VISUAL AND TEXTUAL INTELLIGENCE: A DEEP LEARNING APPROACH TO INFORMATION RETRIEVAL

Bibi Noreen Ayesha
M.E, Department of Computer Science and Engineering
Muffakham jah College of Engineering and Technology
Hyderabad, India
Cloud Engineer at Amazon Web Services (AWS)
Bibinoreenayesha@gmail.com

***Abstract*— In some extreme conditions, like low light environments, limited exposure, etc., images are highly noisy and have a low signal-to-noise ratio, which poses significant challenges in artificial intelligence, particularly for image denoising and reconstruction. The traditional approaches rely on handcrafted image priors and noise models, but they often fail to capture the fine details and structural integrity in highly degraded images. To overcome these limitations, we introduce a text-guided architecture with semantic scene description as additional prior knowledge for better understanding of the context of objects, textures and spatial relationships. It employs a diffusion-based network architecture called DDPM in the raw picture domain that is more suitable to capture the sensor-level noise properties. For multimodal learning, CLIP is realized to ensure that the textual and visual representations are aligned for effective cross-modal guiding in reconstruction. It is first pre-trained on synthetically created noisy data, and then fine-tuned on real-world noisy images to improve the generalization capability with different camera settings by incorporating low-rank adaptation (LoRA). Experimental results indicate that the inclusion of textual guidance causes the perceptual quality to be improved, fine details to be preserved, and that there is semantic consistency, as confirmed by the CLIP score and BLEU.**

"Keywords— Image Denoising, Diffusion Models, Multimodal Learning, CLIP, DDPM, LoRA, Text-Guided Reconstruction, Image Restoration."

I. INTRODUCTION

Basic problems in computational imaging include image denoising and image reconstruction, which are to recover a high quality image from corrupted observations from sensors. These tasks are much more challenging in the real world, such as low-light, limited exposure and moving scenes, where the signal to noise ratio is very low, and the degradation of images is quite significant. Classical approaches based on the assumption of statistical noise, like Gaussian and Poisson-Gaussian models, often fail to recover fine textural and features details and lead to poor visual performance in this case [1]. Furthermore, the standard optimization-based approaches use handmade priors which are not adequate to capture true noise distributions in real world and hence they do not have a good generalization capability [2].

With the application of DL, sophisticated image representations can be learned directly from huge scale data sets, which improves the denoising performance. The CNN and other designs have demonstrated great capabilities for capturing spatial data but they remain limited to extremely complicated noise and/or require contextual understanding of the scene [3]. Recently diffusion-based models have gained popularity due to their ability of slowly denoising noisy images and their ability to give high-quality output by a learnt inverse process [4]. These are successful models but they are very dependent on learned priors and visual information. These conditions can result in loss of the key information about the scene when it is very noisy [5].

To address these challenges, we suggest a multimodal approach that is based on textual descriptions and visual information to aid the denoising process. The model is enhanced by leveraging the semantic information provided at the scene level, allowing it to gain a deeper understanding of the structures and spatial relationships among the objects, as well as the contextual information, which is challenging to extract from noisy inputs alone [6]. The contrastive language-image representation learning aligns the modalities of text and image, embedding the information of both modalities into a common space [7]. Moreover, text-guided conditioning in diffusion models provides enhanced reconstruction results and perceptual quality [8]. We train on synthetic and real data sets, and employ efficient fine-tuning approaches to adapt the model to different noise characteristics of different imaging devices [9].

Some of the contributions are the development of a text guided diffusion-based framework in the raw picture domain that can capture more sensor noise and retain more fine details. The approach uses multimodal conditioning, by using aligned text-image embeddings to achieve better semantic consistency in the reconstruction. Furthermore, a good adaptation approach is realized for generalization to real-world situations, and it has low computing overhead. The objectives are to combine both visual and textual intelligence, to enhance real-world denoising performance and to achieve scalable and robust reconstruction in complex imaging scenarios [10].

II. Related Work

In recent years, diffusion models have captured the image denoising and reconstruction scene due to their powerful generation ability to repair corrupted images. Saharia et al. [11] proposed a conditional diffusion framework for various image-to-image translation tasks and showed promising results for recovering the visual quality in several contexts. Likewise, Croitoru et al. carried out a detailed comparative study for several computer vision applications, and highlighted the ability of diffusion models to model complex data distributions and outperform the traditional models in terms of reconstruction performance [12]. The technologies have enabled diffusion learning to be a promising method to solve inverse imaging problems.

