



Designing Cross-Domain Attribution Analysis into A Transformer for Enhancing 5G Under SNR Conditions

Jiaxuan Xu

¹ New Oriental International Curriculum Center, Guangzhou, Guangdong, China
mouexu@gmail.com

Abstract. The Transformer framework has recently been proven to be feasible for noise analysis in low Signal-to-Noise Ratio (SNR) environments. The signal is heavily overwhelmed by noise (the lower the Signal-to-Noise Ratio, the closer or even greater the noise energy becomes compared to the signal energy), making it difficult for traditional methods to distinguish between "useful signal features" and "noise features"; if multiple modalities are superimposed, noise can further confuse features through "cross-modal coupling". Optimizing the Transformer's multimodal noise cross-domain attribution analysis capability provides clear assistance for low SNR noise analysis. It is one of the key technical paths to address the challenges of "difficult source tracing, inaccurate quantification, and coarse denoising" in low SNR multimodal noise. This work presents four parts of the optimizations for multimodal noise cross-domain attribution analysis based on the basic algorithm, which can help with cross-domain attribution analysis and further enhance the capability of solving complex signals.

Keywords: Transformer, signal, multimodal noise cross-domain attribution analysis

1 Introduction

If we aim to enhance 5G, the establishment of a joint time-space-frequency modeling architecture emerges as the most challenging issue at hand, because as the signal condition becomes increasingly complicated, the noise also becomes more complex, leading to the following problems.

Firstly, Increased bit error rate: A low SNR indicates that the signal strength is relatively weaker compared to the noise, making the signal susceptible to noise interference. This makes it difficult for the receiving end to accurately identify the signal. However, if a deep learning model is employed for channel estimation and equalization, it can significantly reduce the bit error rate compared to traditional algorithms such as Least Squares (LS) and Minimum Mean Square Error (MMSE).

Secondly, decreased data transmission rate: A low SNR can affect the accuracy of modulation and demodulation. For instance, in scenarios with severe signal fading, a 5G network may switch from a high-order modulation method to a low-order one, such as switching from 64 Quadrature Amplitude Modulation (QAM) to Quadrature Phase Shift Keying (QPSK), thereby reducing the transmission rate.

© The Author(s) 2026

S. Zhang (ed.), *Proceedings of the 2025 International Conference on Electronics, Electrical and Grid Technology (ICEEGT 2025)*, Advances in Engineering Research 292,

https://doi.org/10.2991/978-94-6463-986-5_94

Thirdly, Increased synchronization difficulty: The initial synchronization of the downlink in 5G systems relies on the Primary Synchronization Signal (PSS), etc. For example, when the SNR is -10dB, the detection probability of the PSS will significantly decrease, affecting the synchronization performance between the terminal and the base station.

Fourthly, Signal detection and recovery are affected: a Low SNR can make signal detection and recovery difficult, but 5G has adopted some technologies to address this issue. For example, Code Division Multiplexing (CDM) technology can enable the receiving end to perform signal detection and recovery more accurately by reducing interference. At the same time, enhancing the density and accuracy of the Demodulation Reference Signal (DMRS) also helps improve the accuracy of signal detection and channel estimation.

Lastly, some modulation methods still exhibit robustness. The $\pi/2$ -Binary Phase Shift Keying (BPSK) modulation method in 5G systems demonstrates robustness against low signal-to-noise ratios, making it suitable for the uplink Physical Random Access Channel (PRACH) and control signals in mobile networks, such as 5G. It can reduce the nonlinear effects of power amplifiers and improve signal integrity.

All in all, the performance of the 5G signal under low SNR conditions will significantly decrease: the Spectral efficiency will decrease by a factor of 4, and the user peak rate will be reduced by 80% compared to normal conditions.

Three dominating reasons lead to these troublesome performances. Firstly, High-Frequency Propagation Loss (e.g., Millimeter Wave Bands, more than 24GHz), one important physical principle of loss sources is Free-space path loss (FSPL)

For example, in Indoor Non-Line-of-Sight (NLOS), High-frequency loss is 400% higher than that of traditional models [1].

