# SDRCNN: A Single-Scale Dense Residual Connected Convolutional Neural Network for Pansharpening

Yuan Fang, Yuanzhi Cai, *Graduate Student Member, IEEE*, and Lei Fan, *Member, IEEE*

***Abstract*—Pansharpening is a process of fusing a high spatial resolution panchromatic image and a low spatial resolution multispectral (MS) image to create a high-resolution MS image. A novel single-branch, single-scale lightweight convolutional neural network, named SDRCNN, is developed in this article. By using a novel dense residual connected structure and convolution block, SDRCNN achieved a better tradeoff between accuracy and efficiency. The performance of SDRCNN was tested using four datasets from the WorldView-3, WorldView-2, and QuickBird satellites. The compared methods include eight traditional methods (i.e., GS, Gram–Schmidt adaptive, partial replacement adaptive CS, band-related spatial detail, smoothing-filter-based intensity modulation, GLP-CBD, CDIF, and LRTCFPan) and five lightweight deep-learning methods (i.e., pansharpening neural network, PanNet, BayesianNet, DMDNet, and FusionNet). Based on a visual inspection of the pansharpened images created and the associated absolute residual maps, SDRCNN exhibited least spatial detail blurring and spectral distortion, among all the methods considered. The values of the quantitative evaluation metrics were closest to their ideal values when SDRCNN was used. The processing time of SDRCNN was also the shortest among all methods tested. Finally, the effectiveness of each component in the SDRCNN was demonstrated in ablation experiments. All of these confirmed the superiority of SDRCNN.**

***Index Terms*—Convolutional neural network (CNN), deep learning (DL), fusion, multispectral (MS) image, pansharpening, resolution.**

## I. Introduction

WITH the rapid development of earth observation satellites, remotely sensed images have widely been used for various applications, such as object detection and semantic segmentation [1], [2], [3], [4]. However, due to the physical constraints of the existing single sensors, a tradeoff between the spectral resolution and the spatial resolution of an image to be acquired needs to be considered [5], [6]. Therefore, remote sensing satellites typically carry two imaging sensors to capture multispectral (MS) images and panchromatic (PAN) images, respectively. MS images consist of multiple bands at the cost of a relatively low spatial resolution, while PAN images contain finer spatial details in a single image band. To combine their advantages, pansharpening is a typical technique used to fuse an MS image and a PAN image to form a high-resolution multispectral (HRMS) image that shares the same spatial resolution of the PAN image and the same spectral resolution of the MS image. An example of this process is illustrated in Fig. 1. Pansharpening has become a preprocessing step of image enhancement in many remote sensing tasks, such as object detection [7], [8], anomaly detection [9], [10], and agricultural management [11], [12].

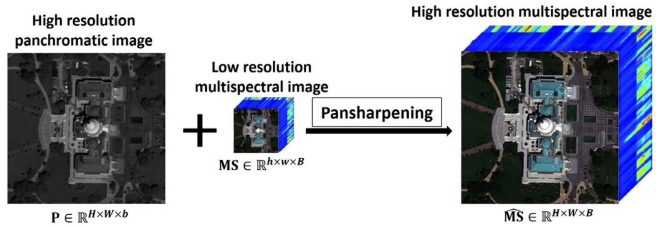


Fig. 1. Example of pansharpening: $H \times W$ and $h \times w$ represent the image size ($h < H$ and $w < W$), $b$ and $B$ represent the number of image bands ($1 = b < B$).

Over the last few decades, various methods have been proposed to achieve pansharpening. They can be divided into four categories, including component substitution (CS) methods, multiresolution analysis (MRA) methods, variational optimization (VO) techniques, and deep-learning (DL) approaches. CS-based methods have widely been used due to their simple principle and easy implementation. They project an interpolated MS image into a transformed domain to separate the spectral and the spatial information. The separated spatial components are then replaced by a PAN image, followed by an inverse transformation to generate the HRMS image. Representative CS-based methods include intensity–hue–saturation [13], principle component analysis (PCA) [14], Gram–Schmidt adaptive (GSA) technique [15], partial replacement adaptive CS (PRACS) [16], and band-related spatial detail (BDSD) scheme [17]. These CS-based methods are effective in increasing spatial resolution, but incur spectral distortion [18]. MRA-based methods inject spatial details extracted from PAN images by spatial filtering into interpolated MS images. Spatial details can be extracted by different decomposition methods. Compared to CS-based methods, this class of methods preserves spectral information well. However, artifacts are prone to occur due to aliasing effects, resulting in spatial distortion [6]. MRA-based methods

Manuscript received 9 March 2023; revised 23 May 2023; accepted 29 June 2023. Date of publication 5 July 2023; date of current version 19 July 2023. This work was supported in part by the Xi'an Jiaotong-Liverpool University Research Enhancement Fund under Grant REF-21-01-003, and in part by the Xi'an Jiaotong-Liverpool University Postgraduate Research Scholarship under Grant PGRS2006010. *(Corresponding author: Lei Fan.)*

The authors are with the Department of Civil Engineering, Design School, Xi'an Jiaotong-Liverpool University, Suzhou 215000, China, and also with the School of Engineering, University of Liverpool, L69 3BX Liverpool, U.K. (e-mail: yuan.fang16@student.xjtlu.edu.cn; yuanzhi.cai19@student.xjtlu.edu.cn; lei.fan@xjtlu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3292320

include wavelet transform [19], smoothing-filter-based intensity modulation (SFIM) [20], additive wavelet luminance proportion [21], and the modulation transfer function generalized Laplacian pyramid with full-resolution regression-based injection model [22]. VO-based methods mainly need to construct and optimize an energy function. After the P+XS [23] method was proposed, such methods have received more attention and research efforts. They can well model the relationships between PAN, MS, and HRMS images and produce a high-quality image fusion. However, compared to CS-based and MRA-based pansharpening methods, VO-based methods are complex in operation and computationally expensive, which limit their wide applications [24], [25], [26].

In recent years, the successful application of DL in computer vision has attracted a lot of attention [27], [28], [29], [30], [31], [32]. In particular, convolutional neural networks (CNNs) stand out for powerful nonlinear modeling capabilities and have made significant progress in pansharpening [33], [34], [35]. As a pioneering attempt, Masi et al. [33] proposed a simple three-layer pansharpening neural network (PNN) that was based on a modification of the super-resolution CNN [36] and led to promising results. Based on PNN, Wei et al. [37] proposed a deeper residual pansharpening network. Yuan et al. [38] introduced multilevel feature extraction and proposed a multiscale multidepth CNN. Shao and Cai [39] proposed a two-branch network called RSIFNN. It was noticed that the majority of DL improve the pansharpening accuracy by utilizing complex network structure designs and/or diverse computational units [40], which are often achieved by using multiscale structures [41], [42], [43], [44], multibranch structures [45], [46], [47], [48], deeper network [49], generative adversarial networks [50], [51], [52], [53], attention mechanism [54], [55], [56], and transformer modules [57], [58], [59]. However, these would make CNNs far larger in size (i.e., number of parameters) than PNN, and consequently lead to reduced computational efficiency [49], [60], [61].

The efficiency of the PAN sharpening methods is important for practical applications, especially considering the huge volume of data captured by satellite. Therefore, it is of interest to develop a lightweight network for higher pansharpening accuracy. To this end, a single-branch, single-scale, lightweight convolutional network named SDRCNN is developed in this article. In SDRCNN, novel convolutional block and dense residual connected structures are designed to recover spatial details from coarse to fine. The effectiveness and the efficiency of SDRCNN are tested using MS images acquired by three satellites, including WorldView-2, WorldView-3, and QuickBird. The major contributions of this article are as follows:

1) Development of a lightweight pansharpening network (i.e., SDRCNN) to achieve a better tradeoff between accuracy and efficiency.
2) A comprehensive fair (i.e., with similar numbers of parameters) evaluation of the performance of various lightweight network structures and the convolutional blocks.

The rest of the article is organized as follows. Section II describes the methodology. Experimental results and the associated discussions are presented in Sections III and IV, respectively. Finally, Section V concludes this article.

## II. Methodology

### A. Overall Architecture of SDRCNN

As illustrated in Fig. 2, the proposed SDRCNN takes a PAN image and an upsampled low-resolution multi-spectral (LRMS) image as the inputs. Different levels of structural details are extracted through a Stem Block and three Residual Blocks, which are arranged in a concatenated manner. Using the proposed dense residual connection mechanism, 52 feature maps are obtained at each of the 3 different network depths (i.e., after the three Residual Blocks). In addition to the Stem Block, reaching these three network depths requires passing one, two, and three identical residual blocks, respectively. A total of 156 feature maps are concatenated and passed through a layer of $1 \times 1$ convolution operations to enable the transfer of different levels of feature information, generating a data cube with the same number of channels as the input MS image. The output data cube is then directly summed with the input upsampled LRMS to produce the final HRMS image. The L1 loss is adopted in the network as the loss function in this article. An elaboration of proposed SDRCNN is presented in Sections II-B, II-C, and II-D.

### B. Preprocessing

The existing CNN-based pansharpening methods usually rely on large-scale training datasets to learn the nonlinear mapping between the input PAN and MS images and the ground-truth HRMS image. However, this process treats pansharpening as an image regression problem in a black-box learning process. Ideally, pansharpening is to inject useful spatial details from the input PAN image into the input MS image. In order to better preserve the spectral information, our network directly maps an input upsampled MS image to the output of the network, which achieves lossless propagation of the spectral information on the one hand, and makes the network more focused on the extraction of spatial features on the other, thus significantly reducing the difficulty of network learning. The ablation experiment results in Section III-E3 show that adding spectral mapping can improve the spectral accuracy of the fused image.

### C. Dense Residual Connection

To further improve the learning capability and efficiency of feature extraction in SDRCNN, we propose a dense residual connection strategy, as shown in Fig. 2. As the key idea of this strategy, three different levels of feature maps are integrated and added to the input upsampled LRMS to obtain HRMS images with more spatial details. In other words, features output from Residual Blocks of different depths are stacked with the feature information output from all previous blocks and fed into the next Residual Block, thus making fuller use of the residual information at different levels to learn features from coarse to fine to enhance the network's ability on reconstructing the spatial information of the fused image. The inputs to the Stem Block and each of the three residual blocks in SDRCNN can be expressed as the following formulae:

$$I_S = \mathrm{PAN} + \mathrm{LRMS}\uparrow_4 \tag{1}$$

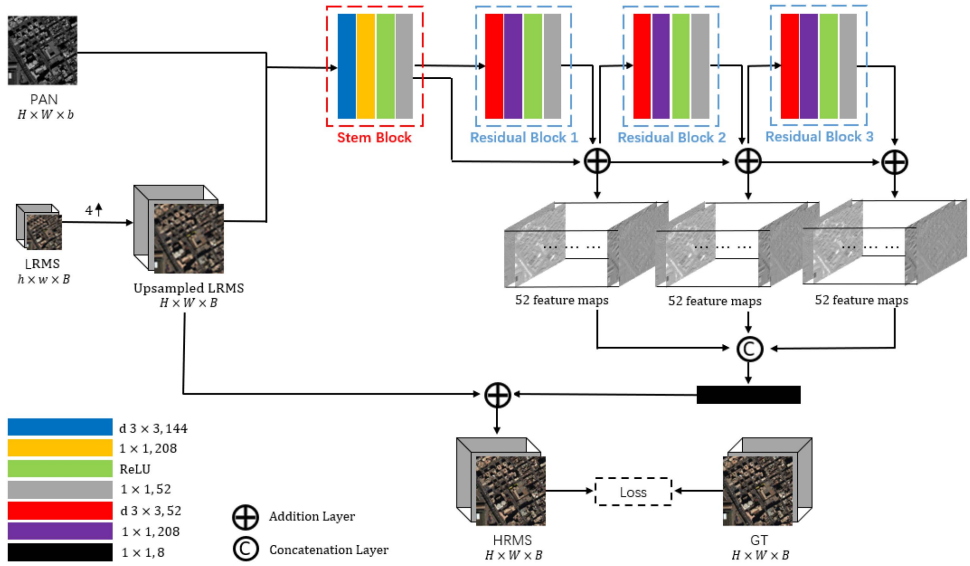


Fig. 2. Overall architecture of proposed SDRCNN.

$$I_R^1 = F_S \tag{2}$$

$$I_R^2 = F_S \oplus F_R^1 \tag{3}$$

$$I_R^3 = (F_S \oplus F_R^1) \oplus F_R^2 \tag{4}$$

where $I_S$ and $F_S$ are the input and output feature maps of the Stem Block, respectively; $I_R^i$ and $F_R^i$ are the input and output feature maps of the $i$th Residual Block, respectively; $\uparrow_4$ represents the operation of upsampling by a factor of 4; $+$ indicates a series connection; and $\oplus$ indicates the operation of adding the values of each element.

