

Received January 3, 2020, accepted January 14, 2020, date of publication January 22, 2020, date of current version February 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2968559

Fusionnet: Multispectral Fusion of RGB and NIR Images Using Two Stage Convolutional Neural Networks

CHEOLKON JUNG¹, (Member, IEEE), KAILONG ZHOU, AND JIAWEI FENG

School of Electronic Engineering, Xidian University, Xi'an 710071, China

Corresponding author: Cheolkon Jung (zhengzk@xidian.edu.cn)

This work was supported in part by the Huawei Innovation Research Program under Grant HF2018095025, in part by the National Natural Science Foundation of China under Grant 61872280, and in part by the International S&T Cooperation Program of China under Grant 2014DFG12780.

ABSTRACT In low light condition, color (RGB) images captured by imaging systems suffer from severe noise causing loss of colors and textures. Near infrared (NIR) images, which tend to ignore interference from external lights, have advantage of capturing invisible information that can not be obtained by regular RGB cameras. In this paper, we propose multispectral fusion of RGB and NIR images using two stage convolutional neural networks (CNNs), called FusionNet. Lack of training data is a huge obstacle to the learning-based fusion. We synthesize noisy RGB images for training by adding multiscale Gaussian noise. We adopt two stage CNNs for RGB-NIR fusion that consists of denoising and fusion. First, we use a compact denoising subnetwork to remove severe noise from the input RGB image. Then, we utilize a fusion subnetwork to recover textures of the denoised RGB image with the help of its corresponding NIR image. We provide a perceptually motivated loss function to ensure color/texture consistency between the input RGB image and the output fusion result. Experimental results show that the proposed method produces natural looking fusion results by successfully recovering colors and textures. Moreover, the proposed method outperforms state-of-the-art fusion methods in terms of visual quality and quantitative measurements.

INDEX TERMS Image fusion, color, convolutional neural network, FusionNet, near infrared.

I. INTRODUCTION

In low light condition, it is a challenging task to take satisfactory photographs. Increasing ISO in a short exposure introduce noise, while long exposure causes motion blur. Although artificial light such as flash can be added to the such scene for sharp and noise free images, the color tones of acquired images differ from those of the no-flash images due to the color temperature difference between the ambient light and the flash [30]. With advances in the sensor technology, accompanying with image acquisition devices has become affordable in recent years. The near-infrared (NIR) imaging is considered as a solution for high quality photographs [7]. Along with the development of monitoring and mobile equipment, researches on NIR images are of practical significance. The NIR images, which provide high-quality photographs even in low light condition, are being treated as an alternative

for natural-looking image fusion. Different from visible light spectrum, NIR light is with the electromagnetic spectrum ranging from 750 nm to 1000 nm. In low light condition, NIR cameras capture information presented at another spectral frequency. Compared with regular color (RGB) images, NIR images are not affected by ambient light only related to the reflection property of the object materials. Therefore, the fusion of RGB and NIR images is able to improve imaging quality in low light condition, which has various applications including visual surveillance, behavior understanding [1], and remote sensing [28].

A. RELATED WORK

The fusion of RGB and NIR images aims at enhancing a noisy RGB image by its corresponding NIR image. Most existing fusion methods [11], [30], [31] are based on regularization that uses the NIR image as a guidance image. The clear edges are inevitably fused by the weight average of two input images during the fusion process. The RGB

The associate editor coordinating the review of this manuscript and approving it for publication was Quan Zou¹.

and NIR images have their own advantages with respect to color and texture. Although RGB remains the original color and conforms to human perception, in low light condition, RGB is not able to capture the details of the dark scene and suffers from severe noise. Different from RGB, NIR illuminates the scene only based on the reflection property of the material. However, NIR lacks color discrimination because the NIR light is out of the range of human visual perception. Conventional fusion approaches use them for fusion such as gradient difference regularization (GDR) [5], [14], [30], multiresolution (MR) [17], [19], [23], [27], and weighted least squares (WLS) [31]. The WLS model utilized the infrared image to determine the weights of the regularization term. The gradient-based method failed in removing the noise and describing the details since they were not suitable for overcoming the serious discrepancies in the edges and brightness between the NIR and RGB. More recently, the state-of-the-art methods based on scale map (SM) [24] and layer decomposition (LD) [22] were introduced to handle the serious discrepancy problem. Different from the traditional methods, the SM handled gradient direction and magnitude discrepancy using the scale map even though it needs to estimate the unknown scale map and the original image at the same time. The SM approach tends to smooth textures with color over-enhancement. LD and transfer based method were effective for the detail description over the conventional fusion and GF [9] methods, and were required to improve visual perception in the fusion results.

