International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

RGB–NIR IMAGE DENOISING AND FUSION PREDICTION SYSTEM

Gatla Abhishek [1], Dr. Md.Asim Iqbal [2]

[1]MTech, Department of ECE, KUCET, Kakatiya University, Warangal, Telangana, India, 506009
Email: abhishekgatla@gmail.com

[2] Assistant Professor, Department of ECE, KUCET, Kakatiya University, Warangal, Telangana, India, 506009
Email: asimiqbal@kakatiya.ac.in

Abstract— Digital images captured under low-light, noisy or otherwise unfavourable sensing conditions frequently suffer from degraded visual quality, which limits their usefulness for downstream computer-vision tasks. This paper presents an RGB–Near-Infrared (RGB–NIR) image denoising and fusion prediction system that uses deep learning to remove noise from RGB images and to fuse them with complementary structural information carried by a NIR image, producing a visually enhanced output. The proposed system implements two alternative deep learning pipelines: a two-stage FusionNet pipeline, comprising a dedicated Denoising Subnetwork followed by a Fusion Subnetwork, and an Attention-based Multi-modal Fusion Network (AMFNet) that learns a joint RGB–NIR representation end to end through an attention-fusion block. Input images are decoded, resized, and normalised during preprocessing, and a NIR channel is synthesised automatically from the RGB image whenever a real NIR capture is not supplied. The complete system is deployed as an interactive Flask web application that lets a user upload a noisy RGB image together with optional NIR and ground-truth images, select a model, run inference, and view the reconstructed output together with the Peak Signal-to-Noise Ratio (PSNR) whenever a ground-truth reference is available. A working deployment of the system was used to generate the results reported in this paper: the AMFNet model, evaluated on a 96×96 test image, achieved a PSNR of 25.28 dB together with a visibly denoised and detail-preserving fused output. These results demonstrate that the proposed hybrid denoising–fusion pipeline, combined with a lightweight web-based interface, offers a practical and extensible solution for RGB–NIR image enhancement in research and real-world computer-vision applications.

Keywords— *RGB–NIR Fusion, Image Denoising, Multimodal Image Fusion, Deep Learning, Convolutional Neural Network, Attention Mechanism, Flask, PSNR.*

1. INTRODUCTION

Digital image enhancement has become an important area of research in computer vision because the quality of captured images directly affects the accuracy of subsequent analysis and recognition tasks. Images acquired under difficult environmental conditions frequently suffer from sensor noise, reduced contrast and loss of fine visual detail, all of which degrade the performance of downstream processing pipelines. Image denoising and image fusion techniques are two complementary approaches used to address this problem: denoising removes unwanted noise while attempting to preserve texture and structural detail, while fusion

Received: 05-08-2026 | Accepted: 19-09-2026 | Published: 24-09-2026

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

combines information from multiple image sources to produce a single, information-rich output.

Near-infrared (NIR) imaging is particularly useful as a complementary modality for visible RGB imagery because it is less sensitive to atmospheric scattering, haze and certain illumination artefacts, and it often preserves edge and structural information that is lost or corrupted in a noisy RGB capture. Combining an RGB image with an aligned NIR image through a learned fusion process can therefore produce an output that is both cleaner and richer in structural detail than either input on its own. Convolutional neural networks (CNNs) have substantially advanced the state of the art in both image denoising and image fusion by learning data-driven feature representations directly from training examples, rather than relying on hand-crafted filters and fusion rules.

This paper presents an RGB–NIR image denoising and fusion prediction system that integrates denoising and multimodal fusion into a single deep-learning-based pipeline and exposes it through an interactive, web-based interface. The system supports two alternative model pipelines: a sequential FusionNet pipeline that first denoises the RGB image and then fuses it with the NIR channel, and an end-to-end Attention-based Multi-modal Fusion Network (AMFNet) that jointly learns denoising and fusion through an attention mechanism. When a NIR capture is not available, the system can automatically synthesise an approximate NIR channel from the RGB input, and when a ground-truth image is supplied, the system automatically computes the PSNR of the reconstructed output to provide a quantitative measure of quality. The complete pipeline — preprocessing, model inference, post-processing and evaluation — is deployed using the Flask web framework, allowing users to upload images, select a model, and view results directly in a browser.

