

Article

A Multi-Branch Multi-Scale Deep Learning Image Fusion Algorithm Based on DenseNet

Yumin Dong ^{1,*}, Zhengquan Chen ^{1,†}, Ziyi Li ¹ and Feng Gao ²¹ School of Computer and Information Science, Chongqing Normal University, Chongqing 401331, China² Science School, Qingdao Technological University, Qingdao 266525, China

* Correspondence: dym@cqnu.edu.cn

† These authors contributed equally to this work.

Abstract: Infrared images have good anti-environmental interference ability and can capture hot target information well, but their pictures lack rich detailed texture information and poor contrast. Visible image has clear and detailed texture information, but their imaging process depends more on the environment, and the quality of the environment determines the quality of the visible image. This paper presents an infrared image and visual image fusion algorithm based on deep learning. Two identical feature extractors are used to extract the features of visible and infrared images of different scales, fuse these features through specific fusion methods, and restore the features of visible and infrared images to the pictures through the feature restorer to make up for the deficiencies in the various photos of infrared and visible images. This paper tests infrared visual images, multi-focus images, and other data sets. The traditional image fusion algorithm is compared several with the current advanced image fusion algorithm. The experimental results show that the image fusion method proposed in this paper can keep more feature information of the source image in the fused image, and achieve excellent results in some image evaluation indexes.

Keywords: deep learning; image fusion; multi-scale image

Citation: Dong, Y.; Chen, Z.; Li, Z.; Gao, F. A Multi-Branch Multi-Scale Deep Learning Image Fusion Algorithm Based on DenseNet. *Appl. Sci.* **2022**, *12*, 10989. <https://doi.org/10.3390/app122110989>

Academic Editor: Manuel Graña

Received: 24 August 2022

Accepted: 27 October 2022

Published: 30 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the development of science and technology, more and more data are collected using sensors, and more and more information is obtained from them. From black and white images and color images taken by optical cameras at the beginning to hyperspectral, infrared images, and radar images taken by various sensors today. With the increase in image types, the use of images is also increasing. However, these images are generated by a single sensor, which has some shortcomings. A single sensor has limitations, so it is difficult to collect all the information about the scene, which leads to no image containing all the information about the scene. Image fusion technology is to fuse the image information collected by different sensors into a new image by some means. The new image contains both the information of the original image and reduces redundant information between many images, which improves the use rate of the image.

Traditional image fusion methods can extract image features and fuse them well, but these algorithms have defects, resulting in noise and poor image quality in the fused image. The appearance of deep learning brings a new research direction for image fusion algorithms. A large number of image fusion algorithms based on deep learning have emerged. The convolutional neural network has a good ability to extract image features. However, the traditional convolutional neural network entering the bottleneck period, researchers have gradually abandoned the convolutional neural network and started to study Transformer. However, some scholars are still studying convolutional neural networks, and proposed a pure convolutional neural network ConeNeXt [1] in 2022. Fewer activation functions and larger convolution kernels are used in ConeNeXt. Although

ConvNeXt does not propose new structures and methods, ConvNeXt reduces the use of activation functions, making ConvNeXt have faster reasoning speed and higher accuracy than Swin Transformer [1]. There is no doubt that convolution neural networks have a good ability to extract image features. Researchers use pre-trained AlexNet [2], VGGNet [3], GoogleNet [4,5], ResNet [6], DenseNet [7], CNN [8], etc. to extract image depth features and restore them to images after fusion. Besides being a tool for extracting image features, neural networks can also be used in end-to-end image fusion. The representative network models are FusionGAN [9], IFCNN [10], PPTFusion [11].

In the self-coding network framework, the network is composed of an encoder and a decoder. The encoder is used to extract image features, and the decoder is used to restore features to images. The encoder and decoder are independent, so the network structure of the self-coding network is very flexible. Because the self-coding network fusion framework is flexible and extensible, a large number of fusion algorithms based on the self-coding networks have been produced. In 2018, researchers proposed the first image fusion algorithm based on self coding network and named it DenseFuse [12,13]. NestFuse [14,15] was proposed in 2020, etc.

This paper uses the idea from the Inception module in GoogleNet for reference to build an encoder to extract features from the input images. The basic structure of the Inception module is shown in the Figure 1. There are three important parts to the Inception module. Firstly, 1×1 convolution kernel for lifting dimensions to reduce the calculation of subsequent feature maps. Secondly, the convolution kernel size and output feature map size of each branch is different, which ensures the multi-scale of the image. Finally, the feature maps of each branch are spliced to obtain all feature maps.

