1. Related Work

1.1. Multi-modality image fusion

In recent years, deep learning has significantly advanced fusion technologies, with Convolutional Neural Networks (CNNs) and Transformers emerging as core architectures. CNNs, as one of the early deep learning applications in image fusion, enable end-to-end learning. Recently, Transformer-based models have demonstrated exceptional capability in capturing long-range dependencies. Unlike CNNs, which focus on local receptive fields, Transformer models model global context, making them highly suitable for complex fusion tasks. CMTFusion [PVL23] utilizes cross-modal Transformers to reduce redundancy and fuse spatial and channel information, while another Transformer-based method achieves fusion in the frequency domain via multi-scale feature extraction. CDDFuse introduces a correlation-driven feature decomposition strategy, combining the Restormer module and Transformer–CNN dual-branch architecture for effective global and local information fusion.

In deep learning-based methods, generative model-based image fusion is also a crucial direction, especially Generative Adversarial Networks (GANs) [LCH22]. DDcGAN, proposed by Ma et al. [MXJ*20], handles multi-resolution infrared–visible image fusion, while Liu et al. introduced a dual-discriminator model for medical image fusion, integrating adversarial networks to guide fusion images while preserving more information related to brain tumor segmentation tasks.

To address the limitations of task-specific solutions, general-purpose fusion methods have emerged [JKJ*20, XMJ*20], aiming to handle multiple fusion tasks with a single model. Recent frameworks like DDFM leverage Bayesian theory, score matching, and pre-trained diffusion models to achieve superior fusion performance.

1.2. Invertible neural networks

Invertible Neural Networks (INNs) are commonly employed in tasks that require the preservation of original data and support for reversible transformations. To better accommodate image processing tasks within INN architectures, Dinh et al. [DSDB16] introduced a general affine coupling layer, known as RealNVP, which incorporates convolutional and multi-scale layers to reduce computational complexity and improve regularization. In the Steg-cINN model, a conditional invertible neural network (cINN) [RLZW22] is used to guide the generation of colored images from grayscale images in a secure manner. The INN framework has also been applied to image recovery tasks. In the field of image fusion, the CDDFuse model leverages INN blocks to achieve lossless information transmission, and introduces LT blocks to strike a balance between fusion quality and computational cost by enabling mutual generation of input and output features. Due to their inherent ability to perform reversible mappings between inputs and outputs, INNs have found widespread applications in various image-related tasks, including image hiding [JDX*21], image rescaling [XZL*20], and image coloring [ALK*19].

1.3. Kolmogorov-Arnold Networks

Kolmogorov–Arnold Networks (KANs) are neural networks inspired by the Kolmogorov–Arnold superposition theorem. By constructing multilayer structures that compose and superpose univariate functions, KANs effectively approximate complex nonlinear relationships with fewer parameters, making them well-suited for high-dimensional and complex tasks [WLW*24].

Due to their fine-grained sequence modeling capability, KANs were initially applied to two-dimensional time series forecasting. U-KAN embeds KAN modules into a U-Net architecture to enable efficient medical image segmentation. ConvKAN incorporates the nonlinear activation functions of KAN into convolutional layers, allowing KANs to adapt to convolutional outputs, effectively reducing parameter counts while maintaining high levels of accuracy. However, the application of KAN in image processing remains underexplored, and effectively leveraging KAN architectures for image fusion still presents a significant challenge. In our research, we propose an EfficientKAN integrated with the Transformer architecture to harness its nonlinear modeling strengths. This enhancement reduces model parameters and computational complexity while enabling adaptive global feature modeling, thereby addressing the Transformer limitation in local feature extraction.

References

[ALK*19] Ardizzone L., Lüth C., Kruse J., Rother C., Köthe U.: Guided image generation with conditional invertible neural networks (2019). *arXiv preprint arXiv:1907.02392* (2019). 1

[DSDB16] Dinh L., Sohl-Dickstein J., Bengio S.: Density estimation using real nvp. *arXiv preprint arXiv:1605.08803* (2016). 1

[JDX*21] Jing J., Deng X., Xu M., Wang J., Guan Z.: Hinet: Deep image hiding by invertible network. In *Proceedings of the IEEE/CVF international conference on computer vision* (2021), pp. 4733–4742. 1

[JKJ*20] Jung H., Kim Y., Jang H., Ha N., Sohn K.: Unsupervised deep image fusion with structure tensor representations. 3845–3858. 1

[LCH22] Lee C.-K., Cheon Y.-J., Hwang W.-Y.: Least squares generative adversarial networks-based anomaly detection. *IEEE Access 10* (2022), 26920–26930. 1

[MXJ*20] 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* (2020), 4980–4995. 1

[PVL23] Park S., Vien A. G., Lee C.: Cross-modal transformers for infrared and visible image fusion. *IEEE Transactions on Circuits and Systems for Video Technology 34*, 2 (2023), 770–785. 1

[RLZW22] Ren Y., Liu T., Zhai L., Wang L.: Hiding data in colors: Secure and lossless deep image steganography via conditional invertible neural networks. *arXiv preprint arXiv:2201.07444* (2022). 1

[WLW*24] Wu Y., Li T., Wang Z., Kang H., He A.: Transukan: Computing-efficient hybrid kan-transformer for enhanced medical image segmentation. *arXiv preprint arXiv:2409.14676* (2024). 1

[XMJ*20] Xu H., Ma J., Jiang J., Guo X., Ling H.: U2fusion: A unified unsupervised image fusion network. 502–518. 1

[XZL*20] Xiao M., Zheng S., Liu C., Wang Y., He D., Ke G., Bian J., Lin Z., Liu T.-Y.: Invertible image rescaling. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16* (2020), Springer, pp. 126–144. 1

Proceedings published by Eurographics - The European Association for Computer Graphics.