverse or real-world datasets, which may improve its generalization capability and robustness in practical scenarios. Additionally, investigating the performance of the model under different array geometries or with varying numbers of sources could provide deeper insights into its limitations and strengths. Finally, a comparative study involving other denoising techniques could offer a more comprehensive evaluation of the proposed framework's advantages.

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

Bibliography

[1] Georgios Chrysanidis, Antonios Argyriou, Le-Nam Tran, Yanming Zhang, and Yanwei Liu. Power-efficient deceptive wireless beamforming against eavesdroppers. In *2025 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6, 2025.

[2] Antonios Argyriou. Passive angle-doppler profile estimation for narrowband digitally modulated wireless signals. In *2022 IEEE 12th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pages 246–250, 2022.

[3] Antonios Argyriou and Shion Andrew. DFT & MVDR Beamforming Algorithms for Transient Search in Radio Interferometers that Experience RFI. *Astrophysical Journal Supplement Series*, 2025.

[4] Christos Ntemkas and Antonios Argyriou. RFI and Jamming Detection in Antenna Arrays with an LSTM Autoencoder. In *2025 IEEE Radar Conference (RadarConf25)*, Krakow, Poland, 2025. IEEE.

[5] Antonios Argyriou. Joint estimation of the number of antennas and aoa of a wireless communication transmitter. In *2022 IEEE International Symposium on Phased Array Systems Technology (PAST)*, pages 1–4, 2022.

[6] Everything RF. Phased array antennas. `https://www.everythingrf.com/community/what-is-phased-array-antenna`, 2025. Last accessed June 17, 2025.

[7] R. J. Mailloux. Phased array theory and technology. *Proceedings of the IEEE*, 70(3):246–291, March 1982.

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

[8] Marija Agatonović, Z. Stankovic, Nebojsa Doncov, Leen Sit, Bratislav Milovanović, Tomas, and Thomas Zwick. Application of artificial neural networks for efficient high-resolution 2d doa estimation. *Radioengineering*, 21:1178–1186, 2012.

[9] N. K. Chauhan and K. Singh. A review on conventional machine learning vs deep learning. In *2018 International Conference on Computing, Power and Communication Technologies (GUCON)*, pages 347–352, Greater Noida, India, 2018.

[10] G. K. Papageorgiou and M. Sellathurai. Direction-of-arrival estimation in the low-snr regime via a denoising autoencoder. In *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1–5, Atlanta, GA, USA, 2020.

[11] Mohit Sewak, Sanjay Sahay, and Hemant Rathore. An overview of deep learning architecture of deep neural networks and autoencoders. *Journal of Computational and Theoretical Nanoscience*, 17:182–188, 2020.

[12] N. Shah and A. Ganatra. Comparative study of autoencoders-its types and application. In *2022 6th International Conference on Electronics, Communication and Aerospace Technology*, pages 175–180, Coimbatore, India, 2022.

[13] Komal Bajaj, Dushyant Kumar Singh, and Mohd. Aquib Ansari. Autoencoders based deep learner for image denoising. *Procedia Computer Science*, 171:1535–1541, 2020.

[14] D. Bank, N. Koenigstein, and R. Giryes. Autoencoders. In L. Rokach, O. Maimon, and E. Shmueli, editors, *Machine Learning for Data Science Handbook*. Springer, Cham, 2023.

[15] Saba Hesaraki. Feature map. `https://medium.com/@saba99/feature-map-35ba7e6c689e`. Last accessed June 17, 2025.

[16] Vanna Winland. Self-attention. `https://www.ibm.com/think/topics/self-attention`. Last accessed June 17, 2025.

[17] A guide to receptive field arithmetic for convolutional neural networks. `https://blog.mlreview.com/a-guide-to-receptive-field-arithmetic-for-convolutional-neural-networks-e0f514068807`. Last accessed June 17, 2025.

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250

[18] Keshav Aggarwal. Transpose convolution explained for up-sampling images. `https://www.digitalocean.com/community/tutorials/transpose-convolution#convolution-operation`. Last accessed June 17, 2025.

[19] Ranjan Kumar Mishra, G. Y. Sandesh Reddy, and Himanshu Pathak. The understanding of deep learning: A comprehensive review. *Mathematical Problems in Engineering*, 2021:15, 2021.

[20] U. Sara, M. Akter, and M. S. Uddin. Image quality assessment through fsim, ssim, mse and psnr—a comparative study. *Journal of Computer and Communications*, 7:8–18, 2019.

Institutional Repository - Library & Information Centre - University of Thessaly
03/09/2026 15:17:47 EEST - 216.73.216.250