Using this dense residual connection mechanism, the Addition Layer following each Residual Block numerically sums the residuals obtained from all blocks before that network depth and forms a set of feature maps (i.e., 52 maps). The output features from a deeper layer would include more spatial information. In total, 156 feature maps from the three levels are concatenated to propagate and integrate the information at different levels through a $1 \times 1$ convolution operation. The generated residual information is summed with the input upsampled LRMS to produce a fused HRMS image. This dense residual connection mechanism allows the proposed single-branch, single-scale SDRCNN to learn diverse (i.e., different levels) features similar to that of complex (e.g., multibranch, multiscale) networks.

### D. Structure of Functional Blocks

Two major types of convolution blocks are used in this article, i.e., Stem Block and Residual Block. Their detailed structures are shown in Fig. 3. The Stem Block is used to integrate the input images of the network into a data cube with predetermined width to obtain the number of channels appropriate to the Residual Block. The Residual Blocks are used for feature extraction and processing.

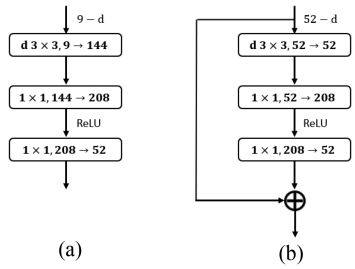


Fig. 3. Illustration of the convolution blocks used in SDRCNN. (a) Stem Block. (b) Residual Block.

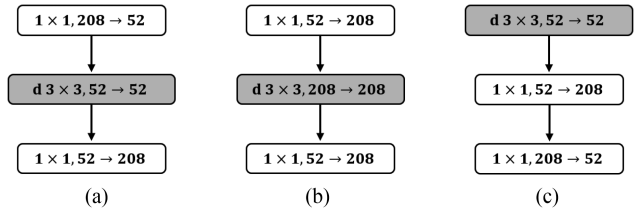


Fig. 4. Process of modifying the block configuration and the associated specifications. (a) Bottleneck block. (b) Inverted bottleneck block. (c) Adopted block with the position of the spatial depthwise convolutional layer being moved to top.

Inspired by the well-known network ResNet [62], the same residual connections are used in our Residual Blocks, which can solve the problem of exploding or disappearing gradients. In order to reduce the number of parameters to be learned for convolutional computation, all the $3 \times 3$ convolutional layers in the network use depthwise separable convolution [63], thus improving the computational efficiency of SDRCNN. Unlike the usual bottleneck configuration [Fig. 4(a)], the inverted bottleneck structure [Fig. 4(b)] was employed to avoid the loss from compressed dimensions during the information transformations

TABLE I
SEVERAL BENCHMARK PANSHARPENING METHODS CONSIDERED FOR COMPARISONS

| **Traditional methods** |
|---|
| GS: Gram–Schmidt transformation [73] |
| GSA: Gram–Schmidt adaptive [15] |
| PRACS: Partial replacement adaptive CS approach [16] |
| BDSD: Band-related spatial detail scheme [17] |
| SFIM: Smoothing filter intensity modulation [20] |
| GLP-CBD: GLP with MTF-matched filter and regression-based injection model [74], [75] |
| CDIF: Context-aware details injection fidelity with adaptive coefficients estimation for variational pansharpening [76] |
| LRTCFPan: Low-rank tensor completion-based framework for pansharpening [77] |
| **DL-based methods** |
| PNN: CNN for pansharpening [33] |
| PanNet: A deep network architecture for pansharpening [65] |
| BayesianNet: Bayesian pansharpening with multiorder Gradient-based deep network constraints [66] |
| DMDNet: Deep multiscale detail networks for multiband spectral image sharpening [67] |
| FusionNet: Deep CNN inspired by traditional CS and MRA methods [60] |

between different dimensional feature spaces. However, because the inverted bottleneck amplifies the intermediate convolutional layers, a direct replacement will lead to an increase in the number of parameters. Therefore, the depthwise separable convolutional layer is moved to the top of a block [Fig. 4(c)] to reduce the computational effort. This block configuration [Fig. 4(c)] is adopted for the Stem Block and the Residual Blocks in SDRCNN.

In addition, the two convolution blocks (i.e., the Stem Block and the Residual Block) used in this article removed the commonly used batch normalization (BN) layers. For most cases, using BN can accelerate the training and lower the sensitivity of the network to initialization. However, our trial experiments indicated that using BN did not consistently improve the pansharpening accuracy. Simultaneously, due to the additional computation and memory usage of executing BN, the BN layers inside the ResNet Block are discarded in our approach. Finally, in each block, only one ReLU activation function layer is set between two $1 \times 1$ layers.

### E. Datasets

Four datasets from WorldView-2, WorldView-3, and QuickBird (three different sensors working in the visible and near-infrared spectral ranges) were used to test our network SDRCNN in this article. WorldView-2 and WorldView-3 provide MS images with eight bands (coastal, blue, green, yellow, red, red edge, near-infrared 1, and near-infrared 2) and single-band PAN images. The spatial resolutions of these MS and PAN images are approximately 1.8 and 0.5 m, respectively, for WorldView-2, 1.2 and 0.3 m, respectively, for WorldView-3. QuickBird provides MS images with four bands (red, green, blue, and near-infrared) and single-band PAN images. The spatial resolutions are approximately 2.44 m for MS images and 0.61 m for PAN images.

### F. Implementation Simulation

*1) Dataset Simulation:* Since there is no ground truth (GT) image as a reference, the Wald's protocol [64] was adopted. According to this protocol, the original MS images were used as the GT images. In addition, the original MS and PAN images were simultaneously blurred and downsampled to produce modified images that were used as input images. Taking the two WorldView-3 datasets (i.e., Tripoli and Rio) as an example, each dataset was simulated with 12 580 samples (also called patch pairs), in which each sample includes PAN ($256 \times 256$ in size), LRMS ($64 \times 64 \times 8$ in size), and GT ($256 \times 256 \times 8$ in size) patches. These 12 580 samples were randomly split into 70%, 20%, and 10% for training, validation, and test, respectively. Furthermore, another eight-band dataset (i.e., WorldView-2) and a four-band dataset (i.e., QuickBird) were also used for performance evaluations in this article.

*2) Benchmark:* To check the performance of our network SDRCNN, several representative pansharpening methods of various categories were implemented for comparisons, as summarized in Table I. The traditional methods were considered because they are still widely used in applications. In terms of DL-based methods, PNN, PanNet [34], and three state-of-the-art lightweight networks with the number of parameters being consistent with PNN were considered. For fair comparisons, PanNet, BayesianNet [65], DMDNet [66], FusionNet [60], and SDRCNN were adjusted to have almost the same number of parameters (i.e., approximately 100 000 parameters) as PNN.

## III. EXPERIMENTAL RESULTS

### A. Reduced-Resolution Assessment

The reduced-resolution assessment measures the similarity between a pansharpened HRMS image and a reference image (i.e., the original MS image). The similarity can be determined by several evaluation indices, such as the spectral angle mapper (SAM) [67], the dimensionless global error in synthesis (ERGAS) [68], the spatial correlation coefficient (SCC) [69], and the Q2n [70] (e.g., Q8 for eight-band datasets and Q4 for four-band datasets). The ideal values are 0 for SAM and ERGAS, and 1 for Q2n and SCC.

As described in Section II-F1, there are two groups of test samples from WorldView-3 datasets (i.e., Tripoli and Rio). Each consisted of 1258 test samples and was processed for comparing SDRCNN with 5 representative CNN-based pansharpening methods (i.e., PNN, PanNet, BayesianNet, DMDNet, and FusionNet). As shown in Table II, SDRCNN achieved the best average quantitative performance in all the metrics, demonstrating that it performed better than the other methods compared. This

TABLE II
AVERAGE VALUES OF THE METRICS FOR PROPOSED SDRCNN AND ALL THE COMPARED CNNS, BASED ON 1258 REDUCED-RESOLUTION SAMPLES: (A) TRIPOLI DATASET AND (B) RIO DATASET

| Method | (a) Tripoli dataset | | | | (b) Rio dataset | | | |
|---|---|---|---|---|---|---|---|---|
| | ERGAS (±std) | SAM (±std) | SCC (±std) | Q8 (±std) | ERGAS (±std) | SAM (±std) | SCC (±std) | Q8 (±std) |
| PNN | 4.6067±0.6705 | 6.4517±1.4093 | 0.9022±0.0402 | 0.8528±0.1306 | 3.0520±0.6450 | 3.9317±0.6641 | 0.8899±0.0580 | 0.8750±0.0860 |
| PanNet | 4.2245±0.6598 | 5.8269±1.1900 | 0.9115±0.0413 | 0.8674±0.1380 | 2.6781±0.4700 | 3.1122±0.5332 | 0.9367±0.0183 | 0.9272±0.0534 |
| BayesianNet | 4.1116±0.6402 | 5.7082±1.1886 | 0.9176±0.0391 | 0.8726±0.1318 | 2.5070±0.5210 | 2.9861±0.5368 | 0.9250±0.0154 | 0.9184±0.0351 |
| DMDNet | 4.1386±0.6417 | 5.7717±1.2069 | 0.9171±0.0381 | 0.8727±0.1314 | 2.5151±0.6464 | 3.0815±0.6339 | 0.9227±0.0529 | 0.9151±0.0698 |
| FusionNet | 4.2349±0.6372 | 5.8100±1.2270 | 0.9139±0.0384 | 0.8666±0.1323 | 2.5275±0.6313 | 2.9938±0.6174 | 0.9241±0.0521 | 0.9181±0.0690 |
| **SDRCNN** | **4.0567±0.5880** | **5.5678±1.1743** | **0.9211±0.0343** | **0.8765±0.1297** | **2.4980±0.5659** | **2.9763±0.6170** | **0.9269±0.0508** | **0.9190±0.0681** |
| Ideal value | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |

The bold values stand for the results obtained by our methods.