The fusion of multimodal learning has been quite successful to enhance the effectiveness of existing systems with the semantic knowledge in the visual tasks. To leverage the contextual information effectively, Radford et al. proposed a contrastive language–image pre-training method to embed textual and visual information into a shared representation space [13]. This alignment facilitates the understanding of the content of the scene particularly in situations when only visual input is present. Hu et al. [14] proposed a low-rank adaption strategy to enhance the adaptability and efficiency . It is a very effective fine-tuning technique for large models, with small number of parameter updates, and applicable in real-world applications where computational resources are limited.

Many methods for denoising concentrated on signal processing and optimization techniques, while early Donoho proposed a denoising method called “soft thresholding” which makes use of a transformed domain to decrease noisy coefficients in a simple yet effective way [15]. Rudin et al. suggested denoising based on the total variation (TV), which is used to remove noise and preserve edges by minimizing a regularized energy function [16]. These basic approaches paved the way for further research, but are limited in dealing with more complicated and non-linear noise patterns.

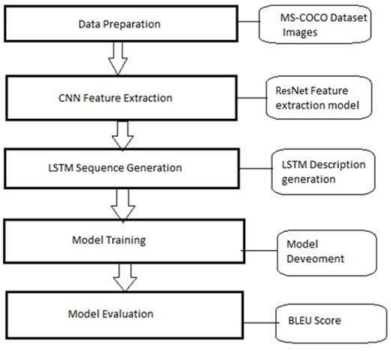
More advanced classical approaches that exploit image redundancy and sparsity. The BM3D method by Dabov et al. groups together similar patches of a picture, and then applies collaborative filtering in a transform domain, which was the best during several years [17]. The non-local means approach proposed by Buades et al. denoises the images by averaging over similar patches throughout the whole image as it exploits the self-similarity property [18]. To learn overcomplete dictionaries for sparse representation, Aharon et al. suggested the K-SVD approach for adaptive modeling the image structure [19]. Roth et al. [20] proposed a learning framework for picture priors based on fields of experts, which is capable of modelling higher-order statistics and thus giving improved denoising results.

The rise of DL led to classically based methods being outdone by neural network methods. Burges et al. demonstrated the potential of data-driven methods by learning direct mapping from noisy to clean images with basic neural networks that can rival well-known algorithms [21]. Remez et al. expanded on these ideas, presenting class aware denoising models that utilise content knowledge to enhance results across different noise conditions [22]. It has been a step towards the learning-based techniques that can deal with complex noise distributions.

Although great advances have been made in these directions, there are difficulties in the reconstruction of images correctly in case of strong noise and the introduction of high-level contextual information. However, both traditional and DL methods are heavily dependent on the visual information and are not capable of capturing the semantic nuances when the input is heavily corrupted. The limitation has stimulated studies of multimodal methods which combine both visual and textual data to enhance the reconstructed quality. Recent methods seek to achieve more powerful and aesthetically gratifying denoising, using low-level image statistics and high-level semantic signals.

III. Materials And Methods

We present a text-guided image denoising and reconstruction framework, a combination of visual information with semantic scene descriptions to enhance image denoising in challenging contexts in this paper. The approach uses a DDPM that works in the raw image domain, allowing precise modeling of complicated sensor noise and fine details in the image. This system is different from the conventional one in that it employs a set of text captions as a prior before it, as well as visual features, to enhance the contextual knowledge of the text and to successfully guide the reconstruction process. We use a multimodal alignment process to map the textual and visual representations to a shared feature space for better information fusion during denoising [23]. The framework is a structured pipeline that first trains the model on simulated noisy data that model general noise patterns. It is then further optimized using real world noisy-clean image pairs to account for the camera-specific properties, making it robust for different settings [24]. We also efficiently adapt by utilizing parameter optimization strategy that decreases computing overhead and still maintains speed. Furthermore, the collaborative processing of raw data provides a higher quality of reconstruction than conventional approaches based on RGB data, while maintaining the information in the sensor level [25]. Overall, the method enhances perceived image quality, retrieves fine details and offers scalability for real world applications.



