applied sciences

MDPI

Article

Multi-Focus Microscopy Image Fusion Based on Swin Transformer Architecture

Han Hank Xia [1,2], Hao Gao [3,*], Hang Shao [2], Kun Gao [2] and Wei Liu [2]

[1] School of Electronic Science and Engineering, Nanjing University, Nanjing 210093, China; xiahannju@gmail.com
[2] Yangtze Delta Region Institute of Tsinghua University, Jiaxing 314006, China; shaohang@zfti.org.cn (H.S.); tongyoung9123@163.com (K.G.); liuwei@zfti.org.cn (W.L.)
[3] Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing 210023, China
* Correspondence: tsgaohao@gmail.com

Abstract: In this study, we introduce the U-Swin fusion model, an effective and efficient transformer-based architecture designed for the fusion of multi-focus microscope images. We utilized the Swin-Transformer with shifted window and path merging as the encoder for extracted hierarchical context features. Additionally, a Swin-Transformer-based decoder with patch expansion was designed to perform the un-sampling operation, generating the fully focused image. To enhance the performance of the feature decoder, the skip connections were applied to concatenate the hierarchical features from the encoder with the decoder up-sample features, like U-net. To facilitate comprehensive model training, we created a substantial dataset of multi-focus images, primarily derived from texture datasets. Our modulators demonstrated superior capability for multi-focus image fusion to achieve comparable or even better fusion images than the existing state-of-the-art image fusion algorithms and demonstrated adequate generalization ability for multi-focus microscope image fusion. Remarkably, for multi-focus microscope image fusion, the pure transformer-based U-Swin fusion model incorporating channel mix fusion rules delivers optimal performance compared with most existing end-to-end fusion models.

Keywords: multi-focus fusion; microscope images; Swin Transformer

Citation: Xia, H.H.; Gao, H.; Shao, H.; Gao, K.; Liu, W. Multi-Focus Microscopy Image Fusion Based on Swin Transformer Architecture. *Appl. Sci.* **2023**, *13*, 12798. https://doi.org/10.3390/app132312798

Academic Editor: Sungho Kim

Received: 25 October 2023
Revised: 22 November 2023
Accepted: 23 November 2023
Published: 29 November 2023

1. Introduction

All optical imaging systems, especially for light microscopy, have a limited depth of field. Three-dimensional objects under investigation are often thicker than the depth of field of the imaging system, meaning that it is impossible to acquire a whole object completely in focus in one single image; only those portions that lie within the depth of field appear in focus and sharp, whereas the remaining regions are blurred by the system's point spread function (PSF) [1]. To overcome this limitation, multi-focus image fusion (MFIF) is an effective way to generate an all-in-focus image from a set of partially focused images, extending the depth of field of cameras. Compared with other methods to extend the depth of field of cameras, multi-focus image fusion tries to ensure the accuracy of the information, which is quite important for medical images. However, it takes some time to collect a set of partially focused images.

MFIF has been applied to various applications, such as micro-image fusion [2], visual sensor networks [3], visual power patrol inspection [4] and optical microscopy [5]. Image fusion technology has been developed for more than 30 years, during which various methods have been published. Conventional MFIM methods can be divided into spatial-based methods and transform-domain-based methods. Spatial-based methods fuse the image in the spatial domain directly, which can be further divided into pixel-based [6], block-based [7] and region-based [8] methods. In contrast, transform-domain-based methods

transform images into another domain firstly. Then, the transformed coefficients are merged by a pre-designed fusion rule. Finally, the fused image is reconstructed by applying the corresponding inverse transform based on the fused coefficients. Many transform-domain-based methods have been proposed so far, such as parse representation (SR) methods [9–11], multi-scale methods [12–16], gradient domain-based methods [17,18] and hybrid methods [19].

In recent years, with the advancement of deep learning, machine learning algorithms have been widely employed for various image fusion tasks, achieving remarkable success in the image fusion field. Various deep learning models, such as CNNs [20,21], GANs [22] and ensemble learning [23], have demonstrated their capability to attain state-of-the-art (SOTA) results in MFIF tasks. Recently, the transformer, a prominent architecture in natural language processing (NLP) [24], has been introduced into the computer vison domain to address the limitation of the long-range dependencies for CNN-based models [25–30]. Furthermore, models based on the Swin Transformer [31] achieved SOTA performance in various tasks, such as image classification, object detection and semantic segmentation. Swin-Unet, a novel pure transformer-based U-shaped encoder-decoder architecture, was proven to have excellent performance and generalization ability for medical image segmentation [32].

In summary, the transformer architecture based on the self-attention mechanism has achieved great success in the natural language processing domain. Recently, many transformer architectures have been introduced into the computer vision domain and achieved SOTA performance for various of computer vision tasks. In these studies, the Swin Transformer is an excellent work, which achieved perfect performance with linear computational complexity concerning image size. Therefore, we propose end-to-end image fusion models based on the Swin Transformer backbone to leverage the capabilities of a transformer for multi-focus image fusion in microscope images in this study. The main contributions of this paper can be summarized as follows:

1. Propose end-to-end models that use the Swin Transformer as a backbone to directly generate the fully focused images from multi-focus microscope images. Unlike existing CNN-based methods, this approach can extract long-range dependencies to generate naturally fully focused images.
2. To facilitate comprehensive model training, we created a substantial dataset of multi-focus images, primarily derived from texture datasets.
3. Propose two types of fusion rules which can be used in multi-focus image fusion, known as simple mix and channel mix. The evaluation results demonstrate that our transformer-based U-shaped models achieve state-of-the-art performance in multi-focus image fusion (MFIF) tasks for microscope images.

2. Related Work

2.1. Traditional MFIF Method

Traditional MFIF methods are typically categorized into two main groups: transform domain methods and spatial domain methods [33]. Transform domain methods mainly operate the decomposition coefficient after image transformation, encompassing three key fusion stages: image transformation, coefficient after image transformation, and inverse transformation reconstruction [34]. According to the application of the image transform, transform domain methods can be further classified into multi-scale decomposition (MSD)-based methods (e.g., Laplacian pyramid [14,35], discrete wavelet transform [36,37], nonsubsampled contourlet transform [38–40], neighbor distance filtering [41,42], empirical mode decomposition [43]), sparse representation (SR)-based methods (orthogonal matching pursuit [44,45]), gradient domain (GD)-based methods (structure tensor [46,47]), methods based on other transform (independent component analysis [48], cartoon-texture decomposition [49]) and methods based on the combination of different transforms (curvelet transform and wallet transform [50]). In the spatial domain methods, as the name suggests, source images are fused in the spatial domain based on the spatial features of images,

which can be divided into block-based methods [51–53], region-based methods [8,54] and pixel-based methods [55,56].

2.2. Convolutional Neural Network

In contrast to the traditional MFIF methods, deep-learning-based MFIF methods, especially the CNN-based MFIF method, can extract deep feathers to generate robust fully focused images with various input. Deep-learning-based MFIF methods can be categorized into decision-based and end-to-end methods [57]. In decision-based methods, firstly, a decision map that indicates the focus level (or activity level) is generated based on deep feathers from the CNN. Subsequently, post-processing steps are performed to generate the fused image according to the decision map. To the best of our knowledge, Li et al. [58] proposed the first CNN-based MFIF method, which utilizes a CNN learning a decision map to generate fused images. Later, serval decision-based MFIF methods are proposed, such as MCNN [21], HF-Seg [59], MMF-Net [60] and SSAN [61]. In end-to-end MFIF methods, the fused image is learned directly through training without post-processing steps. Several encoder–decoder architectures are proposed, such as IFCNN [62] and U2Fusion [63].