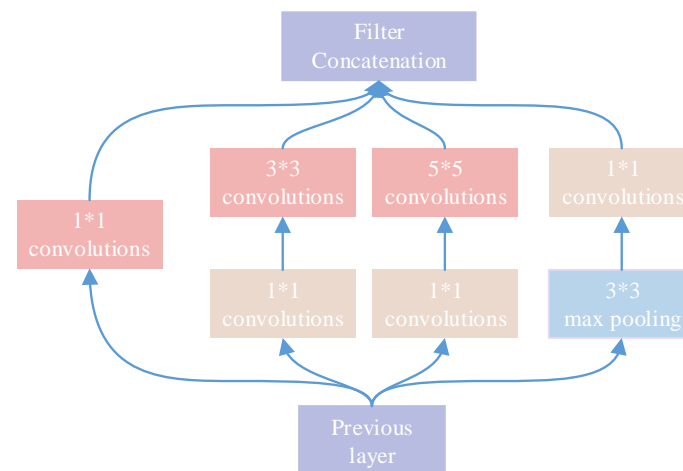


Figure 1. The Architecture of Inception module

This article mainly has the following work and contributions:

- The Inception module is added to the network model to increase the feature extraction ability of the network.
- Dense blocks are added to the branches of the network, and dense modules are used for feature extraction and image generation.
- The use of activation functions is reduced, and activation functions are used only after the first convolution.

2. The Architecture of Network

The network model proposed in this paper is mainly composed of three parts: encoder, fusion strategy, and decoder, as shown in Figure 2.

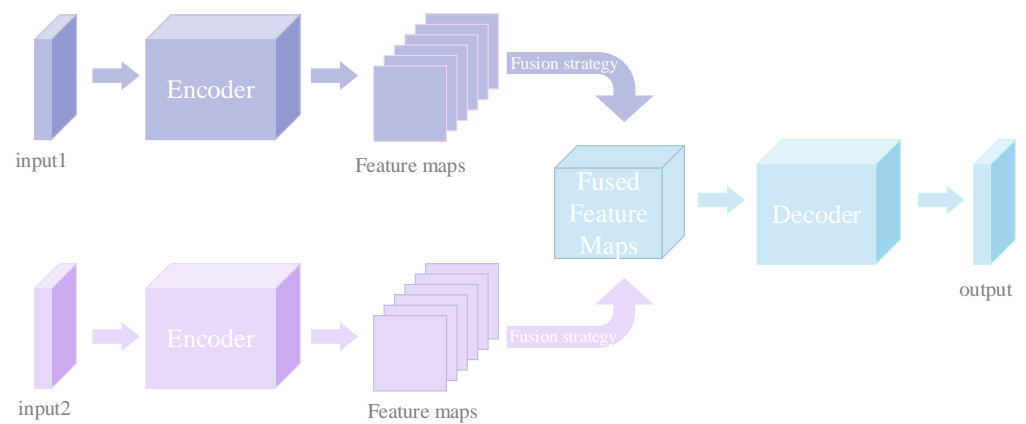


Figure 2. The Architecture of Network

2.1. The Encoder Architecture of Network

The encoder network is used to extract network features. Branch 1 uses a larger convolutional kernel to perform preliminary feature extraction of the input image, and then uses the Inception module to extract the image for multi-scale feature extraction, and then again through the smaller convolutional kernel to obtain feature 1. Branch 2 is structured similarly to Branch 1, but Branch 2 uses two convolution operations and then feeds features into the Inception module to get Feature 2. Branch 3 is similar to the structure of branch 2, convolution is performed again on the structure of branch 2 to obtain feature 3, branch 4 uses dense connections to extract features from the image, and feature 4 is obtained. Splicing features 1, 2, 3, and 4 on dim = 1 so that the resulting features contain the features of the previous 4 branch features. The encoder network is shown in Figure 3. A detailed table of encoder parameters, as shown in Table 1.

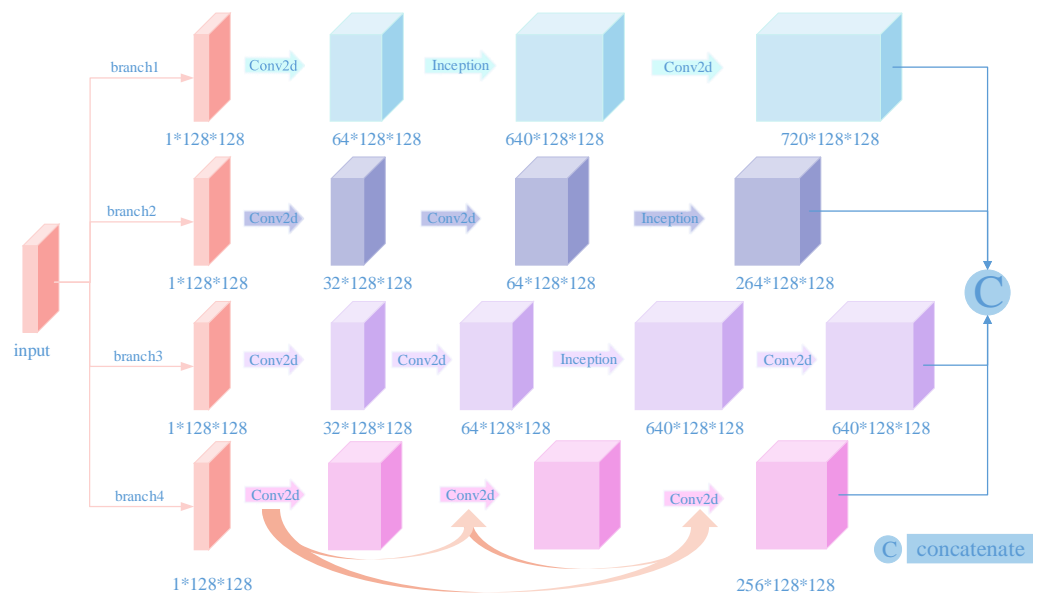


Figure 3. The Encoder Architecture of Network.