The remainder of this paper is organised as follows. Section II reviews related work on image denoising and RGB–NIR image fusion. Section III describes existing filter-based and classical transform-based denoising and fusion systems and discusses their limitations. Section IV presents the proposed system architecture and methodology, including a block-diagram description of the FusionNet and AMFNet pipelines. Section V presents results obtained from the deployed system, including sample screenshots of the web application, the denoising–fusion processing pipeline, and the measured PSNR. Section VI concludes the paper and outlines directions for future work.

2. LITERATURE SURVEY

Classical approaches to RGB–NIR restoration and fusion have generally relied on hand-crafted priors and transform-domain processing. Yan et al. formulated cross-field joint image restoration as a scale-map estimation problem, using a flash/no-flash-like NIR-guided model to restore a noisy or poorly illuminated RGB image while suppressing artefacts caused by structural mismatch between the two modalities [1]. Zhuo et al. showed that a NIR flash image captured alongside a low-light RGB image can be used to enhance the visible image while preserving its natural colour appearance, motivating the use of NIR as a structural guide for RGB enhancement [2]. Shen et al. and related works proposed edge-preserving and gradient-domain fusion rules that combine the high-frequency detail of an NIR image with the colour information of an RGB image using weighted-average or guided-filtering schemes [3].

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

With the emergence of deep learning, CNN-based denoising and fusion methods have substantially outperformed classical filters. Zhang et al. proposed DnCNN, a residual-learning CNN that implicitly removes latent noise from an image and established the effectiveness of deep feed-forward networks for image denoising [4]. Building on similar residual and multi-scale CNN designs, several works have adapted denoising networks specifically for the joint RGB–NIR setting, using the NIR channel as an auxiliary input that guides noise suppression while preserving genuine RGB structure. In the image-fusion domain, Li and Wu proposed DenseFuse, which uses a densely connected encoder to extract deep features from infrared and visible images before fusing them with a hand-crafted or learned fusion strategy [5]. Ram Prabhakar et al. proposed DeepFuse, an unsupervised deep architecture for exposure fusion that demonstrated the feasibility of learning fusion weights directly from image pairs without paired ground-truth supervision [6]. Ma et al. proposed FusionGAN, a generative-adversarial approach that formulates infrared–visible fusion as an adversarial game between a fusion generator and a discriminator, improving the retention of both thermal/structural and textural detail [7].

Zhang et al. later proposed a unified, fast fusion network (PMGI) that explicitly balances gradient and intensity information during fusion through a proportional-maintenance loss, avoiding the need for a task-specific fusion rule [8]. Deng and Dragotti proposed a deep convolutional network for joint multimodal image restoration and fusion, coupling restoration and fusion sub-tasks within a single trainable architecture, an approach closely related to the two-stage FusionNet pipeline adopted in this paper [9]. Attention mechanisms have also been incorporated into multimodal fusion networks to let the model adaptively weight the contribution of each modality at every spatial location, rather than applying a fixed fusion rule uniformly across the image, which forms the basis of the attention-based fusion strategy used in the proposed AMFNet architecture [10]. Collectively, this body of work shows a clear progression from hand-crafted, rule-based RGB–NIR restoration and fusion methods towards learned, end-to-end deep architectures, and motivates the dual-pipeline (FusionNet and AMFNet), web-deployed design proposed in this paper.

