Image Fusion Survey: A Novel Taxonomy Integrating Transformer and Recent Approaches

December 01, 2024

Research Paper Context and framework

This paper explores state-of-the-art image fusion methods, including:

- Multi-Focus, Multi-Exposure, and Multi-Modal techniques.
- Recent and Innovative Architectures.
- Intuitive Comparison Approach and Classification.

Image Fusion Survey: A Novel Taxonomy Integrating Transformer and Recent Approaches

Bernardi Gwendal[1,2], Strubel David[1], Brisebarre Godefroy[1], Garin Jean-François[1], Ardabilian Mohsen[2], Dellandréa Emmanuel[2]

[1] Tiama
[2] Ecole Centrale de Lyon, CNRS, INSA Lyon, Universite Claude Bernard Lyon 1, Université Lumière Lyon 2, LIRIS, UMR5205, 69130 Ecully, France

Abstract. Research progress in multi-modal information fusion, particularly in Image Fusion, has experienced significant advancements over the last decade. By integrating information from multiple sources or modalities, image fusion enables the extraction of comprehensive insights and facilitates more accurate analysis and decision-making processes. The inherent complexity of image fusion, stemming from its unstructured nature, necessitates high levels of abstraction and intricate data representation. The utilization of deep learning, notably CNN and more recently introduced Vision Transformer, has yielded substantial enhancements in image fusion methodologies. This paper presents a comprehensive survey of image fusion methodologies, focusing on recent advancements and introducing a novel taxonomy based on supervised, unsupervised, and task-driven approaches. The survey encompasses recent contributions, including the integration of transformer architectures, which have emerged as powerful tools for image fusion tasks. This classification is supported by a distinction of methods by architecture type (CNN, GAN, Transformer) for a better understanding of the relationships between methods. Through the synthesis of existing literature and the introduction of a new classification paradigm, this survey aims to provide researchers and practitioners with a comprehensive overview of image fusion techniques and guide future research directions in this rapidly evolving field.

Keywords: Image Fusion · Multi-modal · Task-driven · Fusion Transformer

Research Problem

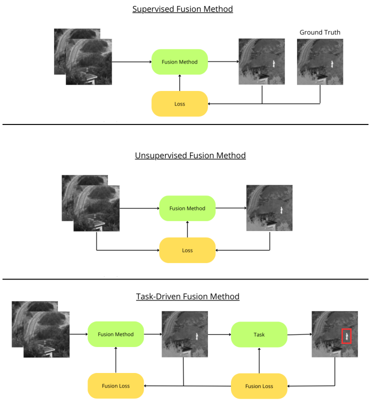
This research aims to propose a taxonomy that better classifies image fusion methods, improving their understanding. The objectives of the paper are as follows :

- Propose a taxonomy that better defines image fusion methods and complements existing taxonomies.
- Highlighting the latest architectures using this approach.
- Unify algorithms across different fields (Multi-Exposure, Multi-Focus, and Multi-Modal) through this classification.

New Taxonomy: Classification Approach

This approach classifies methods based on their learning paradigm rather than their input data. Image fusion methods can be categorized into three learning paradigms:

- Supervised Learning.
- Unsupervised Learning.
- Task-Driven Learning.



Supervised Learning Approach

An example of supervised learning with the IFCNN method:

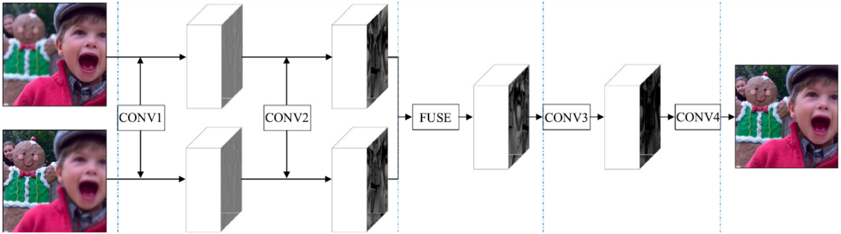


Fig. 1. The proposed general image fusion framework based on convolutional neural network. The above part illustrates the architecture of our image fusion model, and the below part shows a demonstration example for fusing multi-focus images. Please note that the spatial sizes marked in the figure just indicate the ones used in our training phase, and the inputs can be extended to more than two images.

Zhang, Y., Liu, Y., Sun, P., Yan, H., Zhao, X., Zhang, L.: IFCNN: A general image fusion framework based on convolutional neural network. Information Fusion 54, 99–118 (2020)

Unsupervised Learning Approach

An example of unsupervised learning with the FusionGAN method:

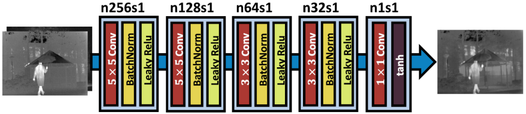


Fig. 3. Network architecture of generator G_{θ_G}. G_{θ_G} is a simple five-layer convolution neural network with 5 convolution layers, 4 batch normalization layers, 4 leaky ReLU activation layers, and 1 tanh activation layer.

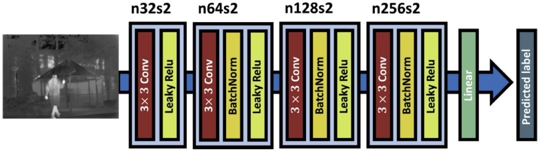


Fig. 4. Network architecture of discriminator D_{θ_D}. D_{θ_D} is a simple five-layer convolution neural network with 4 convolution layers to extract feature maps of input, 1 linear layer to do the classification, 4 batch normalization layers, and 4 leaky ReLU activation layers.