B. MOTIVATION

In this paper, we investigate multispectral fusion of RGB and NIR images to produce natural and realistic results. We propose a perceptually motivated loss function for fusion to recover hidden textures in the input RGB image with help of its corresponding NIR image, thus producing visually pleasing fusion results. As shown in Fig. 1, the NIR image contains clearer textures without color information than its noisy RGB image. With assistance of the NIR image, SM [24] produces a denoised RGB image, but it fails to recover colors completely such as flowers and plants. However, the proposed method successfully removes noise while producing natural colors based on the perception motivated loss function. Datasets of NIR and RGB images play an important role in the deep learning based multispectral fusion that needs a large-scale dataset to learn the NIR and RGB image property. Although many studies have been conducted for NIR-RGB image fusion, deep learning based multispectral fusion is not found due to the lack of available datasets. Since NIR and RGB images are captured from real world scenes, we provide a solution to the deep learning based multispectral fusion in this work. Our dataset synthesis strategy is proved to be used in the training process for multispectral fusion with complete generality. We aim at producing natural-looking fusion results based on multispectral information, and adopt CNNs for the multispectral fusion. In low light condition,

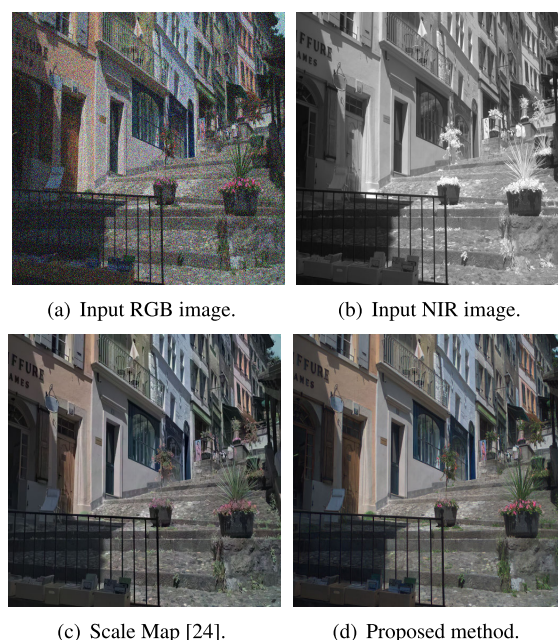


FIGURE 1. RGB-NIR image fusion results.

the captured color image is severely interfered by mess of factors, i.e. distribution of noise gets very complicated. Some naive denoising is disable to remove these noise. Several fusion methods consider the guidance of NIR and conduct weighted averaging on multispectral images, and inevitably fuse them with noise. To cope with such noise, we propose a two stage fusion network that consists of denoising and fusion subnetworks, called FusionNet. The denoising subnetwork, trained on blind noise, generates noise-free RGB image with smooth texture regions. Then, the fusion subnetwork is conducted on the denoised RGB image with the help of its corresponding NIR image. Due to the benefit from powerful generalization ability of CNNs, the proposed method successfully preserves color and high frequency details from multispectral images in the fusion results. FusionNet provides a viable solution to high quality photographs in low light condition that can be used various vision applications such as object recognition and behavior recognition. Fig. 2 illustrates the network architecture of the proposed FusionNet. To the best of our knowledge, this is the first work for multispectral fusion based on deep learning.

Compared to existing methods, main contributions of this paper are as follows:

- We provide a synthesis approach to generate training data in low light condition, which has proved to be suitable for image denoising and multispectral fusion.
- We build a multilevel aggregation neural network to merge features at different levels from the input NIR image and produce natural color and fine details.
- We propose a perceptually motivated feature loss that transfers fine details from the input NIR image to the output fusion result, thus resulting in good textures.