Table 1. The Encoder Architecture of Network.

Branchs	Layers	Kernel_Size	Stride	Padding	Activation
branch 1	layer 1	7 * 7	1	3	Gelu
	layer 2	Inception	-	-	-
	layer 3	3 * 3	1	1	-
branch 2	layer 1	5 * 5	1	2	Relu
	layer 2	1 * 1	1	0	-
	layer 3	Inception	-	-	-
branch 3	layer 1	3 * 3	1	1	Relu
	layer 2	1 * 1	1	0	-
	layer 3	Inception	-	-	-
	layer 4	3 * 3	1	1	-
branch 4	Dense Blocks	-	-	-	-

2.2. The Decoder the Encoder Architecture of Network

The architecture of the decoder is shown in Figure 4. In the design of the decoder, we did not use multiple 3 * 3 convolutional kernels like other image fusion algorithms for multiple feature channel operations, but also added an Inception module on the decoder to reduce the parameters of the network at the same time, as far as possible to retain more features, after the Inception module added a dense connection module, the use of dense connection module can better extract features, this article here will be used to reduce the characteristic channel, achieved good results. Finally, the decoder network reduces the number of feature channels to 1 through a 3 * 3 convolutional kernel, and then connects a Sigmoid activation function to restore it to an image. The specific network parameters are shown in Table 2.

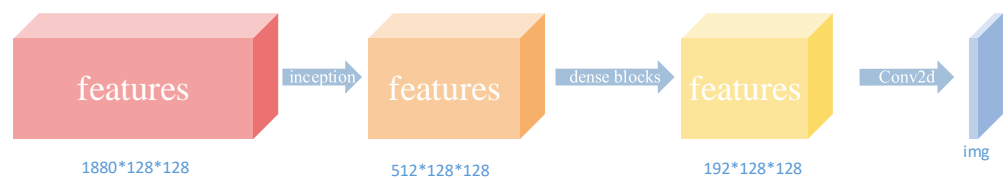


Figure 4. The Decoder Architecture of Network.

Table 2. The Decoder Architecture of Network.

Layers	Kernel_Size	Stride	Padding	Activation
layer1	Inception	-	-	-
layer2	Dense Blocks	-	-	-
layer3	3 * 3	1	1	Sigmoid

2.3. The Loss Function of Network Model

This paper uses the improved SSIM [16,17] function as the loss function of the network, which is mainly used to calculate the structural similarity between images. Self-coding network, the most important feature is an encoder to extract image features, and a decoder to restore the feature map to an image. In the training phase, the main task is to enable the encoder to extract features as much as possible and make the image restored by the decoder close to the source image. Therefore, it is very effective to use SSIM to calculate the error between the source image and the image restored by the decoder. its calculation formula is:

$$SSIM(img_a, img_b) = \frac{2 * \bar{X}_{img_a} * \bar{X}_{img_b} + c_1}{\bar{X}_{img_a}^2 + \bar{X}_{img_b}^2 + c_1} * \frac{2\sigma_{img_a * img_b} + c_2}{\sigma_{img_a}^2 + \sigma_{img_b}^2 + c_2} \tag{1}$$

where \bar{X} represents the average value, σ represents the standard deviation, and the value of SSIM is in $[0, 1]$. the value of SSIM closer the value is to 1, the higher the structural similarity of image A and image B is, and vice versa. So the Loss function is designed as follows:

$$Loss = 1 - SSIM(img_a, img_b) \quad (2)$$

2.4. The Fusion Strategy of Network Model

This paper mainly uses the strategy of averaging the characteristics, and the calculation formula is as follows:

$$F_{img_fuseed}(x, y) = \frac{F_{img_a}(x, y) + F_{img_b}(x, y)}{2} \quad (3)$$

where F_{img_fuseed} represents the fused image, F_{img_a} , F_{img_b} represents the source image a , b , (x, y) represents the corresponding pixel position of the image, and the value of (x, y) depends on the size of the input image.

3. Experiments and Results

This paper uses 80,000 images from the MSCOCO dataset [18], 19-to-multi-exposure images from the Exposure dataset [19,20], 50 pairs of images from the Road dataset [21], and 21 pairs of images from the TNO dataset [18] as training sets and test datasets for the network.

The entire experiment was conducted in an environment of: CPU: AMD R7 1700 and GPU: NVIDIA RTX3060, Memory 32 GB, Pytorch 1.10.1+cu113.

