BAM: Bilateral Activation Mechanism for Image Fusion

Zi-Rong Jin
2018051403016@std.uestc.edu.cn
School of Optoelectronic Science and Engineering, University of Electronic Science and Technology of China

Liang-Jian Deng*
liangjian.deng@uestc.edu.cn
School of Mathematical Sciences, University of Electronic Science and Technology of China

Tian-Jing Zhang
zhangtianjinguestc@163.com
Yingcai Honors College, University of Electronic Science and Technology of China

Xiao-Xu Jin
jinxiaoxu0102uestc@163.com
Yingcai Honors College, University of Electronic Science and Technology of China

ABSTRACT

As the conventional activation functions such as ReLU, LeakyReLU, and PReLU, the negative parts in feature maps are simply truncated or linearized, which may result in unflexible structure and undesired information distortion. In this paper, we propose a simple but effective Bilateral Activation Mechanism (BAM) which could be applied to the activation function to offer an efficient feature extraction model. Based on BAM, the Bilateral ReLU Residual Block (BRRB) that still sufficiently keeps the nonlinear characteristic of ReLU is constructed to separate the feature maps into two parts, i.e., the positive and negative components, then adaptively represent and extract the features by two independent convolution layers. Besides, our mechanism will not increase any extra parameters or computational burden in the network. We finally embed the BRRB into a basic ResNet architecture, called BRResNet, it is easy to obtain state-of-the-art performance in two image fusion tasks, i.e., pansharpening and hyperspectral image super-resolution (HISR). Additionally, deeper analysis and ablation study demonstrate the effectiveness of BAM, the lightweight property of the network, etc. Please find the code from the project page[1].

CCS CONCEPTS

• **Computing methodologies** → **Neural networks**.

KEYWORDS

Activation Function, Bilateral Activation Mechanism, Convolutional Neural Networks, Image Fusion

ACM Reference Format:
Zi-Rong Jin, Liang-Jian Deng, Tian-Jing Zhang, and Xiao-Xu Jin. 2021. BAM: Bilateral Activation Mechanism for Image Fusion. In *Proceedings of the 29th ACM International Conference on Multimedia (MM '21), October 20–24, 2021, Virtual Event, China.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3474085.3475571

*Corresponding author.
[1]https://liangjiandeng.github.io/Projects_Res/bam_mm2021.html

MM '21, October 20–24, 2021, Virtual Event, China

ACM ISBN 978-1-4503-8651-7/21/10…$15.00
https://doi.org/10.1145/3474085.3475571

1 INTRODUCTION

Image fusion aims to fuse the image data that record the same target collected by different sensors through specific techniques to maximize the extraction of the desired information from each sensor. It can improve the spatial and spectral resolution of the original image and finally produce a high-quality image that can be further applied to other high-level vision tasks, such as image segmentation and detection. Recently, convolutional neural networks (CNNs) have demonstrated remarkable superiority in image fusion due to the powerful computing infrastructure and availability of large-scale datasets. The main improvement direction of the latest CNNs-based methods [18, 29, 30, 41, 44] points to the optimization of the network structures. Structural changes, such as deepening of depth, increasing width, and multi-scale convolution operations, are essential to make the CNNs' feature extraction capabilities more powerful. Activation function [25], e.g., Rectified Linear Unit (ReLU), as the important tool which plays a role in activating the nonlinear fitting ability of CNNs, has received more and more attention. In particular, ReLU is anti-symmetric about 0. It activates the positive part of the input, and the partial derivative is 1. At the same time, the negative part of the input is ignored, and the partial derivative is 0. Such an activation mechanism makes the activated unit not have a vanishing gradient at any network depth. However, when the unit is not activated, the gradient is 0, resulting in it remaining inactivated throughout the optimization process. To mitigate potential problems caused by the hard 0 activation of ReLU [24], its generalized versions (i.e., LeakyReLU [22], PReLU [15] and so forth) have been developed. Most of them are devoted to improving the activation performance of the original ReLU by modifying its functional form. It is actually because of the asymmetry of ReLU that some neurons can not be activated and remain in an inhibited state. This model is conducive to image classification, image segmentation, and other tasks. However, in the image fusion task, those parts that are not activated have the latent features we need. Although the existing activation function, such as ReLU, enhances the nonlinear fitting ability of the network, it will cause undesired information distortion.

In this paper, we present a framework from the direction of the activation mechanism extension, expecting to explore and utilize the features that can not be activated while retaining the nonlinearity of the activation function. Thus, we propose a Bilateral Activation

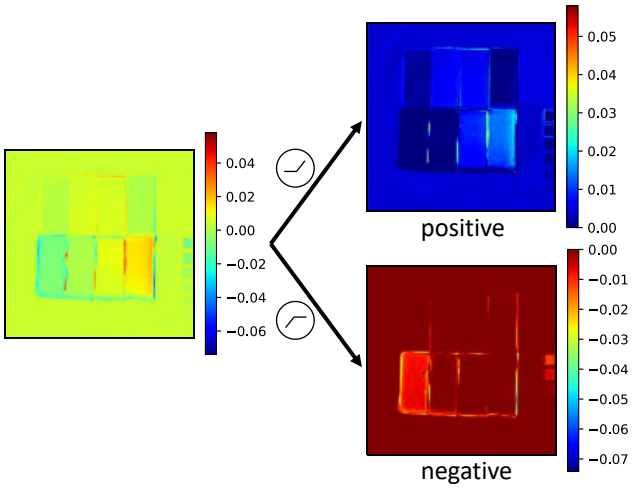

Figure 1: A toy example of the proposed Bilateral Activation Mechanism (BAM) with ReLU as the activation function, by which the feature is separated into two parts, i.e., the positive and negative parts that will be fed to the subsequent convolution layers. Compared with conventional activation functions such as ReLU, the BAM could effectively prevent information distortion, especially for the negative features.

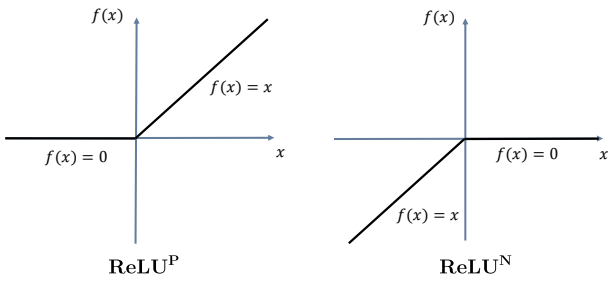

Figure 2: The left one is the diagram of $\mathrm{ReLU}^{\mathrm{P}}$, which is the same as ReLU. The right one is $\mathrm{ReLU}^{\mathrm{N}}$ obtained through rotating ReLU by 180° clockwise.

Mechanism. Take ReLU as an example, BAM with ReLU is able to separate input feature maps into positive and negative parts, as shown in Fig. 1, and then send them to different convolutional layers for feature extraction.