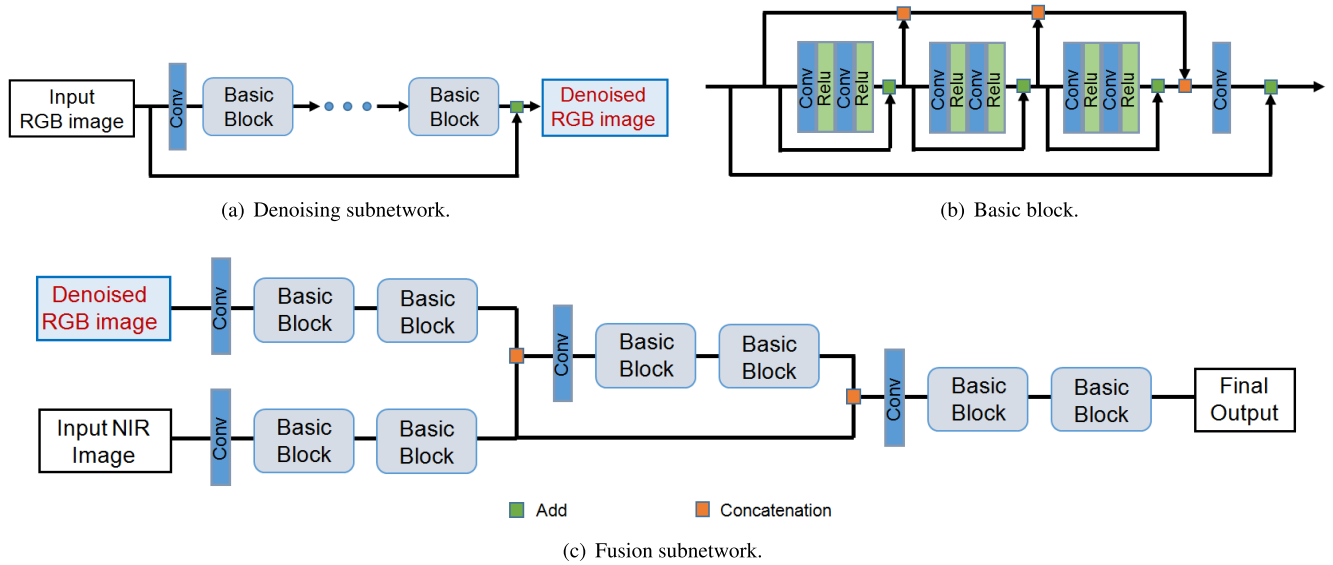


FIGURE 2. Network architecture of the proposed FusionNet. FusionNet consists of two subnetworks: denoising in (a) and fusion in (c). We build the basic block in (b) for FusionNet. The output of the denoising subnetwork is the input of the fusion subnetwork. We utilize the pretrained VGG-16 to extract features. We use the clean RGB image as the ground truth for the color loss and the concatenation of 3-channel RGB and 1-channel NIR images as the ground-truth for the multispectral feature loss.

II. PROPOSED METHOD

A. DATA PREPARATION

Different from other inverse problems [4], [6], [26], multispectral fusion takes advantage of multiple inputs to produce high quality photographs. However, a huge amount of training datasets are necessary, and public available RGB-NIR datasets are rather limited, thus greatly hindering deep learning-based multispectral fusion. Moreover, there is no ground truth available for deep learning-based multispectral fusion. Although there are many real RGB images online, there exist a few datasets containing both RGB and NIR images, and the RGB images in the datasets are of high quality. Thus, the datasets themselves are not able to be used as training data. To overcome the training data insufficiency, we synthesize noisy RGB images for training data by adding noise in clean RGB images, and use the clean RGB images as ground truth. In this work, we provide a reasonable way to generate noisy RGB images by taking noise production of dim light shooting scene into consideration. We synthesize them similar to the images taken in real-world scenes. We use a publicly available RGB-NIR dataset from [3] to synthesize the training dataset for fusion. In this dataset, RGB-NIR image pairs are captured under several scenes in daylight conditions.

1) DATASET FOR DENOISING

RGB-NIR image pairs are limited, but there are a lot of publicly available datasets for image denoising. To train the denoising subnetwork, we use the BSDS500 database [16] as our training dataset. Among total 400 images, we use 200 images for training and 200 images for testing. We train the denoising subnetwork for Gaussian denoising by setting

the noise level to $\sigma \in [30, 60]$. We split the training images into 40×40 patches with stride 21.

2) DATASET FOR FUSION

We generate noisy RGB images by adding Gaussian noise at the same level as denoising. We use the denoised RGB images obtained by the denoising subnetwork as input for the fusion subnetwork. Then, we use the original clean RGB images as the ground truth. We use the same noise distribution for training data to synthesize their noisy ones for training.

B. NETWORK ARCHITECTURE

To remove server noise in low light condition, we build a novel basic block as depicted in Fig. 2(b). In such structure, a gating mechanism [25] is adopted as the convolutional unit. We learn each residual building block to generate multi-level representation considering different receptive fields. Then, we concatenate the multi-level representation as the input of the gate unit based on a convolutional layer. The gate unit accomplish a gating mechanism which learns adaptive weights from the multi-level representation by a 1×1 convolution layer. Next, we add the input to the output of the gate unit and make the shallow representation flow forward. For the denoising subnetwork, we stack basic blocks and adopt residual learning to accelerate the training process as shown in Fig. 2(a). To cope with severe noise, we train the denoising subnetwork in blind noisy condition and adopt a compact structure considering computational complexity. We train the denoising subnetwork in the BSD500 dataset, and then train the fusion subnetwork from the RGB-NIR dataset under the fixed parameters of the denoising subnetwork. To refine the denoised RGB image, we perform the fusion subnetwork with

the help of the input NIR image. In the training phase of the fusion subnetwork, we feed the denoised RGB image into the fusion subnetwork with the input NIR image. The denoising subnetwork produces a denoised RGB image Y_1 From the synthesized noisy RGB image X as follows:

$$Y_1 = D(X) \quad (1)$$

where D denotes the denoising subnetwork, X represents the noisy RGB image, and Y_1 is taken as the input for the fusion subnetwork. The NIR image N is used as the other input for the fusion subnetwork. As illustrated in Fig. 2(c), the denoising subnetwork is cascading to the fusion subnetwork that takes the denoised RGB and NIR images as the input. The final output Y_2 is represented as follows:

$$Y_2 = F(Y_1, N) \quad (2)$$

where F denotes the fusion subnetwork.

C. LOSS FUNCTION

The training process of the denoising subnetwork is independent of the fusion subnetwork. The output of the denoising subnetwork is the input of the fusion subnetwork. We use different training datasets and loss functions for two subnetworks. For the denoising subnetwork, we use a normal pixel-wise loss to reconstruct the denoised RGB image as follows:

$$L_{pixel}(X', X) = \frac{1}{C \times H \times W} \|X, X'\|^2 \quad (3)$$

where $C \times H \times W$ is the image size, and X and X' denote the noisy RGB image and its ground truth, respectively.