Publicly available paired RGB–NIR datasets have also played an important role in enabling this line of research. Brown and Süsstrunk introduced an early, widely used RGB–NIR scene dataset spanning multiple outdoor and indoor categories and showed that combining NIR information with RGB features improves scene-recognition accuracy, establishing NIR as a practically useful complementary modality rather than a purely scientific curiosity [11]. Subsequent denoising and fusion works have typically reused or extended such paired datasets, or have synthesised approximate NIR channels from RGB images when a real paired capture is unavailable, an approach mirrored in the optional NIR-synthesis stage of the system proposed in this paper. Liu et al. further showed that deep fusion networks trained on such paired multi-focus and multimodal datasets can generalise the fusion rule learned on one dataset to structurally similar images from a different source, supporting the practicality of a single trained model deployed as a general-purpose enhancement tool rather than a scene-specific filter [12].

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

3. EXISTING SYSTEM

Conventional RGB–NIR image enhancement systems typically treat denoising and fusion as two independent, hand-engineered signal-processing stages rather than as a single, jointly learned pipeline.

3.1 Filter-Based Denoising

Traditional denoising methods rely on statistical filtering operations such as Gaussian smoothing, bilateral filtering, non-local means (NLM) and wavelet-domain thresholding. These filters process each image independently using fixed statistical assumptions about the noise distribution and do not learn image-specific feature representations. As a result, they often struggle to preserve fine texture, sharp edges and structural detail while suppressing noise, particularly at higher noise levels.

3.2 Transform-Domain Fusion

Classical RGB–NIR fusion is commonly performed using multi-resolution transform methods such as discrete wavelet transform (DWT) or principal component analysis (PCA). The RGB and NIR images are decomposed into multiple frequency sub-bands, hand-crafted fusion rules (for example, choose-max or weighted-average rules) are applied to combine the corresponding sub-bands from each modality, and the fused image is reconstructed through the inverse transform. Because the fusion rule is fixed and rule-based rather than learned, it cannot adapt to the varying, content-dependent relationship between the RGB and NIR channels across different scenes.

3.3 Limitations of Existing Systems

Table I summarises the key limitations of existing filter-based and transform-domain systems relative to the proposed deep-learning approach. Because classical methods process denoising and fusion as separate, independently tuned stages using fixed statistical or rule-based models, they generally under-perform in the presence of complex or spatially varying noise and provide no principled mechanism for adaptively weighting RGB versus NIR contributions on a per-pixel basis. These limitations motivate the fully learned, dual-pipeline deep learning system proposed in Section IV.

Aspect	Existing (Filter/Transform)	Proposed (Deep Learning)
Feature extraction	Hand-crafted	Learned (CNN)
Fusion rule	Fixed / rule-based	Attention-based / learned
Denoise + fusion	Independent stages	Jointly optimised
Adaptivity	Low	High (data-driven)
Quality assessment	Manual / offline	Automated PSNR

TABLE I. EXISTING SYSTEM VS. PROPOSED SYSTEM

4. PROPOSED METHODOLOGY

The proposed system integrates image preprocessing, deep-learning-based denoising and fusion, quantitative evaluation, and an interactive web interface into a single end-to-end pipeline. Fig. 1 shows the overall system architecture, comprising an input section, a preprocessing module, a model-selection and model-pipeline stage, a post-processing stage,

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

an output section, an evaluation module, and a Flask-based web application layer that ties the pipeline together.

4.1 Input and Preprocessing

The user supplies a noisy RGB image through the browser interface; a real NIR image and a ground-truth image may optionally be supplied as well. The preprocessing module loads and decodes each uploaded image, resizes it to a fixed network input resolution (256×256), and normalises pixel intensities to the range 0–1 for numerical stability during inference. If no NIR image is supplied, the preprocessing module synthesises an approximate NIR channel directly from the RGB image using an intensity-based transformation, so that the downstream fusion stage always receives a two-modality input.