Another is that the penetration/diffraction capability sharply decreases when the signal attempts to penetrate the 1.2cm gypsum board and the 15cm concrete wall. Their additional losses reached 5.2 and greater than 80 (signal interruption), respectively[2].

Secondly, Vulnerability of high-order modulation: High-order modulation (such as 256QAM, 1024QAM) enhances spectral efficiency by carrying more bits within a single symbol, but its vulnerability is significantly magnified in low signal-to-noise ratio (SNR) environments. The key to the vulnerability of high-order modulation: High-order modulation is a fundamental trade-off between spectral efficiency and robustness. Thirdly, overhead of multi-antenna system: Multi-antenna systems (such as Massive MIMO) enhance capacity and coverage through spatial degrees of freedom, but their overhead increases rapidly with the number of antennas, becoming a core bottleneck in 5G/6G deployments

To sum up, being immersed in a low SNR condition causes the signal to become weak to the point where the energy/frequency of the signal is overwhelmed by that of the noise.

Traditional methods, such as spread spectrum technology, matched filters, and adaptive filters, have been proven to be replaced or integrated by AI algorithms [3]. Furthermore, integrating these methods into the I scheme approximates Bayesian optimality under non-Gaussian noise [4].

Actually, in 2012, the Convolutional Neural Networks (CNN) started to perform in some noise solution [5], this method began to burst in 2018

But CNN has poor adaptability to dynamic noise, deficiencies in long-range dependency modeling, low efficiency in complex signal processing, low computing energy efficiency, and failure in noise-signal coupling scenarios and graph layers for learning unstructured local relations in a system restricted to the specific resolution and the geometry observed in the training phase [6, 7].

Those problem that causes high error rate are optimized by performing the transformer frame or the combined system of CNN+transformer, which has a much lower Bit Error Rate (BER) than the traditional CNN method [8].

Conventional signals exhibit temporal/frequency domain continuity, whereas noise undergoes random mutations.

However, in our latest research, we have proposed a solution with enhanced adaptability to dynamic noise, which is a framework that integrates CNN and Transformer

Microsoft and Tsinghua University (THU) collaborate to present a successful case study on the application and design of the combined transformer system, which solves the signal drowning in low SNR conditions [9].

They provided a detailed introduction to standard attention, softmax-free attention, Fourier-type attention, and Galerkin-type attention, as well as methods for enhancing model performance through normalization. They proposed an Encoder-Decoder structure, where the input encoder processes sampling points and function values, the query encoder processes query positions, and information is transferred across attention modules.

After that, they raise and introduce the operator transformer (Ofprmer), which is a fully point-based attention architecture, featuring flexibility in operator learning and strong generalization capabilities.

This method has obvious advantages in the aspect of non-stationary noise suppression, low-complexity real-time processing, and cross-scene generalization, and lays the foundation for 6G enhancement.

2 Related work

The transformer enables breaking the technical barrier that signal analysis has long faced—specifically, the uncontrollable propagation of noise and ambiguity in attribution during cross-modal interaction. Optimizing its "cross-domain attribution capability" precisely addresses the core pain points in low SNR scenarios: the confusion between noise and signal, and the interference caused by cross-modal noise superposition.

2.1 Background

Cross-domain attribution analysis of multimodal noise: using self-attention-based "modal alignment" to analyze the transmission path of noise

Multimodal data provides a more comprehensive dimension of information, enabling researchers to overcome the limitations of a single data source and obtain more accurate and real-world answers.

In today's world, AI has a wide range of applications in fields such as intelligent security, medical diagnosis, and autonomous driving. That's all because in the multimodal data, the signal may come from different domains. The propagation and interaction of noise between these domains can affect the consistency and reliability of data [10]. So, cross-domain attribution analysis will be important for understanding the multimodal data of noise. But during manufacturing and research, there are introductions of various types of noise that strongly affect the accuracy of the data and the stability of the system [11].

