Dr. Yuting Wan
State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University
129 Luoyu Road, Wuhan, Hubei, 430079
P. R. China
E-mail: wanyuting@whu.edu.cn

Dear Prof. Zhao-Liang LI,

Revision of our manuscript TGRS-2024-05984.R1

Thank you for your comments and suggestions. We have carefully revised our paper (TGRS-2024-05984.R1) in light of the associate editor's and reviewers' comments. A point-by-point response to the comments is attached to this letter. The major changes have been highlighted by colored text and are summarized as follows:

1. The Section I has been updated to incorporate some latest image fusion methods with detailed explanation. Additionally, we have included a comprehensive review of existing big-model-guided fusion approaches to enhance the reader's understanding. (see the response to reviewer 2).
2. Some relevant references have been supplemented and formatted properly. (see the response to reviewer 2).

For detailed changes, please refer to the response letters attached.
Thank you again for your comments and time.
Best regards,

Dr. Yuting Wan
Post-doctoral of remote sensing

Response to the Comments of the Associate Editor

The revised manuscript has significantly improved compared to the previous version, and all the concerns raised by the reviewers have been effectively addressed. However, there are still a few minor suggestions and comments that could be considered to further enhance the completeness of the work before it is ready for publication.

Thank you for your comments and suggestions. We have carefully revised our paper (TGRS-2024-05984.R1) in light of the associate editor's and reviewers' comments. A point-by-point response to the comments is attached to this letter. The major changes have been highlighted by colored text and are summarized as follows:

1. The Section I has been updated to incorporate some latest image fusion methods with detailed explanation. Additionally, we have included a comprehensive review of existing big-model-guided fusion approaches to enhance the reader's understanding. (see the response to reviewer 2).
2. Some relevant references have been supplemented and formatted properly. (see the response to reviewer 2).

Finally, thank you again for your very helpful suggestions.

Response to the Comments of the Reviewer #1

The authors have addressed all my questions. I believe this version is ready for publishing.

R. Thank you for your careful and helpful review.

We sincerely appreciate your thorough review and constructive feedback on our manuscript. Your detailed questions and insightful comments have been instrumental in refining our work, and we are profoundly grateful for the time and expertise you have invested in this process.

Finally, thank you again for your very helpful suggestions.

Response to the Comments of the Reviewer #2

1. ***The references cited in this paper are insufficient, lacking some of the latest image fusion literature. For example:***

   [1]Color-aware fusion of nighttime infrared and visible images, Engineering Applications of Artificial Intelligence, 2025, 139, 109521

   [2]Laplacian pyramid fusion network with hierarchical guidance for infrared and visible image fusion, IEEE Transactions on Circuits and Systems for Video Technology 33 (9), 4630-4644

   [3]Navigating Uncertainty: Semantic-Powered Image Enhancement and Fusion, IEEE Signal Processing Letters,

   R. Thank you for your suggestion.

   We have added three recommended papers in Section I with detailed analysis. (Page 1, right column and Page 2, left column)

"Recently, many innovative approaches have emerged in the field of image fusion. For example, the color-aware fusion technique [23] effectively improves the recognizability of targets in low-light environments by preserving the color information and the thermal features. In addition, a hierarchical guided fusion network based on Laplace pyramid [24] enhances the complementarity of cross-modal information while preserving image details through multi-scale feature decomposition and hierarchical fusion strategies."

"In addition, recent studies have also proposed a semantic-driven fusion framework [28], which further improves the fusion quality in complex scenes by introducing semantic prior knowledge to guide the feature fusion and provides richer semantic information for target detection."

The related reference is listed as follows. (Page 17, left column)

[23] J. Yao, Y. Zhao, Y. Bu, S. G. Kong, and X. Zhang, "Color-aware fusion of nighttime infrared and visible images," *Engineering Applications of Artificial Intelligence*, vol. 139, pp. 109521, 2025.

[24] J. Yao, Y. Zhao, Y. Bu, S. G. Kong, and J. C. W. Chan, "Laplacian pyramid fusion network with hierarchical guidance for infrared and visible image fusion," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 9, pp. 4630-4644, 2023.

[28] J. Yao, Y. Zhao, S. G. Kong, and X. Zhang, "Navigating uncertainty: Semantic-powered image enhancement and fusion," IEEE Signal Processing Letters, 2024.

2. ***Review the existing big model guided fusion methods.***

   R. Thank you for your comment and helpful suggestion.

   We have additionally added a review of existing big model guided fusion methods. The specific details are as follows. (Page 5, right column)

"Besides, the transformer-based large-scale models for multi-modal infrared and visible image fusion have also attracted significant attention. These approaches [25]-[29] leverage self-attention mechanisms to effectively integrate complementary features across different modalities. Notably, SwinFusion [25] implements a hierarchical attention framework based on swin transformer [26], establishing dual-stream attention modules for cross-domain feature alignment. Furthermore, DATfusion [27] is proposed to preserve important features and global information through the interaction of dual attention with transformer blocks, thereby achieving the multimodal image fusion. In addition, recent studies have also proposed a semantic-driven fusion framework [28], which further improves the fusion quality in complex scenes by introducing semantic prior knowledge to guide the feature fusion and provides richer semantic information for target detection. PromptFusion [29] further explores a prompt-based method for guiding image fusion from Vision-Language Models, which aims to enhance target identification and fusion quality by leveraging semantic prompts."

The related reference is listed as follows. (Page 17, left column)

[25] J. Ma, L. Tang, F. Fan, J. Huang, X. Mei, and Y. Ma, "SwinFusion: Cross-domain long-range learning for general image fusion via swin transformer," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 7, pp. 1200-1217, 2022.

[26] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10012-10022, 2021.

[27] W. Tang, F. He, Y. Liu, Y. Duan, and T. Si, "DATFuse: Infrared and visible image fusion via dual attention transformer," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 7, pp. 3159-3172, 2023.

[28] J. Yao, Y. Zhao, S. G. Kong, and X. Zhang, "Navigating uncertainty: Semantic-powered image enhancement and fusion," *IEEE Signal Processing Letters*, 2024.

[29] J. Liu, X. Li, Z. Wang, Z. Jiang, W. Zhong, W. Fan, and B. Xu, "PromptFusion: Harmonized semantic prompt learning for infrared and visible image fusion," *IEEE/CAA Journal of Automatica Sinica*, 2024.

Finally, thank you again for your very helpful suggestions.

Response to the Comments of the Reviewer #3

I have no further comments.

R. Thank you for your careful and helpful review.

We are deeply grateful for your thorough review and constructive feedback on our manuscript. It is encouraging to know that our responses addressed all your concerns, and we are profoundly thankful for the time and expertise you have invested in this process.

Finally, thank you again for your very helpful suggestions.