For the fusion subnetwork, the pixel-wise loss function is not able to deal with the discrepancy between RGB and NIR images. The pixel-wise loss function is good for reconstruction, but it is inconsistent with human visual perception. Thus, it is difficult to produce satisfactory fusion results by the pixel-wise loss function. We aim to preserve color information of the RGB image while enhancing details with the help of the NIR image. Considering them, we build a perceptual loss function for fusion to recover both colors and details from RGB and NIR images [10], [15]. We use the pretrained VGG16 [20] to extract features from the RGB and NIR images as shown in Fig. 3. The two feature maps are obtained by the second convolution (after activation) prior to the first max pooling layer. The second and third rows present the 1_{st} and 16_{th} channels, respectively. In the right column, the feature maps of the noisy RGB image present rough surface because of the noise. However, those of the NIR image has clear edges with smooth flat regions in the left column. The NIR image shows clear details of the teapot, but the RGB image shows complicated textures without edge information of the teapot. Also, the shadow of the teapot is unique in the NIR image with clear features. To feed the NIR image into VGG16, we duplicate the NIR image to three channels. We transfer details from the NIR image (the details are invisible in the RGB image due to the low light) to the fusion result while keeping color information from the

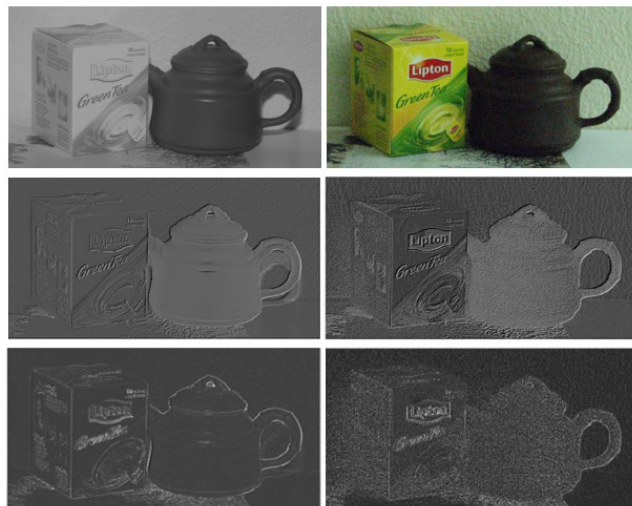


FIGURE 3. Feature maps for Teapot. Left: NIR image. Right: Noisy RGB image. They are extracted by VGG16 at the second layer before the first max-pooling layer.

original RGB image. Thus, we make the fusion result similar to clean RGB and NIR images by a function ϕ through the training process. To keep the original colors from the RGB image ground truth, we compute a perceptual loss as follows:

$$L_{perceptual}(Y_2, Y) = \frac{1}{C \times H \times W} \|\phi(Y_2) - \phi(Y)\|_2^2 \quad (4)$$

We choose the layer output prior to the third downsampling in VGG16 for feature extraction. The perceptual loss optimization in Eq. (4) ensures the color/texture consistency between the input RGB image and the output fusion result. To extract features from the NIR image, we duplicate one single channel NIR image as three channels for the input of VGG16. To produce natural looking fusion results with clear details and textures, we build a multispectral loss as follows:

$$L_{nir}(Y_2, N) = \frac{1}{C \times H \times W} \|\phi(Y_2) - \phi(N)\|_2^2 \quad (5)$$

Combining the above terms, our complete loss function becomes:

$$L = L_{perceptual}(Y_2, Y) + \lambda L_{nir}(Y_2, N) \quad (6)$$

The weighting coefficient λ severely affects the fusion result. As λ is higher, the fusion result is more similar to the NIR image.

III. EXPERIMENTAL RESULTS

For experiments, we train the proposed FusionNet using stochastic Adam optimizer. During the training process, we set learning rate to 0.0001 and batch size to 32 for hyperparameters. The denoising subnetwork shares the same hyperparameters as the fusion subnetwork. We compare the performance of FusionNet with those of multispectral fusion methods such as the weighted least squares (WLS) [31], scale map (SM) [24] and DenseFuse, i.e. CNN-based fusion of

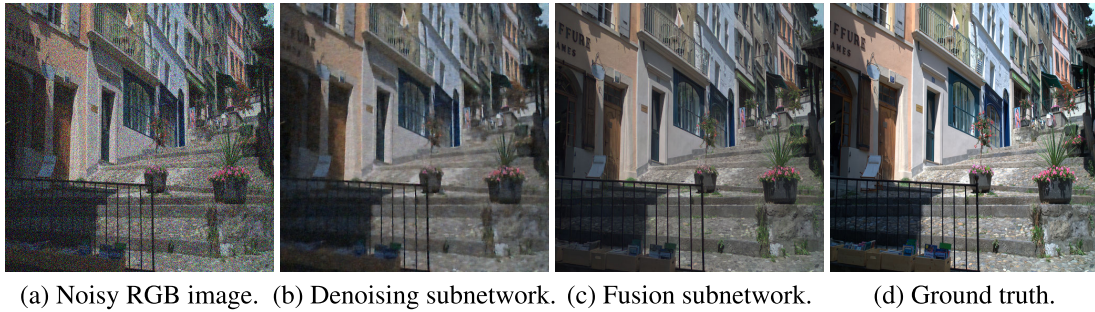


FIGURE 4. Fusion results by FusionNet.