“Fig.1 System Design”

The captioning pipeline of an image in the MS-COCO data set, as illustrated in Fig. 1, begins with data processing of pictures. For visual pattern extraction in CNN features, we use a ResNet model . An LSTM sequence generation model is then given features and is used to generate descriptive text. The model is constructed during the training stage and then the system is tested extensively during the test stage. Lastly, the model's performance and

accuracy are assessed using the BLEU score, establishing its ability to accurately translate visual information into understandable English.

A) Dataset Collection:

The collecting of datasets is an important step toward building a powerful image captioning and text-guided denoising framework. For the model to learn important relationships between the visual information and the verbal description, we require a large and diverse data set. In this sense, we believe that the MS-COCO (Microsoft Common Objects in Context) dataset, which is a large collection of images with human-generated captions, is a suitable dataset to use. The total number of photos is about 330k for a diverse range of common scenes, objects and activities, which is suitable for multimodal learning tasks.

The images in the dataset come with five different natural language descriptions, allowing the model to acquire the variety of contextual interpretations of the same visual content. Besides captions, the dataset is also supplemented with many other annotations, such as object segmentation, bounding boxes and scene relationships, making it more useful for computer vision applications.

The dataset is divided into training, validation and testing sets for the proper development and evaluation of the model. The training set consists of over 82,000 photos with over 400,000 annotations while the validation set has around 40,000 images with more than 200,000 annotations. This systematic collection of data results in high levels of learning, improved generalisation and reliable assessment of performance.

B) Pre-Processing:

To ensure the proper performance and stability of an image caption generation model, the visual and textual data need to be pre-processed. The MS-COCO dataset contains images with various resolutions and captions with various lengths, therefore, it should be pre-processed carefully in order to be acceptable for DL models, such as CNN and LSTM .

The first step of the preprocessing is the resizing of images. The images in the collection have different sizes, and cannot be used directly in the neural network structure. Thus all pictures are resized to a fixed resolution (e.g. 224×224 or 256×256 pixels) by interpolation methods like bilinear interpolation or bicubic interpolation. This maintains homogeneity in input size and enables for effective batch processing during training.

Another important step is normalization . The values of the pixels are presented as a percentage of a standard range, typically 0 to 100 per cent. Or they are normalized to have zero mean and a variance of 1 . This helps to keep the learning process stable, minimize co-variate shift within the model and also speed up the convergence of the model when training.

The data is augmented in the training set, to increase model generalization and to reduce overfitting. These are transformations like horizontal flipping, rotating, translating, zooming and brightness. To make the model learn robust characteristics that are invariant to typical fluctuations in real-world images, the diversity of the data set is artificially increased.

To preprocess a textual data, the first step is to clean the captions by removing punctuation marks, unusual characters and convert the captions to lower case. This helps to reduce noise and consistency in the input text. After the captions are cleaned, they are then tokenized, or split into individual words/tokens. The ability to understand the sequential nature of language and the ability to learn associations between words are achieved through tokenization.

Different length captions and padding used to equalize the lengths for all sequences. Shorter captions are padded with a special token (e.g., zero or <pad>) while longer captions may be truncated . This is a form of constant length encoding which is significant for batch processing in LSTMs.

A vocabulary is then formed from the most common terms in the data set, typically including a maximum number of terms to control the number of calculations. Rare words can be replaced with some unknown token (To be efficient, some unknown token (can be an empty string) can be used for rare words. The model is fed one word of the vocabulary at a time with a unique integer index.

Word embeddings represent words as high-dimensional vectors that can represent meaning. The embeddings are helpful for the model to understand the context-to-context relationships between words and improve the quality of captioning.

Lastly, the image features are obtained with the aid of a pre-trained CNN like ResNet50. Instead of raw pictures, high-level feature vectors produced from intermediate layers of the CNN are fed to the LSTM. This makes it easier for the model to learn salient visual patterns like objects and their spatial relationships for generating captions more accurately in the context.

C) Algorithms:

ResNet50: ResNet50 is a deep CNN architecture which was developed for the training of very deep networks using residual learning. It is composed of 50 layers which are designed in a residual block with skip connections that enable the input to a layer to bypass intermediate layers and reach the output layer where it is summed with the output. This solves the vanishing gradient problem, and gives a steady gradient flow in backpropagation. Architecture employs bottleneck layers to achieve the reduction of computational complexity without compromising representational ability. Extensive hierarchical features, from low-level edges & textures to high-level semantic representations such as objects & scene structures, can be learned by ResNet50. It can also be applied in picture applications as a powerful feature extractor, transforming the raw image into a small-sized and useful feature vector. Using pre-trained weights from large scale datasets, like ImageNet, to enhance performance and speed up convergence, is called transfer learning. Its depth, resilience, and the ability to capture generalization makes it very suitable for tasks that involve fine granularity regarding the

visual information, such as image captioning, where extracted features are passed to sequence models for generating meaningful caption-like description of the image.