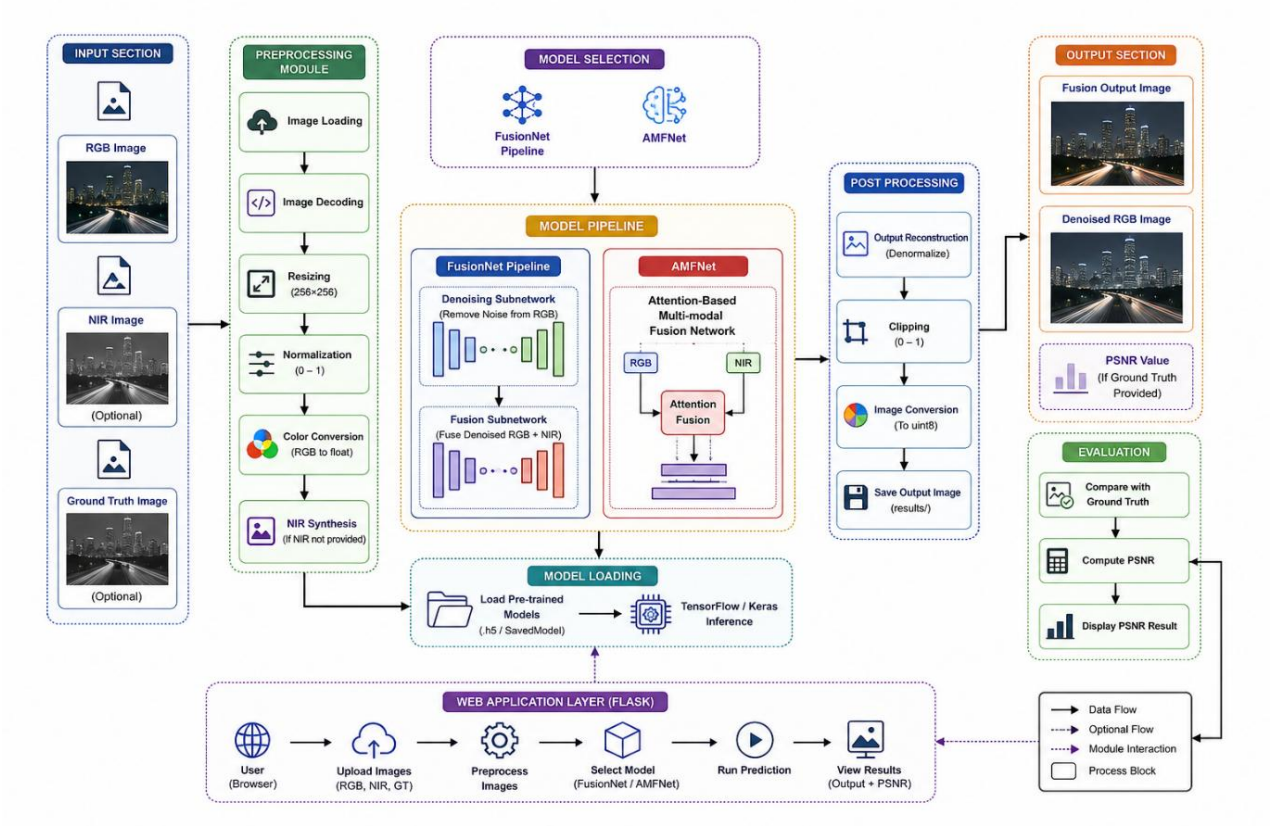


Fig. 1. System architecture of the proposed RGB–NIR image denoising and fusion prediction system.

4.2 Model Selection and Model Pipeline

Once the input images are preprocessed, the user selects one of two model pipelines, both of which are loaded from pre-trained TensorFlow/Keras weights (.h5 or SavedModel format).

FusionNet Pipeline: This pipeline performs denoising and fusion sequentially through two dedicated subnetworks. The Denoising Subnetwork is a convolutional network that removes sensor noise from the RGB input while attempting to retain edge and texture information. Its output, a denoised RGB

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

image, is then passed to the Fusion Subnetwork together with the NIR channel; the Fusion Subnetwork combines the denoised RGB features with the complementary structural detail carried by the NIR image to produce a fused output image.