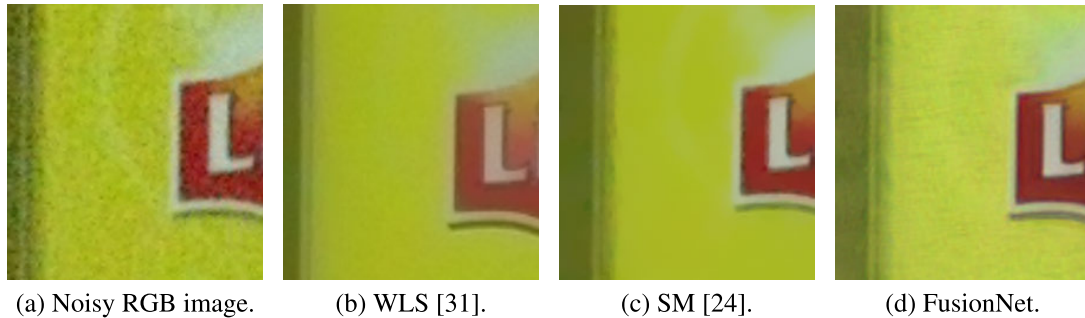


FIGURE 5. Visual comparison between different fusion methods. We provide close up examples for *Teapot* on the fusion results.

visible and infrared images [12] in terms of visual quality and quantitative measurements. We use the NIR and RGB image datasets from [22] and [24] with a synthetic dataset for performance comparison.

