

# Improvement and Application of Fusion Scheme in Automatic Medical Image Analysis

-- Subtitle Is Not Required, Please Write It Here If Your Article Has One

Minjun Liang<sup>1</sup>, Mingyang Wei<sup>1</sup>, Yanan Li<sup>2</sup>, He Tian<sup>1</sup>, Yabei Li<sup>1</sup>,\*

<sup>1</sup> School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan, 454000, P. R. China

<sup>2</sup> Institute of Orthopaedics and Traumatology, The First Affiliated Hospital of Zhejiang Chinese Medical University, Hangzhou, China

\* Corresponding author: Yabei Li (liyabei@home.hpu.edu.cn)

**Abstract:** The research in this paper provides generalization and new ideas for research topics in computer-assisted medicine. The main improvement efforts in deep learning-based multimodal fusion schemes, which provide alternative directions and robust feature fitting performance for fusion schemes, are building complex structures, migrating knowledge or experience, processing and enhancing data, and targeting features for semantic correction based on contextual features. At the application level, the brain, liver, and lungs are the main targets of scientific research, so this paper surveys related work and analyzes the reasons for performance gains. Taken together, deep learning-based image fusion schemes can assist physicians in understanding information about lesion sites, lesion types, and sizes, providing an important basis for developing personalized treatment plans, which is important for improving diagnosis and specifying precise treatment plans. Therefore, the investigation of medical image fusion schemes is promising and beneficial.

**Keywords:** Image Fusion, Medical Image Analysis, Deep Learning, Multimodal image fusion.

## 1. Introduction

Image fusion [1, 2] is a critical task in the medical field to combine medical image data from various modalities into a single image representation. Maintaining the spatial alignment of the fused images is a fundamental requirement for this process, and it should also be ensured that the information affecting the observation, such as image artifacts and noise, is as minimal as possible and that the information that is lacking is supplemented by the remaining information as much as possible. Deep learning techniques have been

widely used and improved in medical image fusion, and from the perspective of human involvement, fusion schemes in the field of medical image analysis can be divided into manual-based fusion and automatic adjustment-based fusion schemes. Manually designed fusion schemes require accurate spatial localization capability, generally considered from the perspectives of spatial alignment (feature vector agreement) and spatial distribution (e.g., color), however, once the image distribution has large differences, these schemes will fail the task due to spatial distortion and spectral distortion. Figure 1 shows the three main ways about how images are fused.

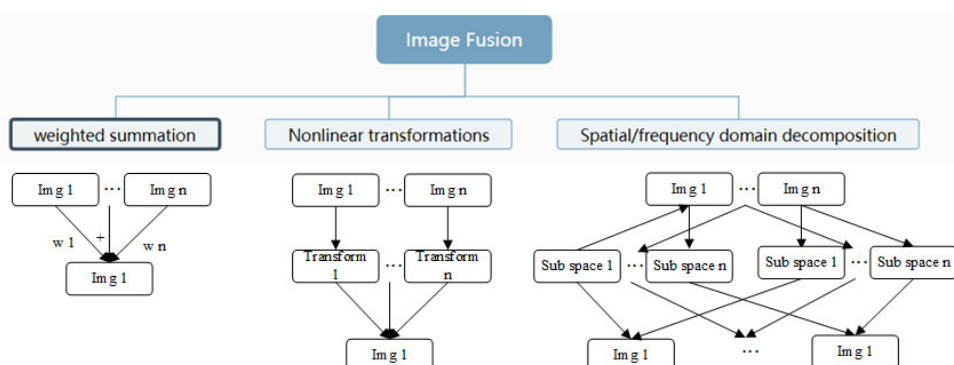


Figure 1. Illustration of common type of image fusion

Maintaining the spatial alignment of the fused images is a fundamental requirement in the medical field, where image fusion is a key task to integrating medical image information from various modalities into a unified image representation. It should also be ensured that the information affecting the observation, such as image artifacts and noise, is as minimal as possible, and that the lacking information is supplemented by each other as much as possible [3]. In the field of medical image analysis, fusion algorithms may be classified into those based on manual fusion and those based on the automated