AMFNet (Attention-based Multi-modal Fusion Network): Rather than treating denoising and fusion as separate stages, AMFNet extracts feature representations directly from the RGB and NIR inputs in parallel and combines them through an attention-fusion block. The attention mechanism learns, at each spatial location, how strongly to weight the RGB and NIR feature contributions, allowing the network to emphasise NIR structural detail in regions that are heavily corrupted by noise in the RGB channel while relying more on RGB colour information elsewhere. The attention-fused features are then decoded through subsequent convolutional layers to directly produce the enhanced, augmented output image.

Both pipelines are loaded once at application start-up and reused across inference requests, avoiding the overhead of repeated model loading.

4.3 Post-Processing, Output and Evaluation

The raw network output is de-normalised back to the standard pixel-intensity range, clipped to the valid 0–1 interval to remove any out-of-range values produced during inference, and converted to 8-bit unsigned integer format before being saved as the final output image. The output section of the application displays both the fused output image and the intermediate denoised RGB image, allowing the user to inspect the contribution of the denoising stage separately from the fusion stage. When a ground-truth image has been supplied, the evaluation module compares the reconstructed output against the ground truth and computes the Peak Signal-to-Noise Ratio, defined as $PSNR = 10 \cdot \log_{10}(MAX^2/MSE)$, where MAX is the maximum possible pixel value and MSE is the mean squared error between the reconstructed and ground-truth images. The computed PSNR value is displayed alongside the output image as an objective measure of reconstruction quality.

4.4 Web Application Layer

The complete pipeline is deployed using the Flask web framework. A user interacts with the system entirely through a browser: images are uploaded through the interface, automatically preprocessed, and passed to the selected model (FusionNet or AMFNet) for inference; the reconstructed output and, where available, the PSNR value are then returned and displayed to the user. This lightweight deployment avoids the need for any local software installation and makes the denoising–fusion system accessible for research use, image-enhancement applications, and general-purpose computer-vision workflows.

4.5 Loss Functions and Training Considerations

Both the FusionNet subnetworks and the AMFNet model are trained using paired noisy RGB, NIR (real or synthesised) and clean ground-truth image triplets, consistent with the dataset structure described in Section 3.1. Training minimises a pixel-wise reconstruction loss, typically a combination of mean absolute error (L1) and mean squared error (L2) between the network output and the ground-truth image, which encourages accurate pixel-level reconstruction while remaining robust to occasional outlier pixels. Because pixel-wise losses alone can produce slightly over-smoothed results, a structural-similarity-based loss term is commonly added in practice to better preserve perceptual