2.3. Self-Attention-Based Backbone

The transformer architecture [24] based on the self-attention mechanism has achieved SOTA performance for various tasks in the natural language processing (NLP) domain [64]. Building upon the transformer's success in the NLP domain, Dosovitskiy et al. [25] introduced the vision transformer (ViT), which has demonstrated excellent results as an alternative to convolutional networks while requiring significantly fewer computational resources for training. Most recently, several excellent works based on ViTs [30] have emerged, notably Liu et al. [31] proposed a hierarchical transformer known as the Swin Transformer. This architecture offers flexibility in modeling at various scales, exhibits linear computational complexity concerning image size and is compatible with a wide range of vision tasks. Building on the Swin Transformer architecture, Cao et al. [33] prosed the pure transformer-based U-shaped encoder-decoder network (Swin-Unet).

3. Methodology

3.1. Overall Architecture

We proposed an end-to-end MFIF module for microscope images based on the Swin Transformer architecture, as presented in Figure 1. The features pyramid of the multi-focus images was encoded by the Swin-Transformer-based model Swin-S [31]. To generate the fused image, a Swin-Transformer-based decoder with patch expansion (Figure 1) was applied to decode the feature pyramid. A Swin Transformer block is composed of a LayerNorm (LN) layer, a multi-head self-attention module, a residual connection and a two-layer MLP with GELU non-linearity. The window-based multi-head self-attention (*W-MSA*) module and the shifted window-based multi-head self-attention (*SW-MSA*) module are applied in the two successive transformer blocks (Figure 2). Based on such a window partitioning mechanism, the Swin Transformer blocks are computed as:

$$\hat{z}^l = W - MSA(LN(z^{l-1})) + z^{l-1} \tag{1}$$

$$z^l = MLP(\hat{z}^l) + \hat{z}^l \tag{2}$$

$$\hat{z}^{l+1} = SW - MSA(LN(z^l)) + z^l \tag{3}$$

$$z^{l+1} = MLP(\hat{z}^{l+1}) + \hat{z}^{l+1} \tag{4}$$

where $\hat{z}^l$ and z^l denote the ouputs of the *(S)W-MSA* module and the MLP module of the l^{th} block, respectively. Self-attention is computed as:

$$Attention(Q, K, V) = SoftMax(\frac{QK^T}{\sqrt{d}} + B)V \quad (5)$$

where $Q, K, V \in \mathbb{R}^{M^2 \times d}$ represent the query, key and value matrices. M^2 and d denote the number of patches in a window and the dimension of the query or key, respectively. And the values in B are taken from the bias matrix $\hat{B} \in \mathbb{R}^{(2M-1)\times(2M+1)}$.

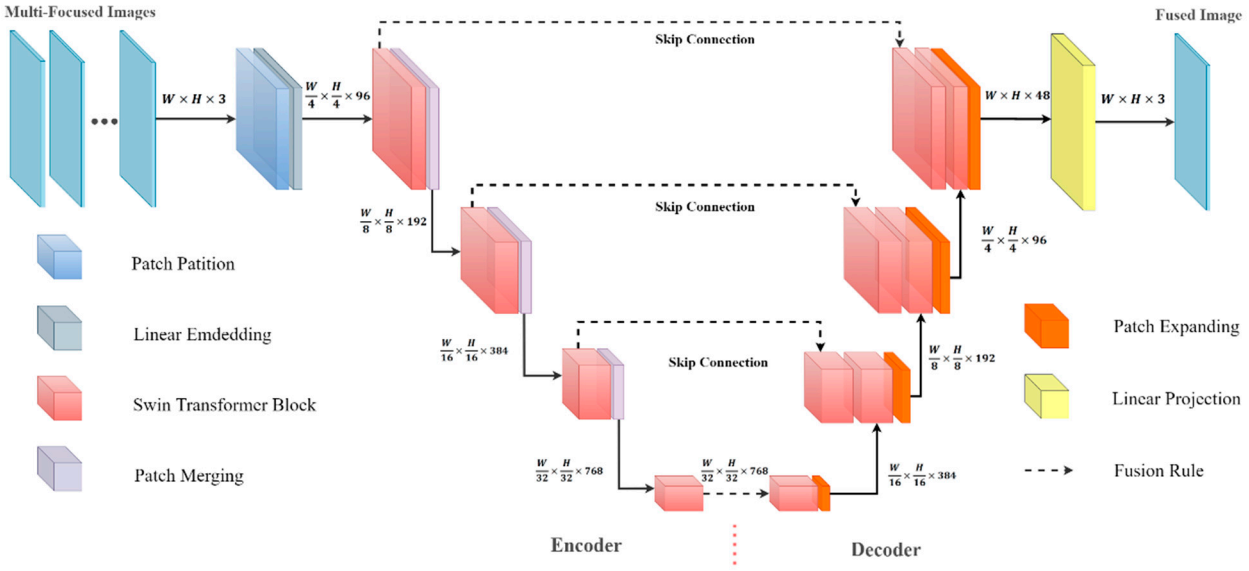

Figure 1. The architecture of the end-to-end MFIF models based on the Swin Transformer block, U-Swin fusion model.

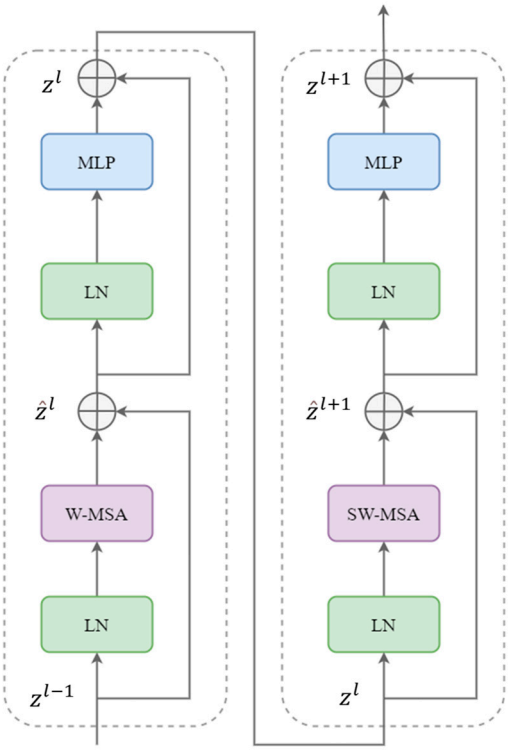

Figure 2. Two successive Swin Transformer blocks.

Initially, the input RGB images were divided into non-overlapping patches with a 4×4 patch and a linear embedding layer and were applied to project raw-value features into an arbitrary dimension, similar to ViTs [25]. To generate hierarchical representations, several Swin Transformer blocks and patch-merging layers were applied to encode the transformed patch tokens. The patch-merging layer concatenates the features of each group of 2×2 neighboring patches to down-sample and increase dimensions. Swin Transformer blocks are applied to learn feature representations without dimension transformation. To generate the fused image, symmetric Swin Transformer blocks and patch expansion were applied to perform up-sampling in U-Swin fusion models, like Swin-Unet [32]. The patch expansion layer reshapes the feature maps into a higher resolution feature map ($2\times$ up-sampling). To fuse the hierarchical features of multi-input images, varieties of fusion rules (tensor max and tensor mean) have been utilized to fuse the transformed path tokens, as presented in the IFCNN [62]. Furthermore, two types of mix fusion rules (simple mix and channel mix) were used in this study, which can be expressed as follows:

$$F = \omega_1 F^{Max} + \omega_2 F^{Mean} \ (Simple\ mix) \tag{6}$$

$$F_i = \omega_{(1,\,i)} F_i^{Max} + \omega_{(2,\,i)} F_i^{Mean} \ (Channel\ mix) \tag{7}$$

where F^{Max} and F^{Mean} denote the fused features from tensor max and tensor mean, respectively, and i represents the different channel of features. ω is not the hyperparameter that can be optimized during training mode. In this study, the channel mix fusion rule was implemented by the depthwise convolution to avoid information exchange between different channels. Inspired by U-net [65], the skip connections were applied to concatenate the hierarchical features from the encode with the up-sample features to enhance overall performance. The shallow features are concatenated with the up-sampling features from deep features to reduce the loss of spatial information caused by down-sampling.