2.2 Basic Principles

Multimodal data consists of various modalities such as vision, audition, and text, and noise exhibits different characteristics in each modality [11]. For example, when using a brainwave signal (a classical SNR signal) and functional magnetic resonance imaging (fMRI), the electromagnetic noise will simultaneously interrupt the signal recorded from the brainwaves.

The key method for our frame consists of self-attention mechanism, optimal transport theory, time-frequency analysis, etc. For the self-attention mechanism, it has been widely used in processing multimodal data to capture the relationships within and between modalities [12]. All in all, self-attention means for each element in a set of Tokens, calculate its association strength (attention weight) with all elements in the group (including itself), and then weight and sum the values of all elements based on these weights to obtain a new representation that integrates the global context. it key for this analyse since its enable help the computer associate each token. For Optimal transport theory, it provides a mathematical framework for aligning and comparing multimodal data in the presence of noise.

2.3 Noise classification and characteristic analysis

All method is based on the premise of the EEG condition. Because the amplitude of EEG signals recorded from the scalp is very small, typically on the order of 10-100 microvolts, this provides an ideal and natural experimental condition.

In the signal analysis using the transformer, every fraction of the signal is endowed with a token. In order to avoid the error caused by the noise as a token, the noise will be inhibited respectively.

Classifying modality-specific noise in visual data into sensor-related noise and illumination-induced noise is the key to the analysis [13]. For example, for the electroencephalogram (EEG) research. What signal we got is the combined signals with brainwave, fMRI, Eye-Tracking, Magnetoencephalography (MEG), and Behavioral. In this case, we tend to design an independent and special system for each, because each of these noises can be classified into Modality-Specific Noise and Cross-Modality Noise. Specific modal noise refers to interference that only affects a certain modal data

and is directly related to the physical characteristics, acquisition principles, or equipment characteristics of that modality. This type of noise does not propagate across modalities and only exists in a single data type. And cross-modal noise refers to interference generated by a single source that simultaneously affects two or more modal data, or cross-modal interference caused by the interaction and synchronization errors between multimodal devices. The impact of such noise is not limited to a single mode, but spreads through shared physical environment, physiological activities, or equipment association. All in all, recognizing each noise and designing each system is a difficulty in designing a system for multimodal signals. All methods are based on the premise of the EEG condition. Because the amplitude of EEG signals recorded from the scalp is very small, typically on the order of 10-100 microvolts, this provides an ideal and natural experimental condition. These include the following types of signals:

EEG (electroencephalogram) has frequency domain characteristics, Time domain characteristics, and Model-driven features

Frequency domain characteristics represent the proportion of the 50/60Hz frequency band in power spectral density (PSD) (quantifying the intensity of power frequency noise);

High frequency band (>30Hz) energy proportion (where electromyographic noise is mostly concentrated).

Time domain characteristics represent Signal variance (noise increases overall fluctuations), kurtosis (eye/muscle electrical noise is often a spike signal with high kurtosis).

Transient mutation rate (measured by threshold detection, such as the proportion of samples exceeding ± 3 standard deviations, reflecting electrode contact noise).

Model-driven features represent the variance contribution rate of noise components separated by Independent Component Analysis (ICA).

Autoencoder reconstruction error (for models trained on clean EEG, the reconstruction residual of noisy samples can characterize the noise intensity).

fMRI (Functional Magnetic Resonance Imaging) with Time series characteristics, Spatial features, and Model-driven features.

Time series characteristics represent the standard deviation of voxel time series (increased noise will increase the standard deviation). Low frequency drift intensity (quantifying physiological noise by comparing the variance ratio of the signal before and after high-pass filtering);

The correlation between head motion parameters (such as inter-frame displacement) and signals (evaluating the impact of motion artifacts).

Spatial features represent spatial smoothness (low spatial autocorrelation in noisy areas, which can be quantified by the difference before and after Gaussian filtering);

Non-brain area signal intensity (such as the average signal of cerebrospinal fluid area, reflecting scanner noise overflow).

Model-driven features represent the Regression residuals based on physiological signals (respiratory/heartbeat recordings) (the larger the residuals, the stronger the unexplained physiological noise);

The energy proportion of noise components extracted by spatially independent component analysis (sICA), such as vascular artifacts and scanner noise.