Ma, J., Yu, W., Liang, P., Li, C., Jiang, J.: Fusiongan: A generative adversarial network for infrared and visible image fusion. Information fusion 48, 11–26 (2019)

Task-Driven Learning Approach

An example of task-driven learning with the IFT method:

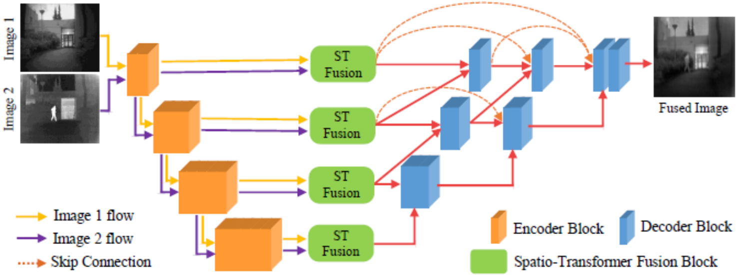


Fig. 2: Overview of the proposed Image Fusion Transformer (IFT) network. Image 1 and Image 2 are passed through the encoder to obtain multi-scale deep features. These extracted deep features are fused using the Spatio-Transformer (ST) fusion block. Finally, the decoder reconstructs the multi-scale fused features to output a fused image.

Vs, V., Valanarasu, J.M.J., Oza, P., Patel, V.M.: Image fusion transformer. 2022 IEEE International Conference on Image Processing (ICIP) pp. 3566–3570 (2022)

Efficient and Representative Image Fusion Methods :

Methods	Categories	Learning paradigm	Dataset / Images	Advantages / Disadvantages
GMFNet [1]	CNN	Task-Driven	MFNet Dataset	+ learning resulting in a robust method - complex method with 3 models for 1 useful output
U2Fusion [7]	CNN	Unsupervised	TNO, RS, Harvard, EMPA HDR, public Dataset	+ generalist methods (multiple fusion problems) - only captures local relationships in images (no long-range relationships)
DDcGAN [4]	CNN, GAN	Unsupervised	TNO	+ unsupervised method, includes multiscale support - possible artifact generation, unstable GAN training
MEF-GAN [8]	CNN, GAN, Attention	Supervised, Unsupervised	HDR-Eye, Fairchild, public Dataset	+ applies attention to GAN - partially based on supervised learning
TarDAL [2]	CNN, GAN	Task-Driven	TNO, INO, RS, M3FD, MS	+ dual path discriminator, task-driven - focus on Infra-red / visible only, uses image processing extraction
SCGRFuse [6]	CNN, Transformer	Task-Driven	MSRS, TNO, RS	+ includes Transformer, task-driven learning - not very generalizable to other contexts, hyperparameters only based on other related literature
IFT [5]	CNN, Transformer	Unsupervised	KAIST, TNO, Harvard and PET Dataset	+ spatial and Transformer path (extract local and long-distance information) - complex architecture
STFNet [3]	CNN, Transformer	Unsupervised	KAIST, LLVIP, M3FD, MSRS, VLIRVDIF	+ feature align network, cross-attention model - need stronger detail constrain, complex architecture

Conclusion of the Research Paper

The following are the conclusions reached by the comparisons of image fusion methods:

- Transformer-based methods offer significant advantages because of their ability to capture long-distance relationships.
- Task-driven methods excel in ease of convergence and loss calculation but require task-specific fusion for effective labelling.
- Hybrid methods can integrate task-driven loss with unsupervised loss for improved fusion performance.
- Taxonomy lets us easily define a method. Example: GMFNet -> Task-Driven Multi-Modal Method.

Future Research Directions

Promising advancements in image fusion research include:

- Enhancing methods using transformer-based architectures.
- Exploring innovative architectures, such as the proposed Mamba framework.
- Optimizing methods to address industrial constraints.
- Improving the explainability of fusion techniques.

Bibliography

- Balit, E., Chadli, A.: Gmfnet: Gated multimodal fusion network for visible-thermal semantic segmentation. Proc. 16th Eur. Conf. Comput. Vis pp. 1–4 (2020)
- Liu, J., Fan, X., Huang, Z., Wu, G., Liu, R., Zhong, W., Luo, Z.: Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection. IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 5802–5811 (2022)
- Liu, Q., Pi, J., Gao, P., Yuan, D.: Stfnet: Self-supervised transformer for infrared and visible image fusion. IEEE Transactions on Emerging Topics in Computational Intelligence (2024)
- Ma, J., Xu, H., Jiang, J., Mei, X., Zhang, X.P.: Ddcgan: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion. IEEE Transactions on Image Processing **29**, 4980–4995 (2020)
- Vs, V., Valanarasu, J.M.J., Oza, P., Patel, V.M.: Image fusion transformer. 2022 IEEE International Conference on Image Processing (ICIP) pp. 3566–3570 (2022)
- Wang, Y., Pu, J., Miao, D., Zhang, L., Zhang, L., Du, X.: Scgrfuse: An infrared and visible image fusion network based on spatial/channel attention mechanism and gradient aggregation residual dense blocks. Engineering Applications of Artificial Intelligence **132**, 107898 (2024)
- Xu, H., Ma, J., Jiang, J., Guo, X., Ling, H.: U2fusion: A unified unsupervised image fusion network. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(1), 502–518 (2020)
- Xu, H., Ma, J., Zhang, X.P.: Mef-gan: Multi-exposure image fusion via generative adversarial networks. IEEE Transactions on Image Processing **29**, 7203–7216 (2020)

Thank you!