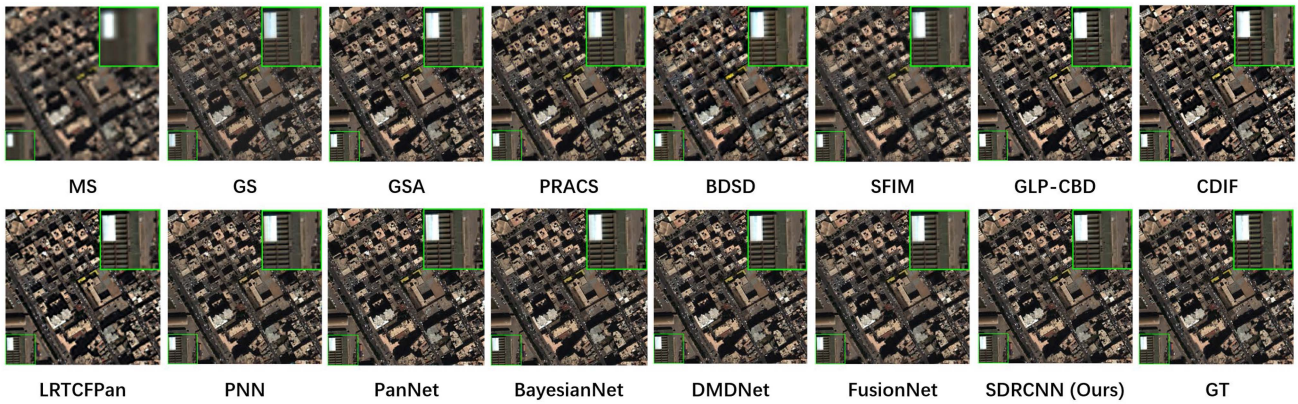


Fig. 5. Visual comparisons of the HRMS images from the compared methods on the reduced-resolution Tripoli dataset (sensor: WorldView-3).

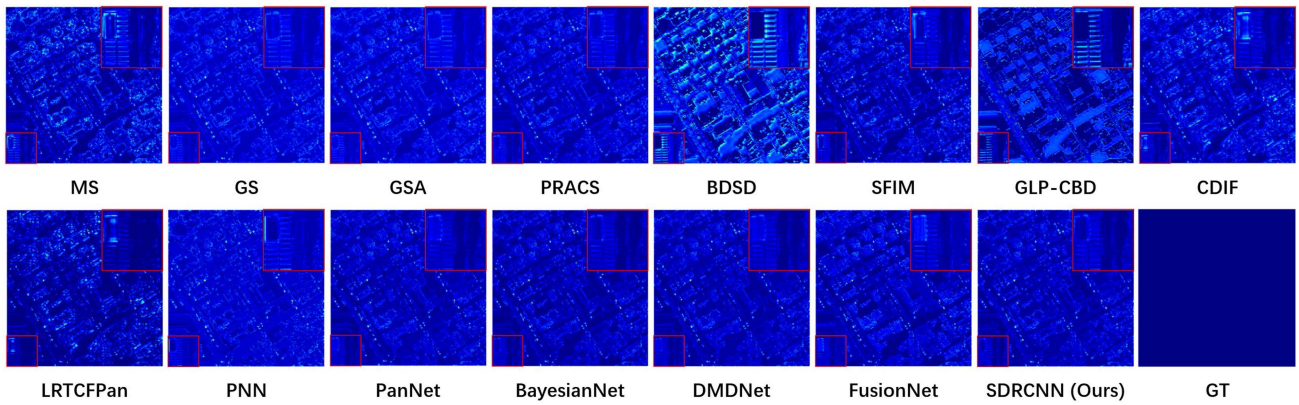


Fig. 6. Corresponding AEMs using the reference (GT) image on the reduced-resolution Tripoli dataset (sensor: WorldView-3).

can be justified as compared to other CNN-based pansharpening methods, SDRCNN utilizes a dense residual connection mechanism to more fully exploit the residual information of different layers and achieves feature extraction for different levels from coarse to fine. In addition, combined with the proposed residual blocks, SDRCNN can obtain more detailed spatial features using a deeper network structure without increasing the number of parameters, thus improving the learning capability of the network.

In addition, two test cases using Rio and Tripoli at a reduced resolution by applying the Wald's protocol stated in Section II-F1 were generated. Figs. 5 and 7 show the visual comparisons of the pansharpened images derived using all the pansharpening methods considered for Tripoli and Rio, respectively. These visualizations were based on the three spectral channels: Red, Green, and Blue. For a better visualization of their differences, a small local area (highlighted by the smaller square) was selected, zoomed in and shown in the larger square at the corner. Figs. 6 and 8 show the corresponding residuals, i.e., absolute error maps (AEMs), between the pansharpened images and GT images for Tripoli and Rio, respectively. The pansharpened images from the traditional methods (i.e., GS,

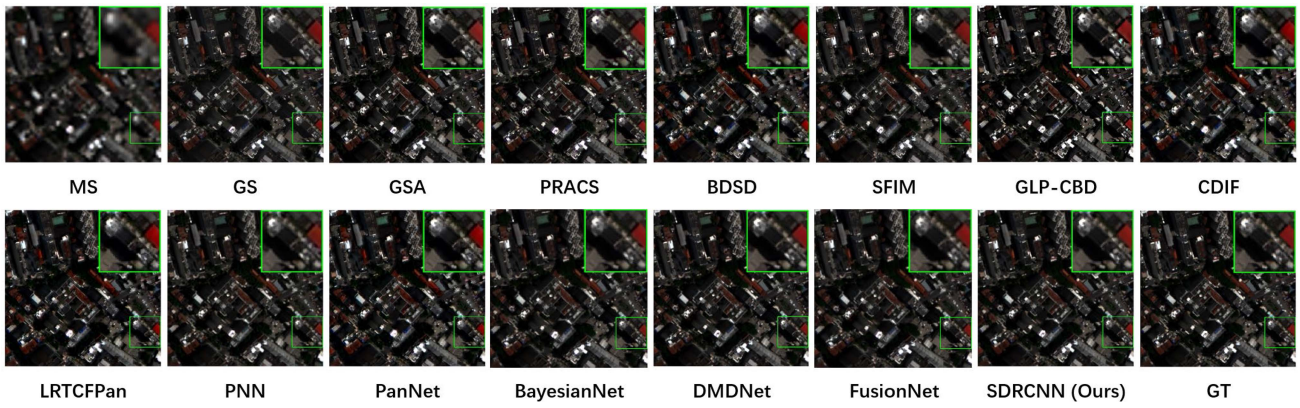


Fig. 7. Visual comparisons of the HRMS images from the compared methods on the reduced-resolution Rio dataset (sensor: WorldView-3).

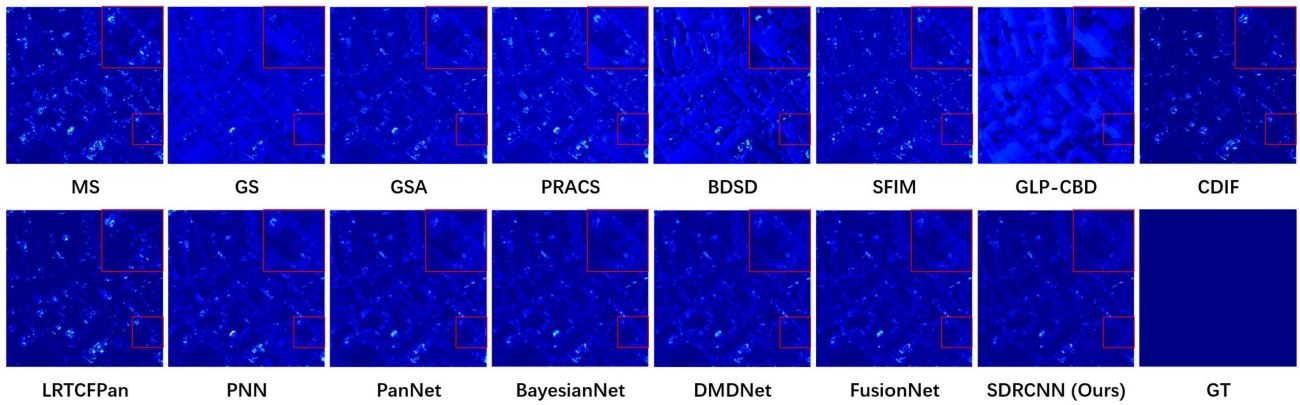


Fig. 8. Corresponding AEMs on the reduced-resolution Rio dataset (sensor: WorldView-3).

GSA, PRACS, BDSD, SFIM, GLP-CBD, CDIF, and LRTCF-Pan) showed some large spatial blurring and spectral distortions (indicated here by the color distortions), particularly near the boundaries of buildings. All the CNN methods considered performed better (spatially and spectrally) than the traditional methods. Due to the size of the individual images shown in Figs. 5 and 7, it is not easy to distinguish the visual differences between the pansharpened images from the CNN methods. However, some differences can be observed in Figs. 6 and 8. For example, in the AEMs of SDRCNN, the color of the zoomed region has a larger area tending to be the same dark blue as GT and showed less spectral residuals at the object boundaries, which suggests a good spectral preservation. Meanwhile, the AEMs of SDRCNN showed less details and textures than the other methods, which suggests that SDRCNN achieved the best spatial preservation. To better show their differences in performance, the values of the quantitative evaluation metrics were calculated and shown in Table III, which confirm that SDRCNN obtained the best performance with the smallest spatial and spectral distortions in the pansharpened images.

Figs. 9 and 10 show the loss curves with the iterations for the considered network structures trained on Tripoli and Rio datasets, respectively. To better evaluate the convergence of the networks, the vertical coordinates of the graphs use the mean value of the training loss for a total of 100 iterations before this iteration, thus reducing the fluctuation of the curves. It is clear that the SDRCNN exhibits the lowest test error after the curve has stabilized, which is consistent with the previous quantitative evaluation. In addition, the convergence rate of all these networks considered is approximately the same.

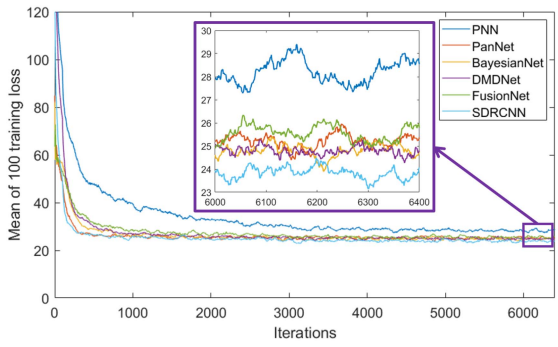


Fig. 9. Training loss curves for considered networks on Tripoli dataset (sensor: WorldView-3).