Eye Tracking with Trajectory characteristics, Event characteristics, and Model-driven features:

Trajectory characteristics refer to Gaze dispersion (like the standard deviation of gaze points in a region of interest, increasing noise will enhance dispersion).

Abnormal scanning speed (normal scanning speed has a range; noise can cause speed jumps).

Event characteristics refer to Blink frequency and duration (blinking beyond the normal range may be accompanied by signal noise).

Signal loss rate (the proportion of invalid data caused by device occlusion or blinking).

Model-driven features refer to the Kalman filter prediction error (using a smooth trajectory model to predict the actual trajectory, the larger the error, the stronger the noise).

The effectiveness of fixation point clustering (Density Based Spatial Clustering of Applications with Noise (DBSCAN) clustering, reflecting trajectory stability).

MEG (magnetoencephalography) with Frequency domain characteristics, Temporal and spatial characteristics, and Model-driven features.

Frequency domain characteristics refer to the Peak power spectrum of environmental interference frequency bands (such as 50Hz power lines and high-frequency electronic equipment noise);

The energy proportion of the cardiac magnetic interference frequency band (~1Hz) (quantifiable when synchronized with the Electrocardiogram (ECG)).

Temporal and spatial characteristics refer to the signal consistency of the sensor array (noise interference often manifests as abnormal fluctuations in local sensors, with low consistency).

Head motion compensation residual (the signal residual corrected by head motion tracking data, reflecting the motion noise that has not been eliminated).

Model-driven features refer to the ratio of baseline noise to measured noise in a magnetic shielding environment (to evaluate the intensity of environmental interference);

The residual energy (spatial distribution characteristics of noise sources) of beamforming spatial filtering.

Behavioral data with Reaction time characteristics, Accuracy features, and Model-driven features

Reaction time characteristics refer to the proportion of outliers in the reaction time (such as the proportion of trials exceeding the mean ± 2.5 standard deviation).

and the Reaction time volatility (standard deviation of reaction time for consecutive trials, increased noise will enhance volatility).

Accuracy features refer to the distribution of types of incorrect attempts (such as the proportion of random errors, reflecting attention noise).

Speed accuracy trade-off outlier (deviation from the subject's own baseline).

Model-driven features refer to the noise parameters of the Drift Diffusion Model (DDM), such as decision noise σ , which directly quantify the noise in behavioral decision-making;

Sequence-dependent residuals (normal behavioral responses exhibit sequence correlation, noise reduces this correlation, and residuals increase)

2.4 Cross-domain attribution analysis method

After we have figured out the type of noises and the principle of analysis, we can now execute the feature extraction.

As this paper mentioned earlier, in the case of EEG, it is affected by noise of different modes. So we have to separate them into 3 cases.

Single modal noise feature extraction: application of time-frequency features and self-attention. Single modal noise has significant modal specificity (such as power frequency interference in EEG and Gaussian noise in images), and its dynamic characteristics need to be captured through time-frequency analysis first, and then the key areas of the noise can be strengthened by a self-attention mechanism.

Time frequency analysis applies to temporal modalities (such as EEG, MEG, audio) and spatiotemporal modalities (such as video, dynamic fMRI). Core methods include the Short-Time Fourier Transform (STFT), the Wavelet Transform, and the Hilbert-Huang Transform (HHT) [13]. Use time-frequency analysis to extract features from noisy signals, which can effectively represent the characteristics of different noises.

The self-attention mechanism can highlight dense noise areas by calculating the correlation weights between data points.