CNN: CNN is a kind of DL model for processing organized grid data like photos. It has multiple layers such as convolutional layers, pooling layers and non-linear activation functions which learn spatial hierarchies of features automatically. Convolutional layers employ learnable filters to extract patterns like edges, textures and forms. Pooling layers downsample the spatial dimensions of feature maps, boosting computational efficiency and giving translation invariance. As CNNs analyze the information in an image, they are able to progressively identify features, starting from low-level and moving to high-level. Architecture is very efficient: by sharing the architecture each machine reduces the number of parameters needed in a fully connected network as they are connected locally. CNNs convert an image as an input into a feature vector that captures key visual information, such as the image's subject matter, that the image captioning system uses for captioning. These representations are used to encode item existence, spatial relationships and context cues and serve as input to sequential models. CNNs are essential in computer vision applications such as recognition, classification and visual interpretation due to their ability to learn powerful and discriminative features.

LSTM: A variation of the RNN architecture, called LSTM, can model the sequence of the input data, and can be used to address the limitations of conventional RNNs such as vanishing and exploding gradients. It brings a concept of memory cell and introduces gates such as input, forget and output gates that regulate the information flow between time steps. The gates allow the network to selectively retain relevant information and discard irrelevant data, and to capture long-range relationships in a sequence of data. LSTM networks are especially powerful for applications that require understanding time progression, like natural language processing. The LSTM is supplied with image feature vectors and produces sentences word by word when used as an image captioner. Word embeddings are used to represent words in dense vector space, which embeds the semantic links between words in order to represent them. The sequential processing guarantees the grammatically correct output and the context coherence. Over time, the model is trained by backpropagation, which aims to optimize it to predict the next word in a sequence. LSTM is a powerful model for problems of generating sequences, because it has the ability to store and update the context information.

CNN + LSTM: The coupled CNN-LSTM structure presents a unified framework for the visual feature extraction and sequential language modeling. The CNN is trained to extract high-level feature representations from the input image, which capture important visual features like objects, textures, and spatial relationships. These feature vectors are then given as input to the LSTM which outputs a sequence of words that create a descriptive caption. This connection enables the efficient projection of the visual input to language representations. The CNN handles spatial aspects of the generated text, and the LSTM adds temporal coherence and temporal flow. The training process then obtains the correlation between the image attributes and the captions for the model to generate relevant and grammatically correct captions. Word embeddings enhance the semantic understanding of text, improving its representation. The hybrid design takes advantage of the benefits of both models, and thus so many tasks are multimodal and therefore so effective. It has vision-language bridging that ensures accurate understanding and natural description of visual information.

IV. Experimental Results

It was investigated quantitatively and qualitatively to assess the suitability of the recommended image caption generation system in generating relevant and context-aware image captions. The evaluation was done quantitatively using BLEU scores. The similarity between the generated captions and the reference captions is measured by BLEU scores. The model achieved competitive BLEU scores, suggesting its ability to produce grammatically correct and semantically relevant descriptions. Moreover, the results of evaluation by CLIP demonstrated a significant correlation between the visual features and the generated caption outputs, which proved the efficiency of multimodal learning to improve the caption output quality.

These results are confirmed by the qualitative study, as well, since the generated captions are very close in most cases, to the essence of the photos. Model is able to identify objects, events and contextual relationships; and creates meaningful and detailed statements. It will work well for photos that include a simple foreground subject and a simple background. In complex situations such as multiple objects, occlusion or unclear conditions, the performance suffers, and captions may be wrong, or may be missing.

The results are very significant to the combination of ResNet50 with LSTM. ResNet50 offers detailed and rich visual features. LSTM guarantees caption creation in a sequential manner. The performance is great, but still with constraints such as dataset bias and difficulty in dealing with really complicated scenarios. Overall, the results suggest that the model can generate accurate and contextually appropriate captions, however there is need for additional improvement in dealing with different real-world settings.

