remote sensing

MDPI

Article

Time-Series FY4A Datasets for Super-Resolution Benchmarking of Meteorological Satellite Images

Jingbo Wei [1,2], Chenghao Zhou [1], Jingsong Wang [3] and Zhou Chen [1,2,*]

[1] School of Mathematics and Computer Sciences, Nanchang University, Nanchang 330031, China
[2] Institute of Space Science and Technology, Nanchang 330031, China
[3] Key Laboratory of Space Weather, National Center for Space Weather, China Meteorological Administration, Beijing 100081, China
* Correspondence: chenzhou@ncu.edu.cn

Abstract: Meteorological satellites are usually operated at high temporal resolutions, but the spatial resolutions are too poor to identify ground content. Super-resolution is an economic way to enhance spatial details, but the feasibility is not validated for meteorological images due to the absence of benchmarking data. In this work, we propose the FY4ASRgray and FY4ASRcolor datasets to assess super-resolution algorithms on meteorological applications. The features of cloud sensitivity and temporal continuity are linked to the proposed datasets. To test the usability of the new datasets, five state-of-the-art super-resolution algorithms are gathered for contest. Shift learning is used to shorten the training time and improve the parameters. Methods are modified to deal with the 16-bit challenge. The reconstruction results are demonstrated and evaluated regarding the radiometric, structural, and spectral loss, which gives the baseline performance for detail enhancement of the FY4A satellite images. Additional experiments are made on FY4ASRcolor for sequence super-resolution, spatiotemporal fusion, and generalization test for further performance test.

Keywords: dataset; Fengyun-4A; FY4A; super resolution; deep learning

check for updates

Citation: Wei, J.; Zhou, C.; Wang, J.; Chen, Z. Time-Series FY4A Datasets for Super-Resolution Benchmarking of Meteorological Satellite Images. *Remote Sens.* **2022**, *14*, 5594. https://doi.org/10.3390/rs14215594

Academic Editor: Carmine Serio

Received: 27 August 2022
Accepted: 31 October 2022
Published: 6 November 2022

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

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

1. Introduction

To observe Earth quickly and timely, new-generation meteorological satellites usually run at geostationary orbits with low spatial resolutions. Typical geostationary satellites include the U.S. GOES-16, Japan's Himawari-8, Europe's MTG-I, and China's Fengyun-4. The low spatial resolution ensures that they can observe clouds and rainfall in a timely manner given the limited physical size and signal-to-noise ratio of the sensors. However, with the continuous demand for improved accuracy in weather forecasting, a higher spatial resolution is desired to discern weather differences across geographic locations. From a software perspective, the need can be partially addressed by super-resolution.

Super-resolution of meteorological images is a reconstruction problem that has been studied for decades. During the imaging process of weather satellites, sunlight reflected from the Earth's surface undergoes atmospheric turbulence, lens blurring and satellite motion before reaching the sensor. Geometric corrections in the post-processing process also lose detail. The super-resolution we wish to perform is not only resolution improvement but also detail recovery, but it is a classical illness-posed inverse problem because a low-resolution image can be obtained by downsampling from an infinite number of different high-resolution images. To improve this problem, constrained models, optimization methods, and prior knowledge can be targeted [1].

There is a long-term need for standardized super-resolution datasets to benchmark various methods under the same conditions. There have been many super-resolution datasets for natural images. Early datasets include Set5, Set14, B100, and Urban100. The development of deep learning calls for far larger data volume, so new datasets are proposed, such as the well-known DIV2K and Flickr2K [2], and Flickr30K [3]. However, for remotely

sensed applications, such datasets are still absent, especially when deep learning is used for remote sensing [4–7].

Remote sensing datasets are usually designed for segmentation, classification, change detection, or object detection. For example, with images from the Google Earth platform, [8] designed the LoveDA dataset for segmentation, [9] designed the S^2UC dataset for urban village classification, and [10] designed the FAIR1M dataset along with the Gaofen-1 images for object detection. The HSI-CD dataset [11] from the EO-1 Hyperion images can be used for hyperspectral change detection.

Unfortunately, there are no available datasets from meteorological satellites for super-resolution benchmark. Very few datasets on weather images are related to cloud segmentation. Ref. [12] designed the 38-Cloud dataset consisting of 8400 patches of 18 scenes for training and 9201 patches of 20 scenes for testing. Patches are extracted from the Landsat 8 Collection 1 Level-1 scenes with four visible bands. Later, they extended the dataset to the 95-Cloud dataset [13] consisting of 34,701 patches of 75 scenes for training. The test sets in 95-Cloud and 38-Cloud are the same. The patch size is 384 × 384 for 38-Cloud and 95-Cloud datasets. Similar work was done by [14], who proposed a new Landsat-8 dataset (WHU cloud dataset) for simultaneous cloud detection and removal. The dataset consists of six cloudy and cloud-free image pairs in different areas. However, these cloud images are time independent and do not show time continuity, and the image size is too small to learn sufficiently for large-scale neural networks.