3.2. Training Data Augmentation

As reported in the previous literature, the multi-focus image dataset can be more easily generated from other types of image datasets, and more importantly, the ground-truth fusion images of the multi-focus images could be obtained simultaneously while generating training data [62]. In this paper, 525 texture images were selected from the Describable Textures Datasets [66], Salzburg Texture Image Database (https://www.wavelab.at/sources/STex/ (accessed on 19 April 2023)) and Original Brodatz's Texture Database (http://multibandtexture.recherche.usherbrooke.ca/original_brodatz.html (accessed on 19 April 2023)). The texture images contain bubbly, fibrous, honeycombed, marble, pleated, veined, etc., images as well as images of objects such as gravel and hair. These textures are similar to the biomedical images observed under a microscope. Initially, the RGB images were resized to 224×224 by cubic interpolation. Depth images for each texture dataset were generated by the Perlin noise algorithm with specified frequencies (2, 4, 8) and random curves (from 1 to 7). As a result, we generated 1575 training datasets. The resized pairs of RGB images (I_s, Figure 3a) and depth images of the datasets (I_d, Figure 3b) were applied to create the multi-focus image dataset as the following procedures:

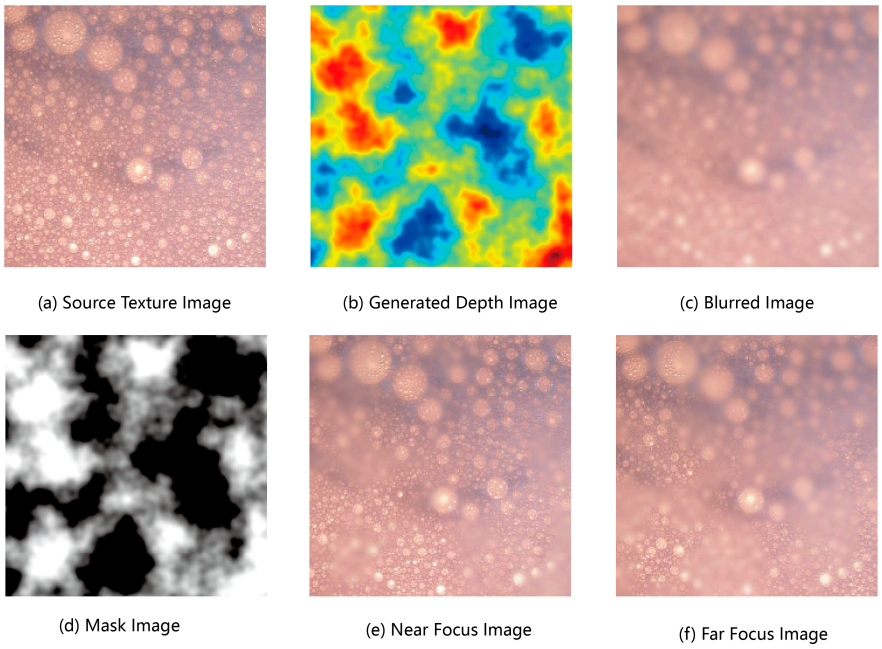

Figure 3. Example of generating multi-focus training dataset according to Section 3.2. (**a**,**b**) are the source image from texture datasets and depth image generated by Perlin noise algorithm, respectively. (**c**) denotes the randomly blurred image. (**d**) shows the mask map according to random threshold depth. (**e**,**f**) are the near-focus image and far-focus image, which are the input images for training models.

1. A completely blurry image (I_b, Figure 3c) is generated by randomly blurring the source RGB image (I_s) with a Gaussian filter, which can be expressed as:

$$I_b = G * I_s \tag{8}$$

Here, * denotes the convolution operation, and G denotes the Gaussian kernel, which is generated with a random kernel radius, kr, from 1 to 15:

$$G(x,y) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{x^2+y^2}{2\sigma^2}} \tag{9}$$

where σ denotes the standard deviation of the Gaussian filter and can be expressed as:

$$\sigma = 0.3 \times (kr - 1) + 0.8 \tag{10}$$

2. The mask map (I_m, Figure 3d) was generated from depth images of datasets (I_d) based on a random threshold depth (d_{th}).

$$I_m(x,y) = \begin{cases} 0, & I_d(x,y) \geq d_{th} \\ 1, & I_d(x,y) < d_{th} \end{cases} \tag{11}$$

$$d_{th} = \gamma \times (max(I_d) - min(I_d)) + min(I_d) \tag{12}$$

Here, γ denotes the depth threshold ratio, which is randomly selected from the range 0.3 to 0.7.

3. The near-focus image I_n (Figure 3e) and the far-focus image I_f (Figure 3f) were generated based on source image (I_s), completely blurred image (I_b) and the mask map (I_m), which can be expressed as the following equation:

$$\begin{cases} I_n = I_s \cdot I_m + I_b \cdot (1 - I_m) \\ I_f = I_b \cdot I_m + I_s \cdot (1 - I_m) \end{cases} \tag{13}$$

Naturally, the source images (I_s) can serve as the ground truth for fused images.

3.3. Training Details

The total of training process can be divided into two parts which are described as the following:

Stage 1: To improve the generation of hierarchical features, the encode part of the fused model was initialized by the pretrained model Swin-S [31], which has been trained on ImageNet-1K dataset [67] for image classification. Subsequently, the fusion model was trained on the training dataset generated above with mean squared loss and the AdamW optimizer [68] for 1000 epochs. It is worth noting that we employed a cosine decay learning rate with an 0.0005 initial learning rate, 0.05 weight decay and 20 epochs of linear warm-up. The batch size during the training period was set to 16.

Stage 2: Mean square error loss is the basic loss function which is used to regularize the prediction close to the ground-truth output. To encourage the network to produce images with greater texture similarity to the ground-truth fusion images, we incorporated a texture loss into the training model with the same hyperparameters as in the previous stage (1000 epochs, 16 batch size and Adam optimizer). The training loss is expressed as:

$$F_{loss} = \omega_1 M_{loss} + \omega_2 T_{loss} \tag{14}$$

Here, M_{loss} and T_{loss} denote the mean square error and textures loss separately. In this study, ω_1 and ω_2 are both set to 1.

4. Experiments

In this section, we conducted extensive experiments to validate the advantage of the proposed fusion models based on various evaluation methods. Firstly, the basic experimental settings are described, and then qualitative and quantitative results are illustrated and discussed for the EDoF Fraunhofer dataset [69].

4.1. Experiment Details

To evaluate the effectiveness of the transformer architecture in multi-focus image fusion, we compared the fusion performance of our Swin-Transformer fusion model with representative end-to-end methods.