correction. Deep learning approaches have been extensively used and enhanced in this area. However, once the image distribution has significant differences, these schemes will fail the task due to spatial distortion and spectral distortion [4]. Manually designed fusion schemes require accurate spatial localization capabilities, which are typically considered in terms of spatial alignment (feature vector agreement) and spatial distribution (e.g., color). These issues can be effectively avoided with the help of some schemes that do not use image space, such as wavelet transforms with null-

frequency transforms [5], the division of image frequency domain features into multiple bands, or sparse representation schemes based on dictionary learning [6]. However, these indirect transform-like schemes require more computational work, are more time-consuming, and are more complex. Therefore, high assured feature quality in the spatial domain and low-cost computation based on the non-spatial domain are the core components of manual-based design approaches. By allowing the fusion mechanism to automatically modify the best matching material depending on the task goals, the deep learning-based automatic adaptation strategy has more flexibility and may ultimately enhance the model's performance. Additionally, deep learning-derived multiple image representation algorithms may be employed as transformation techniques for fusing multimodal medical images, overcoming the problem of multiscale transformations' high computational cost and poor representation of edge information [7]. For instance, CSR convolutional sparse representation [8] solves the issues with sparse coding with limited ability to retain details and sensitivity to position bias [9]. It can obtain a sparse representation of the entire image without independently computing the representation of a set of overlapping image blocks like the traditional standard sparse representation (SR). In conclusion, the deep learning approach offers more research potential, with the main focus being on how to enhance the model's representational and perceptual capacities.

Feature perception in deep learning models refers to the model's capacity to extract and use features at various levels, and the capacity directly indicates the degree to which the model has a thorough, in-depth, and distinctive grasp of data features. As a general rule, improving feature perception is a common technique to increase the segmentation accuracy and robustness of a model. Common operations include deepening the network depth and width (ResNet [10], ResNeXt [11]), increasing the number of branches and utilizing multi-stage features (WideResNet [12], FPN [13], improving existing structures, using 3D structures of data (V-Net [14]), fusing models or data (H-DenseU-Net [15]). As of right now, the majority of deep learning-based image segmentation models are in tested benchmark networks like FCN and U-Net. By applying improvements to one or a combination of the aforementioned techniques, the models used in fusion schemes for medical image ROI, such as classification and segmentation of common brain and organ tissues, can perform even better in downstream tasks. Table 1 shows some examples of modal fusion and its role.

**Table 1.** Some modal fusion methods

No.	Modal	Categories
1	MUC	MRI&&CT
2	MUS	MRI&&SPECT
3	MUP	MRI&&PET
4	CUP	CT&&PET
5	CUS	CT&&SPECT
6	SUP	SPECT&&PET

## 2. Improvement of Multimodal Medical Image Fusion

### 2.1.1. Deep Learning

The densely connected network DenseNet, proposed by Huang et al [16], is one of the most effective methods to solve

gradient disappearance and gradient explosion in deep learning, while the dense connection can also be able to also fully extract information from the intermediate layers during medical image fusion to prevent the loss of significant information. In the process of network implementation to ensure the maximum information flow between the layers of the network, the network cascades the previous feature maps of the same size, adds additional inputs to the current layer, and the obtained feature map's convolution is used as input to the subsequent layers. Because the feature maps are processed in a cascade operation instead of the conventional addition operation, the densely connected network has  $L(L + 1)/2$  connections compared to the conventional L-layer network with L connections.