To fill the gap of super-resolution meteorological datasets, this paper presents two new super-resolution datasets from the visible bands of the Fengyun-4A (or FY4A) [15] satellite designed for cloud detection. One dataset is a single-channel 8-bit quantized dataset, and the other is a 3-channel 16-bit quantized dataset. Low-resolution images in our datasets are accompanied with corresponding 4-times high-resolution images. The size of the high-resolution images is 10,992 × 4368 which is far larger than the size in commonly used datasets. The total scale of our dataset is comparable to that of DIV2K. However, due to the low resolution, there is less structural information in images, which makes the performance of super-resolution algorithms for natural images necessary of being re-evaluated. The single-channel 8-bit quantized dataset can be used to quickly test the effectiveness of the existing super-resolution algorithms, as it requires no modification to existing code. The 3-channel 16-bit quantized dataset is more realistic to real scenarios of remote sensing application pursuing accurate digital numbers for further quantitative analysis.

The proposed datasets differ from existing datasets for two reasons. On the one hand, since the meteorological satellites focus on adverse air conditions, such as cloud, fog, rain, haze, and so on, their need for super-resolution are reasonably different from commonly used datasets. On the other hand, meteorological satellite images have very high temporal resolutions which creates an image sequence giving the chance for the study of spatiotemporal fusion [16] or spatiotemporal-spectral fusion [17]. The framework of generative adversarial network (GAN) can be used for this topic which is investigated in [18]. Consequently, the super-resolution methods for the proposed datasets can be of more variety to make full use of the repeated scanning as time-series images provide far richer information than images in single-patch super resolution.

In order to evaluate the performance when the datasets are used for super-resolution, state-of-the-art algorithms are used. Our experiments try to touch the performance boundaries of the super-resolution algorithms from the perspective of quantitative remote sensing. That is, we try to uncover how much the best super-resolution algorithm can affect the reconstruction error. The conclusion can be used to assess the practical possibilities of super-resolution algorithms on the proposed dataset.

The following contributions are made in our work.

(1) We present two medium-resolution remote sensing datasets that are the first meteorological datasets and are almost temporally continuous.
(2) We validate the performance bounds of existing single-image super-resolution algorithms on the datasets to provide the baseline for performance improvement.

The remainder of the paper is structured as follows. Section 2 introduces the FY4ASR dataset. Section 3 introduces the experimental schemes, including the state-of-the-art super-resolution algorithms, training strategies, 16-bit preprocessing, and metrics for image quality assessment. Section 4 presents the experimental results which are evaluated visually and digitally. Section 5 gives the conclusion.

2. Proposed FY4ASRgray and FY4ASRcolor Datasets

We propose the FY4ASRgray and FY4ASRcolor datasets for benchmarking both time-based and example-based image super-resolution algorithms. These two datasets are captured by the FengYun-4A (or FY4A) satellite launched by China in 2016 and equipped with sensors, such as Advanced Geostationary Radiation Imager (AGRI) [19], Geostationary Interferometric Infrared Sounder (GIIRS), and Lightning Mapping Imager (LMI).

AGRI is the main payload which has a complex double-scanning mirror mechanism enabling both precise and flexible imaging modes. The flexible mode allows for quick scanning at high minute rates with the loss of spatial resolution, while the precise scans slowly for higher spatial resolutions. AGRI in FY4A has 14 channels with the 0.5–1 km resolutions for visual light (450–490 nm, 550–750 nm) or near infrared (750–900 nm) bands, and 2–4 km for infrared (2.1–2.35 µm) bands. AGRI spends 15 min for full-disc scanning to present a global cloud image. On-board black body is available for calibration of infrared bands at very short time intervals.

The proposed two super-resolution datasets differ in bit length and channel number. Images in FY4ASRgray are 8-bit quantized with single channel, while images in FY4ASRcolor are 16-bit quantized with three channels. All images in FY4ASRgray and FY4ASRcolor are paired, where the ground resolutions are 1 km for high-resolution images and 4 km for low-resolution images. Ref. [20] have tried the 8-bit FY4A data for super-resolution, but the 16-bit reconstruction is far challenging.

The images in FY4ASRgray and FY4ASRcolor datasets are all captured by AGRI full disc scanning covering China (region of China, REGC, see Figure 1) with the 5-min time interval for regional scanning. The images were originally quantized using 16 bits, and the valid range of the data is 0 to 4095. However, many super-resolution algorithms are designed for natural images where 16 bits cannot be fed in. At this point, FY4ASRgray can be used to test the effectiveness of these algorithms. However, the quantization accuracy of FY4ASRgray is insufficient and the spectral information is lost due to the single channel. In contrast, FY4ASRcolor serves to reconstruct the rich information more accurately using the multi-channel information for subsequent segmentation and classification applications, which is in line with the purpose of remote sensing.