In Single modal, the transformer performs (Figure 1):

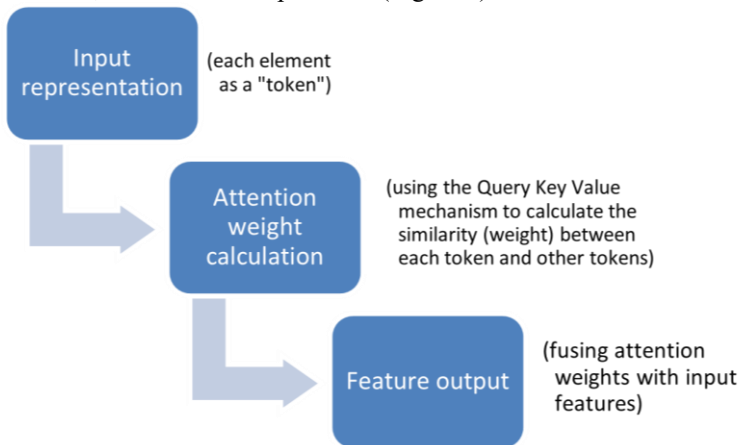


Fig. 1. Transformer operation logic

Multi-modal noise feature fusion: capturing cross-modal correlations. The core of multimodal noise is cross-modal correlation (such as head movements simultaneously affecting EEG and fMRI), which requires aligning time-frequency features and cross-

attention mechanisms to explore the synchronization patterns of noise between modalities.

Cross-modal alignment and fusion of time-frequency features, due to the significant differences in time-frequency resolution between different modalities (such as millisecond-level EEG vs. second-level fMRI), require alignment of the time-frequency dimension before fusing noise features. First of all, the system will define the time domain or the frequency domain. Interpolation of low temporal resolution modalities (such as fMRI) to match the temporal axis of high-resolution modalities (such as EEG), or downsample high-resolution modalities and align low-resolution modalities. For the frequency domain, map the time-frequency characteristics of each modality to a common frequency range (such as 0-50Hz), ignoring modality-specific high frequencies (such as MEG's high-frequency environmental noise that may be unrelated to EEG).

The fusion strategy consists of two steps: feature concatenation and weighted fusion.

Feature concatenation: Concatenate the aligned multimodal time-frequency features (such as the STFT matrix of EEG+wavelet coefficients of fMRI) into a joint matrix, calculate the singular value decomposition (SVD) of the matrix, and take the first k singular values as the compressed features of cross-modal noise.

Secondly, using the correlation coefficient of intermodal noise (such as Pearson correlation between EEG electromyographic noise and fMRI head movement noise) as weights, the time-frequency features are weighted and summed to highlight strongly correlated noise. For example, for the Noise fusion of EEG and eye tracking, the input token for EEG is the "time window feature containing eye movement artifacts", and the input token for eye tracking is the "time window feature of blink events".

Then, Cross attention assigns high weights to time overlapping windows, and the fused features can quantify the synchronization strength (noise correlation) of "eye movement artifacts blink".

All in all, the core logic for extracting noise features from multimodal data is: first, capturing the time-frequency dynamics of single-mode noise using time-frequency analysis. Second, using self-attention to enhance the local correlation of intra-modal noise. Last, using modal alignment and cross attention to mine cross-modal synchronization patterns of noise.

3 Method

To transform the classic Transformer signal processing framework to support "multimodal noise cross-domain attribution analysis", structural adjustments need to be made at four levels: input processing, attention mechanism, feature interaction, and analysis module, targeting the three core objectives of noise tracking, modal alignment, and attribution quantification.

3.1 Input layer modification

To retain noise features and enhance modal distinguishability, we should create a mutually independent input layer, allowing for a more accurate analysis.

Dual-channel input for modal separation. Building a Parallel construction of "signal channel" and "noise channel" for each modal signal, for the signal channel, using filtering and detrending to extract the feature of the signal. For the noise channel, separating the part that is caused by noise through residual calculation.

Noise prior feature embedding. Extract frequency domain features from known physical characteristics of noise through the Fourier transform, embed them as "noise prior vectors" into the input sequence, and guide the model to focus on the characteristic features of noise.

3.2 Attention mechanism transformation:

In cross-domain attribution analysis, introducing a "noise-guided cross-domain alignment attention" can enhance the accuracy of modal feature alignment by leveraging noise information, reduce false correlations, improve the model's robustness against noise, and facilitate fine-grained feature interaction through dynamic weighting, providing reliable support for attribution.