TABLE III
QUALITY METRICS FOR ALL THE COMPARED APPROACHES ON THE REDUCED-RESOLUTION TRIPOLI AND RIO DATASETS, RESPECTIVELY: (A) TRIPOLI DATASET AND (B) RIO DATASET

| Method | (a) Tripoli dataset | | | | (b) Rio dataset | | | |
|---|---|---|---|---|---|---|---|---|
| | ERGAS | SAM | SCC | Q8 | ERGAS | SAM | SCC | Q8 |
| GS | 5.3307 | 6.9476 | 0.9310 | 0.8842 | 4.0774 | 4.4937 | 0.9451 | 0.9435 |
| GSA | 4.3501 | 6.1909 | 0.9352 | 0.9294 | 3.2516 | 3.9121 | 0.9462 | 0.9668 |
| PRACS | 4.4824 | 6.5096 | 0.9320 | 0.9228 | 3.3292 | 4.4638 | 0.9377 | 0.9668 |
| BDSD | 6.2745 | 8.5998 | 0.8795 | 0.8780 | 4.5744 | 5.9080 | 0.9367 | 0.9450 |
| SFIM | 4.9953 | 6.5628 | 0.9113 | 0.9076 | 3.5697 | 3.9780 | 0.9299 | 0.9606 |
| GLP-CBD | 4.8034 | 7.1610 | 0.9249 | 0.9308 | 5.1089 | 4.7943 | 0.9311 | 0.9367 |
| CDIF | 4.8266 | 6.8797 | 0.9341 | 0.9334 | 4.4216 | 4.0562 | 0.9516 | 0.9531 |
| LRTCFPan | 4.7632 | 6.6310 | 0.9350 | 0.9355 | 4.1438 | 3.9058 | 0.9558 | 0.9592 |
| PNN | 4.7168 | 6.9136 | 0.9285 | 0.9206 | 3.1055 | 4.2022 | 0.9443 | 0.9690 |
| PanNet | 4.1971 | 6.1086 | 0.9388 | 0.9374 | 2.6961 | 3.4805 | 0.9565 | 0.9759 |
| BayesianNet | 4.1126 | 6.0088 | 0.9434 | 0.9385 | 2.5537 | 3.3158 | 0.9601 | 0.9790 |
| DMDNet | 4.1425 | 6.0966 | 0.9424 | 0.9385 | 2.5582 | 3.4520 | 0.9600 | 0.9789 |
| FusionNet | 4.2815 | 6.1144 | 0.9380 | 0.9332 | 2.6283 | 3.3242 | 0.9595 | 0.9776 |
| **SDRCNN** | **4.0994** | **5.9123** | **0.9437** | **0.9404** | **2.5502** | **3.3057** | **0.9604** | **0.9796** |
| Ideal value | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |

The bold values stand for the results obtained by our methods.

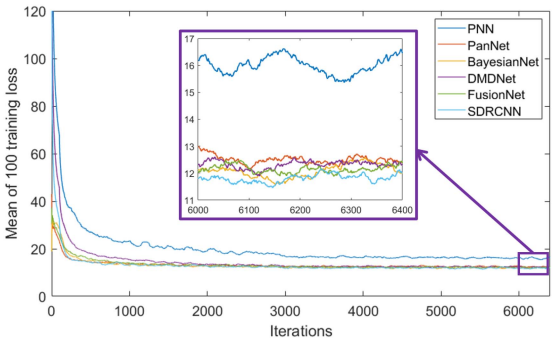


Fig. 10. Training loss curves for considered networks on Rio dataset (sensor: WorldView-3).

TABLE IV
AVERAGE VALUES OF QNR, $D_s$, AND $D_\lambda$ WITH THE RELATED STANDARD DEVIATIONS (STD) FOR 50 FULL-RESOLUTION WORLDVIEW-3 (TRIPOLI) SAMPLES

| Method | QNR (±std) | $D_\lambda$ (±std) | $D_s$ (±std) |
|---|---|---|---|
| GS | 0.8601±0.0714 | 0.0505±0.0345 | 0.1073±0.0542 |
| GSA | 0.8829±0.0525 | 0.0461±0.0377 | 0.0948±0.0385 |
| PRACS | 0.9223±0.0392 | 0.0457±0.0178 | 0.0741±0.0272 |
| BDSD | 0.9116±0.0426 | 0.0443±0.0224 | 0.0824±0.0218 |
| SFIM | 0.8982±0.0347 | 0.0492±0.0368 | 0.0716±0.0277 |
| GLP-CBD | 0.8831±0.0559 | 0.0527±0.0357 | 0.0842±0.0340 |
| CDIF | 0.9271±0.0321 | 0.0430±0.0218 | 0.0501±0.0186 |
| LRTCFPan | 0.9273±0.0307 | 0.0428±0.0306 | 0.0498±0.0182 |
| PNN | 0.9267±0.0356 | 0.0428±0.0235 | 0.0514±0.0175 |
| PanNet | 0.9292±0.0364 | 0.0427±0.0230 | 0.0487±0.0201 |
| BayesianNet | 0.9335±0.0294 | 0.0421±0.0233 | 0.0444±0.0174 |
| DMDNet | 0.9326±0.0432 | 0.0422±0.0244 | 0.0445±0.0179 |
| FusionNet | 0.9308±0.0342 | 0.0426±0.0219 | 0.0461±0.0194 |
| **SDRCNN** | **0.9368±0.0274** | **0.0418±0.0212** | **0.0442±0.0168** |
| Ideal value | 1 | 0 | 0 |

The bold values stand for the results obtained by our methods.

### B. Full-Resolution Assessment

To confirm the results obtained at reduced resolutions, a full-resolution analysis involving the original MS and PAN images was also needed. In this case, however, there are no GT images. For the full-resolution assessments where GT images are absent, the widely used metrics in the literature include the quality with no reference (QNR) index, the spectral distortion $D_\lambda$ index, and the spatial distortion $D_s$ index [71], which were also adopted in this article. The ideal values of QNR, $D_\lambda$, and $D_s$ are 1, 0, and 0, respectively. In this case, the spectral quality of a fused image was referenced to the original MS image, while the spatial detail of a higher quality fused image should be more similar to that of the original PAN image. Table IV reports the average performances of 50 full-resolution examples. On the indices considered, their values obtained for SDRCNN were closest to the ideal values with a small standard deviation, confirming the superiority of SDRCNN against the other methods compared.

### C. Assessment on WorldView-2 and QuickBird Datasets

To study the generalization of SDRCNN, the proposed network was also tested using the data from another eight-band sensor (i.e., the Washington DC dataset acquired by WorldView-2) and a four-band sensor (i.e., the QuickBird dataset). For a standard evaluation, the training and test data were generated according to the Wald's protocol. Visual comparisons of the results on the WorldView-2 dataset are shown in Figs. 11 and 12. It was observed that the fused image derived by SDRCNN was the most qualified one, evidenced not only by sharper and clearer edges, but also by unobservable ghosting and blurring. A visualization of the output HRMS images on QuickBird was not provided as the difference could not be seen clearly with a naked eye. The values of the quality indices for the HRMS images obtained are reported in Table V for the two datasets, which show that SDRCNN performed best among the methods considered. This suggests the generalization of our approach.

### D. Visual Analysis of Dense Residual Connection

It was expected that the dense residual connection mechanism proposed in this article extracted different levels of features at different network depths and thus enhanced the spatial details

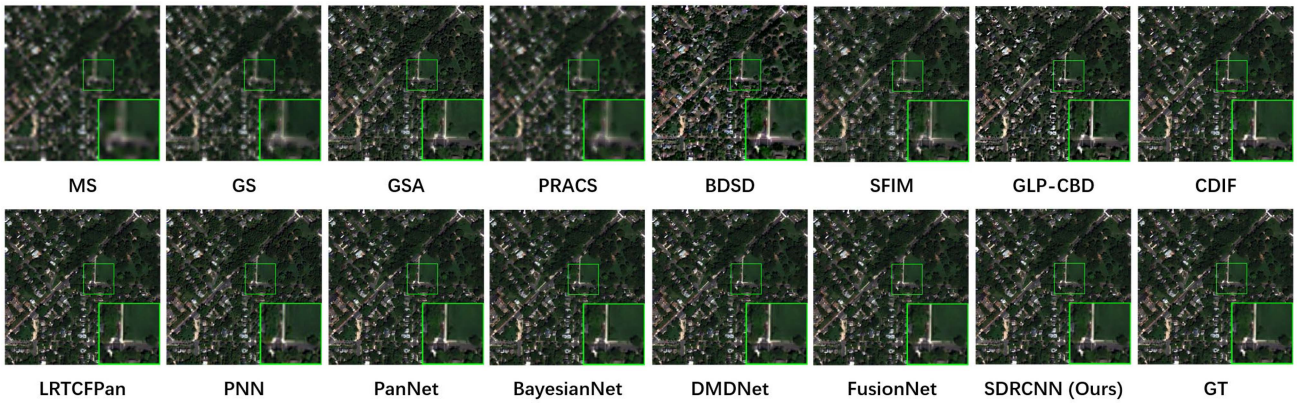


Fig. 11. Visual comparisons of the HRMS images from the compared methods on the reduced-resolution Washington DC dataset (sensor: WorldView-2).

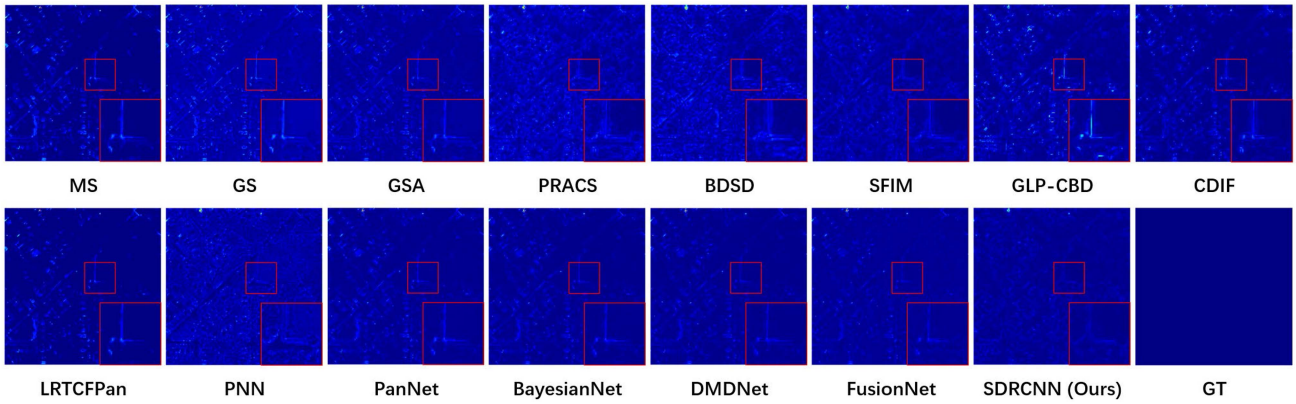


Fig. 12. Corresponding AEMs on the reduced-resolution Washington DC dataset (sensor: WorldView-2).