We mainly focus on two image fusion tasks, hyperspectral image super-resolution (HISR) [4, 11, 28] and remote sensing image pansharpening [7, 8, 31]. The first one (HISR) is to obtain a high-resolution hyperspectral image (HR-HSI) by fusing a low-resolution hyperspectral image (LR-HSI) and a high-resolution multispectral image (HR-MSI). And pansharpening yields a high-resolution multispectral image (HR-MSI) by fusing a low-resolution multispectral image (LR-MSI) and a high-resolution panchromatic image (HR-PANI). Whether it is HISR or pansharpening, the difficulty mainly lies in achieving competitive spatial and spectral preservation. Therefore, both precisely modeling the nonlinear relationship between images and fully exploring image features are of critical importance. In our work, the bilateral activation mechanism is applied to a residual block with ReLU as the activation function. Thus bilateral ReLU residual block (BRRB) is constructed without increasing the number of parameters. Furthermore, we embed the BRRB into a simple ResNet [16], called BRResNet, to implement two image fusion tasks. Experiments demonstrate that BRResNet can easily surpass other advanced methods. The main contributions can be summarized as follows:

- A novel bilateral activation mechanism (BAM) is designed to avoid the neuron inactivation problem caused by a peculiar form of the activation function, e.g., ReLU. Not only the nonlinearity of ReLU is retained, but also the features of the input can be fully utilized.
- As a mechanism, BAM provides a more efficient feature extraction mode without increase the computational burden. Also, it has many variants and can be used as a substitution to replace any structure like "Activation + Convolution", giving us more flexibility in designing the network structure.
- A BRResNet with BAM is proposed, which achieves state-of-the-art performance in two fusion tasks. Especially, the given BRResNet holds a large margin among other CNNs-based methods in terms of the parameters, thus can be viewed as a lightweight network.

2 RELATED WORKS AND MOTIVATION

In this section, we will first introduce a common form of several activation functions and present their similarities and differences. Then, the motivations of this paper will be detailed.

2.1 Related Works

As mentioned above, ReLU will ignore the negative elements of the input and cause it to remain inactive. And its generalized versions, such as LeakyReLU, and PReLU, are all changed in their basic form. Thus they can be unified into the following mathematical expression:

$$\mathcal{A}(x) = \begin{cases} x\,, x \geq 0 \\ \alpha x, x < 0 \end{cases}, \tag{1}$$

where $\mathcal{A}(\cdot)$ represents the activation function, and the corresponding common derivative form can be expressed as follows:

$$\mathcal{A}'(x) = \begin{cases} 1\,, x \geq 0 \\ \alpha, x < 0 \end{cases}, \tag{2}$$

where $\mathcal{A}(\cdot)$ represents the corresponding common derivative of the activation function, x is the input, and α is the coefficient. For ReLU, α is set to zero, thus in the process of backpropagation, the gradient for parameters of the inactive unit is zero, which means it can not be updated during the training process. For LeakyReLU, α is set as a small value, which offers a small, non-zero gradient to the negative parts of the input. Besides, in PReLu, α is adaptively learned by networks. It is more flexible but still inhibits the negative parts.

2.2 Motivation

The nonlinear nature of ReLU is reflected in the fact that all negative values are zero, which brings more possibilities for the network to extract features, but the remaining features of the negative part

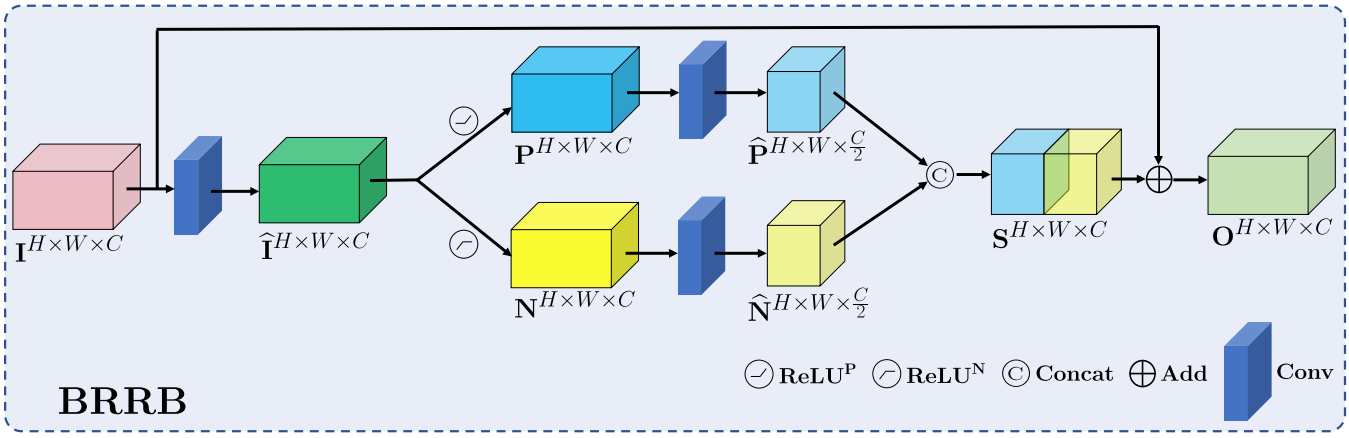

Figure 3: The illustration of BRRB.

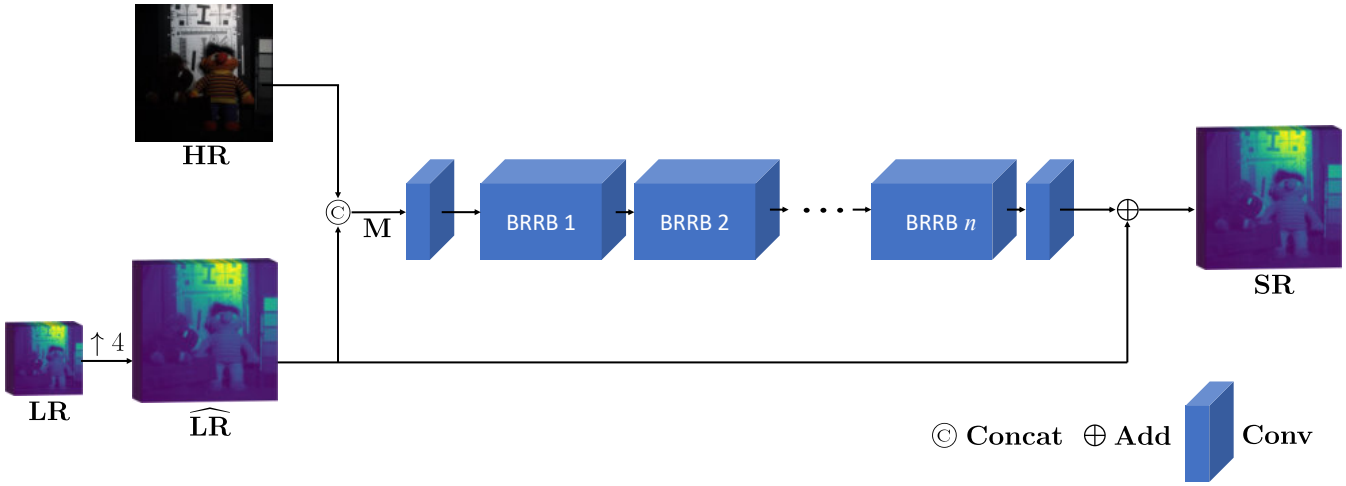

Figure 4: The flowchart of the proposed overall network architecture, i.e., BRResNet that mainly contains n BRRBs, for image fusion tasks such as HISR and pansharpening.

will be directly discarded. In order to extract the residual features in the negative part, LeakyReLU multiplies the negative part with a coefficient α so that the information in the negative part can be preserved. However, although the features of the negative part become better extracted, the nonlinearity of the activation function decreases, further causing the nonlinear fitting ability of the network to decrease. PReLU regards α in the activation function as a trainable parameter in the network. This can indeed balance the relationship between the feature extraction of the negative part and the nonlinearity of the activation function, but in essence, the feature extraction of the negative part is strengthened by weakening the nonlinearity of the activation function.

Therefore, to avoid information loss such as the negative features in ReLU and the parameter tuning of α in LeakyReLU and PReLU, we develop a simple and effective BAM to make full use of both positive and negative features, aiming to reduce the spatial distortion in pixel-wise tasks, e.g., image fusion.