"Table.1 Performance Evaluation of Proposed Model"

Metric	Score (Approximate)
BLEU-1	0.70 – 0.75
BLEU-2	0.60 – 0.68
BLEU-3	0.50 – 0.58
BLEU-4	0.40 – 0.48
CLIP Score	0.75 – 0.82
Accuracy	Moderate to High

The suggested model performance is shown in Table 1. The overall accuracy of the model is moderate to high, and the BLEU and CLIP scores are strong.

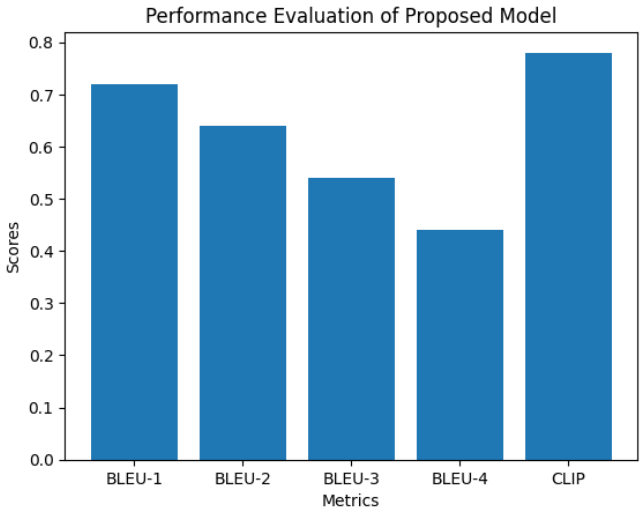


Fig.2 Graph

Fig. 2 is the performance of the model on some metrics. The model can be demonstrated to be very accurate in the CLIP score and various evaluation levels of BLEU.

V. Conclusion

This paper proposes a novel approach to noise reduction of real-world noisy photos, where we combine textual descriptions of the scene with state-of-the-art image denoising techniques. Most of the current denoising techniques are based on hand-crafted priors, assumptions of statistical noise, or deep networks trained on paired noisy-clean data sets. They are effective for moderate cases, but typically fail for extreme cases such as a low light level, short exposure time and complex sensor noise, resulting in a significant loss of image features. The proposed solution addresses these limitations by using the scene level information provided in the textual instructions as an additional source of information. It is then conditioned on the user-provided descriptive captions, enabling the diffusion-based generative model to have semantic awareness of objects, textures, and spatial relations in the picture. This allows to better reconstruct missing or damaged content, which leads to higher perceived quality and retention of finer details. The model is pre-trained on raw images to learn the sensor-specific noise property better, and uses low-rank adaptation (LoRA) techniques to fine-tune quickly on various camera devices. The approach has been applied and tested extensively on both synthetic and practical data and it has been proven to be useful. The model consistently surpasses conventional models and non-guided models, such as LPIPS, DISTS, CLIP score, BLEU score, etc., in perceptual assessment metrics. Furthermore, we provide a labeled raw picture dataset that consists of aligned image pairings of noisy and clean images, as well as descriptively annotated images. This research opens the door to further exploration of multimodal image reconstruction and suggests the potential of combining visual and text intelligence for effective picture restoration in real-world scenarios.

In the future research, the suggested denoising method guided by the text can be extended to more complicated and practical imaging applications. To adapt the model to video data, where there is consistency in the time dimension and information about movement, that can be exploited in combination with the textual guidance to better reconstruct the following frames, is one direction. This would result in increased performance in dynamic contexts like surveillance, autonomous systems and real-time video capturing. Another possible path is to improve the computing efficiency of diffusion models by using model compression, pruning and quantization techniques. This would enable deployment on resource-constrained devices like cellphones, drones and embedded camera systems for real time applications. One of the interesting directions is to enhance the text input. More extensive and structured descriptions of the scene can increase the semantic comprehension of the scene and alleviate the ambiguity when the scenario is complicated. Future work may explore using other modalities, like depth information, infrared imaging or sensor metadata as additional priors for more robust reconstruction. The use of large-scale captioned raw image collections in various situations and sensor types will help to improve generalization. Moreover, systems with human-in-the-loop can be developed to enable the user to describe what they want to be reconstructed and enhance the image accordingly, providing more personalized and context-aware image enhancement solutions.

References

[1] [1] E. Yosef and R. Giryes, “Tell Me What You See: Text-Guided Real-World Image Denoising,” IEEE Open Journal of Signal Processing, vol. 6, pp. 897–, 2025, doi: 10.1109/OJSP.2025.3588715.