TABLE V
QUALITY METRICS FOR ALL THE COMPARED APPROACHES ON THE REDUCED-RESOLUTION WORLDVIEW-2 AND QUICKBIRD DATASETS, RESPECTIVELY: (A) WORLDVIEW-2 (WASHINGTON DC) DATASET AND (B) QUICKBIRD DATASET

| | (a) WorldView-2 (Washington DC) dataset | | | | (b) QuickBird dataset | | | |
|---|---|---|---|---|---|---|---|---|
| Method | ERGAS (±std) | SAM (±std) | SCC (±std) | Q8 (±std) | ERGAS (±std) | SAM (±std) | SCC (±std) | Q4 (±std) |
| GS | 4.5344±0.6049 | 6.3690±0.9200 | 0.8911±0.0420 | 0.8671±0.0727 | 1.6975±0.4013 | 2.2697±0.6203 | 0.9326±0.0200 | 0.7996±0.0611 |
| GSA | 3.6619±0.5000 | 6.0020±1.1338 | 0.8962±0.0341 | 0.9153±0.0715 | 1.4612±0.3640 | 1.9387±0.5798 | 0.9287±0.0229 | 0.8489±0.0720 |
| PRACS | 4.4238±0.5433 | 6.2247±1.1091 | 0.8516±0.0468 | 0.8770±0.0714 | 1.4128±0.3789 | 1.7971±0.5889 | 0.9313±0.0222 | 0.8449±0.0581 |
| BDSD | 5.0711±0.9168 | 8.3883±1.8437 | 0.8727±0.0373 | 0.8826±0.0722 | 1.7734±0.4468 | 2.3778±0.7730 | 0.9248±0.0193 | 0.8407±0.0554 |
| SFIM | 4.1077±2.4432 | 6.0334±1.0645 | 0.8839±0.0427 | 0.9014±0.0669 | 1.3934±0.3139 | 1.7945±0.5521 | 0.9319±0.0149 | 0.8474±0.0591 |
| GLP-CBD | 5.1367±0.7480 | 7.2313±1.2616 | 0.8468±0.0446 | 0.8895±0.0698 | 2.7821±0.6702 | 2.9029±0.9949 | 0.8922±0.0162 | 0.7647±0.0452 |
| CDIF | 3.7402±0.4158 | 6.0043±0.8574 | 0.9026±0.0116 | 0.9172±0.0711 | 1.6243±0.2547 | 1.9273±0.4177 | 0.9322±0.0105 | 0.8471±0.0620 |
| LRTCFPan | 3.7248±0.4389 | 5.7269±0.8634 | 0.9101±0.0230 | 0.9255±0.0726 | 1.5722±0.2566 | 1.8947±0.4203 | 0.9386±0.0142 | 0.8533±0.0623 |
| PNN | 3.6821±0.4414 | 6.1188±0.9339 | 0.8946±0.0239 | 0.9159±0.0730 | 1.5126±0.3558 | 1.9939±0.6281 | 0.9239±0.0184 | 0.8350±0.0533 |
| PanNet | 3.1710±0.3783 | 5.2083±0.8784 | 0.9304±0.0128 | 0.9330±0.0722 | 1.2476±0.2438 | 1.8423±0.4215 | 0.9412±0.0140 | 0.8766±0.0587 |
| BayesianNet | 2.9683±0.4076 | 4.8895±0.8464 | 0.9390±0.0095 | 0.9396±0.0713 | 1.0312±0.2416 | 1.3643±0.4021 | 0.9639±0.0096 | 0.8964±0.0610 |
| DMDNet | 3.0248±0.4132 | 4.9262±0.8720 | 0.9380±0.0102 | 0.9392±0.0706 | 1.0471±0.2464 | 1.4030±0.4195 | 0.9631±0.0100 | 0.8900±0.0607 |
| FusionNet | 3.0311±0.4229 | 4.9161±0.8587 | 0.9384±0.0097 | 0.9394±0.0719 | 1.0316±0.2414 | 1.3702±0.4058 | 0.9631±0.0102 | 0.8958±0.0612 |
| **SDRCNN** | **2.9563±0.4017** | **4.7633±0.8472** | **0.9404±0.0090** | **0.9404±0.0648** | **1.0309±0.2405** | **1.3514±0.3987** | **0.9649±0.0092** | **0.8970±0.0607** |
| Ideal value | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |

The bold values stand for the results obtained by our methods.

of output images. To explore its effectiveness, the feature maps obtained at the three depths (i.e., three addition layers after Residual Blocks) in SDRCNN were visualized. For 52 feature maps obtained from each addition layer, it was considered as a data cube with 52 channels. PCA was performed using the singular value decomposition method to achieve a dimensionality reduction of the data, resulting in four principle component bands. For visualization, the principle component values were rescaled to the range [0, 1]. Fig. 13 shows an example of a $256 \times 256$ size region taken from the Tripoli dataset of WorldView-3. The subplots in the first, second, and third rows correspond to the visualized feature maps obtained from the addition layer after Residual Block 1, 2, and 3, respectively. It was observed that the feature maps in the first row included relatively coarse (i.e.,

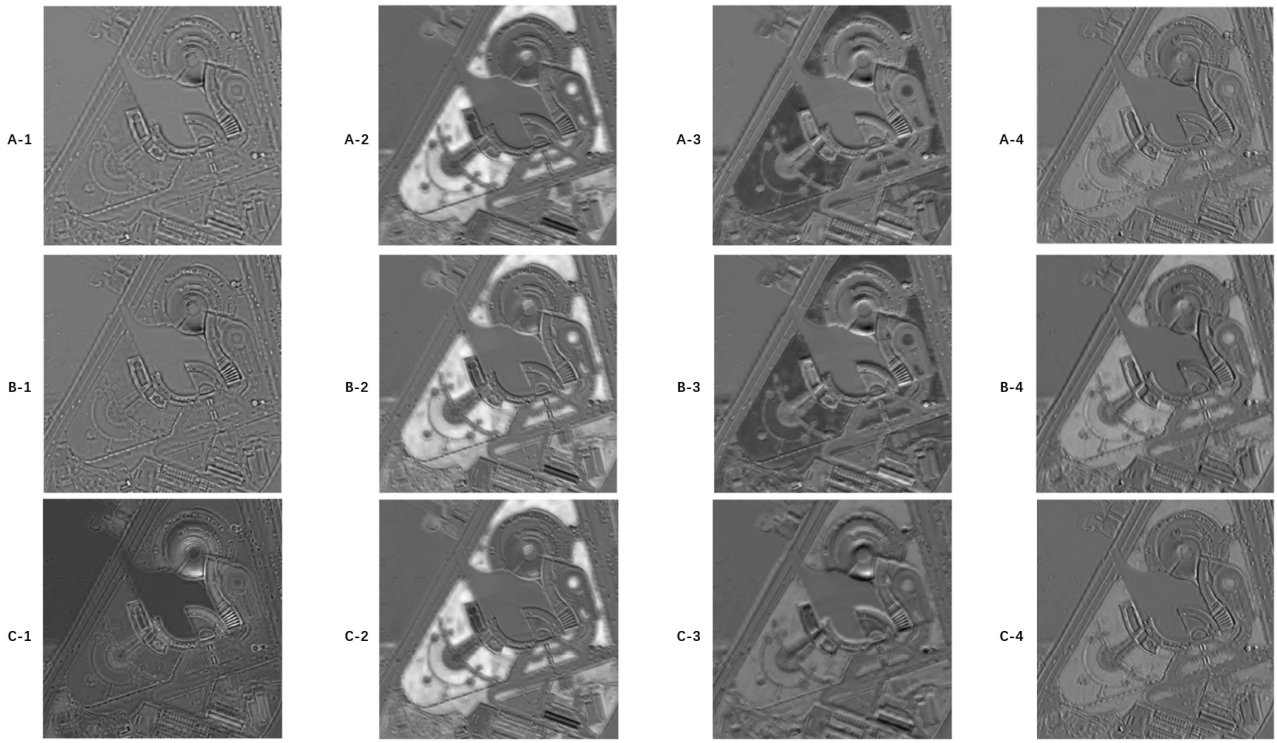


Fig. 13. Visual comparisons of the feature maps from three network depths of SDRCNN on the reduced-resolution Tripoli dataset (sensor: WorldView-3). A, B, and C correspond to the addition layers after the residual block 1, 2, and 3, respectively; the numbers 1, 2, 3, and 4 correspond to the four principle component bands generated by the dimensionality reduction of the data.

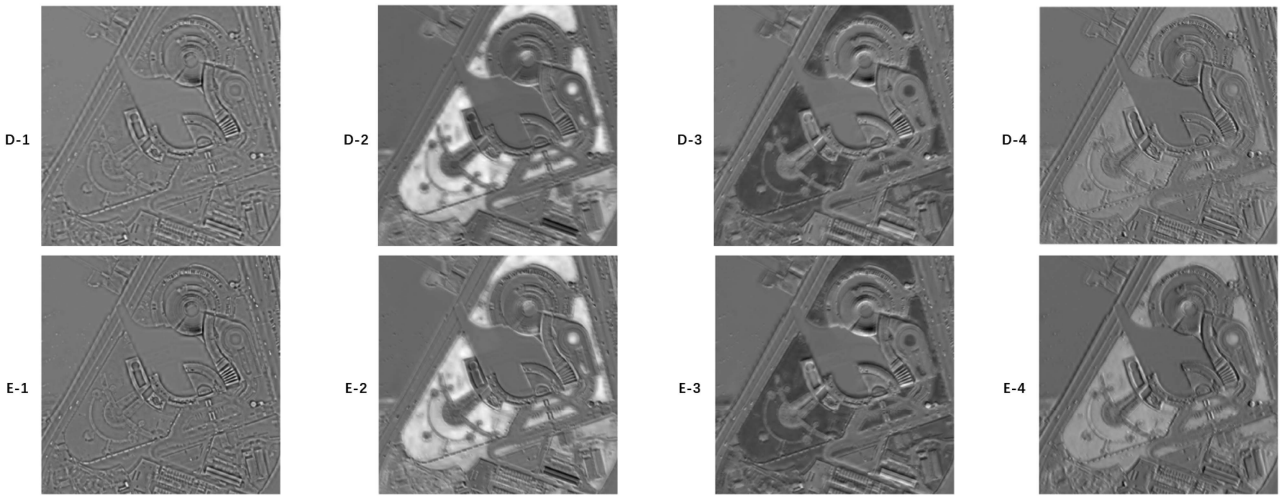


Fig. 14. Corresponding AEMs of the feature maps from two adjacent addition layers on the reduced-resolution Tripoli dataset (sensor: WorldView-3). D corresponds to the residual of the shallower two addition layers, and E corresponds to the residual of the deeper two addition layers, the numbers 1, 2, 3, and 4 correspond to the four principle component bands generated by the dimensionality reduction of the data.

larger objects) contour information. Although the maps in the second row showed similar edge information as those in the first row, these edges appeared to be darker and clearer, making the delineation between different objects or regions clearer. From the maps in the third row, more detailed (i.e., smaller objects) contour information and some building textures were observed. Overall, as the network depth increased, features extracted by SDRCNN varied from coarse to more detailed. In other words, the feature maps generated at a deeper level were found to have included more detailed spatial information. In order to better show the differences between the feature maps generated at the three depths in SDRCNN, Fig. 14 shows the AEMs obtained by subtracting the four principle component bands of the previous addition layer from those of the deeper addition layer. It was seen from the maps in the first row of Fig. 14 that the second addition layer included more spatial features than the first layer mainly in the edges of buildings. The maps associated with the third addition layer showed more detailed texture information

TABLE VI
QUANTITATIVE COMPARISON OF THE NETWORKS COMBINING BASIC BLOCK OF SDRCNN AND **STRUCTURES** OF BENCHMARK METHODS (DL-BASED) ON 1258 TEST SAMPLES FOR TRIPOLI (WORLVIEW-3) DATASET

| **Structure** type of network | ERGAS (±std) | SAM (±std) | SCC (±std) | Q8 (±std) | Time (s) |
|---|---|---|---|---|---|
| PNN | 4.4544±0.6509 | 5.8912±1.2248 | 0.9095±0.0384 | 0.8615±0.1312 | 301.89 |
| PanNet | 4.1083±0.6057 | 5.6603±1.1909 | 0.9183±0.0349 | 0.8745±0.1315 | 338.14 |
| BayesianNet | 4.1292±0.6235 | 5.7045±1.1915 | 0.9173±0.0360 | 0.8740±0.1310 | 380.22 |
| DMDNet | 4.1083±0.6057 | 5.6603±1.1909 | 0.9183±0.0349 | 0.8745±0.1315 | 338.14 |
| FusionNet | 4.1031±0.5989 | 5.6822±1.2028 | 0.9187±0.0354 | 0.8740±0.1324 | 345.34 |
| **SDRCNN** | **4.0567±0.5880** | **5.5678±1.1743** | **0.9211±0.0343** | **0.8765±0.1297** | **327.46** |
| Ideal value | 0 | 0 | 1 | 1 | |

The bold values stand for the results obtained by our methods.