We conducted comparisons with the SOTA unsupervised deep learning fusion model, called FusionDN [70], which employs a unified densely connected network to fuse images. Additionally, our models applied the same fusion rules as the IFCNN [62], which is a representative fusion framework based on the convolutional neural network. Therefore, we compared our fusion models with the IFCNN framework with the elementwise-maximum, elementwise-mean and elementwise-sum fusion rules, referred to as IFCNN-MAX, IFCNN-MEAN and IFCNN-SUM, respectively. Furthermore, we used the fusion result from four existing fusion models for validation: an unsupervised adversarial network with adaptive and gradient joint constrains (MMF-GAN) [71], a fast unified network based on proportional maintenance of gradient and intensity (PMGI) [72], a novel unified and unsupervised end-to-end image fusion network (U2Fusion) [63] and a novel general image fusion framework based on cross-domain long-range learning and the Swin Transformer (Swin fusion) [73]. Moreover, we conducted a comparison with recent SOTA fusion models, a novel memory unit architecture for image fusion (MUFusion) [74] and one of the first zero-shot models for image fusion (ZMFF) [75].

In this study, the complex wavelet extended-depth-of-field method, which remains the traditional gold standard method for image fusion [76], was utilized to generate the ground truth for multi-focus microscope image fusion. To perform image fusion with this method, we used ImagJ (v1.54b) software [77] with the extended-depth-of-field plugin (http://bigwww.epfl.ch/demo/edf/ (accessed on 25 May 2023)).

To evaluate the performance of the different models, we employed three widely used metrics: mean square error (*MSE*), peak signal-to-noise ratio (*PSNR*) and structure similarity index (*SSIM*). *MSE* and *PSNR* are commonly used metrics for objective image quality assessment, which can be expressed as follows:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[GT(i,j) - F(i,j)]^2 \tag{15}$$

$$PSNR = 10 \cdot log_{10}(\frac{255^2}{MSE}) \tag{16}$$

MSE can quantify the discrepancy between the ground truth and fused image at the pixel level, while *PSNR* represents the ratio between effective information and noise in the fused image. However, *MSE* and *PSNR* do not account for the spatial arrangement of the pixels, which is essential for image quality assessment [78]. Moreover, *MSE* and *PSNR* are easily dominated by local outliers. To overcome the weakness of *MSE/PSNR*, a new metric, named the structural similarity (*SSIM*) index, has been developed to consider location knowledge during a quality assessment experiment [79], which can be expressed as:

$$SSIM = L(GT, F) \cdot C(GT, F) \cdot S(GT, F) \tag{17}$$

Here, L, C and S correspond to luminance, contrast and structural similarity, respectively.

4.2. Experiment Results for EDoF Fraunhofer Dataset

Many end-to-end fusion models are limited to the fusion of only two images, which poses a challenge when dealing with multi-focus microscope images that consist of a series of z-slices. To ensure a fair evaluation of fusion model performance, we selected two microscope images from the EDoF Fraunhofer dataset. Since ZMFF is a zero-shot image fusion model and the model should be trained for each fusion step, we evaluated the results for different training epochs (600 epochs, 900 epochs and 1300 epochs). The comparative fusion examples generated by various models are presented in Figure 4. Significantly, the fused image generated by Fusion DN exhibits conspicuous errors along its edges, indicating a notable shortfall in edge fidelity. Furthermore, the fused images resulting from PMGI and U2Fusion display a discernible reduction in luminance, while PMGI manifests pronounced and undesirable shadow artifacts. In contrast, the images produced by MFF-GAN, Swin Fusion and MUFusion exhibit higher luminance when compared to both the ground truth generated from the complex wavelet extended-depth-of-field method and the source input images. Surprisingly, the ZMFF model achieved satisfactory performance for multi-focus microscope image fusion after 500 epochs of training without priors learned from large-scale datasets. Notably, IFCNN and U-Swin, each employing distinct fusion rules, excel in generating fusion images that are universally regarded as superior in terms of quality. This subjective assessment underscores their prominence in achieving the highest levels of image fusion quality.

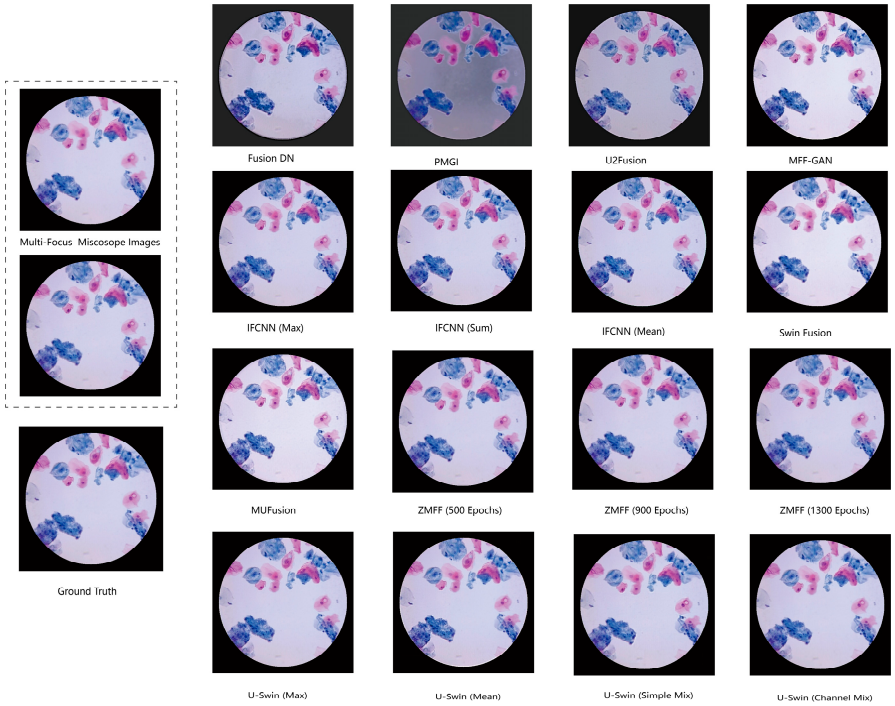

Figure 4. The comparison example of two selected multi-focus fused images from EDoF Fraunhofer dataset.

Table 1 presents the quantitative evaluation results for the EDoF Fraunhofer dataset. The values highlighted in bold font represent the best results in their respective evaluation metric columns. These values represent the mean and standard deviation of the evaluation metrics across the entire dataset. The statistics evaluation results listed in Table 1 clearly indicate that our proposed U-Swin fusion model, particularly when employing channel mix fusion, achieves the best performance in terms of mean squared error (*MSE*) and ranks second in structural similarity (*SSIM*). Notably, the fused images from MFF-GAN and Swin fusion exhibit relatively lower peak signal-to-noise ratios (*PSNR*s), which correlates with comparable performance in *SSIM*. In contrast, MUFusion cannot achieve optimal performance for multi-focus microscope image fusion. Surprisingly, the ZMFF model achieved comparable performance to U-Swin (mean) and IFCNN (mean) without prior training from large-scale datasets. However, Since ZMFF needs to be retrained every time image fusion is performed, this will take a lot of time. It is worth noting that the other fusion models, with the exception of IFCNN and ZMFF, struggle to generate satisfactory fusion images. This difficulty is often attributed to the characteristics of microscope images, which typically exhibit low contrast and repetitive structures that pose challenges for feature extraction. The majority of fusion models, with the notable exceptions of IFCNN and the U-Swin fusion model, conduct the fusion process within the Y component of the YCrCb color space. This approach, while effective for many types of images, may not be optimally suited for microscope images. The Y component primarily encodes the brightness information, which is important for overall image quality but may not adequately capture the nuances of microscope images. As a result, these models tend to yield less favorable results when assessed using metrics like mean squared error (*MSE*) and peak signal-to-noise ratio (*PSNR*). The limitations arise from the fact that microscope images often exhibit unique characteristics such as low contrast, repetitive structures and lower signal-to-noise ratios, which demand more sophisticated fusion techniques for accurate and high-quality results.