Although convolutional neural network-based multimodal fusion algorithms typically produce the best results, some medical images lose their semantic content after a fusion because CNN algorithms typically only use the results of the last layer as the extracted information and ignore the feature information corresponding to the intermediate layers. To address this issue, Li Hui et al. [17] proposed a deep learning architecture based on DenseNet that consists of an encoding network and a decoding network. The encoding network uses convolutional layers and dense blocks to extract features from images, and the fused images are then reconstructed by corresponding fusion strategies and decoding networks. This architecture successfully avoids the issue of information loss in CNN methods and achieves very high accuracy. The coding network used to extract the visual characteristics has the drawback of only taking into account one scale. A DenseNet-based method called MSDNet [18] that added a multiscale mechanism to DenseFuse and used three filters of various sizes to extract features from the coding network to address this issue also came from Li. By widening the coding network, more image details were extracted. The DenseNet technique is used to avoid gradient growth and gradient disappearance in the medical image fusion system proposed by Zhao et al. [19] based on dense blocks and adversarial generative networks. In conclusion, the medical picture fusion deep learning approach has two problems that the dense connection module fixes. The concerns of gradient disappearance and gradient expansion that emerge in deep learning are addressed first. The second is to use the last layer of extracted features in the CNN model to solve the problem of information loss in the intermediate layers caused by fusion. The design of the network structure's dense connections also has a regularization effect that might greatly reduce the overfitting phenomenon. Using DenseNet thick connection blocks as an effective enhancement technique for medical fusion, it is feasible to combine several models.

### 2.2. Transfer Learning

Although deep learning can automatically extract features, it needs a high number of data samples to train the model, and the tiny amount of medical data makes it impossible to do so. As a result, the trained model has poor generalization performance. Transfer learning allows for the application of already-trained models to unrelated but related fields. In deep learning, parameter migration is a common form of transfer learning that helps models fit data more accurately and generalize well, even with sparse data [20]. For instance, Sabine et al [21] used the VGG pre-trained in ImageNet and 16 networks for medical image fusion without prior training in image modalities, and the model can be effectively

generalized. Hermes et al [22] proposed a CNN-based MRI and CT image fusion method in the shear domain Shearlet, which employs migration learning from the spatial domain to the frequency domain. Pre-trained networks may successfully speed up learning, and employing previous knowledge can cut down on network training time and greatly raise the quality of fused pictures in the absence of datasets, according to Y Wang et al. [23].

### 2.3. Data Enhancement

The lack of medical image datasets can also be addressed through data augmentation, and current data augmentation techniques [24] can be used to enhance the existing datasets by rotating, scaling, or cropping them to produce new relevant data from the existing data and give the trained neural network better generalization capabilities. To some extent, the overfitting issue brought on by short training samples during deep learning training can be mitigated by the conventional technique of data augmentation through slicing [25]. Other data augmentation techniques exist, including CutMax [26], Mixup [27], and randomly modified luminance and affine transformation. A data augmentation technique was utilized to transform the 120w natural photos in imageNet into medical images with the same intensity and texture distribution as the dataset for the multilayer cascade network provided by Liang et al [28]. Since the medical industry has strict requirements for image realism and integrity, Liu et al. [29] proposed a deep convolutional neural network and data enhancement for image fusion for diagnosing fungal keratitis. They claimed that clipping after the angular rotation will lose a part of the image, and nonlinear enhancement methods such as noise addition and brightness adjustment will change the pixel values of the original image. As a result, image flipping is used to increase the normal image. However, the class imbalance in medical data across multiple modalities is frequently severe, and GAN development has led to further advances in data production [30]. Data unevenness may be solved using GAN-based modal transformation mapping. Data augmentation can help with the issue of the lack of medical image datasets, but because medical images are a specialized field with small samples, there aren't many efficient data augmentation techniques. Finding ways to obtain sufficiently rich and high-quality medical image data is still a promising area for future research.

### 2.4. Attentional Mechanisms

Vaswani [31] first suggested the Transformer model architecture, which is now extensively used in natural language processing (NLP) [32]. Transformer does away with conventional RNN and CNN in favor of a completely Attention-based network structure. Transformer-based models' success can be due to their superior capacity, when compared to recurrent neural networks and convolutional neural networks, to collect distant information.