TABLE VII
QUANTITATIVE COMPARISON OF THE NETWORKS COMBINING STRUCTURE OF SDRCNN AND **BASIC BLOCKS** OF BENCHMARK METHODS (DL-BASED) ON 1258 TEST SAMPLES FOR TRIPOLI (WORLVIEW-3) DATASET

| **Block** type of network | ERGAS (±std) | SAM (±std) | SCC (±std) | Q8 (±std) | Time (s) |
|---|---|---|---|---|---|
| PNN | 4.2668±0.6933 | 5.8839±1.2231 | 0.9107±0.0423 | 0.8668±0.1313 | 349.54 |
| PanNet | 4.1978±0.6653 | 5.7783±1.1909 | 0.9124±0.0392 | 0.8724±0.1332 | 352.61 |
| BayesianNet | 4.2065±0.6674 | 5.7811±1.1996 | 0.9119±0.0411 | 0.8705±0.1329 | 417.49 |
| DMDNet | 4.1207±0.6276 | 5.7087±1.1937 | 0.9160±0.0378 | 0.8737±0.1311 | 793.61 |
| FusionNet | 4.2470±0.6439 | 5.8031±1.2143 | 0.9107±0.0386 | 0.8670±0.1323 | 355.99 |
| **SDRCNN** | **4.0567±0.5880** | **5.5678±1.1743** | **0.9211±0.0343** | **0.8765±0.1297** | **327.46** |
| Ideal value | 0 | 0 | 1 | 1 | |

The bold values stand for the results obtained by our methods.

of the objects/buildings than the second layer. These visual observations were reasonable and as expected. By connecting the residual information from the shallow layers and feeding it into the deeper convolutional layers for feature extraction, the residuals from different layers can be more fully utilized to obtain more detailed spatial features than the shallower layers of the network, thus improving the learning ability of the network for feature extraction.

### *E. Ablation Study*

Ablation studies were conducted to explore the effect of each component of SDRCNN. The WorldView-3 Tripoli dataset was used for these ablation studies.

*1) Network Structure of SDRCNN:* To verify the effectiveness of the single-branch single-scale network structure with a dense residual connectivity mechanism of SDRCNN, this ablation experiment replaced the structure of SDRCNN with those of other comparison networks (i.e., PNN, PanNet, BayesianNet, DMDNet, and FusionNet) for comparison, respectively. Table VI shows the average results of the quantitative analysis of these combined networks and SDRCNN on 1258 test data from the Tripoli dataset. For a fair comparison, the number of parameters for all networks is 100 K (K stands for $10^3$). The results show that the averaging performance of all metrics decreases when the network structure of SDRCNN was replaced with any other structure, thus demonstrating the effectiveness of the SDRCNN structure. It is worth noting that the quantification results corresponding to the network structures of PanNet and DMDNet in Table VI are the same, which is because of the same structures of these two networks. In addition, during this experiment, we found that the network framework of FusionNet is also the same as these two, but with a deeper structure than them by stringing one more basic block together. The last column of Table VI shows the corresponding testing times for each network. Since the PNN structure is much shallower than the others, it is reasonable that it has the shortest testing time. However, the testing times corresponding to the other structures are all longer than SDRCNN, thus further indicating that the SDRCNN structure is more computationally efficient.

*2) Residual Block of SDRCNN:* To demonstrate the effectiveness of the proposed residual block, this ablation experiment retains the network structure of the SDRCNN and replaces the blocks in the SDRCNN with the basic blocks of other comparison networks (i.e., PNN, PanNet, BayesianNet, DMDNet, and FusionNet), respectively. Table VII shows the test results of these combined networks and SDRCNN (with the same 100 K parameters) on the Tripoli dataset. It is obvious that the accuracy of all metrics decreases when the basic blocks in the SDRCNN are replaced with blocks from other networks, thus proving that better pansharpening accuracy can be obtained with blocks of SDRCNN. In addition, SDRCNN takes the shortest testing time, which further proves that the blocks of SDRCNN have higher efficiency.

*3) Spectral Mapping:* The spectral mapping refers to the operation of directly adding an upsampled MS image to an output image, and has been widely used in CNNs for pansharpening tasks. To verify whether it can improve the pansharpening accuracy of CNN, two networks, PNN and SDRCNN, were considered in this section for experiments. Table VIII shows the values of the accuracy metrics for PNN and SDRCNN with and without the spectral mapping on the 50 testdata. Compared with PNN, the fusion accuracy of the network with the spectral mapping were higher in all the metrics. In addition, the accuracy of pansharpening decreases when the spectral mapping of SDRCNN was removed. These suggest that the spectral mapping could enhance the performance of CNN for a higher pansharpening accuracy.

*4) Removal of the BN Layer:* In the existing CNNs for pansharpening, there is no agreement on whether the BN layer should be retained or not. To demonstrate the effectiveness of

TABLE VIII
QUALITY METRICS FOR ALL THE ABLATION STUDIES ON 50 TEST SAMPLES FOR TRIPOLI (WORLDVIEW-3) DATASET

| Method | ERGAS | SAM | SCC | Q8 | |
|---|---|---|---|---|---|
| PNN (without MS image connection in its original network) | 5.0635 | 7.0424 | 0.9152 | 0.9099 | Section III-E-3) |
| PNN with MS image connection | 4.7168 | 6.9136 | 0.9285 | 0.9206 | |
| SDRCNN without MS image connection | 4.1478 | 5.9468 | 0.9487 | 0.9396 | |
| **SDRCNN (with MS image connection by default)** | **4.0945** | **5.9078** | **0.9512** | **0.9432** | |
| SDRCNN with BN | 4.2110 | 6.0690 | 0.9388 | 0.9359 | Section III-E-4) |
| **SDRCNN (without BN by default)** | **4.0945** | **5.9078** | **0.9512** | **0.9432** | |
| SDRCNN with additional ReLU | 4.1369 | 5.9401 | 0.9410 | 0.9394 | Section III-E-5) |
| **SDRCNN (without additional ReLU by default)** | **4.0945** | **5.9078** | **0.9512** | **0.9432** | |
| Ideal value | 0 | 0 | 1 | 1 | |

The bold values stand for the results obtained by our methods.

TABLE IX
QUANTITATIVE COMPARISONS WITH THE BENCHMARK METHODS (PANNET, BAYESIANNET, DMDNET, AND FUSIONNET) UNDER DIFFERENT PARAMETER NUMBERS ON 1258 TEST SAMPLES FOR TRIPOLI (WORLVIEW-3) DATASET

| Number of Parameters | Network | ERGAS (±std) | SAM (±std) | SCC (±std) | Q8 (±std) |
|---|---|---|---|---|---|
| ≈ 50K | PanNet | 4.2444±0.6677 | 5.8561±1.1970 | 0.9104±0.0416 | 0.8658±0.1400 |
| | BayesianNet | 4.1758±0.6461 | 5.7610±1.2007 | 0.9141±0.0395 | 0.8695±0.1328 |
| | DMDNet | 4.2375±0.6592 | 5.8354±1.2175 | 0.9103±0.0400 | 0.8697±0.1315 |
| | FusionNet | 4.2724±0.6352 | 5.8222±1.2279 | 0.9124±0.0386 | 0.8649±0.1325 |
| | **SDRCNN** | **4.1331±0.5998** | **5.6504±1.1967** | **0.9164±0.0348** | **0.8742±0.1306** |
| ≈ 100K | PanNet | 4.2245±0.6598 | 5.8269±1.1900 | 0.9115±0.0413 | 0.8674±0.1380 |
| | BayesianNet | 4.1116±0.6402 | 5.7082±1.1886 | 0.9176±0.0391 | 0.8726±0.1318 |
| | DMDNet | 4.1386±0.6417 | 5.7717±1.2069 | 0.9171±0.0381 | 0.8727±0.1314 |
| | FusionNet | 4.2349±0.6372 | 5.8100±1.2270 | 0.9139±0.0384 | 0.8666±0.1323 |
| | **SDRCNN** | **4.0567±0.5880** | **5.5678±1.1743** | **0.9211±0.0343** | **0.8765±0.1297** |
| ≈ 200K | PanNet | 4.2025±0.6454 | 5.8012±1.1854 | 0.9126±0.0405 | 0.8679±0.1381 |
| | BayesianNet | 4.0869±0.6360 | 5.6513±1.1719 | 0.9184±0.0388 | 0.8732±0.1329 |
| | DMDNet | 4.1091±0.6227 | 5.6791±1.1841 | 0.9180±0.0371 | 0.8738±0.1316 |
| | FusionNet | 4.1917±0.6273 | 5.7821±1.2134 | 0.9163±0.0385 | 0.8678±0.1325 |
| | **SDRCNN** | **4.0051±0.5893** | **5.4926±1.1530** | **0.9245±0.0335** | **0.8784±0.1307** |
| Ideal value | | 0 | 0 | 1 | 1 |

The bold values stand for the results obtained by our methods.

removing BN as described in Section II-D, experiments were conducted by plugging BN into our SDRCNN model. The BN operation was deployed after each convolution operation. Table VIII shows the average results of SDRCNN before and after the BN inclusion on 50 testdata using the WorldView-3 dataset. It was seen that the performance of SDRCNN with BN decreased. In addition, BN operations consumed additional computation and storage resources. As such, it is beneficial to exclude BN for higher pansharpening accuracy and improved computational and storage budget.

*5) Reduced Activation Function:* Similar to the problem of BN layer, there is no conclusion on the effect of activation function on the performance of CNN for pansharpening. In this section, the effectiveness of the reduced activation functions (i.e., ReLU) described in Section II-D was investigated by inserting more activation functions into our SDRCNN model. The additional activation functions were deployed after each depthwise convolution operation and concentration operation. Table VIII shows the average results of SDRCNN before and after the insertion of more activation functions on the 50 testdata. It was seen that better average results were obtained for the SDRCNN model retaining only one activation function between the two $1 \times 1$ layers.

*6) Effect of the Number of Parameters:* To check the effects of the network size on the performance of the networks (i.e., the benchmark methods PanNet, BayesianNet, DMDNet, and FusionNet, and our method SDRCNN), three network sizes with approximately 50 K, 100 K, and 200 K parameters were tested using the Tripoli dataset (containing 1258 test samples) from WorldView-3. Table IX shows the results of the quantity evaluation of the five CNN methods for the three network sizes considered. It can be seen that in each network size tested, SDRCNN achieved the closest average results to the ideal values and the smallest standard deviations in all quality metrics. In addition, as expected, increasing the network size can further improve the accuracy of SDRCNN.

## IV. DISCUSSION

Based on the experimental results of this article, the CNN-based methods exhibited better performances in pansharpening than the classical CS and MRA methods. This is mainly due to the fact that CNN-based methods exploit large-scale data during the training phase. In this section, the convolution kernel size, the network complexity, and the testing time were discussed.

