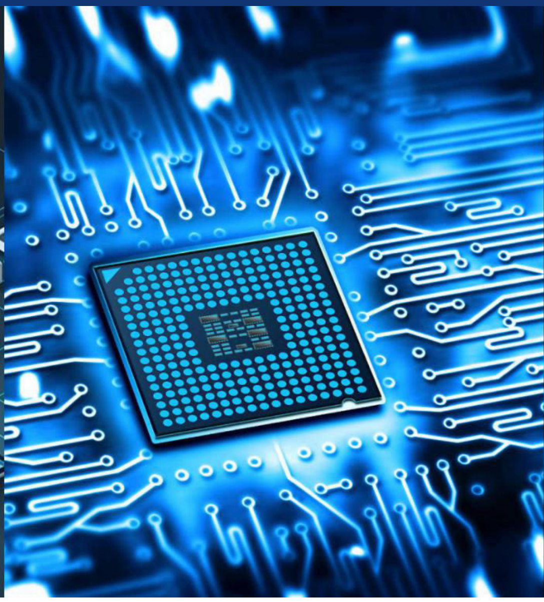
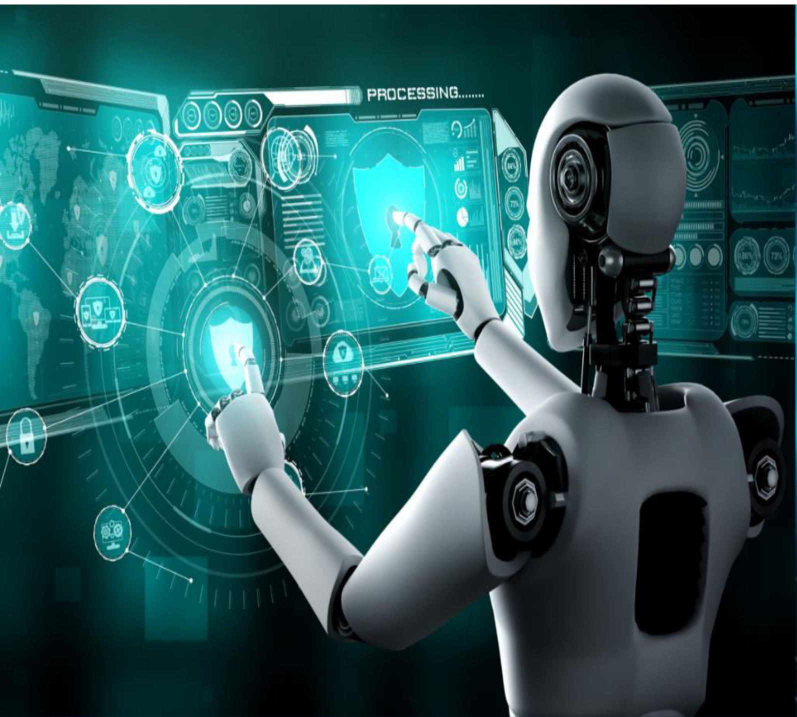
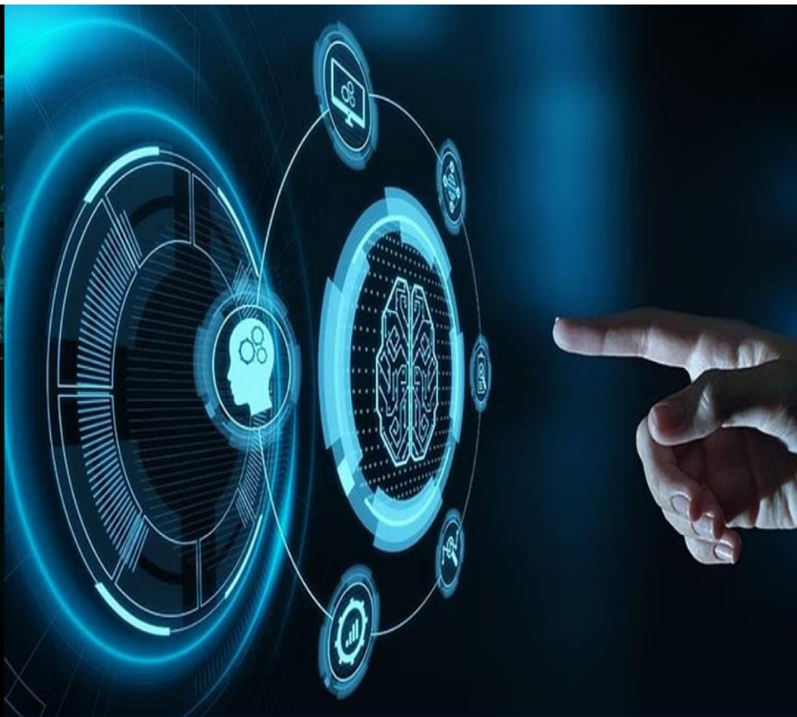
ISSN(O): 2320-9801
ISSN(P): 2320-9798