Vibashan et al. [33] proposed a new network for image fusion (IFT) based on CNN and Transformer, and developed a SpatioTransformer (ST) fusion strategy, which handles both local and remote dependencies. The disadvantage of CNN-based fusion techniques is that they cannot extract remote dependencies in images, which results in the loss of some global contextual information that may be useful for fusing images. A CNN branch and a transformer branch are included in the ST fusion technique to fuse local and global characteristics. Additionally, it was utilized to combine

medical pictures from MRI and PET. When compared to USFusion and DDcGAN, the combined images showed greater intensity fluctuations and brighter colors. A novel architecture for multi-exposure image fusion based on Transformer was presented by Linhao Qu et al. [34], which employs self-supervised multitask learning and does not need ground and truth fusion pictures. Based on the properties of multi-exposure pictures, three self-supervised reconstruction tasks are developed, and the Transformer and CNN modules are merged to address the problem of distant dependencies in CNN-based architectures. Linhao Qu et al. [35] developed a unified Transformer-based image fusion framework using self-supervised learning and proposed three destructive reconstruction tasks for multimodal image fusion, multi-exposure image fusion, and multi-focus image fusion based on pixel intensity nonlinear transform, luminance transform, and noise transform, respectively. The collected MRI and CT images are contrasted and detailed using USFusion, IFCNN, and PGMI deep learning algorithms, which maintain the best texture and functional information.

### 2.5. Multi-scale Transformation Method Improvement

An important direction for medical picture fusion is the optimization of multiscale decomposition. Traditional multiscale transformations are used with deep learning-based medical image fusion techniques like CNN, GAN, etc. to dissect the high and low-frequency components of pictures, and the images are then fused using various fusion rules. Translation invariance issues plague earlier multiscale decompositions like curvilinear and contour waves, and to address these issues, the non-downsampled contour wave NSCT was introduced in [36] for picture fusion. To address the issue of considerable intensity changes across several multimodal medical pictures at the same place, Wang et al. [37] suggested medical image fusion based on convolutional neural networks and non-downsampled contour waves (NSCT). This was made possible by the translation invariance of NSCT. Goyal et al. [38] used convolutional neural networks and fractional order total generalized variance (FOTGA) in the NSCT domain for multimodal medical image fusion and denoising, first extracting features from noisy source images using the non-subsampled contour transform (NSCT). They hypothesized that the acquired medical images might be corrupted by noise due to sensor transmission errors. Then, a concatenated convolutional neural network is utilized to weigh and fuse the important characteristics from the two multimodal pictures. The noise information of the fused pictures is decreased by using the fractional order total generalized variation (FOTGV) approach for noise reduction. Shibu et al. [39] suggested a unique method for merging medical pictures that rely on convolutional neural networks and multi-scale decomposition with sparse representation, in contrast to Wang et al. The original picture was split into low- and high-frequency layers using L0 smoothing filters, and the high-frequency layers were combined using CNN. The low-frequency layers were combined using NSCT-based sparse representations (NSCT and SR), and the combined image was then rebuilt.

It uses NSCT decomposition to enhance the visual quality. This decomposition, nevertheless, is limited to a certain number of directional components. The translation invariance issue is also addressed by the non-downsampled shear wave transform (NSST), which has been proposed as a means of

overcoming these restrictions. Using the non-subsampled shear wave transform (NSST) and the curvilinear wave transform, for instance, Abas et al. [40] suggested a novel convolutional neural network-based technique that divides the source picture into low-frequency and high-frequency components first. Second, a weight map produced by a concatenated convolutional neural network (SCNN), which is formed from several feature maps that include information on pixel activity from various sources, is used to combine the low-frequency and high-frequency coefficients. Finally, multi-scale inverse transform (MST) is employed to recreate the fused pictures. Experiments show that the suggested strategy performs well in terms of visual quality and objective assessment, and may efficiently maintain complex structural information.

### 3. Application of Multimodal Medical Image Fusion

Deep learning-based medical image fusion methods are mainly applied to the diagnosis of brain diseases represented by brain tumors and Alzheimer's disease, as well as liver diseases and lung diseases.