### A. Convolution Kernel Size

Currently, some researchers argued that increasing the receptive field can improve the pansharpening performance of CNN models, and as such they used larger convolutional kernel sizes in their proposed methods [66]. This idea came from an observation of the behavior of CNNs applied in some other image processing areas. In the proposed SDRCNN model, we also considered the behavior of increasing the convolutional kernel. Initially, we

tried using convolutional kernel sizes that are larger than 3 × 3, for example, 7 × 7 and 11 × 11. However, the reduction in network width or depth was necessary to control the network with a consistent number of parameters. In order not to change the width or depth of the network, we considered the use of the atrous convolution to explore the effect of increasing the receptive field. It was found that increasing the receptive field using the atrous convolution did not significantly improve the pansharpening performance of the CNN. It is possible that the limitations of the atrous convolution, such as grid effects, adversely affected the fusion results, which may lead to inaccurate results. In addition, it was also thought that a larger receptive field could improve the performance of CNNs in some other image processing domains because these image processing tasks are usually highly dependent on the relationship between a pixel and its surrounding pixels, such as semantic segmentation and object detection. However, for pansharpening, in the process of spatial and spectral information transfer and fusion, more attention should be paid to the relationship of features between images. In addition, larger convolution kernels usually bring a larger set of parameters, which often increase the computational cost and learning difficulty. Therefore, increasing the size of the convolution kernel or receptive field may not be suitable for improving the performance of the CNN model in the pansharpening task.

### B. Network Complexity

SDRCNN is simpler than DMDNet because SDRCNN does not need to compute high-pass filtered images from the input PAN and MS images and uses smaller convolution kernels and a single-scale network structure. FusionNet replicates the input PAN image to the same number of bands as the input MS image and then feeds those into the network, which increases the workload of subsequent processing and feature extractions on the input images. In addition, SDRCNN has fewer layers than FusionNet. However, given that SDRCNN uses the depthwise separable convolutions, it is difficult to directly compare its network complexity with FusionNet. Furthermore, the structure of PNN is a simple three-layer network without any jump connections. However, it is too shallow to extract enough image features from this simple network. Therefore, SDRCNN not only improves the accuracy of pansharpening without increasing the network size, but also avoids unnecessary increases in network complexity. PanNet has a similar network structure and input data preprocessing requirement as DMDNet. BayesianNet is a multibranch network, which is more complex than the single-branch single-scale SDRCNN.

### C. Testing Time

Fig. 15 reports the testing times for PanNet, BayesianNet, DMDNet, FusionNet, and SDRCNN on the Tripoli test set (containing 1258 test samples) from WorldView-3. DMDNet took more than twice as long as the other four networks, and the time spent on the testing increased notably with increasing number of parameters. As the number of parameters increases, SDRCNN always takes the least amount of time, and the efficiency of SDRCNN becomes more evident when the number of parameters reaches 200K.

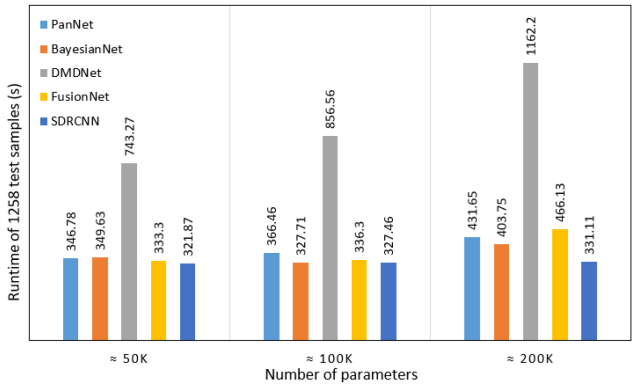


Fig. 15. Testing times for SDRCNN, DMDNet, and FusionNet under different parameter numbers for 1258 test samples from Tripoli (WorldView-3) dataset.

## V. Conclusion

Developing highly accurate and efficient pansharpening methods can be very valuable. In this article, a novel single-branch, single-scale lightweight CNN architecture named SDRCNN was developed to this end. It uses a dense residual connected structure and novel convolution blocks to achieve a better tradeoff between accuracy and efficiency. The performances of SDRCNN were compared to eight traditional methods (i.e., GS, GSA, PRACS, BDSD, SFIM, GLP-CBD, CDIF, and LRTCF-Pan) and five lightweight CNN-based pansharpening methods (i.e., PNN, PanNet, BayesianNet, DMDNet, and FusionNet), using the reduced-resolution and the full-resolution tests. SDRCNN achieved highest pansharpening accuracies (i.e., SAM, ERGAS, SCC, Q2n, QNR, $D_s$, and $D_\lambda$) on all four datasets tested. SDRCNN exhibited least spatial detail blurring and spectral distortions in visual comparisons using the pansharpened images and the associated AEMs. Furthermore, the ablation study confirmed the effectiveness of the dense residual connected structure and the convolutional block used in SDRCNN. Finally, it is demonstrated that SDRCNN is the most efficient among the compared lightweight networks. All these results demonstrated the superiority of SDRCNN against the traditional and lightweight CNN-based methods compared, in preserving both spatial and spectral information in fused HRMS images.

## References

[1] W. Ma et al., "Feature split-merge-enhancement network for remote sensing object detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jan. 2022, Art. no. 5616217, doi: 10.1109/TGRS.2022.3140856.

[2] M. H. R. Sales, S. de Bruin, C. Souza, and M. Herold, "Land use and land cover area estimates from class membership probability of a random forest classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jun. 2022, Art. no. 4402711, doi: 10.1109/TGRS.2021.3080083.

[3] Y. Cai, L. Fan, P. M. Atkinson, and C. Zhang, "Semantic segmentation of terrestrial laser scanning point clouds using locally enhanced image-based geometric representations," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Mar. 2022, Art. no. 5702815, doi: 10.1109/TGRS.2022.3161982.

[4] Y. Cai, L. Fan, and Y. Fang, "SBSS: Stacking-based semantic segmentation framework for very high-resolution remote sensing image," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Jan. 2023, Art. no. 5600514, doi: 10.1109/TGRS.2023.3234549.

[5] Q. Xu, Y. Zhang, and B. Li, "Recent advances in pansharpening and key problems in applications," *Int. J. Image Data Fusion*, vol. 5, no. 3, pp. 175–195, Jul. 2014, doi: 10.1080/19479832.2014.889227.

[6] X. Meng, H. Shen, H. Li, L. Zhang, and R. Fu, "Review of the pansharpening methods for remote sensing images based on the idea of meta-analysis: Practical discussion and challenges," *Inf. Fusion*, vol. 46, pp. 102–113, Mar. 2019, doi: 10.1016/j.inffus.2018.05.006.

[7] Y. Qu, H. Qi, B. Ayhan, C. Kwan, and R. Kidd, "DOES multispectral /hyperspectral pansharpening improve the performance of anomaly detection?," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 6130–6133, doi: 10.1109/IGARSS.2017.8128408.

[8] A. Mohammadzadeh, A. Tavakoli, and M. J. Valadan Zoej, "Road extraction based on fuzzy logic and mathematical morphology from pan-sharpened IKONOS images," *Photogramm. Rec.*, vol. 21, no. 113, pp. 44–60, Mar. 2006, doi: 10.1111/j.1477-9730.2006.00353.x.

[9] P. Du, S. Liu, J. Xia, and Y. Zhao, "Information fusion techniques for change detection from multi-temporal remote sensing images," *Inf. Fusion*, vol. 14, no. 1, pp. 19–27, Jan. 2013, doi: 10.1016/j.inffus.2012.05.003.

[10] F. Bovolo, L. Bruzzone, L. Capobianco, A. Garzelli, S. Marchesi, and F. Nencini, "Analysis of the effects of pansharpening in change detection on VHR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 53–57, Jan. 2010, doi: 10.1109/LGRS.2009.2029248.

[11] E. Ibarrola-Ulzurrun, C. Gonzalo-Martin, and J. Marcello-Ruiz, "Influence of pansharpening techniques in obtaining accurate vegetation thematic maps," in *Proc. Earth Resour. Environ. Remote Sens./GIS Appl. VII*, 2016, pp. 311–330, doi: 10.1117/12.2241501.

[12] A. Jenerowicz and M. Woroszkiewicz, "The pan-sharpening of satellite and UAV imagery for agricultural applications," in *Proc. Remote Sens. Agriculture, Ecosyst., Hydrol. XVIII*, 2016, pp. 565–575, doi: 10.1117/12.2241645.

[13] T.-M. Tu, P. S. Huang, C.-L. Hung, and C.-P. Chang, "A fast intensity–hue–saturation fusion technique with spectral adjustment for IKONOS imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 309–312, Oct. 2004, doi: 10.1109/LGRS.2004.834804.

[14] V. P. Shah, N. H. Younan, and R. L. King, "An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1323–1335, May 2008, doi: 10.1109/TGRS.2008.916211.

[15] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS + pan data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3230–3239, Oct. 2007, doi: 10.1109/TGRS.2007.901007.

[16] J. Choi, K. Yu, and Y. Kim, "A new adaptive component-substitution-based satellite image fusion by using partial replacement," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 295–309, Jan. 2011, doi: 10.1109/TGRS.2010.2051674.

[17] G. Vivone, "Robust band-dependent spatial-detail approaches for panchromatic sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6421–6433, Sep. 2019, doi: 10.1109/TGRS.2019.2906073.

[18] Y. Yang, H. Lu, S. Huang, and W. Tu, "Pansharpening based on joint-guided detail extraction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 389–401, Oct. 2021, doi: 10.1109/JSTARS.2020.3032472.

[19] Y. Kim, C. Lee, D. Han, Y. Kim, and Y. Kim, "Improved additive-wavelet image fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 263–267, Mar. 2011, doi: 10.1109/LGRS.2010.2067192.

[20] J. G. Liu, "Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details," *Int. J. Remote Sens.*, vol. 21, no. 18, pp. 3461–3472, Jan. 2000, doi: 10.1080/014311600750037499.

[21] X. Otazu, M. Gonzalez-Audicana, O. Fors, and J. Nunez, "Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2376–2385, Oct. 2005, doi: 10.1109/TGRS.2005.856106.

[22] G. Vivone, R. Restaino, and J. Chanussot, "Full scale regression-based injection coefficients for panchromatic sharpening," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3418–3431, Jul. 2018, doi: 10.1109/TIP.2018.2819501.

[23] C. Ballester, V. Caselles, L. Igual, J. Verdera, and B. Rougé, "A variational model for P+XS image fusion," *Int. J. Comput. Vis.*, vol. 69, no. 1, pp. 43–58, Apr. 2006, doi: 10.1007/S11263-006-6852-X.

[24] Q. Wei, N. Dobigeon, J. Y. Tourneret, J. Bioucas-DIas, and S. Godsill, "R-FUSE: Robust fast fusion of multiband images based on solving a Sylvester equation," *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1632–1636, Nov. 2016, doi: 10.1109/LSP.2016.2608858.

[25] Z. Y. Zhang, T. Z. Huang, L. J. Deng, J. Huang, X. le Zhao, and C. C. Zheng, "A framelet-based iterative pan-sharpening approach," *Remote Sens.*, vol. 10, no. 4, Apr. 2018, Art. no. 622, doi: 10.3390/RS10040622.

[26] L. J. Deng, M. Feng, and X. C. Tai, "The fusion of panchromatic and multispectral remote sensing images via tensor-based sparse modeling and hyper-Laplacian prior," *Inf. Fusion*, vol. 52, pp. 76–89, Dec. 2019, doi: 10.1016/J.INFFUS.2018.11.014.

[27] N. Meng, T. Zeng, and E. Y. Lam, "Spatial and angular reconstruction of light field based on deep generative networks," in *Proc. IEEE Int. Conf. Image Process.*, 2019, pp. 4659–4663, doi: 10.1109/ICIP.2019.8803480.

[28] Y. Zhu, C. H. Yeung, and E. Y. Lam, "Digital holographic imaging and classification of microplastics using deep transfer learning," *Appl. Opt.*, vol. 60, no. 4, Feb. 2021, Art. no. A38, doi: 10.1364/AO.403366.