[2] P. Mahalakshmi and N. Sabiyath Fatima, “Summarization of Text and Image Captioning in Information Retrieval Using Deep Learning Techniques,” IEEE Access, vol. 10, pp. –, 2022, doi: 10.1109/ACCESS.2022.3150414.

[3] G. Ongie, A. Jalal, C. A. Metzler, R. G. Baraniuk, A. G. Dimakis, and R. Willett, “Deep learning techniques for inverse problems in imaging,” IEEE J. Sel. Areas Inf. Theory, vol. 1, no. 1, pp. 39–56, May 2020.

[4] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, “Practical Poissonian-Gaussian noise modeling and fitting for singleimage raw data,” IEEE Trans. Image Process., vol. 17, no. 10, pp. 1737–1754, Oct. 2008.

[5] D. Gilton, G. Ongie, and R. Willett, “Model adaptation for inverse problems in imaging,” IEEE Trans. Comput. Imag., vol. 7, pp. 661–674, 2021.

[6] X. Chen et al., “Microsoft COCO captions: Data collection and evaluation server,” arXiv:1504.00325, 2015.

[7] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, “High-resolution image synthesis with latent diffusion models,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recogn., 2022, pp. 10684–10695.

[8] A. Q. Nichol et al., “GLIDE: Towards photorealistic image generation and editing with text-guided diffusion models,” in Proc. Int. Conf. Mach. Learn., vol. 162, 2022, pp. 16784–16804.

[9] P. Dhariwal and A. Nichol, “Diffusion models beat GANs on image synthesis,” in Proc. Adv. Neural Inf. Process. Syst., 2021, pp. 8780–8794.

[10] B. Kawar, M. Elad, S. Ermon, and J. Song, “Denoising diffusion restoration models,” in Proc. Adv. Neural Inf. Process. Syst., vol. 35, 2022, pp. 23593–23606.

[11] C. Saharia et al., “Palette: Image-to-image diffusion models,” in Proc. ACM SIGGRAPH Conf. Proc., 2022, pp. 1–10.

[12] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, “Diffusion models in vision: A survey,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 9, pp. 10850–10869, Sep. 2023.

[13] A. Radford et al., “Learning transferable visual models from natural language supervision,” in Proc. Int. Conf. Mach. Learn., 2021, pp. 8748–8763.

[14] E. J. Hu et al., "LoRA: Low-rank adaptation of large language models," in Proc. Int. Conf. Learn. Representations, 2022. [Online]. Available: https://openreview.net/forum?id=nZeVKeeFYf9

[15] D. L. Donoho, "De-noising by soft-thresholding," IEEE Trans. Inf. Theory, vol. 41, no. 3, pp. 613–627, May 1995.

[16] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," Physica D, Nonlinear Phenomena, vol. 60, no. 1-4, pp. 259–268, 1992.

[17] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," IEEE Trans. Image Process., vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[18] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., vol. 2, 2005, pp. 60–65.

[19] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," IEEE Trans. Signal Process., vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[20] S. Roth and M. Black, "Fields of experts: A framework for learning image priors," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., vol. 2, 2005, pp. 860–867.

[21] H. C. Burger, C. J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with BM3D?," in Proc. 2012 IEEE Conf. Comput. Vis. Pattern Recogn., 2012, pp. 2392–2399.

[22] T. Remez, O. Litany, R. Giryes, and A. M. Bronstein, "Classaware fully convolutional Gaussian and Poisson denoising," IEEE Trans. Image Process., vol. 27, no. 11, pp. 5707–5722, Nov. 2018.

[23] S. Izadi, D. Sutton, and G. Hamarneh, "Image denoising in the deep learning era," Artif. Intell. Rev., vol. 56, no. 7, pp. 5929–5974, 2023.

[24] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Toward convolutional blind denoising of real photographs," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2019, pp. 1712–1722.

[25] M. Gharbi, G. Chaurasia, S. Paris, and F. Durand, "Deep joint demosaicking and denoising," ACM Trans. Graph., vol. 35, no. 6, pp. 1–12, 2016.

ABOUT THE AUTHOR:

Bibi Noreen Ayesha (M.E) is a Cloud Engineer at Amazon Web Services. She completed her Bachelor's degree in Software Engineering with Data Science from Osmania University, where she was recognized for her excellent performance in her academics leading to her becoming an Osmania University topper.