International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 8.771

Volume 14, Issue 3, March 2026

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Modal Feature Disentanglement and Contribution Estimation for Multimodality Image Fusion (MFDCE-Fuse)

D. Sridhar[1], T.Sirisha[2], K.Hema[3], G.Mokshagna[4], M.Prudhvi Narayana[5]

U.G. Student, Department of ECE, SVIET Engineering College, Nandamuru, Pedana, Andhra Pradesh, India[1,2,3,4]

Associate Professor, Department of ECE, SVIET Engineering College, Nandamuru, Pedana, Andhra Pradesh, India[5]

ABSTRACT: Multimodality image fusion (MMIF) tasks aim to fuse complementary information from different modalities, such as salient objects and texture details, to improve image quality and information comprehensiveness. Most current MMIF methods adopt a "black-box" decoder to generate fused images, which leads to insufficient interpretability and difficulty in training. To deal with these problems, MMIF is converted into a modality contribution estimation task through a novel self-supervised fusion network named MFDCE-Fuse.

KEYWORDS: Multimodality image fusion,,Black Box, Disentanglement, Visible-Infrared Fusion

I. INTRODUCTION

The core challenge in multimodality image fusion is effectively integrating distinct features—like the thermal radiation of infrared images and the high-resolution spatial details of visible images—without losing critical data. Traditional methods often struggle with transparency. MFDCE-Fuse addresses this by focusing on modal feature disentanglement and explicit contribution estimation

II. RELATED WORK

The MFDCE-Fuse framework is built upon several innovative components designed to enhance both performance and interpretability:

Contrast-Learning Autoencoder: The model integrates the strengths of CNNs and Swin Transformers to capture long-range global features alongside local texture details.

Contrastive Reconstruction Loss: This loss is designed to promote the uniqueness and non-redundancy of the captured features.

Feature Disentangled Representation: To prevent modal redundant features from interfering with contribution estimation, a framework based on contrastive constraints is used to obtain modal-common and modal-private features.

Modality Contribution Estimation: The contribution of modal images to the fusion is evaluated through the proportion of modal-private features, enhancing process interpretability.

III. METHODOLOGY

To ensure the high quality of the fused image and the integrity of the features, the network utilizes specific loss functions:

Weighted Perceptual Loss: Constructed to guarantee the visual quality of the output.
Feature Disentanglement Contrastive Loss: Specifically designed to ensure that private features remain intact during the disentanglement process.

IV. EXPERIMENTAL VALIDATION

The MFDCE-Fuse model has been validated through qualitative and quantitative experiments across multiple domains:

Visible-Infrared Fusion (VIF): Successfully merges thermal and visual data.