3 THE PROPOSED METHODS

This section will first introduce how BAM works when ReLU is used as the activation function. Then, the Bilateral ReLU Residual Block (BRRB) and BRResNet will be detailed.

3.1 Bilateral Activation Mechanism

By Fig. 1, it is clear that after the feature map passing the ReLU (also called **ReLU$^{\mathrm{P}}$**), there still exist obvious negative features containing abundant details by the ReLU that rotated by 180°, i.e., **ReLU$^{\mathrm{N}}$**. If only taking conventional ReLU, the negative features will be discarded, weakening the ability of feature representation and extraction. In this work, we novelly develop a simple BAM that takes

ReLU as the activation function to make use of both positive and negative features in the network. Please see Fig. 2.

From Fig. 3, consider an input feature map $\widehat{\mathbf{I}}$, which contains the positive part $\mathbf{P}$ and negative part $\mathbf{N}$, the proposed BAM with ReLU will separate $\mathbf{P}$ and $\mathbf{N}$ first. The separation process can be expressed as follows:

$$\mathbf{P}_{ijk} = max\left\{\widehat{\mathbf{I}}_{ijk}\ ;\ 0\right\}, \tag{3}$$

$$\mathbf{N}_{ijk} = min\left\{\widehat{\mathbf{I}}_{ijk}\ ;\ 0\right\}, \tag{4}$$

where (i, j, k) denotes the coordinate in the feature map. Then, $\mathbf{P}$ and $\mathbf{N}$ will be sent to different paths for feature extraction. More details about BAM with ReLU can refer to Fig. 2.

3.2 Bilateral ReLU Residual Block

After defining the BAM, we here present how to embed the BAM into a common network architecture. The classical ResNet is choosen as the basic architecture because of its high performance and efficiency. Especially, Bilateral ReLU Residual Block (BRRB) embedded with BAM is proposed to replace the residual block in ResNet [16]. Assume that the input feature map is $\mathbf{I} \in \mathbb{R}^{H\times W\times C}$, where H and W is the size in spatial dimension, C is the channels of the input feature, $\mathbf{I}$ will be sent to the convolutional layer to exact its shallow feature $\widehat{\mathbf{I}} \in \mathbb{R}^{H\times W\times C}$. Then the exacting feature map $\widehat{\mathbf{I}}$ will pass through the BAM with ReLU to separate the the positive part $\mathbf{P} \in \mathbb{R}^{H\times W\times C}$ and negative part $\mathbf{N} \in \mathbb{R}^{H\times W\times C}$. Next, the separated two parts will pass through two independent convolutional layers to obtain the potential feature $\widehat{\mathbf{P}} \in \mathbb{R}^{H\times W\times \frac{C}{2}}$ and $\widehat{\mathbf{N}} \in \mathbb{R}^{H\times W\times \frac{C}{2}}$ respectively. After that, $\widehat{\mathbf{P}}$ and $\widehat{\mathbf{N}}$ will be concatenated as $\mathbf{S} \in \mathbb{R}^{H\times W\times C}$. Finally, $\mathbf{S}$ will be added with $\mathbf{I}$ as the final output feature map $\mathbf{O} \in \mathbb{R}^{H\times W\times C}$. It is worth noting that the process from $\widehat{\mathbf{I}}$ to $\mathbf{S}$ can be seen as a basic module that can replace the "Activation + Convolution" structure in any networks. More details about BRRB can refer to Fig. 3.

3.3 The Overall Network and Loss Function

In this section, we proposed an overall network architecture for the task of image fusion, in which we embed the BRRB into a simple ResNet [16], called BRResNet. Especially, the BRResNet that contains five BRRBs is used to solve the pansharpening problem in this paper. Let $\mathbf{HR} \in \mathbb{R}^{H\times W\times b}$ and $\mathbf{LR} \in \mathbb{R}^{\frac{H}{4}\times\frac{W}{4}\times B}$ represent the high-resolution image and low-resolution image respectively, where B and b denotes the bands in $\mathbf{LR}$ and $\mathbf{HR}$. More information about $\mathbf{HR}$ and $\mathbf{LR}$ can refer to Sec. 4.2.2 and Sec. 4.3.2. Firstly, $\mathbf{LR}$ is upsampled to the same size as $\mathbf{HR}$. Then, the upsampled $\mathbf{LR}$, represented as $\widehat{\mathbf{LR}} \in \mathbb{R}^{H\times W\times B}$, will concatenate with $\mathbf{HR}$ as $\mathbf{M} \in \mathbb{R}^{H\times W\times(b+B)}$. Next, $\widehat{\mathbf{LR}}$ will be fed into the BRResNet and the output will be added with $\widehat{\mathbf{LR}}$ as the final fused image $\mathbf{SR} \in \mathbb{R}^{H\times W\times B}$. The overall process can be expressed as follows:

$$\mathrm{SR} = \mathcal{F}_\theta(\widehat{\mathbf{LR}}; \mathbf{HR}) + \widehat{\mathbf{LR}}, \tag{5}$$

where $\mathcal{F}_\theta(\cdot)$ represents the BRResNet with its parameters θ. More details about BRResNet can refer to Fig. 4.

To depict the distance between $\mathbf{SR}$ and the ground-truth ($\mathbf{GT}$) image, we adopt the mean square error (MSE) as our loss function in the training process. The loss function can be expressed as follows:

$$\mathcal{L}(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left\|\mathcal{F}_\theta(\widehat{\mathbf{LR}}^{(i)}, \mathbf{HR}^{(i)}) + \widehat{\mathbf{LR}} - \mathbf{GT}^{(i)}\right\|_F^2, \tag{6}$$

where N denotes the amount of training examples, and $\|\cdot\|_F$ is the Frobenius norm.

Table 1: Average quantitative comparisons on 11 CAVE examples (Red: the best; Blue: the second best).

Method	*PSNR*	*SAM*	*ERGAS*	*SSIM*
FUSE [35]	39.72 ± 3.52	5.83 ± 2.02	4.18 ± 3.08	0.975 ± 0.018
GLP-HS [27]	37.81 ± 3.06	5.36 ± 1.78	4.66 ± 2.71	0.972 ± 0.015
CSTF [19]	42.14 ± 3.04	9.92 ± 4.11	3.08 ± 1.56	0.964 ± 0.027
CNN-FUS [10]	42.66 ± 3.46	6.44 ± 2.31	2.95 ± 2.24	0.982 ± 0.007
SSRNet [42]	45.28 ± 3.13	4.72 ± 1.76	2.06 ± 1.30	0.990 ± 0.004
ResTFNet [21]	45.35 ± 3.68	3.76 ± 1.31	1.98 ± 1.62	0.993 ± 0.003
MHFNet [36]	46.32 ± 2.76	4.33 ± 1.48	1.74 ± 1.44	0.992 ± 0.006
BRResNet	47.85 ± 3.56	2.96 ± 0.89	1.50 ± 1.18	0.995 ± 0.003
Ideal value	∞	**0**	**0**	**1**

Table 2: Average quantitative comparisons on 10 Harvard examples.