Table 1. Quantitative evaluation results for EDoF Fraunhofer dataset for all images. † means the evaluation results from our proposed model. The values highlighted in bold font represent the best results in their respective evaluation metric columns.

	MSE	**PSNR**	**SSIM**
FusionDH	1129.10 ± 515.22	18.07 ± 2.03	0.6034 ± 0.0115
PMGI	5185.17 ± 1817.41	11.31 ± 1.80	0.5086 ± 0.0359
U2Fusion	586.70 ± 204.62	20.73 ± 1.63	0.6543 ± 0.0053
MFF-GAN	137.00 ± 24.77	26.85 ± 0.90	0.9690 ± 0.0085
IFCNN (Max)	4.99 ± 0.81	41.21 ± 0.67	**0.9844 ± 0.0022**
IFCNN (Sum)	9.03 ± 1.41	38.62 ± 0.66	0.9779 ± 0.0027
IFCNN (Mean)	12.31 ± 1.93	37.28 ± 0.66	0.9768 ± 0.0030
Swin Fusion	173.41 ± 48.72	25.96 ± 1.48	0.9785 ± 0.0032
MUFusion	150.60 ± 44.95	26.53 ± 1.22	0.8366 ± 0.0068
ZMFF (600 Epochs)	24.89 ± 7.93	34.32 ± 1.08	0.9701 ± 0.0049
ZMFF (900 Epochs)	17.18 ± 5.43	35.97 ± 1.33	0.9749 ± 0.0037
ZMFF (1300 Epochs)	14.59 ±3.54	36.66 ± 1.34	0.9761 ± 0.0033
U-Swin (Max) †	5.36 ± 2.05	41.17 ± 1.76	0.9829 ± 0.0030
U-Swin (Mean) †	15.78 ± 3.60	36.26 ± 0.96	0.9651 ± 0.0070
U-Swin (Simple Mix) †	7.88 ± 2.67	39.44 ± 1.61	0.9745 ± 0.0046
U-Swin (Channel Mix) †	**2.74 ± 0.63**	**43.85 ± 0.93**	0.9839 ± 0.0020

Figure 5 provides a visual illustration of the comparison between multi-fused images for all slices within the EDoF Fraunhofer dataset. In cases where the fusion models are designed to accommodate only double-image fusion, generating a fully focused image involved a step-by-step process. It is worth noting that ZMFF cannot obtain effective results as the number of fused images increases. This may be due to the fact that for a series of multi-focus microscopic images, the multi-focus images have a high degree of similarity, which increases the difficulty of zero-shot model optimization. The results produced by the FusionDN, PMGI and U2Fusion methods exhibit a noticeable and substantial misalignment when compared against the ground truth. With an increase in the number of fused images, the cumulative fused error becomes more pronounced for models designed to support only the fusion of two images. As a result of these pronounced disparities, there appears to be no necessity for a quantitative evaluation of these models. The imagery generated by MFF-GAN and MUFusion prominently showcases synthetic textures that are readily discernible to the observer. Conversely, our proposed U-Swin fusion model has demonstrated exceptional performance, particularly for models utilizing channel mix fusion rules, as emphasized in Table 2.

Table 2. Quantitative evaluation results for EDoF Fraunhofer dataset for two-image fusion. † means the evaluation results from our proposed model. The values highlighted in bold font represent the best results in their respective evaluation metric columns.

	MSE	**PSNR**	**SSIM**
MFF-GAN	217.62 ± 60.80	24.95 ± 1.36	0.7340 ± 0.0284
IFCNN (Max)	7.15 ± 1.38	39.66 ± 0.80	**0.9772 ± 0.0030**
IFCNN (Mean)	12.711 ± 1.93	37.28 ± 0.66	0.9768 ± 0.0030
Swin Fusion	208.57 ± 50.52	25.10 ± 1.28	0.9660 ± 0.0051
MUFusion	345.51 ± 159.47	23.17 ± 1.91	0.7405 ± 0.0293
U-Swin (Max) †	8.60 ± 3.09	39.09 ± 1.70	0.9762 ± 0.0040
U-Swin (Mean) †	54.10 ± 8.09	30.84 ± 0.64	0.8911 ± 0.0132
U-Swin (Simple Mix) †	8.91 ± 2.73	38.84 ± 1.41	0.9708 ± 0.0053
U-Swin (Channel Mix) †	**6.28 ± 1.59**	**40.29 ± 1.07**	0.9649 ± 0.0036

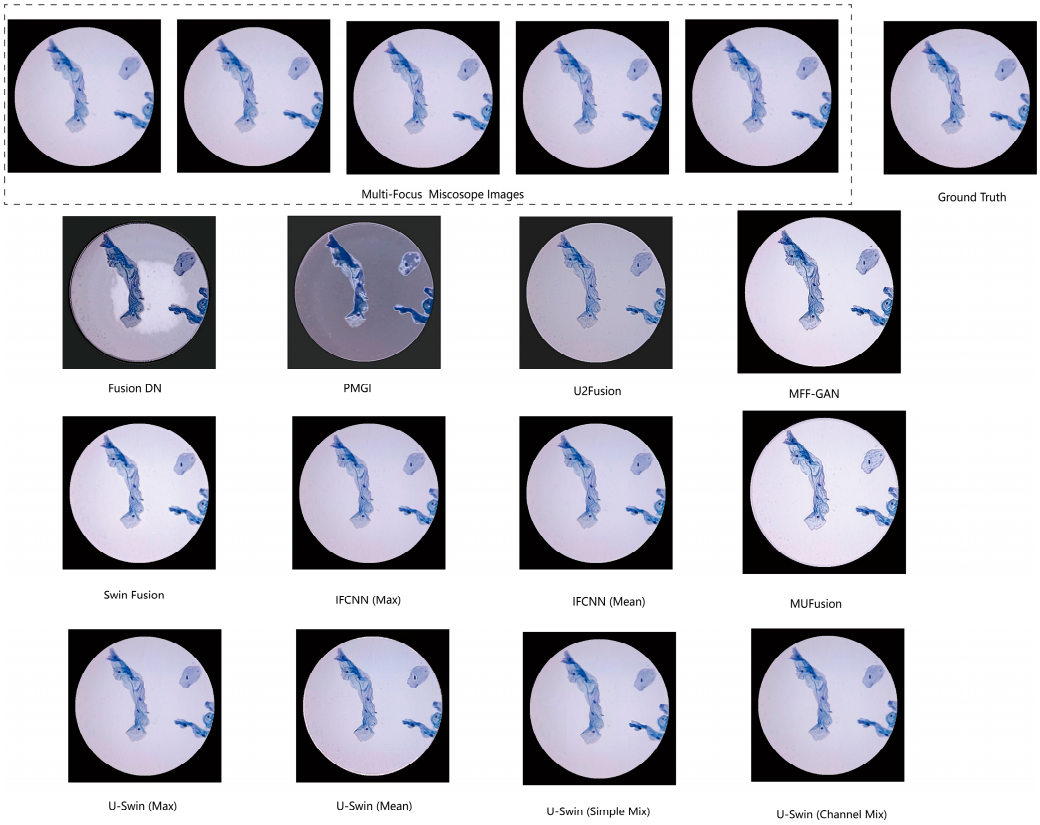

Figure 5. The comparison example of multi-focus fused images from EDoF Fraunhofer dataset.