#### 3.1. Multimodal Fusion of Brain Diseases

For the detection and surgical diagnosis of brain disorders including Alzheimer's disease and brain malignancies, multimodal medical image fusion is frequently employed. Lei et al. [41], for instance, fused MRI-PET images using conventional correlation analysis (CCA) to detect Alzheimer's disease (AD) with a diagnostic accuracy of 96.93%. To combine visual feature imaging MD maps from structural MRI and diffusion tensor to discriminate AD from MCI diseases, Ahmed [42] suggested a multimodal image fusion approach based on multinuclear learning (MKL). To detect brain cancers, Algani et al. [43] fused MRI-CT images using a CNN method.

For the diagnosis of Alzheimer's disease, Vu et al. [44] suggested multimodal picture fusion based on convolutional neural networks and sparse self-coding. To identify an effective convolutional filter, PET and MRI image data were trained and evaluated using coefficient autoencoders. A three-layer neural network with a softmax function was then utilized to achieve classification. According to the findings, there was a 90% accuracy rate between AD patients and healthy controls. Ma et al. [45] suggested a dual discriminator-based adversarial generative network (DDcGAN) method for combining high-resolution MRI and PET scans with low-resolution positron emission tomography images of the brain. On publicly accessible datasets, qualitative and quantitative tests showed that DDcGAN was visibly The qualitative and quantitative tests using publicly available datasets show that DDcGAN outperforms FusionGAN in terms of visualization and quantitative measures, and the fused pictures can be identified in the brain with reasonable accuracy.

To combine brain MRI and PET images, Zhang et al. [46] suggested an end-to-end multimodal brain image fusion architecture. To begin with, an auto-encoder was used to extract the source pictures' characteristics. Then it is suggested to combine the image features using an information preservation weighted channel space attention model (ICS), establishing an adaptive weight based on the information preservation degree of the features. The fused medical

pictures are finally rebuilt using the decoder model. By using an improved attention model and encoder-decoder structure, the method efficiently increases the quality and decreases the fusion time of the fused images. The fused images also contain functional and structural information about the brain that can be used to diagnose brain diseases more effectively and accurately.

Abirami et al. [47] proposed a GAN-based fusion of positron emission computed tomography (PET) and magnetic resonance imaging (MRI) images, and the fused images were able to segment brain tumors accurately. Additionally, tumor areas could be accurately localized in a single fused image using information obtained from PET and MRI images, and this method improved the accuracy of diagnosing tumors. This method shortens the time needed to identify and locate tumors while increasing the accuracy of tumor diagnosis. The structural similarity index and mutual information are two measures used to assess the effectiveness of the GAN-based model. The approach produced mutual information of 2.8059 and a structural similarity index of 0.8551. To extract high-level features and fuse MRI and CT images, Reddy et al. [48] developed a fusion method utilizing a convolutional neural network (CNN) with a pyramidal generating kernel. With the addition of non-local Euclidean median filtering adaptive angular covariance, FCM clustering based on Gaussian kernel, and segmenting brain tumors from fused pictures, the tumor segmentation method for cancer analysis and detection has become extremely accurate and effective.

#### 3.2. Diagnosis of Liver and Lung Diseases

Liver diseases such as tumors not only cause changes in the physical morphology of tissues, but also increase the intensity of local metabolism, and the combination of functional and structural medical images can improve the diagnostic accuracy of these diseases. Li et al [49] proposed an attention-guided discriminative and adaptive fusion approach based on a deep learning architecture to address the difficulty of better multimodal medical feature fusion between larger modalities through complementary information representation problem by introducing discriminative feature learning loss to reduce the distance of features of the same class and increase the distance of features of different classes of tumors in a single modality. Finally, an adaptive weighting strategy is designed to increase the contribution of modalities with relatively low loss values and reduce the influence of modalities with larger loss values on the final loss function, and the results show that the proposed method can effectively classify clinical hepatocellular carcinoma.