[29] N. Meng, H. K.-H. So, X. Sun, and E. Y. Lam, "High-dimensional dense residual convolutional neural network for light field reconstruction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 3, pp. 873–886, Mar. 2021, doi: 10.1109/TPAMI.2019.2945027.

[30] Y. Cai, H. Huang, K. Wang, C. Zhang, L. Fan, and F. Guo, "Selecting optimal combination of data channels for semantic segmentation in city information modelling (CIM)," *Remote Sens.*, vol. 13, no. 7, Apr. 2021, Art. no. 1367, doi: 10.3390/RS13071367.

[31] Y. Cai, L. Fan, and C. Zhang, "Semantic segmentation of multispectral images via linear compression of bands: An experiment using RIT-18," *Remote Sens.*, vol. 14, no. 11, Jun. 2022, Art. no. 2673, doi: 10.3390/RS14112673.

[32] Y. Cai, J. Aryal, Y. Fang, H. Huang, and L. Fan, "OSTA: One-shot task-adaptive channel selection for semantic segmentation of multichannel images," 2023, *arXiv:2305.04766*.

[33] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, Jul. 2016, Art. no. 594, doi: 10.3390/rs8070594.

[34] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley, "PanNet: A deep network architecture for pan-sharpening," in *Proc. IEEE Int. Conf. Comput. Vis* ., 2017, pp. 1753–1761, doi: 10.1109/ICCV.2017.193.

[35] X. Liu, Y. Wang, and Q. Liu, "Remote sensing image fusion based on two-stream fusion network," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Berlin, Germany: Springer Verlag, 2018, pp. 428–439, doi: 10.1007/978-3-319-73603-7_35.

[36] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016, doi: 10.1109/TPAMI.2015.2439281.

[37] Y. Wei, Q. Yuan, H. Shen, and L. Zhang, "Boosting the accuracy of multispectral image pansharpening by learning a deep residual network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1795–1799, Oct. 2017, doi: 10.1109/LGRS.2017.2736020.

[38] Q. Yuan, Y. Wei, X. Meng, H. Shen, and L. Zhang, "A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 978–989, Mar. 2018, doi: 10.1109/JSTARS.2018.2794888.

[39] Z. Shao and J. Cai, "Remote sensing image fusion with deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1656–1669, May 2018, doi: 10.1109/JSTARS.2018.2805923.

[40] L. J. Deng et al., "Machine learning in pansharpening: A benchmark, from shallow to deep networks," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 3, pp. 279–315, Sep. 2022, doi: 10.1109/MGRS.2022.3187652.

[41] D. Lei, Y. Huang, L. Zhang, and W. Li, "Multibranch feature extraction and feature multiplexing network for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, May 2022, Art. no. 5402613, doi: 10.1109/TGRS.2021.3074624.

[42] L. He, J. Zhu, J. Li, A. Plaza, J. Chanussot, and Z. Yu, "CNN-based hyperspectral pansharpening with arbitrary resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Dec. 2022, Art. no. 5518821, doi: 10.1109/TGRS.2021.3132997.

[43] Y. Chi, J. Li, and H. Fan, "Pyramid-attention based multi-scale feature fusion network for multispectral pan-sharpening," *Appl. Intell.*, vol. 52, no. 5, pp. 5353–5365, Mar. 2022, doi: 10.1007/S10489-021-02732-5/FIGURES/10.

[44] W. Tu, Y. Yang, S. Huang, W. Wan, L. Gan, and H. Lu, "MMDN: Multi-scale and multi-distillation dilated network for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, May 2022, Art. no. 5410514, doi: 10.1109/TGRS.2022.3179449.

[45] K. Zhang, A. Wang, F. Zhang, W. Wan, J. Sun, and L. Bruzzone, "Spatial-spectral dual back-projection network for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Apr. 2023, Art. no. 5402216, doi: 10.1109/TGRS.2023.3266799.

[46] L. Jian, S. Wu, L. Chen, G. Vivone, R. Rayhana, and D. Zhang, "Multi-scale and multi-stream fusion network for pansharpening," *Remote Sens.*, vol. 15, no. 6, Mar. 2023, Art. no. 1666, doi: 10.3390/RS15061666.
[47] M. Gong, H. Zhang, H. Xu, X. Tian, and J. Ma, "Multi-patch progressive pansharpening with knowledge distillation," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Mar. 2023, Art. no. 5401115, doi: 10.1109/TGRS.2023.3254053.
[48] S. Peng, Q. Gao, D. Zhu, Y. Lu, and D. Sun, "PSCF-Net: Deeply coupled feedback network for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Mar. 2023, Art. no. 5401812, doi: 10.1109/TGRS.2023.3261386.
[49] T.-J. Zhang, L.-J. Deng, T.-Z. Huang, J. Chanussot, and G. Vivone, "A triple-double convolutional neural network for panchromatic sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–14, Mar. 2022, doi: 10.1109/TNNLS.2022.3155655.
[50] Y. Wu, Y. Li, S. Feng, and M. Huang, "Pansharpening using unsupervised generative adversarial networks with recursive mixed-scale feature fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 3742–3759, Mar. 2023, doi: 10.1109/JSTARS.2023.3259014.
[51] Q. Xu, Y. Li, J. Nie, Q. Liu, and M. Guo, "UPanGAN: Unsupervised pansharpening based on the spectral and spatial loss constrained generative adversarial network," *Inf. Fusion*, vol. 91, pp. 31–46, Mar. 2023, doi: 10.1016/J.INFFUS.2022.10.001.
[52] H. Zhou, Q. Liu, D. Weng, and Y. Wang, "Unsupervised cycle-consistent generative adversarial networks for pan sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Mar. 2022, Art. no. 5408814, doi: 10.1109/TGRS.2022.3166528.
[53] T. Benzenati, Y. Kessentini, and A. Kallel, "Pansharpening approach via two-stream detail injection based on relativistic generative adversarial networks," *Expert Syst. Appl.*, vol. 188, Feb. 2022, Art. no. 115996, doi: 10.1016/J.ESWA.2021.115996.
[54] H. Zhang, H. Wang, X. Tian, and J. Ma, "P2Sharpen: A progressive pansharpening network with deep spectral transformation," *Inf. Fusion*, vol. 91, pp. 103–122, Mar. 2023, doi: 10.1016/J.INFFUS.2022.10.010.
[55] W. Zhu, J. Li, Z. An, and Z. Hua, "Mutiscale hybrid attention transformer for remote sensing image pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Jan. 2023, Art. no. 5400416, doi: 10.1109/TGRS.2023.3239013.
[56] Z. Xiong, N. Liu, N. Wang, Z. Sun, and W. Li, "Unsupervised pansharpening method using residual network with spatial texture attention," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Apr. 2023, Art. no. 5402112, doi: 10.1109/TGRS.2023.3267056.
[57] X. Su, J. Li, and Z. Hua, "Transformer-based regression network for pansharpening remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Feb. 2022, Art. no. 5407423, doi: 10.1109/TGRS.2022.3152425.
[58] J. Qu, S. Hou, W. Dong, S. Xiao, Q. Du, and Y. Li, "A dual-branch detail extraction network for hyperspectral pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Nov. 2022, Art. no. 5518413, doi: 10.1109/TGRS.2021.3130420.
[59] W. Zhang, J. Li, and Z. Hua, "Attention-based multistage fusion network for remote sensing image pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Oct. 2022, Art. no. 5405416, doi: 10.1109/TGRS.2021.3113984.
[60] L.-J. Deng, G. Vivone, C. Jin, and J. Chanussot, "Detail injection-based deep convolutional neural networks for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6995–7010, Aug. 2021, doi: 10.1109/TGRS.2020.3031366.
[61] L. Chen, Z. Lai, G. Vivone, G. Jeon, J. Chanussot, and X. Yang, "ArbRPN: A bidirectional recurrent pansharpening network for multispectral images with arbitrary numbers of bands," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Nov. 2022, Art. no. 5406418, doi: 10.1109/TGRS.2021.3131228.
[62] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[63] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1800–1807, doi: 10.1109/CVPR.2017.195.
[64] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogramm. Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, 1997.
[65] P. Guo, P. Zhuang, and Y. Guo, "Bayesian pan-sharpening with multiorder gradient-based deep network constraints," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 950–962, Mar. 2020, doi: 10.1109/JSTARS.2020.2975000.
[66] X. Fu, W. Wang, Y. Huang, X. Ding, and J. Paisley, "Deep multiscale detail networks for multiband spectral image sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 5, pp. 2090–2104, May 2021, doi: 10.1109/TNNLS.2020.2996498.
[67] R. H. Yuhas, A. F. H. Goetz, J. W. Boardman, A. F. H. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop, Volume 1: AVIRIS Workshop*, 1992, pp. 147–149.
[68] L. Wald, "Data fusion: Definitions and architectures—Fusion of images of different spatial resolutions," 2002. Accessed: Dec. 07, 2022. [Online]. Available: https://hal-mines-paristech.archives-ouvertes.fr/hal-00464703
[69] J. Zhou, D. L. Civco, and J. A. Silander, "A wavelet transform method to merge Landsat TM and SPOT panchromatic data," *Int. J. Remote Sens.*, vol. 19, no. 4, pp. 743–757, Jan. 1998, doi: 10.1080/014311698215973.
[70] A. Garzelli and F. Nencini, "Hypercomplex quality assessment of multi/hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 662–665, Oct. 2009, doi: 10.1109/LGRS.2009.2022650.
[71] G. Vivone et al., "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, May 2015, doi: 10.1109/TGRS.2014.2361734.
[72] C. A. Laben and B. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," U.S. Patent 6,011,875, Jan. 2000.
[73] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, "Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2300–2312, Oct. 2002, doi: 10.1109/TGRS.2002.803623.
[74] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3012–3021, Oct. 2007, doi: 10.1109/TGRS.2007.904923.
[75] J. L. Xiao, T. Z. Huang, L. J. Deng, Z. C. Wu, and G. Vivone, "A new context-aware details injection fidelity with adaptive coefficients estimation for variational pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Feb. 2022, Art. no. 5408015, doi: 10.1109/TGRS.2022.3154480.
[76] Z. C. Wu, T. Z. Huang, L. J. Deng, J. Huang, J. Chanussot, and G. Vivone, "LRTCFPan: Low-rank tensor completion based framework for pansharpening," *IEEE Trans. Image Process.*, vol. 32, pp. 1640–1655, Feb. 2023, doi: 10.1109/TIP.2023.3247165.

**Yuan Fang** received the B.E. degree in civil engineering, in 2021, from the University of Liverpool, Liverpool, U.K., where she is currently working toward the Ph.D. degree in civil engineering.

Her research interests include deep learning and data fusion of satellite images.

**Yuanzhi Cai** (Graduate Student Member, IEEE) received the B.E. degree in civil engineering, in 2020, from the University of Liverpool, Liverpool, U.K., where he is currently working toward the Ph.D. degree in civil engineering.

His research interests include the classification and segmentation of remote sensing data.

**Lei Fan** (Member, IEEE) received the Ph.D. degree in civil engineering from the University of Southampton, Southampton, U.K., in 2014.

He is currently an Assistant Professor with the Department of Civil Engineering, Xi'an Jiaotong Liverpool University, Suzhou, China. His research interests include lidar and photogrammetry techniques, point cloud, machine learning, deformation monitoring, semantic segmentations, monitoring of civil engineering structures, and geohazards.