Method	*PSNR*	*SAM*	*ERGAS*	*SSIM*
FUSE [35]	42.06 ± 2.94	3.23 ± 0.91	3.14 ± 1.52	0.977 ± 0.009
GLP-HS [27]	40.14 ± 3.22	3.52 ± 0.96	3.74 ± 1.44	0.966 ± 0.012
CSTF [19]	42.97 ± 3.33	3.30 ± 1.25	2.43 ± 1.06	0.972 ± 0.021
CNN-FUS [10]	43.61 ± 4.73	3.32 ± 1.17	2.78 ± 1.64	0.978 ± 0.016
SSRNet [42]	44.40 ± 3.49	2.61 ± 0.72	2.39 ± 1.02	0.985 ± 0.007
ResTFNet [21]	44.47 ± 4.04	2.56 ± 0.68	2.21 ± 0.87	0.985 ± 0.008
MHFNet [36]	43.10 ± 3.94	2.76 ± 0.77	3.28 ± 1.54	0.977 ± 0.009
BRResNet	45.74 ± 3.86	2.39 ± 0.66	1.94 ± 0.69	0.986 ± 0.009
Ideal value	∞	**0**	**0**	**1**

4 EXPERIMENTS

This section reports the main results of BRResNet in HISR and pansharpening, where the effectiveness of BAM is demonstrated by comparing with the existing state-of-the-art methods.

4.1 Baseline Methods

HISR is a classic task in the field of image fusion. The methods developed in recent years can be divided into traditional methods and deep learning (DL)-based approaches. Competitive traditional methods including FUSE [35], the coupled sparse tensor factorization (CSTF) [19] method and the CNN Denoiser (CNN-FUSE) [10]. Many DL-based methods based on CNN have emerged, pushing the task of HISR to a new era, including SSRNet [42], ResTFNet [21], and MHFNet [36].

Similarly, previous works for pansharpening can also be classified as traditional methods and DL-based methods [31]. Typical traditional methods are the smoothing filter-based intensity modulation (SFIM) [20], the generalized Laplacian pyramid (GLP) [1] with MTF-matched filter [3] and regression-based injection model (GLP-CBD) [5], and the band-dependent spatial-detail with local parameter estimation (BDSD) [14]. Advanced DL-based methods are PanNet [38], DiCNN1 [17], and DMDNet [12].

4.2 Results for HISR

In this section, we will introduce the implementation of training, then, datasets and evaluation indicators will be shown, and finally,

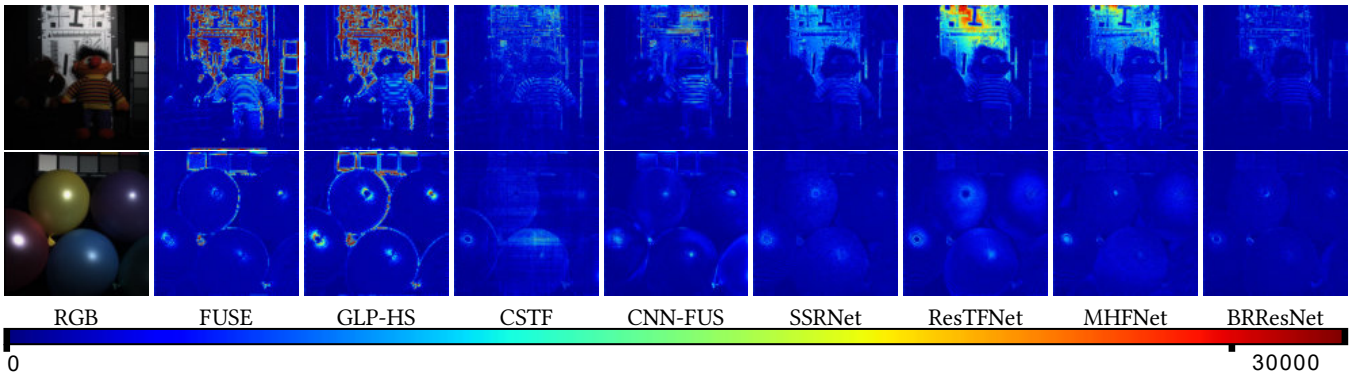

Figure 5: AEMs comparison for HISR on CAVE dataset that is 15 bits and the maximum value is 65535.

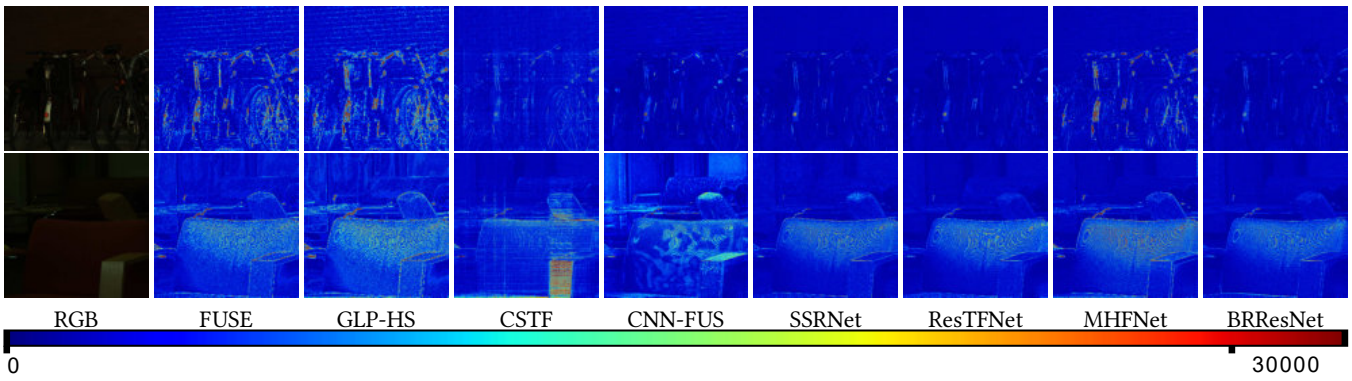

Figure 6: AEMs comparison for HISR on Harvard dataset that is 15 bits and the maximum value is 65535.

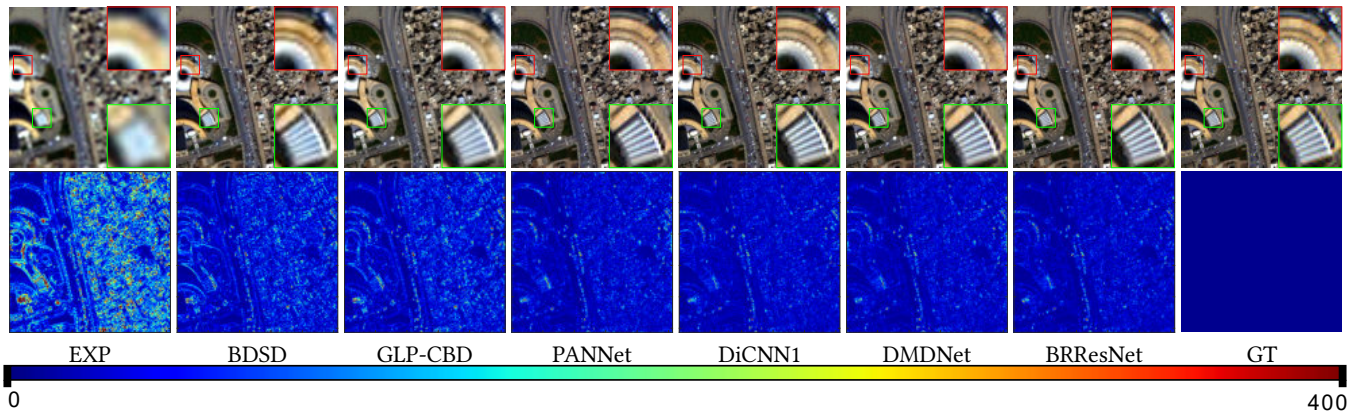

Figure 7: Qualitative comparison on a reduced WV3 data.

the HISR results compared with state-of-the-art methods will be presented.