Fu et al [50] proposed a lung information fusion model based on 3D CT images and serum biomarkers for diagnosing squamous cell carcinoma, adenocarcinoma, inflammation, and other benign lung nodule types. By constructing a resolution 3D multi-classification deep learning model (Mr-Mc) and a multilayer perceptron machine learning model (MLP) for diagnosing multiple pathological types of lung nodules, and also fusing Mr-Mc and MLP using transfer learning, a multimodal information fusion model is constructed that can classify multiple pathological types of lung nodules. The results showed that the average accuracy of the constructed Mr-Mc model can reach 0.805 and the average accuracy of the MLP model can reach 0.887. The fusion model was validated on a dataset containing 64 samples with an average accuracy of 0.906. Zhang et al [51] proposed an adaptive dynamic loss function for multi-scale

expansion network of pathological images in different scales. Weighting is performed, and an ant colony algorithm based on maximum information coefficient correlation is designed for unsupervised feature selection by fusing image features and patient differential genes considering complementary information between different medical modality data. The results show that the incorporation of pathological image information and genetic information plays an important role in the classification of lung cancer subtypes. The algorithm can converge quickly compared with other feature selection methods. Combining pathological images and gene expression matrices for cancer diagnosis can improve the diagnostic accuracy of specific patients with an accuracy rate of 95.62%.

## 4. Conclusion

While relying on exploring and enhancing existing deep learning model structures can enhance performance in medical image analysis, this approach has a drawback in that it can only explore the best model states in the sparse medical image data and cannot deeply refer to related image data or structural data to achieve the benefit of synthesizing data from multiple sources. Researchers can learn new approaches by studying multimodal fusion techniques, such as shifting their focus from mining model structures to mining how data and models are integrated. By weighting, nonlinear processing, and combining, they may often enhance each other's information and eliminate noise.

In conclusion, the following issues with existing multimodal data fusion strategies still exist:

1. Data heterogeneity: Data from several modalities may have unique qualities and features, such as text and picture data, which need particular processing to be fused successfully. The model could over-rely on the data from one of the modalities and overlook the information from the other modalities as a result of variances in the quality and reliability of the data.

2. Feature extraction: Because distinct datasets may exhibit varying distributions and properties, it is necessary to develop effective feature extraction techniques.

3. Semantic inconsistency: Although multimodal data fusion combines the semantic data from several modalities, there may be a semantic discrepancy between various modalities.

4. A lack of large-scale datasets: The development of multimodal data fusion models is constrained by the difficulty of acquiring and annotating multimodal datasets.

Future work must concentrate on the study of feature learning and joint training to successfully fuse the features of various modalities and resolve the semantic inconsistency between them to overcome these issues. The design of medical multimodal fusion will also enable computer-aided diagnosis technology to influence the results of super manual tests as a result of the creation of deep learning model structures and methodologies.

## References

- [1] S.-Q. Deng, L.-J. Deng, X. Wu, R. Ran, D. Hong, and G. Vivone, "PSRT: Pyramid Shuffle-and-Reshuffle Transformer for Multispectral and Hyperspectral Image Fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1-15, 2023.
- [2] L. Tang, J. Yuan, H. Zhang, X. Jiang, and J. Ma, "PIAFusion: A progressive infrared and visible image fusion network based on illumination aware," *Information Fusion*, vol. 83, pp. 79-92, 2022.
- [3] B. Meher, S. Agrawal, R. Panda, and A. Abraham, "A survey on region based image fusion methods," *Information Fusion*, vol. 48, pp. 119-132, 2019.
- [4] A. P. James and B. V. Dasarathy, "Medical image fusion: A survey of the state of the art," *Information fusion*, vol. 19, pp. 4-19, 2014.
- [5] R. Singh and A. Khare, "Fusion of multimodal medical images using Daubechies complex wavelet transform—A multiresolution approach," *Information fusion*, vol. 19, pp. 49-60, 2014.
- [6] H. Yin, Y. Li, Y. Chai, Z. Liu, and Z. Zhu, "A novel sparse-representation-based multi-focus image fusion approach," *Neurocomputing*, vol. 216, pp. 216-229, 2016.
- [7] Y. Liu, X. Chen, Z. Wang, Z. J. Wang, R. K. Ward, and X. Wang, "Deep learning for pixel-level image fusion: Recent advances and future prospects," *Information Fusion*, vol. 42, pp. 158-173, 2018.
- [8] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *2010 IEEE Computer Society Conference on computer vision and pattern recognition*, 2010: IEEE, pp. 2528-2535.
- [9] B. Wohlberg, "Efficient algorithms for convolutional sparse representations," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 301-315, 2015.
- [10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
- [11] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1492-1500.
- [12] S. Zagoruyko and N. Komodakis, "Wide residual networks," *arXiv preprint arXiv:1605.07146*, 2016.
- [13] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117-2125.
- [14] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*, 2016: Ieee, pp. 565-571.
- [15] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE transactions on medical imaging*, vol. 37, no. 12, pp. 2663-2674, 2018.
- [16] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700-4708.
- [17] H. Li and X.-J. Wu, "DenseFuse: A fusion approach to infrared and visible images," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2614-2623, 2018.
- [18] X. Song, X.-J. Wu, and H. Li, "MSDNet for medical image fusion," in *International conference on image and graphics*, 2019: Springer, pp. 278-288.
- [19] C. Zhao, T. Wang, and B. Lei, "Medical image fusion method based on dense block and deep convolutional generative adversarial network," *Neural Computing and Applications*, vol. 33, no. 12, pp. 6595-6610, 2021.