Allocate exclusive heads (such as 1/3 of the total number of heads) in multi-head attention, forcing it to only learn the correlations between noise channels. For example, the input only receives the noise channel features of each modality. The mask design shields the self-attention in the same mode through the mask matrix (only cross-modal noise interaction is allowed), forcing the model to learn the transfer relationship of "modal A noise \rightarrow modal B noise". Additionally, the noise interaction weight matrix $W^{\text{noise}} \in R^{(N \times M)}$ at the output end, which shows the cross-modal correlation strength of noise directly.

Besides, designing a special Dynamic timing alignment mechanism is also significant in the case of delay. Calculate the time difference $\Delta t = |i - j|$ for the i -th time step of modality A and the j -th time step of modality B, encode it as a vector, and add it to the attention score calculation:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T + \text{OFFset}(\Delta t)}{\sqrt{d_k}}\right) \quad (1)$$

3.3 Feature interaction layer transformation:

In cross-modal attribution analysis, adding the "noise transmission path tracking module" can precisely trace the trajectory of noise in multimodal data interaction and feature propagation, locate the source of noise and interfering nodes, and then strip away false associations induced by noise, reduce attribution misguidance, while

supplementing the explanation of the impact of noise on feature weights and enhancing the interpretability of attribution.

Constructing a Noise Transmission Graph, for example: Sing modalities as nodes (such as EEG, Electromyography (EMG), and Electrooculography (EOG)) and the weights of noisy interactive attention heads as edge weights, construct a dynamic directed graph:

Direction of edges: from the weight matrix W_{noise} , the "row mode" points to the "column mode" (e.g., $W[EEG, EMG]$ represents the transfer of EEG noise to EMG noise);

Edge weight: Take the average or peak value of W_{noise} and quantify the transmission strength.

Additionally, adding a filter for the limitation on weight can eliminate random associations to a point that the intervention will be reduced, for example, if $W[EMG \rightarrow EEG] > 0.7$ and $W[EEG \rightarrow EOG] > 0.5$, then the path $EMG \rightarrow EEG \rightarrow EOG$ is conserved.

3.4 Output layer modification

In cross-domain attribution analysis, the core function of adding a "noise attribution quantification module" is to precisely measure, through quantitative calculation, the degree of interference and the proportion of deviation contribution caused by noise in different modalities and different propagation links to the cross-modal attribution results. This not only provides a data basis for identifying key noise interference sources, but also lays a quantitative foundation for subsequent targeted optimization of attribution logic and improvement of attribution accuracy.

Since the classic transformer output only focuses on signal reconstruction or classification results, it is necessary to add attribution analysis output. Firstly, adding the Calculation of the noise source contribution. For example, calculate the contribution of each source mode for the targeted signal(the EEG needed to be denoised). The contribution is separated into 3 types:

1. Direct edge weights from other modalities to EEG (such as the weight from EMG to EEG)
2. Indirect contribution: The product of weights transmitted through intermediate modalities (e.g., the weight from EOG to EMG to EEG is $W[EOG \rightarrow EMG] \times W[EMG \rightarrow EEG]$)
3. Total contribution: The sum of the normalized direct contribution and the indirect contribution, which indicates the proportion of each modality's impact on the noise of the target modality.

Quantitative analysis first, followed by a visual presentation.

Then the contribution can be transferred into Visualized attribution results: Output a dynamic thermodynamic diagram or a Contribution degree time series curve.

3.5 Loss function modification

Integrating "noise attribution supervision" can guide the model to actively identify and avoid noise interference when learning cross-modal associations by constructing noise-related supervision signals (such as labeling noise-disturbed modal regions and setting thresholds for attribution bias caused by noise). At the same time, it standardizes the judgment logic of "noise impact" in the attribution process, reduces the model's misjudgment of key associations due to noise, and ultimately improves the accuracy and reliability of cross-modal attribution results.

Basically, this step is to add attribution loss to the original loss.