© 2026 IJIRCCE | Volume 14, Issue 3, March 2026 DOI: 10.15680/IJIRCCE.2026.1403129
www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|
International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

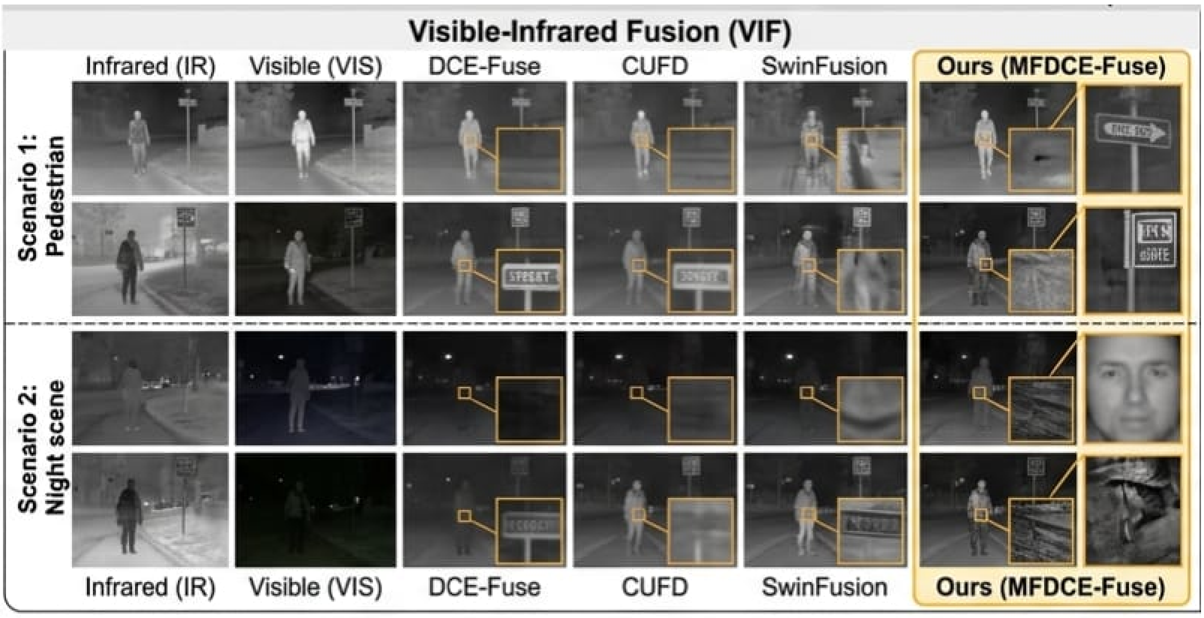


Fig: Visible-Infrared Fusion

Medical Image Fusion (MIF): Demonstrates effective generalization in clinical imaging tasks.