3.1. Training Network

In the training stage, this paper ignores the fusion strategy, mainly using the encoder to extract image features, using the decoder to restore the image. In this paper, the Adam optimizer is selected as the optimizer of the network, and the batchsize = 6 of the training, the image size is $128 * 128$, the learning rate is 0.0001, and the number of iterations is 20 times, The MSCOCO dataset is divided into 5000 images, 20,000 images, and 60,000 image samples, and the network model is trained to obtain the model.

3.2. Image Fusion

In the image fusion stage, two identical encoders extract two different source images to get two feature maps. The two feature maps are merged into one feature map through the fusion strategy. The decoder is used to restore the feature map to an image and output $320 * 320$ fused images.

3.3. Evaluation of Experimental Results and Image Quality

Common ways to evaluate images are Entropy (EN) [22], Mutual information (MI) [23], Structural similarity (SSIM) [24], Multi-scale SSIM [25], Visual information fidelity (VIF) [26], Spatial Frequency (SF) [27], Image Quality (Quality, $Q^{ab/f}$) [28], Noise ($N^{ab/f}$) [29], Definition (DF) [30], Standard Deviation (SD). Except for the image noise method, the lower the value obtained, the better, the higher the value obtained by other methods.

3.4. The Road Dataset Experiments

This paper uses the above method to evaluate the image generated by the fusion of the model proposed in this paper. At present, the more advanced algorithms in the field of image fusion, IFCNN, DenseFuse [12] CBF [31], CNN [8,32], DeepDecFusion [23], MEF-GAN [33], FusionGAN [9], DualBranchFusion [14], are compared, and the experimental results are shown in Table 3:

Table 3. Objectively evaluate the classical and latest fusion algorithms.

Model	SF	EN	Q^{ablf}	SSIM	MS-SSIM	N^{ablf}	MI	VIF	SD	DF
CNN	15.02189	7.27064	0.57479	0.68032	0.91450	0.02794	14.54127	0.76436	44.95776	6.93236
IFCNN	15.06840	6.97300	0.51500	0.70457	0.87985	0.03152	13.94600	0.62485	35.81602	7.04094
MEFGAN	11.91592	7.16128	0.19600	0.47106	0.55244	0.08239	14.32256	0.60731	69.20427	5.17243
Densefuse	9.58711	6.69415	0.35699	0.72426	0.85074	0.00183	13.38829	0.35059	30.82709	4.62602
FusionGAN	8.63996	7.17533	0.27373	0.61422	0.73517	0.01682	14.35067	0.42558	42.30396	3.92426
DeepDec-Fusion	10.78846	6.75995	0.38301	0.69378	0.82358	0.00945	13.51990	0.38246	31.83394	4.81932
Dual-Branch-Fusion	28.80572	7.08938	0.34721	0.59138	0.66963	0.10966	14.17876	0.51860	44.41719	11.78575
Ours	15.35358	7.31037	0.44858	0.66609	0.91937	0.16426	14.62074	0.72256	47.34091	7.20294

The red number represents the best, and the blue number represents the second best.

From Table 3, it can be seen that the network model proposed in this paper is the best in 3 of 10 evaluation indexes. Another 4 indicators are suboptimal. It can be seen from Figure 5 that the part of the red frame in the Infrared Image image cannot see the texture of the wall, but the texture of the image can be clearly seen in the Visible Image. Through the experimental comparison, it is found that in addition to the overexposure of the images obtained by the MEF-GAN model, the proposed model, and other comparable models can retain more texture information of the background wall. However, the brightness of the images after the fusion of the DeepDecFusion and Densefuse models is not as high as that of the proposed model. In addition, in the blue frame, the model proposed in this paper retains the contour information of flowerbed plants to the greatest extent.