4.3. Experiment Results for Various Color Spaces

In order to assess the effectiveness of our proposed model based on the Swin Transformer backbone, we explored various color spaces (RGB, YCrCb, HSV, LAB) for multi-focus microscope image fusion. We evaluated the performance of the fusion model for the two selected microscope images from the EDoF Fraunhofer dataset, as discussed in Section 4.2. In this section, we conducted qualitative and quantitative evaluations of the U-Swin fusion model with the channel mix fusion rule. The proposed model shown in previous sections fused the microscope images in RGB color spaces, treating the three channels of the multi-focus microscope images as the model inputs. Figure 6 and Table 3 demonstrate that our proposed model with the channel mix fusion rule could achieve satisfactory performance in various color spaces. The outcome suggests that the proposed model (U-Swin fusion model) achieves state-of-the-art (SOTA) results for the RGB color space and is applicable to other color spaces as well.

Table 3. Quantitative evaluation results for the U-Swin fusion model with channel mix fusion rule for various color spaces. The values highlighted in bold font represent the best results in their respective evaluation metric columns.

	MSE	**PSNR**	**SSIM**
RGB	**2.74 ± 0.63**	**43.85 ± 0.93**	**0.9839 ± 0.0020**
YCrCb	4.39 ± 0.66	41.75 ± 0.63	0.9526 ± 0.0025
HSV	6.40 ± 1.27	40.16 ± 0.86	0.9623 ± 0.0052
LAB	4.96 ± 0.78	41.23 ± 0.67	0.9628 ± 0.0021

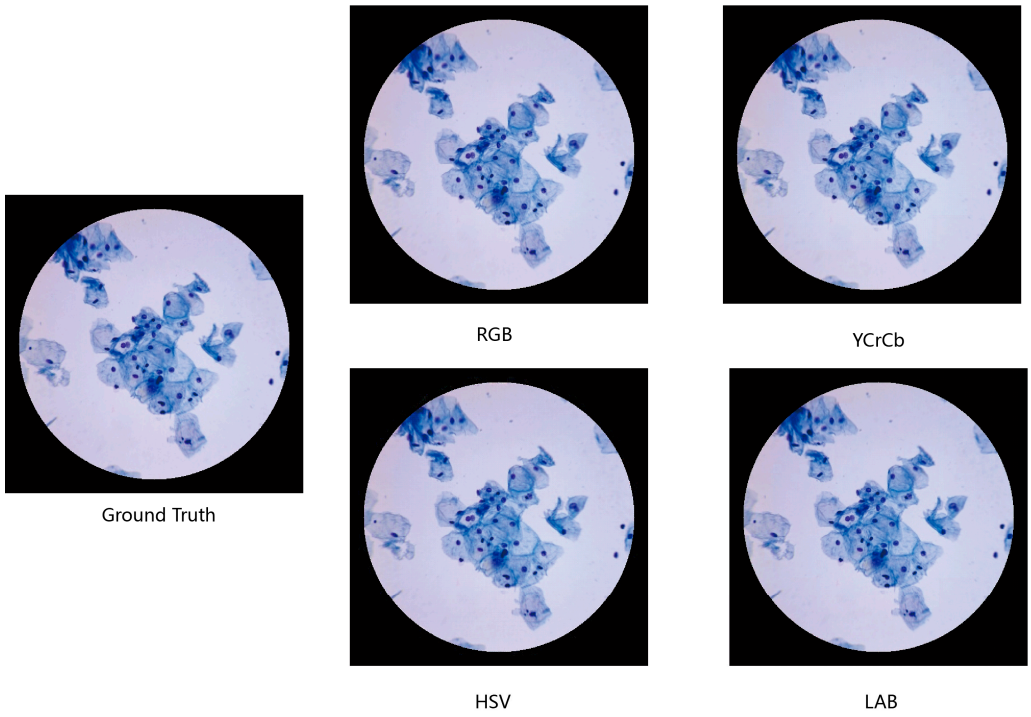

Figure 6. The comparison example of the U-Swin fusion model with channel mix fusion rule for various color spaces.

5. Conclusions

In this paper, we present a groundbreaking end-to-end microscope image fusion model, built on the Swin Transformer backbone, known as the U-Swin fusion model. Leveraging the innate capabilities of transformers to capture long-range dependency features, our innovative transformer-based models consistently produce fusion images that either match or surpass the state-of-the-art (SOTA) image fusion algorithms. Additionally, they exhibit remarkable generalization ability for fusing multi-focus microscope images. Notably, the pure transformer-based U-Swin fusion model, incorporating channel mix fusion rules, attains superior performance in numerous evaluation metrics compared to the majority of existing end-to-end fusion models. This work lays a pioneering foundation for applying transformer-based networks within the realm of microscope image fusion. Despite the extensive experimental results that validate the advantages of our proposed models, there remain several avenues for improvement in pursuit of even more robust image fusion models. For instance, exploring complex transformer backbones and fusion rules holds promise for further enhancing the model's performance in the future.