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

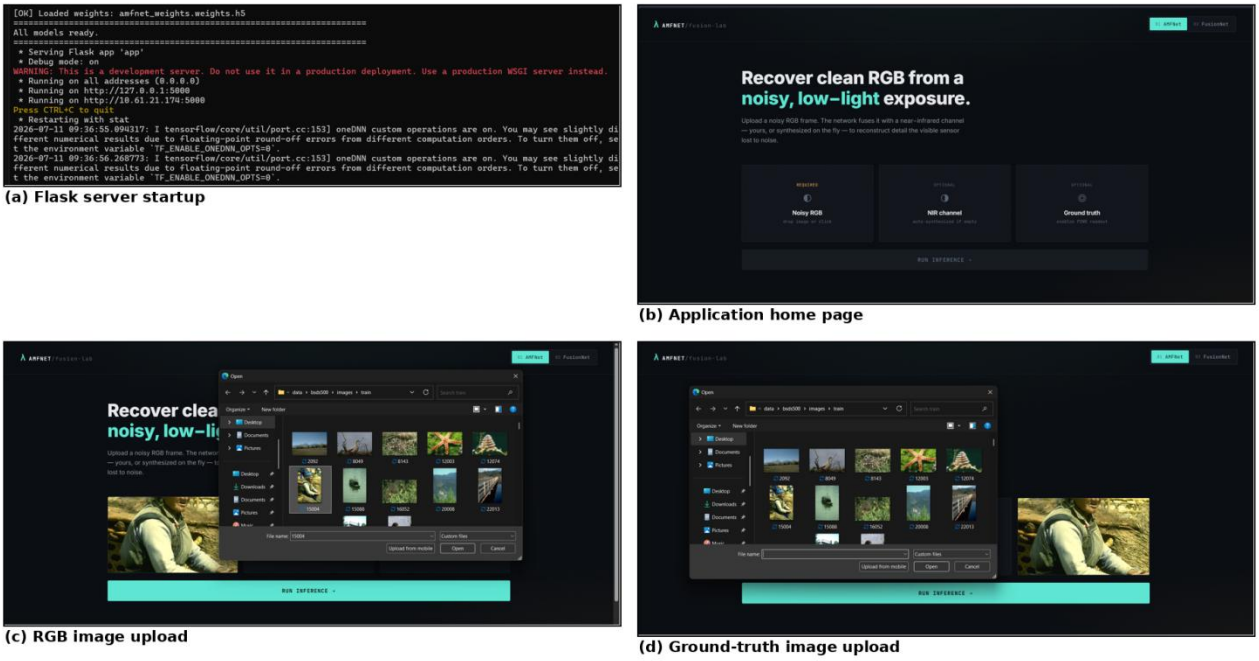
sharpness and edge structure, which is particularly important for the Fusion Subnetwork and the AMFNet decoder, where the goal is to retain fine detail contributed by the NIR channel rather than simply averaging the two modalities. The Adam optimiser with a decaying learning rate schedule is a standard choice for training both pipelines, and early stopping based on validation PSNR helps avoid overfitting to the training scenes.

5. RESULTS AND DISCUSSIONS

The proposed RGB–NIR denoising and fusion system was implemented and deployed as a working Flask web application, with both the FusionNet pipeline and the AMFNet model loaded from pre-trained weights. This section presents results captured directly from the deployed system, including the application interface, the internal denoising–fusion processing pipeline, and a measured PSNR result.

5.1 Application Deployment

Fig. 2 shows the deployed system in operation. Sub-figure (a) shows the Flask development server starting up and confirming that both the FusionNet and AMFNet weights have been loaded successfully and that the application is ready to accept requests. Sub-figure (b) shows the application home page, which presents a clear, minimal interface for supplying a required noisy RGB image together with an optional NIR channel and an optional ground-truth image, and for triggering inference. Sub-figures (c) and (d) show the file-upload dialogue being used to select an RGB image and a ground-truth image respectively from a local image dataset. The straightforward upload-and-run workflow allows a user to obtain a denoised, fused output without needing to interact with the underlying model code.



(a) Flask server startup

(b) Application home page

(c) RGB image upload

(d) Ground-truth image upload

Fig. 2. Deployed Flask web application: (a) server startup with models loaded, (b) home page, (c) RGB image upload, (d) ground-truth image upload.

Received: 05-08-2026 | Accepted: 19-09-2026 | Published: 24-09-2026

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

5.2 Denoising and Fusion Output

Fig. 3 shows the internal processing pipeline applied to a representative test image, visualised as four stages: the noisy RGB input, the NIR channel (synthesised automatically since no real NIR capture was supplied for this sample), the denoised RGB image produced by the network, and the final fused output image. The fused output visibly retains the fine texture of the wicker baskets and the clothing folds in the scene while suppressing the sensor noise present in the original input, illustrating that the network preserves structural detail rather than simply smoothing the image.