Generate supervised signals through the "noise injection experiment": inject noise of known intensity into modality A, and measure the noise enhancement in the target modality B, which serves as the "true contribution" of A to B;

Minimize the Mean Squared Error (MSE) between the predicted contribution degree of the model and the actual contribution degree: $L_{contri} = MSE$

4 Conclusion

These 4 optimizations can effectively work on the basic algorithm to make the calculation for the analysis of the complicated signal faster and solve the Complex signal discrimination under low SNR conditions. However, how much success it can achieve and the reasons that might influence the capability and the efficiency should be tested by a large number of parameters.

What's more, this paper raises some of the opposition to multimodal noise cross-domain attribution analysis, which could solve the problem that the current transformer frame might face in signal analysis. Optimizing the cross-domain attribution analysis capability of Transformer for multimodal noise has a clear help for low SNR noise analysis, and is one of the key technical paths to solve the "difficult traceability, accurate quantification, and rough denoising" of low SNR multimodal noise. When optimizing, it is important to focus on "feature stability modeling under low SNR", "dynamic cross-domain noise tracking", and "linkage between attribution results and denoising steps" - these three directions have the highest matching degree with pain points in low SNR scenarios and can maximize the optimization value.

References

1. Deng, S, J., Samimi, M, K.; Rappaport, T, S.: 28 GHz and 73 GHz millimeter-wave indoor propagation measurements and path loss models. In: 2015 IEEE International Conference on Communication Workshop (ICCW), pp. 1244-1250. IEEE, London (2015)
2. Hu, J, H., Al-jzari, A., Salous, S.: Indoor Channel Characterization Based on Directional Measurements at 140 GHz. In: 2024 18th European Conference on Antennas and Propagation (EuCAP), pp. 1-5. IEEE, Glasgow (2024)
3. Rahmani, M., Zhao, J, B., Bashar, M., Cumanan, K., Burr, A., Tafazolli, R.: Securing 5G NR Networks: Innovative Artificial Noise Methods for Protecting Cell-Free Massive

- MIMO. In: 2025 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1-7. IEEE, Milan (2025)
4. Challita, U., Ryden, H., Tullberg, H.: When Machine Learning Meets Wireless Cellular Networks: Deployment, Challenges, and Applications. *IEEE Communications Magazine*. 58(6), 12-18 (2020)
 5. Xie, J, Y., Xu, L, L., Chen, E, H.: Image denoising and inpainting with deep neural networks. In: 26th International Conference on Neural Information Processing Systems, pp. 341-349. University of Science and Technology of China, Newyork (2012)
 6. Belbute-Peres, F, D, A., Economon, T, D., Kolter, J, Z.: Combining differentiable pde solvers and graph neural networks for fluid flow prediction. In: Proceedings of the 37th International Conference on Machine Learning, pp. 2402-2411 (2020)
 7. Lakovlev, V., Heinonen, M., Lähdesmäki, H.: Learning continuous-time PDEs from sparse data with graph neural networks. In: International Conference on Learning Representations, (2021)
 8. Zhao, Y, C., Wang, G, T., Tang, C, X., Luo, C., Zeng, W, J., Zha, Z, J.: A Battle of Network Structures: An Empirical Study of CNN, Transformer, and MLP
 9. Li, Z, J., Meidani, K., Farimani, A, B.: Transformer for Partial Differential Equations' Operator Learning. *Transactions on Machine Learning Research*. (2023)
 10. Yang, M, X., Huang, Z, Y., Hu, P., Li; T, H., Lv, J, C., Peng, X.: Learning with Twin Noisy Labels for Visible-Infrared Person Re-Identification. In: CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 18-24 IEEE, New Orleans (2022)
 11. Cai, R., Dong, J, F., Liang, T, X., Liang, Y, H., Wang, Y, B., Yang, X.: Cross-Lingual Cross-Modal Retrieval With Noise-Robust Fine-Tuning. *IEEE Transactions on Knowledge and Data Engineering*. 36(11), 5860 – 5873 (2024)
 12. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A, N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 6000 – 6010 (2017)
 13. Yang, M, X., Huang, Z, Y., Hu, P., Li; T, H., Lv, J, C., Peng, X.: Learning with Twin Noisy Labels for Visible-Infrared Person Re-Identification. In: CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 18-24 IEEE, New Orleans (2022)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