Author Contributions: Conceptualization, H.G. and H.S.; Methodology, H.H.X.; Formal analysis, H.H.X. and K.G.; Investigation, W.L.; Data curation, H.H.X.; Writing—original draft, H.H.X.; Writing—review & editing, H.G.; Funding acquisition, H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Key R&D Program of Zhejiang Province (2021C01016) and National Natural Science Foundation of China (61827805). And the APC was funded by Jiaxing Key Laboratory of Visual Big Data and Artificial Intelligence, Zhejiang Province, China.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Aguet, F.; Van De Ville, D.; Unser, M. Model-Based 2.5-D Deconvolution for Extended Depth of Field in Brightfield Microscopy. *IEEE Trans. Image Process.* **2008**, *17*, 1144–1153. [CrossRef] [PubMed]
2. Zhi-guo, J.; Dong-bing, H.; Jin, C.; Xiao-kuan, Z. A Wavelet Based Algorithm for Multi-Focus Micro-Image Fusion. In Proceedings of the Third International Conference on Image and Graphics (ICIG'04), Hong Kong, China, 18–20 December 2004; pp. 176–179. [CrossRef]
3. Sujatha, K.; Shalini Punithavathani, D. Optimized Ensemble Decision-Based Multi-Focus Imagefusion Using Binary Genetic Grey-Wolf Optimizer in Camera Sensor Networks. *Multimed. Tools Appl.* **2018**, *77*, 1735–1759. [CrossRef]
4. Chen, Z.; Wang, D.; Gong, S.; Zhao, F. Application of Multi-Focus Image Fusion in Visual Power Patrol Inspection. In Proceedings of the 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 25–26 March 2017; pp. 1688–1692. [CrossRef]
5. Song, Y.; Li, M.; Li, Q.; Sun, L. A New Wavelet Based Multi-Focus Image Fusion Scheme and Its Application on Optical Microscopy. In Proceedings of the 2006 IEEE International Conference on Robotics and Biomimetics, Kunming, China, 17–20 December 2006; pp. 401–405. [CrossRef]
6. Liu, Y.; Liu, S.; Wang, Z. Multi-Focus Image Fusion with Dense SIFT. *Inf. Fusion* **2015**, *23*, 139–155. [CrossRef]
7. De, I.; Chanda, B. Multi-Focus Image Fusion Using a Morphology-Based Focus Measure in a Quad-Tree Structure. *Inf. Fusion* **2013**, *14*, 136–146. [CrossRef]
8. Li, M.; Cai, W.; Tan, Z. A Region-Based Multi-Sensor Image Fusion Scheme Using Pulse-Coupled Neural Network. *Pattern Recognit. Lett.* **2006**, *27*, 1948–1956. [CrossRef]
9. Yang, Y.; Yang, M.; Huang, S.; Ding, M.; Sun, J. Robust Sparse Representation Combined with Adaptive PCNN for Multifocus Image Fusion. *IEEE Access* **2018**, *6*, 20138–20151. [CrossRef]
10. Zhang, Q.; Shi, T.; Wang, F.; Blum, R.S.; Han, J. Robust Sparse Representation Based Multi-Focus Image Fusion with Dictionary Construction and Local Spatial Consistency. *Pattern Recognit.* **2018**, *83*, 299–313. [CrossRef]
11. Zhang, Q.; Liu, Y.; Blum, R.S.; Han, J.; Tao, D. Sparse Representation Based Multi-Sensor Image Fusion for Multi-Focus and Multi-Modality Images: A Review. *Inf. Fusion* **2018**, *40*, 57–75. [CrossRef]
12. Amin-Naji, M.; Aghagolzadeh, A. Multi-Focus Image Fusion Using VOL and EOL in DCT Domain. *arXiv* **2017**, arXiv:1710.06511.
13. Amin-Naji, M.; Aghagolzadeh, A. Multi-Focus Image Fusion in DCT Domain Using Variance and Energy of Laplacian and Correlation Coefficient for Visual Sensor Networks. *J. AI Data Min.* **2018**, *6*, 233–250.
14. Kou, L.; Zhang, L.; Zhang, K.; Sun, J.; Han, Q.; Jin, Z. A Multi-Focus Image Fusion Method via Region Mosaicking on Laplacian Pyramids. *PLoS ONE* **2018**, *13*, e0191085. [CrossRef] [PubMed]
15. Wang, H. Multi-Focus Image Fusion Algorithm Based on Focus Detection in Spatial and NSCT Domain. *PLoS ONE* **2018**, *13*, e0204225. [CrossRef] [PubMed]
16. Bavirisetti, D.P.; Xiao, G.; Zhao, J.; Dhuli, R.; Liu, G. Multi-Scale Guided Image and Video Fusion: A Fast and Efficient Approach. *Circuits Syst. Signal Process.* **2019**, *38*, 5576–5605. [CrossRef]
17. Zhou, Z.; Li, S.; Wang, B. Multi-Scale Weighted Gradient-Based Fusion for Multi-Focus Images. *Inf. Fusion* **2014**, *20*, 60–72. [CrossRef]
18. Paul, S.; Sevcenco, I.S.; Agathoklis, P. Multi-Exposure and Multi-Focus Image Fusion in Gradient Domain. *J. Circuits Syst. Comput.* **2016**, *25*, 1650123. [CrossRef]
19. Liu, Y.; Liu, S.; Wang, Z. A General Framework for Image Fusion Based on Multi-Scale Transform and Sparse Representation. *Inf. Fusion* **2015**, *24*, 147–164. [CrossRef]
20. Li, H.; Nie, R.; Zhou, D.; Gou, X. Convolutional Neural Network Based Multi-Focus Image Fusion. In Proceedings of the 2nd International Conference on Algorithms, Computing and Systems, Beijing, China, 27–29 July 2018; pp. 148–154. [CrossRef]
21. Du, C.; Gao, S. Image Segmentation-Based Multi-Focus Image Fusion through Multi-Scale Convolutional Neural Network. *IEEE Access* **2017**, *5*, 15750–15761. [CrossRef]
22. Guo, X.; Nie, R.; Cao, J.; Zhou, D.; Mei, L.; He, K. FuseGAN: Learning to Fuse Multi-Focus Image via Conditional Generative Adversarial Network. *IEEE Trans. Multimed.* **2019**, *21*, 1982–1996. [CrossRef]
23. Amin-Naji, M.; Aghagolzadeh, A.; Ezoji, M. Ensemble of CNN for Multi-Focus Image Fusion. *Inf. Fusion* **2019**, *51*, 201–214. [CrossRef]
24. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
25. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image Is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:2010.11929.
26. Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. Learning Transferable Visual Models from Natural Language Supervision. In Proceedings of the International Conference on Machine Learning, Virtual, 18–24 July 2021; pp. 8748–8763.
27. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. Transunet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv* **2021**, arXiv:2102.04306.

28. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-End Object Detection with Transformers. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 213–229.
29. Touvron, H.; Cord, M.; Douze, M.; Massa, F.; Sablayrolles, A.; Jégou, H. Training Data-Efficient Image Transformers & Distillation through Attention. In Proceedings of the International Conference on Machine Learning, Virtual, 18–24 July 2021; pp. 10347–10357.
30. Bao, H.; Dong, L.; Piao, S.; Wei, F. Beit: Bert Pre-Training of Image Transformers. *arXiv* **2021**, arXiv:2106.08254.
31. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
32. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; pp. 205–218.
33. Liu, Y.; Wang, L.; Cheng, J.; Li, C.; Chen, X. Multi-Focus Image Fusion: A Survey of the State of the Art. *Inf. Fusion* **2020**, *64*, 71–91. [CrossRef]
34. Zhou, Y.; Yu, L.; Zhi, C.; Huang, C.; Wang, S.; Zhu, M.; Ke, Z.; Gao, Z.; Zhang, Y.; Fu, S. A Survey of Multi-Focus Image Fusion Methods. *Appl. Sci.* **2022**, *12*, 6281. [CrossRef]
35. Burt, P.J.; Adelson, E.H. Merging Images through Pattern Decomposition. In *Applications of Digital Image Processing VIII*; Optica Publishing Group: Washington, DC, USA, 1985; Volume 575, pp. 173–181. [CrossRef]
36. Li, H.; Manjunath, B.; Mitra, S.K. Multisensor Image Fusion Using the Wavelet Transform. *Graph. Models Image Process.* **1995**, *57*, 235–245. [CrossRef]
37. Tian, J.; Chen, L. Adaptive Multi-Focus Image Fusion Using a Wavelet-Based Statistical Sharpness Measure. *Signal Process.* **2012**, *92*, 2137–2146. [CrossRef]
38. Yang, B.; Li, S.; Sun, F. Image Fusion Using Nonsubsampled Contourlet Transform. In Proceedings of the Fourth International Conference on Image and Graphics (ICIG 2007), Chengdu, China, 22–24 August 2007; pp. 719–724. [CrossRef]
39. Li, X.; Li, H.; Yu, Z.; Kong, Y. Multifocus Image Fusion Scheme Based on the Multiscale Curvature in Nonsubsampled Contourlet Transform Domain. *Opt. Eng.* **2015**, *54*, 073115. [CrossRef]
40. Kong, W.; Lei, Y.; Zhao, R. Fusion Technique for Multi-Focus Images Based on NSCT–ISCM. *Optik* **2015**, *126*, 3185–3192. [CrossRef]
41. Zhao, H.; Shang, Z.; Tang, Y.Y.; Fang, B. Multi-Focus Image Fusion Based on the Neighbor Distance. *Pattern Recognit.* **2013**, *46*, 1002–1011. [CrossRef]
42. Zhou, F.; Li, X.; Li, J.; Wang, R.; Tan, H. Multifocus Image Fusion Based on Fast Guided Filter and Focus Pixels Detection. *IEEE Access* **2019**, *7*, 50780–50796. [CrossRef]
43. Chen, S.; Su, H.; Zhang, R.; Tian, J.; Yang, L. Improving Empirical Mode Decomposition Using Support Vector Machines for Multifocus Image Fusion. *Sensors* **2008**, *8*, 2500–2508. [CrossRef] [PubMed]
44. Yang, B.; Li, S. Multifocus Image Fusion and Restoration with Sparse Representation. *IEEE Trans. Instrum. Meas.* **2009**, *59*, 884–892. [CrossRef]
45. Chen, L.; Li, J.; Chen, C.P. Regional Multifocus Image Fusion Using Sparse Representation. *Opt. Express* **2013**, *21*, 5182–5197. [CrossRef] [PubMed]
46. Piella, G. Image Fusion for Enhanced Visualization: A Variational Approach. *Int. J. Comput. Vis.* **2009**, *83*, 1–11. [CrossRef]
47. Hong, R.; Wang, C.; Ge, Y.; Wang, M.; Wu, X.; Zhang, R. Salience Preserving Multi-Focus Image Fusion. In Proceedings of the 2007 IEEE International Conference on Multimedia and Expo, Beijing, China, 2–5 July 2007; pp. 1663–1666.
48. Mitianoudis, N.; Stathaki, T. Pixel-Based and Region-Based Image Fusion Schemes Using ICA Bases. *Inf. Fusion* **2007**, *8*, 131–142. [CrossRef]
49. Zhang, Y.; Chen, L.; Zhao, Z.; Jia, J. Multi-Focus Image Fusion Based on Cartoon-Texture Image Decomposition. *Optik* **2016**, *127*, 1291–1296. [CrossRef]
50. Li, S.; Yang, B. Multifocus Image Fusion by Combining Curvelet and Wavelet Transform. *Pattern Recognit. Lett.* **2008**, *29*, 1295–1301. [CrossRef]
51. Li, S.; Kwok, J.T.; Wang, Y. Combination of Images with Diverse Focuses Using the Spatial Frequency. *Inf. Fusion* **2001**, *2*, 169–176. [CrossRef]
52. Bai, X.; Zhang, Y.; Zhou, F.; Xue, B. Quadtree-Based Multi-Focus Image Fusion Using a Weighted Focus-Measure. *Inf. Fusion* **2015**, *22*, 105–118. [CrossRef]
53. Guo, D.; Yan, J.; Qu, X. High Quality Multi-Focus Image Fusion Using Self-Similarity and Depth Information. *Opt. Commun.* **2015**, *338*, 138–144. [CrossRef]
54. Huang, Y.; Li, W.; Gao, M.; Liu, Z. Algebraic Multi-Grid Based Multi-Focus Image Fusion Using Watershed Algorithm. *IEEE Access* **2018**, *6*, 47082–47091. [CrossRef]
55. Yang, B.; Li, S. Multi-Focus Image Fusion Based on Spatial Frequency and Morphological Operators. *Chin. Opt. Lett.* **2007**, *5*, 452–453.
56. Liu, Y.; Jin, J.; Wang, Q.; Shen, Y.; Dong, X. Novel Focus Region Detection Method for Multifocus Image Fusion Using Quaternion Wavelet. *J. Electron. Imaging* **2013**, *22*, 023017. [CrossRef]
57. Zhang, X. Deep Learning-Based Multi-Focus Image Fusion: A Survey and a Comparative Study. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 4819–4838. [CrossRef] [PubMed]