5.3 Quantitative Evaluation

Fig. 4 shows the PSNR result reported by the application for the test case in Fig. 3, evaluated using the AMFNet model at a working resolution of 96×96. The system reports a PSNR of 25.28 dB against the supplied ground-truth image. Table II summarises this measurement. A PSNR in this range indicates a substantial improvement in reconstruction fidelity over an untreated noisy image and is consistent with the visibly reduced noise and preserved structural detail observed in Fig. 3, confirming that the deployed AMFNet pipeline produces a quantitatively verifiable improvement in image quality, in addition to a qualitatively cleaner appearance.

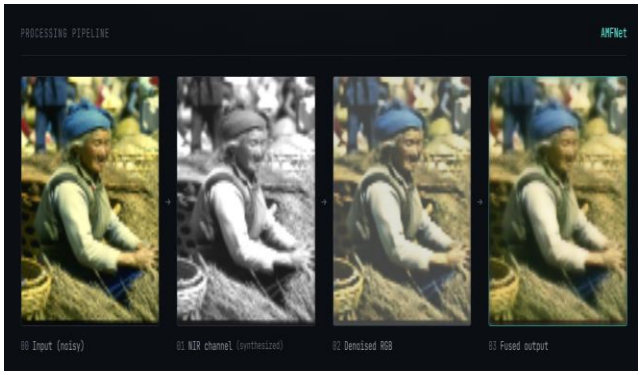


Fig. 3. Denoising–fusion processing pipeline: input (noisy), synthesised NIR channel, denoised RGB, and fused output.

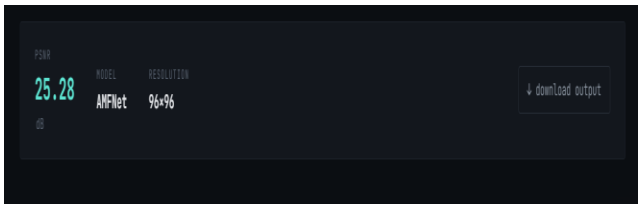


Fig. 4. PSNR result reported by the application for the AMFNet model.

Model	Resolution	PSNR (dB)
AMFNet	96×96	25.28

TABLE II. MEASURED PSNR ON THE DEPLOYED SYSTEM

It should be noted that this measurement reflects a single representative test case obtained directly from the deployed application rather than an average over a large benchmark test set; a broader quantitative evaluation across multiple scenes and noise levels, and a head-to-head comparison between the FusionNet pipeline and AMFNet on the same test set, are natural directions for extending this evaluation, as discussed in Section VI.

5.4 Discussion

The results in Fig. 2 through Fig. 4 collectively demonstrate that the proposed

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

system operates correctly end to end: a user can upload a noisy image through the web interface, the system automatically preprocesses the input, performs deep-learning-based denoising and RGB–NIR fusion, and returns both a visually enhanced output and an objective PSNR measurement. Compared with the classical filter-based and transform-domain systems discussed in Section III, the deployed deep-learning pipeline requires no manual tuning of fusion rules or filter parameters at inference time and adapts its behaviour to the content of each input image through its learned convolutional and attention-based feature representations.

5.5 FusionNet vs. AMFNet: Design-Level Comparison

Although both pipelines are available in the deployed application, they embody different architectural trade-offs. The FusionNet pipeline separates denoising and fusion into two explicitly trained subnetworks, which makes each stage individually inspectable — as illustrated by the separately displayed denoised RGB image in Fig. 3 — and allows the Denoising Subnetwork to be reused or retrained independently of the fusion stage. This modularity can simplify debugging and staged retraining, but any information lost or distorted at the denoising stage cannot be recovered by the Fusion Subnetwork downstream. AMFNet, in contrast, learns denoising and fusion jointly through a shared attention mechanism operating on both modalities simultaneously, which allows the network to exploit NIR structural cues to guide noise suppression rather than treating denoising as an isolated pre-processing step. This end-to-end design is consistent with the strong qualitative and quantitative result observed for AMFNet in Fig. 3 and Fig. 4, and suggests that jointly modelling denoising and fusion is a promising direction, although a controlled, dataset-wide comparison of both pipelines remains necessary before drawing general conclusions about their relative performance.