- [20] K. Bousmalis, G. Trigeorgis, N. Silberman, D. Krishnan, and D. Erhan, "Domain separation networks," *Advances in neural information processing systems*, vol. 29, 2016.
- [21] F. Lahoud and S. Süsstrunk, "Zero-learning fast medical image fusion," in *2019 22th International Conference on Information Fusion (FUSION)*, 2019: IEEE, pp. 1-8.
- [22] H. Hermessi, O. Mourali, and E. Zagrouba, "Convolutional neural network-based multimodal image fusion via similarity learning in the shearlet domain," *Neural Computing and Applications*, vol. 30, no. 7, pp. 2029-2045, 2018.
- [23] Y. Wang, C. Wu, L. Herranz, J. van de Weijer, A. Gonzalez-Garcia, and B. Raducanu, "Transferring gans: generating images from limited data," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 218-234.
- [24] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," *arXiv preprint arXiv:1712.04621*, 2017.
- [25] Y. Guo et al., "Segmenting hippocampus from infant brains by sparse patch matching with deep-learned features," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2014: Springer, pp. 308-315.
- [26] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "Ccnet: Criss-cross attention for semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 603-612.
- [27] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.
- [28] X. Liang, P. Hu, L. Zhang, J. Sun, and G. Yin, "MCFNet: Multi-layer concatenation fusion network for medical images fusion," *IEEE Sensors Journal*, vol. 19, no. 16, pp. 7107-7119, 2019.
- [29] Z. Liu et al., "Automatic diagnosis of fungal keratitis using data augmentation and image fusion with deep convolutional neural network," *Computer Methods and Programs in Biomedicine*, vol. 187, p. 105019, 2020.
- [30] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125-1134.
- [31] A. Vaswani et al., "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [32] Z. Dai, Z. Yang, Y. Yang, J. Carbonell, Q. V. Le, and R. Salakhutdinov, "Transformer-xl: Attentive language models beyond a fixed-length context," *arXiv preprint arXiv:1901.02860*, 2019.
- [33] V. VS, J. M. J. Valanarasu, P. Oza, and V. M. Patel, "Image Fusion Transformer," *arXiv preprint arXiv:2107.09011*, 2021.
- [34] L. Qu, S. Liu, M. Wang, and Z. Song, "TransMEF: A Transformer-Based Multi-Exposure Image Fusion Framework using Self-Supervised Multi-Task Learning," *arXiv preprint arXiv:2112.01030*, 2021.
- [35] L. Qu et al., "TransFuse: A Unified Transformer-based Image Fusion Framework using Self-supervised Learning," *arXiv preprint arXiv:2201.07451*, 2022.
- [36] A. L. Da Cunha, J. Zhou, and M. N. Do, "The nonsubsampling contourlet transform: theory, design, and applications," *IEEE transactions on image processing*, vol. 15, no. 10, pp. 3089-3101, 2006.
- [37] Z. Wang, X. Li, H. Duan, Y. Su, X. Zhang, and X. Guan, "Medical image fusion based on convolutional neural networks and non-subsampling contourlet transform," *Expert Systems with Applications*, vol. 171, p. 114574, 2021.