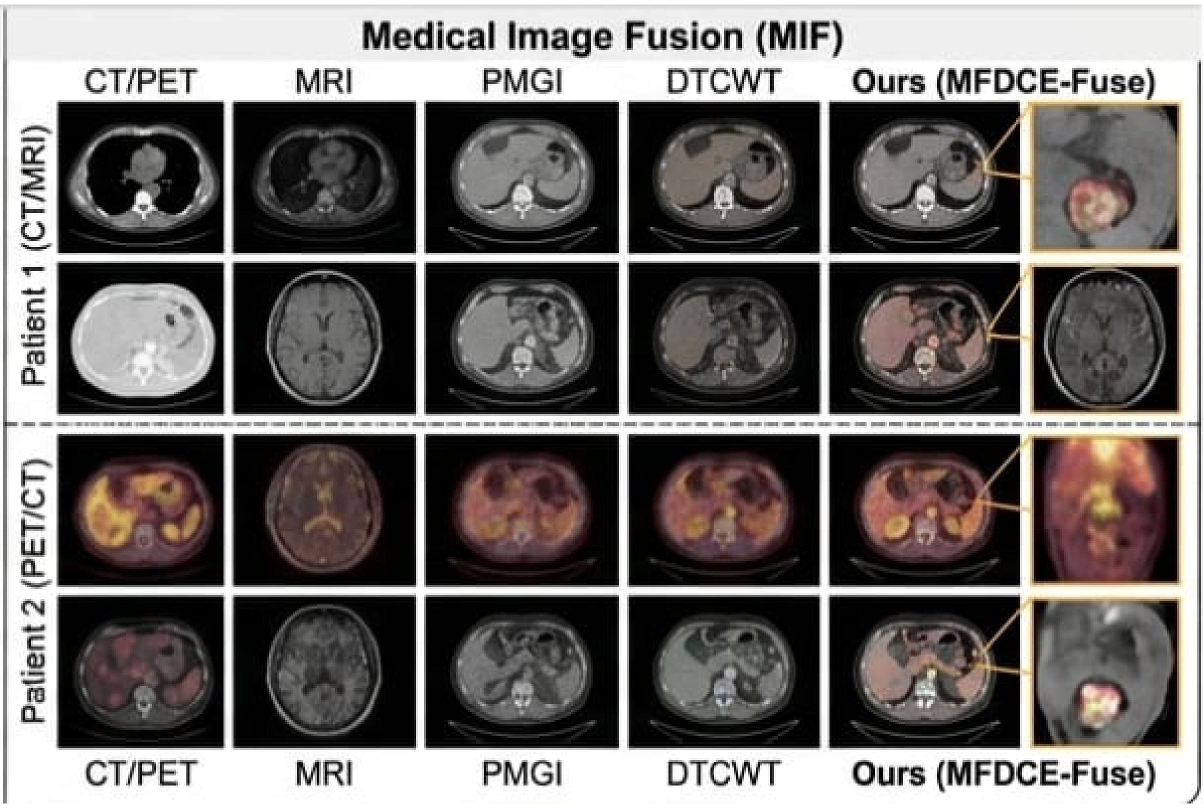


Fig: Medical Image Fusion

© 2026 IJIRCCE | Volume 14, Issue 3, March 2026 DOI: 10.15680/IJIRCCE.2026.1403129

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. CONCLUSION

By shifting the focus from a "black-box" architecture to a contribution estimation task, MFDCE-Fuse provides a highly interpretable and effective solution for multimodality image fusion. The integration of CNN-Transformer architectures and contrastive learning ensures that both global context and local details are preserved, resulting in superior image quality.

REFERENCES

[1] Zhao, Z., Bai, H., Zhang, J., et al. (2023). CDDFuse: Correlation-Driven Dual-Branch Feature Decomposition for Multi-Modality Image Fusion. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 5906-5916.

[2] Xu, H., Ma, J., Jiang, J., et al. (2020). U2Fusion: A Unified Unsupervised Image Fusion Network. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(1), 502- 518.

[3] Liu, J., Lin, R., Wu, G., et al. (2023). Coconet: Coupled contrastive learning network with multilevel feature ensemble for multimodality image fusion. International Journal of Computer Vision, 132(5), 1748-1775.

[4] Li, H., Wu, X. J., & Kittler, J. (2021). RFN-nest: An end-to-end residual fusion network for infrared and visible images. Information Fusion, 73, 72-86.

[5] Zhao, Z., Xu, S., Zhang, C., et al. (2020). DIDFuse: Deep image decomposition for infrared and visible image fusion. arXiv preprint arXiv:2003.09210.

[6] Ma, J., Yu, W., Liang, P., et al. (2019). FusionGAN: A generative adversarial network for infrared and visible image fusion. Information Fusion, 48, 11-26.

[7] Li, J., Huo, H., Li, C., et al. (2020). AttentionFGAN: Infrared and visible image fusion using attention-based generative adversarial networks. IEEE Transactions on Multimedia, 23, 1383-1396.

[8] Zhang, X., & Demiris, Y. (2023). Visible and infrared image fusion using deep learning IEEE Transactions on Pattern Analysis and Machine Intelligence, 1-20.

[9] Zhao, Z., Bai, H., Zhu, Y., et al. (2023). Equivariant Multi-Modality Image Fusion. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[10] Liu, Y., Chen, X., Ward, R. K., et al. (2016). Image fusion with convolutional sparse representation. IEEE Signal Processing Letters, 23(12), 1882-1886.