4.2.1 Training Details and Parameters. All DL-based methods are fairly trained on the same dataset on NVIDIA GeForce GTX 2080Ti. Besides, we set 1000 epochs for the BRResNet training under the Pytorch framework, while the learning rate is set to 1×10^{-4}, the channels of the BRRB is set to 64. Adam optimizer is used for training with the batch size 32 while β_1 and β_2 are set to 0.9 and 0.999, respectively. For the compared approaches, we use the source codes provided by the authors or re-implement the code with the default parameters in the corresponding papers.

4.2.2 Datasets and Evaluation Metrics. In this work, we adopt two widely used datasets: CAVE dataset [39] and Harvard dataset [6]. CAVE dataset includes 32 scenes with the size of 512×512 (31 bands in total), with full spectrum resolution reflectance data from 400nm to 700nm at 10nm increments. Harvard dataset [6] includes 77 indoor and outdoor scenes with the size 1392×1040 (31 bands in total), with full spectrum resolution reflectance data in the range of 420nm to 720nm at 10nm increments. We have simulated a total of 3920 HR-MSI/LR-HSI/GT image pairs (80%/20% as training/testing dataset) with the size 64×64×3, 16×16×31, and 64×64×31 for CAVE dataset, and 3920 HR-MSI/LR-HSI/GT image pairs (80%/20% as training/testing dataset) with the size 64×64×3, 16×16×31, and 64×64×31 for Harvard dataset, respectively. The process of CAVE data generation contains the following three steps: 1) Crop 3920 overlapping patches from the original dataset as GT; 2) Apply a Gaussian blur with the kernel size of 3×3 and standard deviation of 0.5 to GT patches, and then the blurred patches are downsampled to generate LR-HSI patches; 3) Use the spectral response function of Nikon D700 camera [9, 10, 36, 37] to generate MSI patches. Besides, to evaluate the performance of HISR, we adopt the following indicators, SAM, ERGAS, the peak signal-to-noise ratio (PSNR) and the structure similarity (SSIM) [34].

Table 3: Average quantitative comparisons on 1258 reduced resolution WV3 examples.

Method	*SAM*	*ERGAS*	*SCC*	*Q8*
BDSD [14]	6.9997 ± 2.8530	5.1670 ± 2.2475	0.8712 ± 0.0798	0.8126 ± 0.1234
GLP-CBD [5]	5.2861 ± 1.9582	4.1627 ± 1.7748	0.8904 ± 0.0698	0.8540 ± 0.1144
PanNet [38]	4.0921 ± 1.2733	2.9524 ± 0.9778	0.9495 ± 0.0461	0.8942 ± 0.1170
DiCNN1 [17]	3.9805 ± 1.3181	2.7367 ± 1.0156	0.9517 ± 0.0472	0.9097 ± 0.1117
DMDNet [12]	3.9714 ± 1.2482	2.8572 ± 0.9663	0.9527 ± 0.0447	0.9000 ± 0.1142
BRResNet	3.5881 ± 1.2185	2.4618 ± 0.9306	0.9612 ± 0.0444	0.9183 ± 0.1099
Ideal value	**0**	**0**	**1**	**1**

Table 4: Average quantitative comparisons on 36 full resolution WV3 examples.

Method	*QNR*	D_λ	D_s
BDSD [14]	0.9368 ± 0.0416	0.0170 ± 0.0137	0.0473 ± 0.0320
GLP-CBD [5]	0.9107 ± 0.0518	0.0323 ± 0.0243	0.0597 ± 0.0325
PanNet [38]	0.9605 ± 0.01551	0.0215 ± 0.0098	0.0184 ± 0.0074
DiCNN1 [17]	0.9454 ± 0.0268	0.0181 ± 0.0135	0.0374 ± 0.0159
DMDNet [12]	0.9595 ± 0.0155	0.0201 ± 0.0098	0.0209 ± 0.0073
BRResNet	0.9671 ± 0.0095	0.0179 ± 0.0063	0.0152 ± 0.0050
Ideal value	**1**	**0**	**0**

4.2.3 Comparison with State-of-the-art. This section will report the comparison of the results on the CAVE dataset and the Harvard dataset produced by our BRResNet and several advanced methods. Quantitative evaluation results of these approaches for the CAVE dataset are summarized in Table 1, while these approaches for the Harvard dataset are summarized in Table 2. The advantage of the BRResNet could be shown across the board in terms of all assessment metrics. Fig. 5 and Fig. 6 depict a part of the testing fusion outcomes through the absolute error maps (left bottom area of the original image). Visual examination reveals that the proposed approach produces a superior visual impression. Compared to the other results, the output of BRResNet is superior in terms of texture/edge details retention and global intensity.

4.3 Results for Pansharpening

This section will first introduce the training implementation, then datasets and evaluation indicators will be described, and finally, our pansharpening results will be presented.

4.3.1 Training Details and Parameters. We conduct 1000 epochs training under the Pytorch framework, and the learning rate is fixed as 1×10^{-4} during the training process. For the parameters of BRResNet, the number of the BRResBlock is set to 5, while the number of channels of the BRRB is set to 32. The rest of the settings and parameters are the same as that in Sec. 4.2.1.

4.3.2 Datasets and Evaluation Metrics. To benchmark the effectiveness of BRResNet for pansharpening, we adopt a wide range of datasets including 8-band datasets captured by WorldView-3 (WV3), 4-band datasets captured by GaoFen-2 (GF2) and QuickBird (QB) satellites. As the ground truth (GT) images are not available, Wald's protocol [2] is performed to ensure the baseline image generation. All the source data can be download from the public website. For WV3-data, we obtain 12580 HR-PANI/LR-MSI/GT image pairs (70%/20%/10% as training/validation/testing dataset) with the size 64×64×1, 16×16×8, and 64×64×8, respectively; For GF2 data, we use 10000 HR-PANI/LR-MSI/GT image pairs (70%/20%/10% as training/validation/testing dataset) with the size 64×64×1, 16×16×4, and 64×64×4, respectively; For QB data, 20000 HR-PANI/LR-MSI/GT image pairs (70%/20%/10% as training/validation/testing dataset) with the size 64×64×1, 16×16×4, and 64×64×4 were adopted.

The quality evaluation is conducted both at reduced and full resolutions. For reduced resolution test, the spectral angle mapper (SAM) [40], the relative dimensionless global error in synthesis (ERGAS) [33], the spatial correlation coefficient (SCC) [43], and quality index for 4-band images (Q4) and 8-band images (Q8) [13] are used to assess the quality of the results. In addition, to assess the performance of all involved methods on full resolutions, the QNR, the D_λ, and the D_s [32] indexes are applied.

4.3.3 Comparison with State-of-the-art. This section will compare the results on various datasets obtained by our BRResNet and several competitive methods (including traditional techniques and DL-based methods).

Evaluation on 8-band reduced resolution dataset. We compare the proposed method with recent state-of-the-art pansharpening methods on the quantitative performance on 1258 WV3 testing datasets. The results of compared methods and BRResNet are reported in Table 3. It can be observed that BRResNet achieves a transcendence performance. Also, we compare the related approaches on the Rio-dataset (WV3), whose visual results are shown in Fig. 7. We can observe that BRResNet not only provides unambiguous results, but also the color and brightness of the proposed method's results are clearly closer to the LR-MSI (refer to EXP).