A. COMPARISON WITH STATE-OF-THE-ARTS

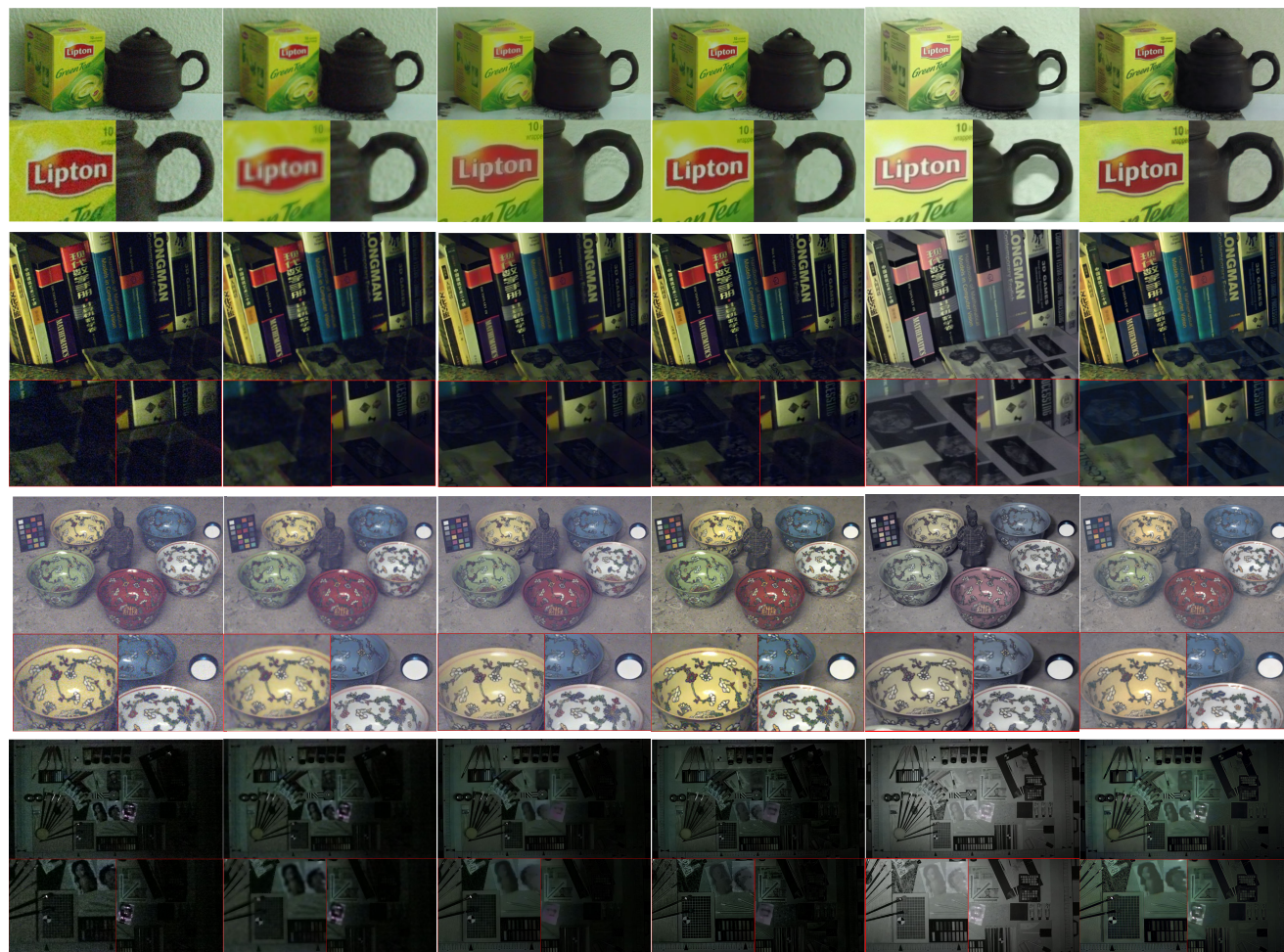
1) VISUAL COMPARISON

Fig. 4 shows the stepwise results of FusionNet on a synthetic image. The denoising subnetwork removes noise (Fig. 4(b)), and the fusion subnetwork refines the denoised result (Fig. 4(c)). Compared with the ground truth, the fusion result gets a little dark because the feature of the NIR image decreases brightness in fusion. Fig. 5 shows close up examples of the fusion results by WLS [31], SM [24] and DenseFuse [12]. The input RGB image is corrupted by severe noise, and three fusion methods remove noise differently in the results. WLS produces water washed colors along the letter area with blurry edge. SM produces an over-smoothed appearance that textures on the tea box are wiped. However, FusionNet produces a natural-looking fusion result that contains more high frequency details than them. FusionNet successfully restores clear edges from the noisy RGB image while reproducing a natural color tone in fusion. Fig. 6 shows more fusion results by different fusion methods. Note that we use a real image pair for fusion captured in low light condition (see the last row of the figure). In the first row, noise is successfully removed, guided filtering (GF) [9] obviously introduces blur in the fusion result causing an unnatural appearance. Such blur results in image quality degradation.

WLS [31] produces a natural looking appearance in fusion. However, a color spill appears along the white letters because WLS is not able to successfully consider the discrepancy between RGB and NIR images. SM [24] produces a good fusion appearance with oversmoothing effects. Some details such as the wall and pot-lid are lost. Although DenseFuse [12] restores the image details well, it causes severe color distortion in the fusion results. FusionNet recovers colors well while successfully preserving textures. The fine textures on the paper box, pot-lid and wall are also presented well, thus leading to the natural looking results. The second row presents an image captured in low light condition, and thus contains very dark area with a narrow dynamic range. Compared with the others, FusionNet successfully enhances the dark area and shows the hidden textures in the area. The same results appear in the last row. FusionNet produces fine details in dark regions better than the others. The binding label on the book (see the lower right corner of the image) is clearly presented, and it is very dark and blurry in the original image. In the third row, SM produces over-enhanced colors for the patterns of the bowls that are somewhat different from its original ones. The proposed FusionNet retains fine details and natural colors in the fusion results. The results demonstrate that FusionNet can be successfully applied to multispectral fusion of RGB and NIR images.

2) QUANTITATIVE MEASUREMENTS

For the publicly available dataset captured in real world scenes [22], [24], the ground truth is not available. Thus,



(a) Input RGB image. (b) GF [9]. (c) WLS [31]. (d) SM [24]. (e) DenseFuse [12]. (f) FusionNet.

FIGURE 6. Fusion results by different methods. We compare FusionNet with GF, WLS and SM. FusionNet produces clear textures with natural colors in the fusion results. Especially, FusionNet has capability of recovering hidden textures in dark regions (2rd and 4th rows).

TABLE 1. Performance comparison between different methods on the four test images of Fig. 6 in terms of spatial frequency (SF) [13] and entropy (EN) [8]. Bold numbers represent the best performance.

Image	Metric	GF [9]	WLS [31]	SM [24]	DenseFuse [12]	FusionNet
1st row	EN	7.4050	7.3973	7.3573	7.4603	7.4907
	SF	6.6075	9.8448	10.2686	11.4209	12.1204
2nd row	EN	6.7891	6.8811	6.8566	6.4232	7.2351
	SF	10.4813	12.3035	15.3309	20.4412	17.5626
3rd row	EN	6.8537	6.9064	7.1620	7.2181	7.8650
	SF	10.2313	16.9927	20.4356	17.9212	19.6622
4th row	EN	5.9922	5.9574	5.9164	6.1891	6.6446
	SF	8.4901	15.1465	19.5108	19.3173	20.0355
Average	EN	6.7600	6.7856	6.8231	6.8227	7.3089
	SF	8.9525	13.5719	16.3865	17.2752	17.3451

we cannot evaluate the reference-based quantitative measurements [8], [18] for quality assessment such as PSNR and SSIM [29]. Thus, we evaluate the blind image quality measure (BIQE) on the fusion results [2], [21]. The BIQE does not

need subjective tests. We use four test images of Fig. 6 for the BIQE evaluation. Table 2 shows the BIQE evaluation results (the lower the better). FusionNet outperforms the others in average BIQE score, which indicates that FusionNet achieves