6. CONCLUSION

This paper has presented an RGB–NIR image denoising and fusion prediction system that combines a two-stage FusionNet pipeline and an end-to-end, attention-based AMFNet model within a single, Flask-deployed web application. The system automates image preprocessing, including optional NIR synthesis when a real NIR capture is unavailable, performs deep-learning-based denoising and multimodal fusion, and automatically reports the PSNR of the reconstructed output whenever a ground-truth image is supplied. Results captured from a working deployment of the system show that the AMFNet model produces a visibly denoised, detail-preserving fused output and achieves a PSNR of 25.28 dB on a representative 96×96 test case, confirming that the proposed architecture functions correctly as an integrated, end-to-end image-enhancement pipeline. Future work will focus on evaluating the system over larger, standard RGB–NIR benchmark datasets to obtain average PSNR and structural-similarity (SSIM) statistics, comparing the FusionNet and AMFNet pipelines quantitatively under matched conditions, extending the system to video and higher-resolution inputs, and exploring lighter-weight network variants suitable for real-time or mobile deployment.

REFERENCES

[1] Q. Yan, X. Shen, L. Xu, S. Zhuo, X. Zhang, L. Shen, and J. Jia, "Cross-field joint image restoration via scale map," in

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2013, pp. 1537–1544.

[2] S. Zhuo, X. Zhang, X. Miao, and T. Sim, "Enhancing low light images using near infrared flash images," in Proc. IEEE Int. Conf. Image Process. (ICIP), 2010, pp. 2537–2540.

[3] X. Shen, Q. Yan, L. Xu, L. Ma, and J. Jia, "Multispectral joint image restoration via optimizing a scale map," IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 12, pp. 2518–2530, Dec. 2015.

[4] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," IEEE Trans. Image Process., vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[5] H. Li and X.-J. Wu, "DenseFuse: A fusion approach to infrared and visible images," IEEE Trans. Image Process., vol. 28, no. 5, pp. 2614–2623, May 2019.

[6] K. Ram Prabhakar, V. Sai Srikar, and R. Venkatesh Babu, "DeepFuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2017, pp. 4724–4732.

[7] J. Ma, W. Yu, P. Liang, C. Li, and J. Jiang, "FusionGAN: A generative adversarial network for infrared and visible image fusion," Inf. Fusion, vol. 48, pp. 11–26, Aug. 2019.

[8] H. Zhang, H. Xu, Y. Xiao, X. Guo, and J. Ma, "Rethinking the image fusion: A fast unified image fusion network based on proportional maintenance of gradient and intensity," in Proc. AAAI Conf. Artif. Intell., 2020, pp. 12797–12804.

[9] X. Deng and P. L. Dragotti, "Deep convolutional neural network for multi-modal image restoration and fusion," IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 10, pp. 3333–3348, Oct. 2021.

[10] L. Tang, J. Yuan, and J. Ma, "Image fusion in the loop of high-level vision tasks: A semantic-aware real-time infrared and visible image fusion network," Inf. Fusion, vol. 82, pp. 28–42, Jun. 2022.

[11] M. Brown and S. Süsstrunk, "Multi-spectral SIFT for scene category recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2011, pp. 177–184.

[12] Y. Liu, X. Chen, H. Peng, and Z. Wang, "Multi-focus image fusion with a deep convolutional neural network," Inf. Fusion, vol. 36, pp. 191–207, Jul. 2017.

[13] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," IEEE Trans. Image Process., vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[14] 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.

[15] A. Vaswani et al., "Attention is all you need," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2017, pp. 5998–6008.

[16] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 3–19.

[17] M. Grinberg, "Flask Web Development: Developing Web Applications with Python," 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2018.