58. Liu, Y.; Chen, X.; Peng, H.; Wang, Z. Multi-Focus Image Fusion with a Deep Convolutional Neural Network. *Inf. Fusion* **2017**, *36*, 191–207. [CrossRef]
59. Peña, F.A.G.; Fernández, P.D.M.; Ren, T.I.; Vasconcelos, G.C.; Cunha, A. A Multiple Source Hourglass Deep Network for Multi-Focus Image Fusion. *arXiv* **2019**, arXiv:1908.10945.
60. Ma, H.; Liao, Q.; Zhang, J.; Liu, S.; Xue, J.-H. An α-Matte Boundary Defocus Model-Based Cascaded Network for Multi-Focus Image Fusion. *IEEE Trans. Image Process.* **2020**, *29*, 8668–8679. [CrossRef] [PubMed]
61. Guo, X.; Meng, L.; Mei, L.; Weng, Y.; Tong, H. Multi-Focus Image Fusion with Siamese Self-Attention Network. *IET Image Process.* **2020**, *14*, 1339–1346. [CrossRef]
62. Zhang, Y.; Liu, Y.; Sun, P.; Yan, H.; Zhao, X.; Zhang, L. IFCNN: A General Image Fusion Framework Based on Convolutional Neural Network. *Inf. Fusion* **2020**, *54*, 99–118. [CrossRef]
63. Xu, H.; Ma, J.; Jiang, J.; Guo, X.; Ling, H. U2Fusion: A Unified Unsupervised Image Fusion Network. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 502–518. [CrossRef]
64. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2018**, arXiv:1810.04805.
65. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18. pp. 234–241.
66. Cimpoi, M.; Maji, S.; Kokkinos, I.; Mohamed, S.; Vedaldi, A. Describing Textures in the Wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3606–3613.
67. Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Li, F.-F. Imagenet: A Large-Scale Hierarchical Image Database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255. [CrossRef]
68. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
69. Albuquerque, T.; Rosado, L.; Cruz, R.; Vasconcelos, M.J.M.; Oliveira, T.; Cardoso, J.S. Rethinking Low-Cost Microscopy Workflow: Image Enhancement Using Deep Based Extended Depth of Field Methods. *Intell. Syst. Appl.* **2023**, *17*, 200170. [CrossRef]
70. Xu, H.; Ma, J.; Le, Z.; Jiang, J.; Guo, X. Fusiondn: A Unified Densely Connected Network for Image Fusion. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 12484–12491.
71. Zhang, H.; Le, Z.; Shao, Z.; Xu, H.; Ma, J. MFF-GAN: An Unsupervised Generative Adversarial Network with Adaptive and Gradient Joint Constraints for Multi-Focus Image Fusion. *Inf. Fusion* **2021**, *66*, 40–53. [CrossRef]
72. Zhang, H.; Xu, H.; Xiao, Y.; Guo, X.; Ma, J. Rethinking the Image Fusion: A Fast Unified Image Fusion Network Based on Proportional Maintenance of Gradient and Intensity. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 12797–12804.
73. Ma, J.; Tang, L.; Fan, F.; Huang, J.; Mei, X.; Ma, Y. SwinFusion: Cross-Domain Long-Range Learning for General Image Fusion via Swin Transformer. *IEEE/CAA J. Autom. Sin.* **2022**, *9*, 1200–1217. [CrossRef]
74. Cheng, C.; Xu, T.; Wu, X.-J. MUFusion: A General Unsupervised Image Fusion Network Based on Memory Unit. *Inf. Fusion* **2023**, *92*, 80–92. [CrossRef]
75. Hu, X.; Jiang, J.; Liu, X.; Ma, J. ZMFF: Zero-Shot Multi-Focus Image Fusion. *Inf. Fusion* **2023**, *92*, 127–138. [CrossRef]
76. Forster, B.; Van De Ville, D.; Berent, J.; Sage, D.; Unser, M. Complex Wavelets for Extended Depth-of-Field: A New Method for the Fusion of Multichannel Microscopy Images. *Microsc. Res. Tech.* **2004**, *65*, 33–42. [CrossRef]
77. Schneider, C.A.; Rasband, W.S.; Eliceiri, K.W. NIH Image to ImageJ: 25 Years of Image Analysis. *Nat. Methods* **2012**, *9*, 671–675. [CrossRef]
78. Wang, Z.; Bovik, A.C. *Modern Image Quality Assessment*; Springer: Berlin/Heidelberg, Germany, 2006.
79. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]