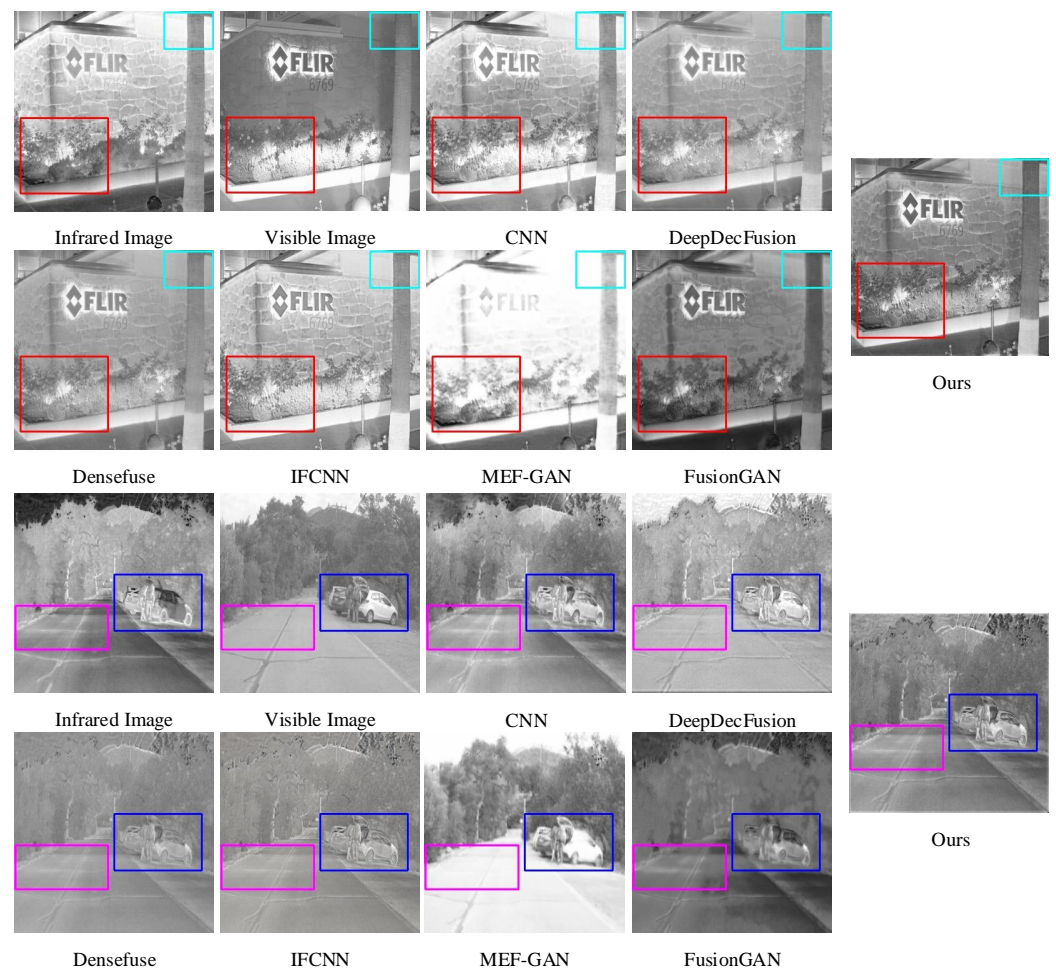


Figure 5. The Result of Road Datasets.

Besides, in the blue rectangle, the car outline in DeepDecFusion and FusionGAN is not clear, and the person's outline is not clear. Densefuse, IFCNN, CNN, DeepDecFusion, and Ours have clear car outlines. In the purple rectangle, the Infrared image can see a crack. In contrast, in the Visible image it is almost invisible, MEF-GAN, closer to the Visible image, Densefuse, DeepDecFusion, IFCNN, and FusionGAN can see the crack inconspicuously, but CNN and Ours can clearly see the crack.

In summary, the network model proposed in this paper retains the various information of Infrared Images and Visible Images to a certain extent.

3.5. The Other Experiments

3.5.1. TNO Dataset

In the infrared and visible image fusion dataset, there are 21 pairs of different infrared and visible image images in TNO dataset. In this paper, these images are converted into grayscale images for experiments and comparison. The specific experimental results are as follows:

The red font in Table 4 indicates that the network model proposed in this paper has the highest score in 4 of the 7 evaluation indicators. In Figure 6, the clouds in the sky cannot be seen in the red box of the visible light image, but the trees in the white box can be clearly seen. In the red box of the infrared image, you can see many clouds and contour information, but you can't see trees. Other images are obtained by various image fusion algorithms. There are many noises in CBF. IFCNN and DenseFuse retain more details of trees and clouds. The images fused by the network model proposed in this paper retain both the cloud layer information in the red box of the infrared image and the tree information in the white box of the visible image, and there is no noise in the fused image to affect the quality of the fused image.

Table 4. The result of TNO datasets.

Model	SF	EN	MI	VIF	SD	DF	SSIM
Densefuse	5.79225	6.17638	12.35276	0.28447	22.55032	2.78489	0.74928
CBF	13.59145	6.85749	13.71498	0.71849	35.91254	6.78595	0.59957
IFCNN	11.49526	6.59729	13.19457	0.59228	31.61534	5.79491	0.73158
Ours	11.12396	7.08364	14.16728	0.83255	39.79600	5.42090	0.70152