- [38] S. Goyal, V. Singh, A. Rani, and N. Yadav, "Multimodal image fusion and denoising in NSCT domain using CNN and FOTGV," *Biomedical Signal Processing and Control*, vol. 71, p. 103214, 2022.
- [39] D. S. Shibu and S. S. Priyadharsini, "Multi scale decomposition based medical image fusion using convolutional neural network and sparse representation," *Biomedical Signal Processing and Control*, vol. 69, p. 102789, 2021.
- [40] A. İ. Abas, H. E. Koçer, and N. A. Baykan, "Medical image fusion with convolutional neural network in multiscale transform domain," *TURKISH JOURNAL OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCES*, 2021.
- [41] B. Lei, S. Chen, D. Ni, and T. Wang, "Discriminative learning for Alzheimer's disease diagnosis via canonical correlation analysis and multimodal fusion," *Frontiers in aging neuroscience*, vol. 8, p. 77, 2016.
- [42] O. B. Ahmed, J. Benois-Pineau, M. Allard, G. Catheline, C. B. Amar, and A. s. D. N. Initiative, "Recognition of Alzheimer's disease and Mild Cognitive Impairment with multimodal image-derived biomarkers and Multiple Kernel Learning," *Neurocomputing*, vol. 220, pp. 98-110, 2017.
- [43] A. D. Algarni, "Automated medical diagnosis system based on multi-modality image fusion and deep learning," *Wireless Personal Communications*, vol. 111, no. 2, pp. 1033-1058, 2020.
- [44] T. D. Vu, H.-J. Yang, V. Q. Nguyen, A.-R. Oh, and M.-S. Kim, "Multimodal learning using convolution neural network and Sparse Autoencoder," in *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 2017: IEEE, pp. 309-312.
- [45] J. Ma, H. Xu, J. Jiang, X. Mei, and X.-P. Zhang, "DDcGAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion," *IEEE Transactions on Image Processing*, vol. 29, pp. 4980-4995, 2020.
- [46] Z. Shi, C. Zhang, D. Ye, P. Qin, R. Zhou, and L. Lei, "MMI-Fuse: Multimodal Brain Image Fusion With Multiattention Module," *IEEE Access*, vol. 10, pp. 37200-37214, 2022.
- [47] R. Nandhini Abirami, P. Durai Raj Vincent, K. Srinivasan, K. S. Manic, and C.-Y. Chang, "Multimodal Medical Image Fusion of Positron Emission Tomography and Magnetic Resonance Imaging Using Generative Adversarial Networks," *Behavioural Neurology*, vol. 2022, 2022.
- [48] M. G. Reddy, P. V. N. Reddy, and P. R. Reddy, "Segmentation of fused MR and CT images using DL-CNN with PGK and NLEM filtered AACGK-FCM," *Biomedical Signal Processing and Control*, vol. 68, p. 102618, 2021.
- [49] S. Li, Y. Xie, G. Wang, L. Zhang, and W. Zhou, "Attention guided discriminative feature learning and adaptive fusion for grading hepatocellular carcinoma with Contrast-enhanced MR," *Computerized Medical Imaging and Graphics*, vol. 97, p. 102050, 2022.
- [50] Y. Fu et al., "Fusion of 3D lung CT and serum biomarkers for diagnosis of multiple pathological types on pulmonary nodules," *Computer Methods and Programs in Biomedicine*, vol. 210, p. 106381, 2021.
- [51] Y. Zhang, J. Zhao, Y. Qiang, X. Yang, W. Wu, and L. Jia, "Improved heterogeneous data fusion and multi-scale feature selection method for lung cancer subtype classification," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 1, p. e6535, 2022.