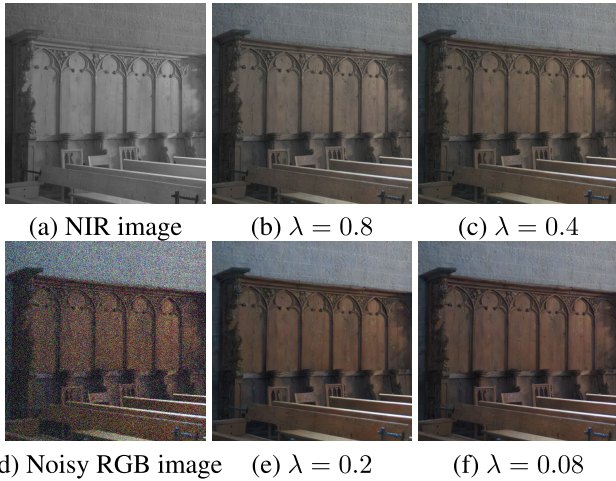


FIGURE 7. Fusion results according to different λ .

TABLE 2. BIQE comparison between different methods on the four test images of Fig. 6. The lower the BIQE is, the better the performance is. Bold numbers represent the best performance.

Method	1st row	2nd row	3rd row	4th row	Average
GF [9]	52.10	32.05	25.18	36.58	36.48
WLS [31]	30.53	27.12	22.98	35.24	28.97
SM [24]	32.72	25.24	19.61	40.28	29.46
Densefuse [12]	29.24	27.33	20.08	27.24	25.97
FusionNet	27.99	26.05	19.72	26.89	25.67

TABLE 3. Runtime comparison among different fusion methods (unit: sec/pair). For tests, we use a PC with Intel Core i7-7700 CPU and GTX1080 GPU running Ubuntu 16.04 and Keras.

Method	GF [9]	WLS [31]	SM [24]	DenseFuse [12]	FusionNet
Time	0.22	7.66	13.15	1.71	1.32

the best visual quality in fusion. Moreover, FusionNet successfully recovers natural colors and fine details, which is more consistent with human visual perception than the others. To further evaluate the performance of FusionNet, we evaluate two measures for comparison: spatial frequency (SF) [13] and entropy (EN) [8]. The SF measure is used to evaluate edge preservation, while the EN measure is used to measure the amount of information. As shown in Table 1, FusionNet yields the highest average scores in most cases in terms of both SF and EN. This is because FusionNet produces fine details and natural colors in the fusion results.

3) COMPUTATIONAL COMPLEXITY

Table 3 shows average runtime of WLS [31], SM [24], DenseFuse [12] and FusionNet on the five test images of Figs. 6 and 7. The unit of the runtime is sec/pair, while the image size is from 512×512 to 1492×1101 . We use a PC with Intel Core i7-7700 CPU and GTX1080 GPU running Ubuntu 16.04 and Keras for tests. As shown in the table, FusionNet achieves the second best runtime among them.

TABLE 4. PSNR and SSIM evaluations for Indoor, Urban and Old-building in the dataset [3] according to different λ .

Dataset	Metric	$\lambda = 0.8$	$\lambda = 0.4$	$\lambda = 0.2$	$\lambda = 0.08$
Indoor	PSNR	33.92	33.85	34.60	34.44
	SSIM	0.5294	0.5548	0.5791	0.5463
Urban	PSNR	33.19	33.12	33.60	33.56
	SSIM	0.53820	0.5550	0.5700	0.5502
Old-Building	PSNR	33.60	33.54	34.06	33.98
	SSIM	0.5279	0.5429	0.5624	0.5387

B. ABLATION STUDY

We perform ablation study on the proposed loss function according to different λ in Eq. (6). Fig. 7 shows the fusion results by different weights of the NIR image. It can be observed that as λ decreases the color gets deeper. As the weight of L_{nir} increases, the NIR image contributes to the fusion result more. Thus, the fusion result contains more details with lower color saturation. If λ is equal to 0.2, the fusion result is optimal and contains most features of the NIR image without color reduction. Table 4 shows PSNR and SSIM evaluations for Indoor, Urban and Old-building in the dataset [3] according to different λ . As λ decreases, the performance increases until λ is 0.2. However, if λ is smaller than 0.2, the performance declines. Therefore, we choose 0.2 for λ in our experiments.

IV. CONCLUSION

We have proposed FusionNet for multispectral fusion of RGB and NIR images based on two stage CNNs. Unlike the traditional fusion methods based on filtering or regularization, FusionNet is a data-driven approach and achieves multispectral fusion based on a concurrent network that consists of denoising and fusion subnetworks. FusionNet successfully recovers colors and textures by using a perception motivated loss function, and produces natural looking fusion results. Experimental results demonstrate that FusionNet outperforms state-of-the-art methods in terms of visual quality and quantitative measurements.

ACKNOWLEDGMENT

The authors would like to thank Dr. C. Su and Dr. H. Liu in Huawei Technologies for their helpful advice and discussion.