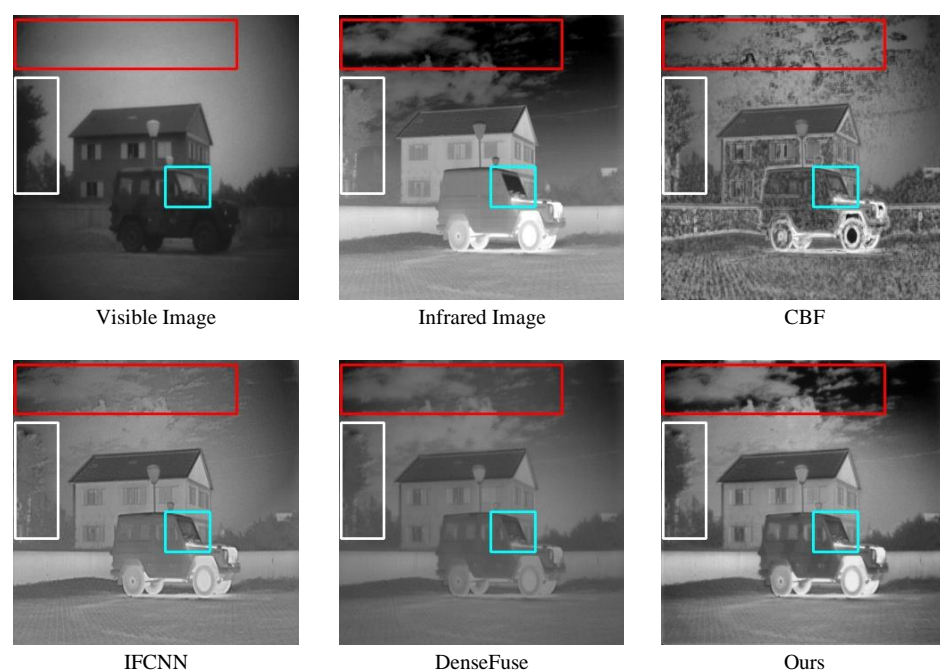


Figure 6. The Result of TNO Datasets.

Therefore, the network model proposed in this paper can better retain more information in infrared images and infrared images.

Besides, to prove the effectiveness of the proposed network model, we also fuse and test other images, such as images with many exposures and images with different focuses. The specific experimental results are as follows:

3.5.2. Exposure Dataset

In multiple Exposure datasets, Exposure dataset has a total of 19 color images with different exposures. In this paper, these color images are converted into grayscale images for experiments and comparison. The specific experimental results are as follows:

The red font in Table 5 indicates that this indicator has the highest score. The network model proposed in this paper has the highest score in 4 of the 7 evaluation indicators. In Figure 7, the background in the blue box of image A is under-exposed, while the background in the red box is under-exposed. In general, image A is under-exposed. The background in the blue box of image B is exposed, while the red box is over-exposed. In general, image B is over-exposed. Other images are obtained through various image fusion algorithms. CNN, Densefuse, and IFCNN images are over-exposed. However, the images fused by the network model proposed in this paper can be seen in the blue box, and also conform to the distribution of light sources. In the red box, you can see the grid under the light without over-exposure to the light effect.

In conclusion, the network model proposed in this paper can better keep more information in the under-exposed and over-exposed images. Thus, the model proposed in this paper is effective and has achieved good results in other areas.

Table 5. The Result of Exposure Datasets.

Model	SF	EN	MI	VIF	SD	DF	SSIM
Densefuse	14.34006	6.70196	13.40393	1.23703	38.99311	5.83722	0.59582
CNN	23.24008	6.52622	13.05245	1.81842	48.16440	9.65854	0.60547
IFCNN	25.97269	6.82564	13.65127	2.18844	47.34595	10.77438	0.59788
Ours	24.03095	7.09933	14.19865	2.33959	53.42360	10.29697	0.54275

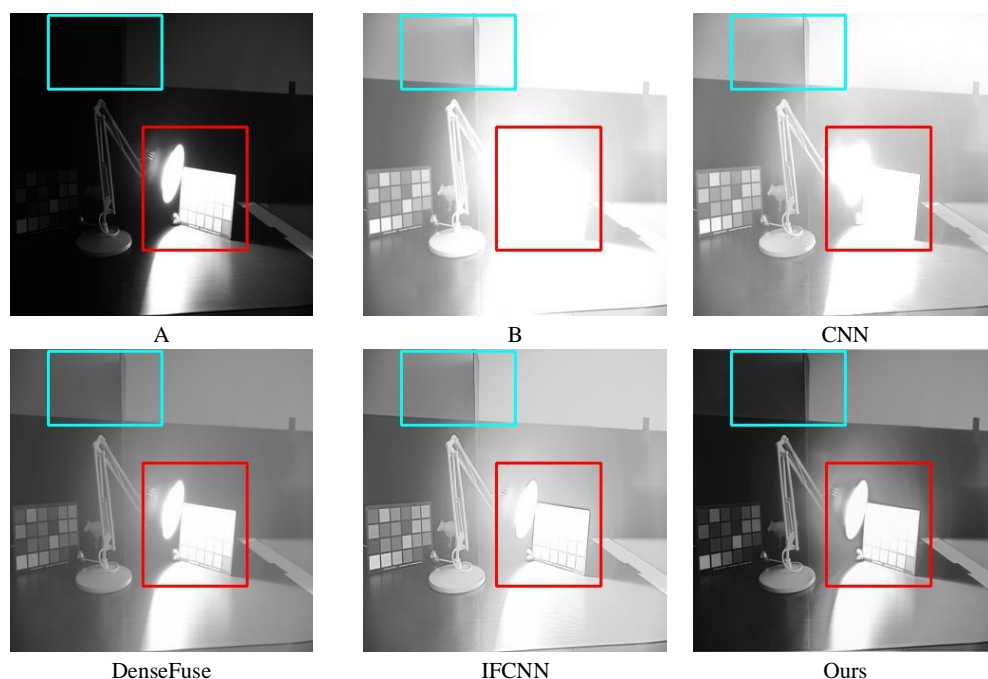


Figure 7. The Result of Exposure Datasets.