Evaluation on 8-band full resolution dataset. We further perform a full-resolution test experiment on the WV3 dataset with 50 pairs. The quantitative results are reported in Table 4, and the visual results are shown in Fig. 10. Again, our method significantly outperforms existing techniques in all quantitative indicators. Furthermore, the compared DL-based techniques PanNtT, DiCNN1, and DMDNet no longer outperform all standard unsupervised approaches, as they did in the previous section. This is because the

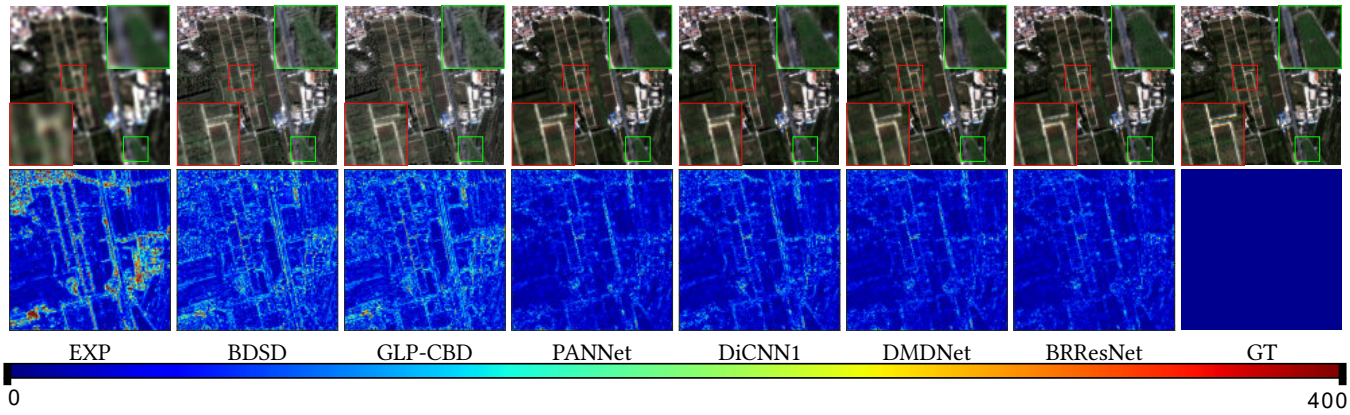

Figure 8: Qualitative comparison on a reduced GF2 data.

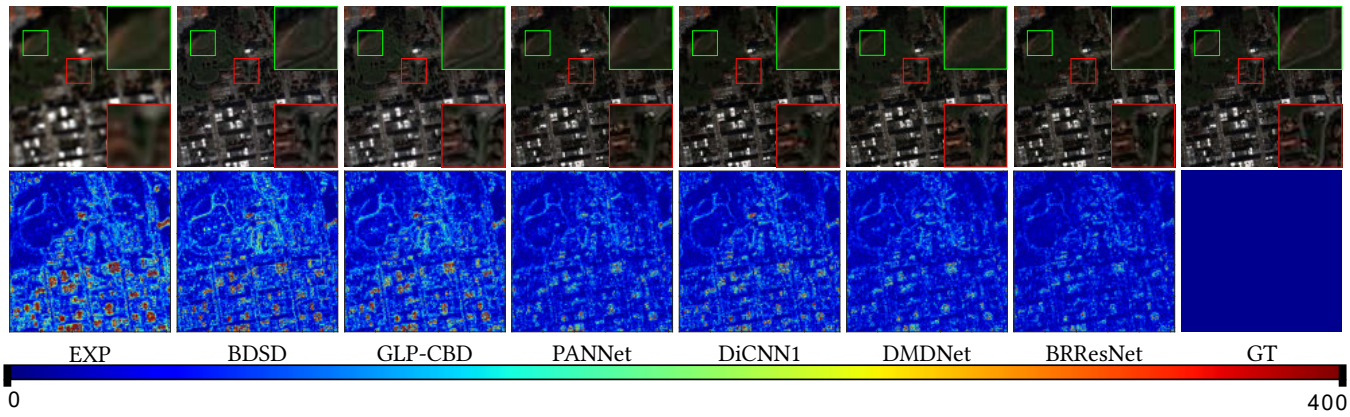

Figure 9: Qualitative comparison on a reduced QB data.

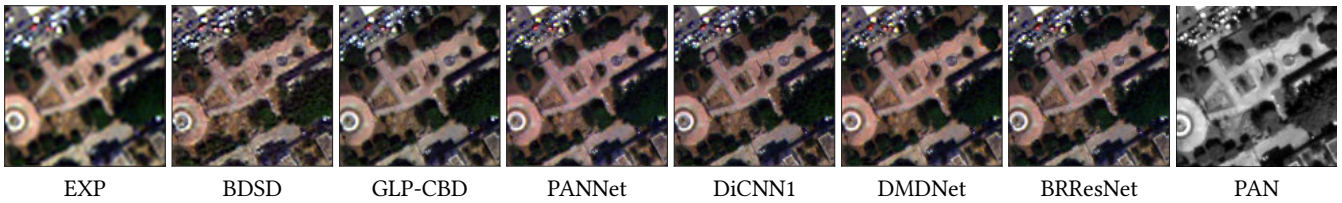

Figure 10: Qualitative comparison on a full resolution WV3 data.

training and testing samples have similar spectral and spatial responses. Traditional deep learning algorithms can only work successfully in constant training-testing scenarios. BRResNet, on the other hand, mitigates this flaw to some extent.

Evaluation on 4-band reduced resolution dataset. In order to prove the wide applicability of BRResNet, we also conducted experiments on the 4-band GF2 and QB datasets. Similarly, the comparison of quantified indicators is shown in Table 6 and Table 5, the visual results are shown in Fig. 8 and Fig. 9. Other competing approaches produce some ambiguity and residual more or less, but our proposed method can generate results that are closest to the GT image, which further proves that BRResNet has a certain generalization ability.

Table 5: Average quantitative comparisons on 81 reduced resolution GF2 examples.

Method	*SAM*	*ERGAS*	*SCC*	*Q4*
BDSD [14]	2.3074 ± 0.2923	2.0704 ± 0.6097	0.8769 ± 0.0516	0.8763 ± 0.0417
GLP-CBD [5]	2.2744 ± 0.7335	2.0461 ± 0.6198	0.8728 ± 0.0527	0.8773 ± 0.0406
PanNet [38]	1.3954 ± 0.3262	1.2239 ± 0.2828	0.9558 ± 0.0123	0.9469 ± 0.0222
DiCNN1 [17]	1.4948 ± 0.3814	1.3203 ± 0.3544	0.9459 ± 0.0223	0.9445 ± 0.0212
DMDNet [12]	1.2968 ± 0.2923	1.1281 ± 0.2670	0.9645 ± 0.0101	0.9530 ± 0.0219
BRResNet	1.2129 ± 0.2923	1.0298 ± 0.2532	0.9686 ± 0.0094	0.9627 ± 0.0175
Ideal value	**0**	**0**	**1**	**1**

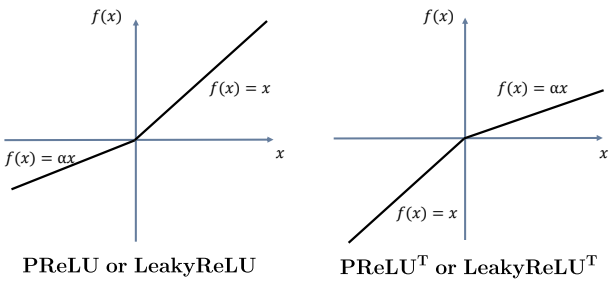

Figure 11: The left one is the diagram of PReLU or LeakyReLU, the right one is the transformed one that denoted as PReLU^T or LeakyReLU^T, the difference between PReLU and LeakyReLU is, PReLU regards α as a trainable parameter while LeakyReLU regards α as an immutable hyperparameter.

Table 6: Average quantitative comparisons on 48 reduced resolution QB examples.