REFERENCES

- [1] N. Bakalos, A. Voulodimos, A. Doulamis, and N. Doulamis, "A deep-narma filter for unusual behavior detection from visual, thermal and wireless signals," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 4005–4009
- [2] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015.
- [3] M. Brown and S. Susstrunk, "Multi-spectral sift for scene category recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 177–184.
- [4] C. Dong, Y. Deng, C. C. Loy, and X. Tang, "Compression artifacts reduction by a deep convolutional network," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 576–584.
- [5] C. Feng, S. Zhuo, X. Zhang, L. Shen, and S. Susstrunk, "Near-infrared guided color image dehazing," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 2363–2367.

- [6] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1256–1272, Jun. 2017.
- [7] C. Fredembach and S. Süsstrunk, "Colouring the near-infrared," in *Proc. Color Imag. Conf.*, 2008, pp. 176–182.
- [8] M. B. A. Haghighat, A. Aghagolzadeh, and H. Seyedarabi, "A non-reference image fusion metric based on mutual information of image features," *Comput. Electr. Eng.*, vol. 37, no. 5, pp. 744–756, Sep. 2011.
- [9] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
- [10] J. Johnson, A. Alahi, and F. F. Li, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 694–711.
- [11] S.-H. Lee, H.-M. Park, and S.-Y. Hwang, "Motion deblurring using edge map with blurred/noisy image pairs," *Opt. Commun.*, vol. 285, no. 7, pp. 1777–1786, Apr. 2012.
- [12] H. Li and X.-J. Wu, "DenseFuse: A fusion approach to infrared and visible images," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2614–2623, May 2019.
- [13] S. Li and B. Yang, "Multifocus image fusion by combining curvelet and wavelet transform," *Pattern Recognit. Lett.*, vol. 29, no. 9, pp. 1295–1301, Jul. 2008.
- [14] W. Li, J. Zhang, and Q.-H. Dai, "Robust blind motion deblurring using near-infrared flash image," *J. Vis. Commun. Image Represent.*, vol. 24, no. 8, pp. 1394–1413, Nov. 2013.
- [15] F. Luan, S. Paris, E. Shechtman, and K. Bala, "Deep photo style transfer," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4990–4998.
- [16] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.
- [17] V. Petrovic and C. Xydeas, "Gradient-based multiresolution image fusion," *IEEE Trans. Image Process.*, vol. 13, no. 2, pp. 228–237, Feb. 2004.
- [18] A. Saha and Q. J. Wu, "Full-reference image quality assessment by combining global and local distortion measures," *Signal Process.*, vol. 128, pp. 186–197, Nov. 2016.
- [19] L. Schaul, C. Fredembach, and S. Süsstrunk, "Color image dehazing using the near-infrared," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 1629–1632.
- [20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015.
- [21] C.-H. Son, "Near-infrared fusion via a series of transfers for noise removal," *Signal Process.*, vol. 143, pp. 20–27, Feb. 2018.
- [22] C.-H. Son and X.-P. Zhang, "Layer-based approach for image pair fusion," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2866–2881, Jun. 2016.
- [23] M. Tico and K. Pulli, "Image enhancement method via blur and noisy image fusion," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 1521–1524.
- [24] Q. Yan, X. Shen, L. Xu, S. Zhuo, X. Zhang, L. Shen, and J. Jia, "Cross-field joint image restoration via scale map," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1537–1544.
- [25] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4549–4557.
- [26] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
- [27] Q. Zhang and B.-L. Guo, "Multifocus image fusion using the nonsubsampled contourlet transform," *Signal Process.*, vol. 89, no. 7, pp. 1334–1346, Jul. 2009.
- [28] W. Zhao, L. Jiao, W. Ma, J. Zhao, J. Zhao, H. Liu, X. Cao, and S. Yang, "Superpixel-based multiple local CNN for panchromatic and multispectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 4141–4156, Jul. 2017.
- [29] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [30] S. Zhuo, D. Guo, and T. Sim, "Robust flash deblurring," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2440–2447.
- [31] S. Zhuo, X. Zhang, X. Miao, and T. Sim, "Enhancing low light images using near infrared flash images," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 2537–2540.



CHEOLKON JUNG (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electronic engineering from Sungkyunkwan University, South Korea, in 1995, 1997, and 2002, respectively. From 2002 to 2007, he was a Research Staff Member with the Samsung Advanced Institute of Technology (Samsung Electronics), South Korea. From 2007 to 2009, he was a Research Professor with the School of Information and Communication Engineering, Sungkyunkwan University. Since 2009, he has been a Professor with the School of Electronic Engineering, Xidian University, China. His main research interests include image and video processing, computer vision, pattern recognition, machine learning, computational photography, video coding, virtual reality, information fusion, multimedia content analysis and management, and 3D TV.



KAILONG ZHOU received the B.S. degree in electronic engineering from Xidian University, China, in 2017, where he is currently pursuing the M.S. degree. His research interests include image processing and deep learning.



JIawei FENG received the B.S. degree in electronic engineering from Chang'an University, China, in 2016. He is currently pursuing the M.S. degree with Xidian University, China. His research interests include image processing and machine learning.

...