3.6. Ablation Experiments

The ablation experiment part of this paper is mainly to replace the Inception module and Denseblocks module on each branch of this article with a convolutional kernel with a convolutional kernel size of 3 * 3, and the other parts are not different from this network model.

The red numbers in Table 6 indicate the optimal in this column. The value of gray background is the score of the network model proposed in this paper on the datasets. The values without background are the scores of ablation experiments on the datasets. With a total of 30 indicators on all data, our model received 24 highest scores. All of the modules in the network model in this article, as well as the Denseblocks, are useful and work well on most of the datasets.

Table 6. The result of ablation experiments.

Model	SF	EN	Q^{ablf}	SSIM	MS-SSIM	N^{ablf}	MI	VIF	SD	DF
Exposure	23.57095	6.82891	0.39976	0.52032	0.83796	0.11205	13.65782	2.76957	55.54952	9.43259
Exposure ¹	24.03095	7.09933	0.54653	0.54275	0.91675	0.06365	14.19865	2.33959	53.42360	10.29697
Road	11.44829	6.88796	0.39165	0.69868	0.85715	0.00744	13.77593	0.43175	34.13524	5.24988
Road ¹	15.35358	7.31037	0.44858	0.66609	0.91937	0.16426	14.62074	0.72256	47.34091	7.20294
Tno	8.57529	6.32894	0.34281	0.71874	0.85294	0.00911	12.65789	0.35355	27.28184	3.89286
Tno ¹	11.12396	7.08364	0.45683	0.70152	0.93059	0.16616	14.16728	0.83255	39.79600	5.42090

¹ The network model proposed in this paper.

4. Conclusions

Deep learning has achieved good results in many fields; this paper organically combines deep learning with image fusion, and proposes an image fusion algorithm based on deep learning. Tables 4–6 show that our model has achieved good results on various datasets.

The multi-branch, multi-scale deep learning image network model proposed in this paper extracts image features by adding multiple branches and introducing the Dense-Blocks module in the design of the encoder. The activation function is not used after each convolution in the entire network, and in the selection of activation functions, this article does not use all of them one activation function but uses multiple activation functions. In the design of the loss function, we use the SSIM values between images as a loss function for the network. In the design of the decoder, we use Dense Blocks for dimensionality reduction processing of images. The model proposed in this paper has largely preserved the original image after the fusion and is very realistic and natural. Experiments have proved that the network model proposed in this paper has achieved the best results on most of the datasets.

In the future, we will try more methods of image fusion and study how to reduce the super parameters of the network. Following this, we will improve the network efficiency and strive to solve more problems in the field of image fusion.

Author Contributions: Conceptualization, Z.C.; methodology, Z.C.; software, Z.C.; validation, Y.D., Z.C., F.G. and Z.L.; formal analysis, Z.C.; data curation, Z.C. and Z.L.; writing—original draft preparation, Z.C.; writing—review and editing, Z.C. and Y.D.; visualization, Z.C. and Z.L.; supervision, Y.D.; funding acquisition, Y.D. and F.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China grant No. 61772295, 61572270, the PHD foundation of Chongqing Normal University (No. 19XLB003), the Science and Technology Research Program of Chongqing Municipal Education Commission (KJZD.M202-000501), Chongqing Technology Innovation and Application Development Special General Project (cstc-2020jscxlyjsAX0002) and Chongqing Technology Foresight and Institutional Innovation Project (cstc2021-jsyj-yzys-bAX0011).

Data Availability Statement: The data we use can be obtained from GitHub, https://github.com/thfylsty/ImageFusion_DeepDecFusion/tree/main/testdata, Our code: <https://github.com/Cevey/MDPI-Applsci>.

Conflicts of Interest: We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and company that could be construed as influencing the position presented in the manuscript entitled “Multi-branch Multi-scale Deep Learning Image Fusion Algorithm Based On DenseNet”.

Abbreviations

The following abbreviations are used in this manuscript:

EN	Entropy
MI	Mutual information
SSIM	Structural similarity
MS-SSIM	Multi-scale SSIM
VIF	Visual information fidelity
SF	Spatial Frequency
$Q^{ab/f}$	Image Quality
$N^{ab/f}$	Noise
DF	Definition
SD	Standard Deviation.