Method	*SAM*	*ERGAS*	*SCC*	*Q4*
BDSD [14]	7.6708 ± 1.9110	7.4661 ± 0.9912	0.8512 ± 0.0622	0.8132 ± 0.1361
GLP-CBD [5]	7.3983 ± 1.7826	7.2965 ± 0.9316	0.8543 ± 0.0643	0.8191 ± 0.1283
PanNet [38]	5.3144 ± 1.0175	5.1623 ± 0.6815	0.9296 ± 0.0586	0.8834 ± 0.1399
DiCNN1 [17]	5.3071 ± 0.9958	5.2310 ± 0.5412	0.9224 ± 0.0507	0.8821 ± 0.1432
DMDNet [12]	5.1197 ± 0.9399	4.7377 ± 0.6487	0.9350 ± 0.0653	0.8908 ± 0.1464
BRResNet	4.5990 ± 0.7882	3.9480 ± 0.2521	0.9541 ± 0.0486	0.9109 ± 0.1367
Ideal value	**0**	**0**	**1**	**1**

From the above experiments, it is clear that apart from the improvement of HISR performance, the proposed BAM also has a favorable performance for pansharpening. We believe that BAM can also achieve satisfactory results for more vision tasks.

4.4 Disscusions

In this section, we discuss the effectiveness of the proposed method from the following three aspects. First, how activation function may affect training results is investigated. Then, we study the performance of BAM with different activation functions. And finally, a comparison of parameters is presented.

Discussion on activation function. We provide experiments that compare three ResNet variants over the Tripoli dataset (A sample from the WV3 datasets). The three ResNet variants use ReLU, LeakyReLU, and PReLU as the activation functions respectively, denoted as ResNet-ReLU, ResNet-LReLU, ResNet-PReLU. As shown in Table 7, ResNet with PReLU has the best results, indicating that the negative part is necessary for feature extraction, but it is not feasible to fix the derivative, like LeakyReLU. It is worth mentioning that BAM can also be used for some of the latest and effective activation functions, such as Mish [23] and Swish [26].

Discussion on BAM. BAM can be applied to different activation functions. Similarly, we select ReLU, LeakyReLU, and PReLU for comparison, denoted as ResNet-BReLU, ResNet-BLReLU, and ResNet-BPReLU respectively. The training and test procedures are the same as the above experiment. The illustration of BPReLU and BLeak-yReLU is shown in Fig. 11. Among them, the coefficient on the negative interval of LeakyReLU is set to 0.2, and the initialization coefficient on the negative interval of PReLU is set to 0.2. The

Table 7: Quantitative comparisons of discussion study on Tripoli dataset (A sample from the WV3 datasets).

Method	*SAM*	*ERGAS*	*SCC*	*Q8*
ResNet-ReLU	4.1367	3.0077	0.9658	0.9667
ResNet-LeakyReLU	4.1845	3.1040	0.9631	0.9533
ResNet-PReLU	4.0843	2.9601	0.9674	0.9561
ResNet-BReLU (our)	4.0124	2.9555	0.9677	0.9562
ResNet-BLeakyReLU (our)	4.0745	2.9694	0.9675	0.9559
ResNet-BPReLU (our)	4.0370	2.9455	0.9679	0.9564
Ideal value	**0**	**0**	**1**	**1**

Table 8: The number of parameters (NoPs). The first two lines are the pansharpening experiment on WV3 dataset, and the last two lines are the HISR experiment.

Method	PanNet	DiCNN1	DMDNet	FusionNet	BRResNet
NoPs	2.5×10^5	1.8×10^5	3.2×10^5	2.3×10^5	0.97×10^5
Method	––	**SSRNet**	**ResTFNet**	**MHFNet**	**BRResNet**
NoPs	––	0.3×10^5	22.6×10^5	36.3×10^5	4.1×10^5

experimental results over the Tripoli dataset are shown in Table 7. It is clear that BAM can enhance the feature extraction ability of the network and obtain the most competitive results on multiple indicators. This is due to the ability of BRRB to receive bilateral contextual information streams, allowing it to extract and apply more extensive representations for image fusion.

Discussion on the number of parameters. The number of parameters (NoPs) of all the compared DL-based methods for two tasks are presented in Table 8. For pansharpening, the BRResNet has only 0.97 million parameters, much less than the other competitive methods. This is due to the fact that the bilateral activation method improves the representation capabilities of the network without increasing any extra parameters. Moreover, for HISR, although the NoPs of the BRResNet is not the least, it is able to achieve a satisfying trade-off between computational burden and performance. Our proposed method has demonstrated remarkable efficacy in experiments while requiring a tolerable amount of computing, indicating that such a BAM is efficient.

5 CONCLUSION

In this work, we introduce a simple but effective Bilateral Activation Mechanism (BAM) that not only retains the nonlinearity of the activation function but also avoids information distortion caused by inactivation. Moreover, a network with residual structure using BAM with ReLU (BRResNet) is proposed, which significantly improves the efficiency of feature extraction in image fusion tasks. Besides, a wide range of experiments confirms that BRResNet exceeds other advanced methods easily with fewer parameters. Finally, through analysis and discussion, BAM can be applied to different activation functions to replace any “Activation + Convolution" structures, thus providing more flexible variants for designing neural networks.

6 ACKNOWLEDGE

This work is supported by NSFC (61702083), Key Projects of Applied Basic Research in Sichuan Province (Grant No. 2020YJ0216), and National Key Research and Development Program of China (Grant No. 2020YFA0714001).

REFERENCES

[1] Bruno Aiazzi, Luciano Alparone, Stefano Baronti, and Andrea Garzelli. 2002. Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis. *IEEE Transactions on Geoscience and Remote Sensing* 40, 10 (2002), 2300–2312.

[2] Bruno Aiazzi, Luciano Alparone, Stefano Baronti, and Andrea Garzelli. 2002. Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis. *IEEE Transactions on Geoscience and Remote Sensing* 40, 10 (2002), 2300–2312.

[3] B Aiazzi, L Alparone, S Baronti, A Garzelli, and M Selva. 2006. MTF-tailored multiscale fusion of high-resolution MS and Pan imagery. *Photogrammetric Engineering & Remote Sensing* 72, 5 (2006), 591–596.

[4] Naveed Akhtar, Faisal Shafait, and Ajmal Mian. 2015. Bayesian sparse representation for hyperspectral image super resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3631–3640.

[5] Luciano Alparone, Lucien Wald, Jocelyn Chanussot, Claire Thomas, Paolo Gamba, and Lori Mann Bruce. 2007. Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest. *IEEE Transactions on Geoscience and Remote Sensing* 45, 10 (2007), 3012–3021.

[6] Ayan Chakrabarti and Todd Zickler. 2011. Statistics of real-world hyperspectral images. In *CVPR 2011*. IEEE, 193–200.

[7] Liang-Jian Deng, Gemine Vivone, Weihong Guo, Mauro Dalla Mura, and Jocelyn Chanussot. 2018. A variational pansharpening approach based on reproducible kernel Hilbert space and heaviside function. *IEEE Transactions on Image Processing* 27, 9 (2018), 4330–4344.

[8] Liang-Jian Deng, Gemine Vivone, Cheng Jin, and Jocelyn Chanussot. 2020. Detail Injection-Based Deep Convolutional Neural Networks for Pansharpening. *IEEE Transactions on Geoscience and Remote Sensing* (2020).

[9] Renwei Dian and Shutao Li. 2019. Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization. *IEEE Transactions on Image Processing* 28, 10 (2019), 5135–5146.

[10] Renwei Dian, Shutao Li, and Xudong Kang. 2020. Regularizing Hyperspectral and Multispectral Image Fusion by CNN Denoiser. *IEEE Transactions on Neural Networks and Learning Systems* (2020), 1–12.