References

- Liu, Z.; Mao, H.; Wu, C.Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A ConvNet for the 2020s. *arXiv* **2022**, arXiv:2201.03545.
- Krizhevsky, A.; Sutskever, I.; Hinton, G. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*; Curran Associates, Inc.: Red Hook, NY, USA, 2012.
- Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556
- Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Rabinovich, A. *Going Deeper with Convolutions*; IEEE Computer Society: Washington, DC, USA, 2014.
- Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016.
- He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
- Iandola, F.; Moskewicz, M.; Karayev, S.; Girshick, R.; Keutzer, K. Densenet: Implementing efficient convnet descriptor pyramids. *arXiv* **2014**, arXiv:1404.1869.
- 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](#)]
- Ma, J.; Wei, Y.; Liang, P.; Chang, L.; Jiang, J. FusionGAN: A generative adversarial network for infrared and visible image fusion. *Inf. Fusion* **2019**, *48*, 11–26. [[CrossRef](#)]
- 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](#)]
- Fu, Y.; Xu, T.; Wu, X.; Kittler, J. PPT Fusion: Pyramid Patch Transformer for a Case Study in Image Fusion. *arXiv* **2021**, arXiv:2107.13967.
- Hui, L.; Wu, X.J. Densefuse: A fusion approach to infrared and visible images. *IEEE Trans. Image Process.* **2018**, *28*, 2614–2623.
- Hui, L.; Wu, X.J.; Kittler, J. Infrared and visible image fusion using a deep learning framework. In Proceedings of the International Conference on Pattern Recognition 2018, Beijing, China, 18–20 August 2018.
- Fu, Y.; Wu, X.J. A Dual-branch Network for Infrared and Visible Image Fusion. *arXiv* **2021**, arXiv:2101.09643.
- Li, H.; Wu, X.J.; Durrani, T. Nestfuse: An infrared and visible image fusion architecture based on nest connection and spatial/channel attention models. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 9645–9656. [[CrossRef](#)]
- Zhou, W.; 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.
- Kumar, N.; Hoffmann, N.; Oelschlägel, M.; Koch, E.; Kirsch, M.; Gumhold, S. Structural Similarity based Anatomical and Functional Brain Imaging Fusion. In Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention and International Workshop on Mathematical Foundations of Computational Anatomy, Shenzhen, China, 17 October 2019; Springer: Cham, Switzerland, 2019.
- Fu, Y.; Wu, X.J.; Durrani, T. Image fusion based on generative adversarial network consistent with perception. *Inf. Fusion* **2021**, *72*, 110–125. [[CrossRef](#)]

19. Prabhakar, K.R.; Srikar, V.S.; Babu, R.V. Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
20. Nejati, M.; Samavi, S.; Shirani, S. Multi-focus image fusion using dictionary-based sparse representation. *Inf. Fusion* **2015**, *25*, 72–84. [[CrossRef](#)]
21. Xu, H.; Ma, J.; Le, Z.; Jiang, J.; Guo, X. FusionDn: A unified densely connected network for image fusion. In Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020.
22. Roberts, J.W.; Van Aardt, J.A.; Ahmed, F.B. Assessment of image fusion procedures using entropy, image quality, and multispectral classification. *J. Appl. Remote Sens.* **2008**, *2*, 023522.
23. Fu, Y.; Wu, X.J.; Kittler, J. Effective method for fusing infrared and visible images. *J. Electron. Imaging* **2021**, *30*, 033013. [[CrossRef](#)]
24. Xydeas, C.S.; Pv, V. Objective image fusion performance measure. *Mil. Tech. Cour.* **2000**, *56*, 181–193.
25. Qu, G.; Zhang, D.; Yan, P. Information measure for performance of image fusion. *Electron. Lett.* **2002**, *38*, 313–315. [[CrossRef](#)]
26. Liu, Y.; Chen, X.; Wang, Z.; Wang, Z.J.; Ward, R.K.; Wang, X. Deep learning for pixel-level image fusion: Recent advances and future prospects. *Inf. Fusion* **2018**, *42*, 158–173. [[CrossRef](#)]
27. Eskicioglu, A.M.; Fisher, P.S. Image quality measures and their performance. *IEEE Trans. Commun.* **1995**, *43*, 2959–2965.
28. Xu, H.; Ma, J.; Jiang, J.; Guo, X.; Ling, H. U2fusion: A unified unsupervised image fusion network. *IEEE Trans. Pattern Anal. Machine Intell.* **2020**, *4*, 502–518.
29. Hui, L.A.; Xjw, A.; Jk, B. Rfn-nest: An end-to-end residual fusion network for infrared and visible images. *Inf. Fusion* **2021**, *73*, 72–86.
30. Li, H.; Wu, X.J.; Kittler, J. Mdlatlr: A novel decomposition method for infrared and visible image fusion. *IEEE Trans. Image Process.* **2020**, *29*, 4733–4746 [[CrossRef](#)]
31. Kumar, B.K.S. Multifocus and multispectral image fusion based on pixel significance using discrete cosine harmonic wavelet transform. *Signal Image Video Process.* **2013**, *7*, 1125–1143. [[CrossRef](#)]
32. Liu, Y.; Chen, X.; Cheng, J.; Peng, H.; Wang, Z. Infrared and visible image fusion with convolutional neural networks. *Int. J. Wavelets Multiresolut. Inf. Process.* **2018**, *16*, 1850018. [[CrossRef](#)]
33. Xu, H.; Ma, J.; Zhang, X.P. Mef-gan: Multi-exposure image fusion via generative adversarial networks. *IEEE Trans. Image Process.* **2020**, *29*, 7203–7216. [[CrossRef](#)]