[11] Weisheng Dong, Fazuo Fu, Guangming Shi, Xun Cao, Jinjian Wu, Guangyu Li, and Xin Li. 2016. Hyperspectral image super-resolution via non-negative structured sparse representation. *IEEE Transactions on Image Processing* 25, 5 (2016), 2337–2352.

[12] Xueyang Fu, Wu Wang, Yue Huang, Xinghao Ding, and John Paisley. 2020. Deep multiscale detail networks for multiband spectral image sharpening. *IEEE Transactions on Neural Networks and Learning Systems* (2020).

[13] Andrea Garzelli and Filippo Nencini. 2009. Hypercomplex Quality Assessment of Multi-/Hyper-Spectral Images. *IEEE Geoscience and Remote Sensing Letters* 6, 4 (2009), 662–665.

[14] Andrea Garzelli, Filippo Nencini, and Luca Capobianco. 2007. Optimal MMSE pan sharpening of very high resolution multispectral images. *IEEE Transactions on Geoscience and Remote Sensing* 46, 1 (2007), 228–236.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE International Conference on Computer Vision*. 1026–1034.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 770–778.

[17] Lin He, Yizhou Rao, Jun Li, Jocelyn Chanussot, Antonio Plaza, Jiawei Zhu, and Bo Li. 2019. Pansharpening via detail injection based convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12, 4 (2019), 1188–1204.

[18] Zheng Hui, Xinbo Gao, Yunchu Yang, and Xiumei Wang. 2019. Lightweight image super-resolution with information multi-distillation network. In *Proceedings of the 27th ACM International Conference on Multimedia*. 2024–2032.

[19] Shutao Li, Renwei Dian, Leyuan Fang, and Jose M Bioucasdias. 2018. Fusing Hyperspectral and Multispectral Images via Coupled Sparse Tensor Factorization. *IEEE Transactions on Image Processing* 27, 8 (2018), 4118–4130.

[20] JG Liu. 2000. Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details. *International Journal of Remote Sensing* 21, 18 (2000), 3461–3472.

[21] Xiangyu Liu, Qingjie Liu, and Yunhong Wang. 2020. Remote sensing image fusion based on two-stream fusion network. *Information Fusion* 55 (2020), 1–15.

[22] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. 2013. Rectifier nonlinearities improve neural network acoustic models. In *Proc. ICML*, Vol. 30. Citeseer, 3.

[23] Diganta Misra. 2019. Mish: A self regularized non-monotonic neural activation function. *arXiv preprint arXiv:1908.08681* 4 (2019), 2.

[24] Vinod Nair and Geoffrey E Hinton. 2010. Rectified linear units improve restricted boltzmann machines. In *ICML*.

[25] Prajit Ramachandran, Barret Zoph, and Quoc V Le. 2017. Searching for activation functions. *arXiv preprint arXiv:1710.05941* (2017).

[26] Prajit Ramachandran, Barret Zoph, and Quoc V. Le. 2017. Searching for Activation Functions. *CoRR* abs/1710.05941 (2017).

[27] Massimo Selva, Bruno Aiazzi, Francesco Butera, Leandro Chiarantini, and Stefano Baronti. 2015. Hyper-sharpening: A first approach on SIM-GA data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8, 6 (2015), 3008–3024.

[28] Miguel Simoes, José Bioucas-Dias, Luis B Almeida, and Jocelyn Chanussot. 2014. A convex formulation for hyperspectral image superresolution via subspace-based regularization. *IEEE Transactions on Geoscience and Remote Sensing* 53, 6 (2014), 3373–3388.

[29] Karasawa Takumi, Kohei Watanabe, Qishen Ha, Antonio Tejero-De-Pablos, Yoshitaka Ushiku, and Tatsuya Harada. 2017. Multispectral object detection for autonomous vehicles. In *Proceedings of the on Thematic Workshops of ACM Multimedia 2017*. 35–43.

[30] Youbao Tang, Xiangqian Wu, and Wei Bu. 2016. Deeply-supervised recurrent convolutional neural network for saliency detection. In *Proceedings of the 24th ACM International Conference on Multimedia*. 397–401.

[31] Gemine Vivone, Luciano Alparone, Jocelyn Chanussot, Mauro Dalla Mura, Andrea Garzelli, Giorgio A Licciardi, Rocco Restaino, and Lucien Wald. 2014. A critical comparison among pansharpening algorithms. *IEEE Transactions on Geoscience and Remote Sensing* 53, 5 (2014), 2565–2586.

[32] Gemine Vivone, Luciano Alparone, Jocelyn Chanussot, Mauro Dalla Mura, Andrea Garzelli, Giorgio A. Licciardi, Rocco Restaino, and Lucien Wald. 2015. A critical comparison among pansharpening algorithms. *IEEE Transactions on Geoscience and Remote Sensing* 53, 5 (2015), 2565–2586.

[33] Lucien Wald. 2002. Data fusion: definitions and architectures: Fusion of images of different spatial resolutions. *Presses des MINES* (2002).

[34] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13, 4 (2004), 600–612.

[35] Qi Wei, Nicolas Dobigeon, and Jean-Yves Tourneret. 2015. Fast fusion of multi-band images based on solving a Sylvester equation. *IEEE Transactions on Image Processing* 24, 11 (2015), 4109–4121.

[36] Qi Xie, Minghao Zhou, Qian Zhao, Zongben Xu, and Deyu Meng. 2020. MHF-Net: An Interpretable Deep Network for Multispectral and Hyperspectral Image Fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PP, 1–1. https://doi.org/10.1109/TPAMI.2020.3015691

[37] Ting Xu, Ting-Zhu Huang, Liang-Jian Deng, Xi-Le Zhao, and Jie Huang. 2020. Hyperspectral Image Super-resolution Using Unidirectional Total Variation with Tucker Decomposition. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (2020).

[38] Junfeng Yang, Xueyang Fu, Yuwen Hu, Yue Huang, Xinghao Ding, and John Paisley. 2017. PanNet: A deep network architecture for pan-sharpening. In *Proceedings of the IEEE International Conference on Computer Vision*. 5449–5457.

[39] Fumihito Yasuma, Tomoo Mitsunaga, Daisuke Iso, and Shree K Nayar. 2010. Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum. *IEEE Transactions on Image Processing* 19, 9 (2010), 2241–2253.

[40] Roberta H Yuhas, Alexander FH Goetz, and Joe W Boardman. 1992. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm. In *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, Vol. 1. 147–149.

[41] Huanrong Zhang, Zhi Jin, Xiaojun Tan, and Xiying Li. 2020. Towards Lighter and Faster: Learning Wavelets Progressively for Image Super-Resolution. In *Proceedings of the 28th ACM International Conference on Multimedia*. 2113–2121.

[42] Xueting Zhang, Wei Huang, Qi Wang, and Xuelong Li. 2020. SSR-NET: Spatial-Spectral Reconstruction Network for Hyperspectral and Multispectral Image Fusion. *IEEE Transactions on Geoscience and Remote Sensing* (2020), 1–13. https://doi.org/10.1109/TGRS.2020.3018732

[43] J Zhou, DL Civco, and JA Silander. 1998. A wavelet transform method to merge Landsat TM and SPOT panchromatic data. *International Journal of Remote Sensing* 19, 4 (1998), 743–757.

[44] Qiang Zhou, Shifeng Chen, Jianzhuang Liu, and Xiaoou Tang. 2011. Edge-preserving single image super-resolution. In *Proceedings of the 19th ACM International Conference on Multimedia*. 1037–1040.