FY4ASRgray uses the second band of ARGI spanning the spectral range 550–750 nm for fog and cloud detection. The images were captured between 26 and 27 August 2021, and preprocessed at level 1 including radiometric and geometric corrections. The Level-1 images were then enhanced and quantized to 8-bit integer types ranging from 0 to 255 and stored using lossy JPEG format.

FY4ASRcolor datasets uses the first three bands of AGRI, namely blue (450–490 nm), red-green (550–750 nm), and visible NIR (750–900 nm). The images were captured on 16 September 2021. All the bands are in 16-bit data format after radiometric and geometric correction for Level-1 preprocessing, and stored using lossless TIFF format. Figure 2 presents some patches from the FY4ASRcolor datasets that are non-linearly enhanced. Since the actual quantization is 12 bits, each digital number ranges from 0 to 4095, but the vast majority of values are between 50 and 2100. Considering that many existing algorithms have codes that limit the input within the range 0 to 255, the input and output of existing algorithms needs to be modified to accommodate 16 bits when using their codes.

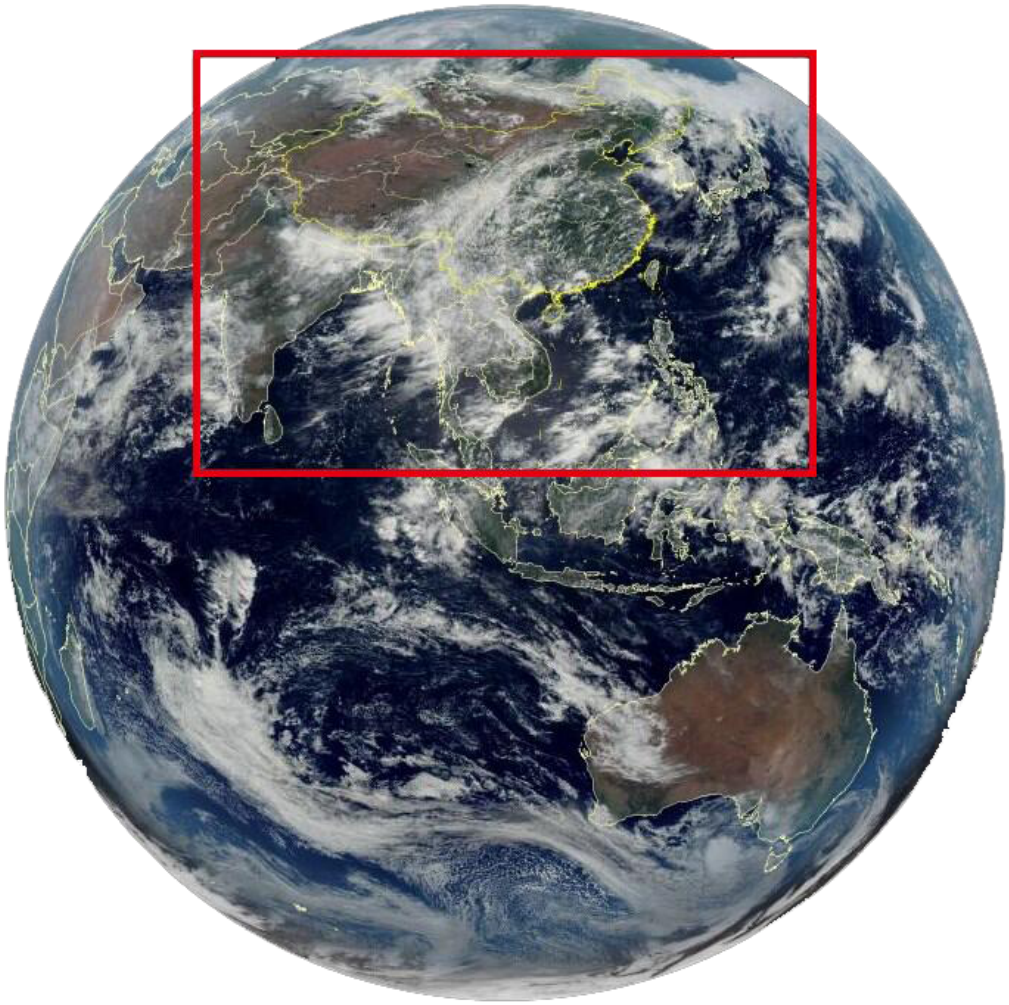


Figure 1. REGC range from the FY4A AGRI 105° scan (marked within the rectangle box).

In terms of data scale, our FY4ASRgray and FY4ASRcolor datasets are comparable to widely used large-scale natural image datasets. FY4ASRgray contains 130 pairs of images, while FY4ASRcolor contains 165 pairs of images. The size of the high resolution image is 10,992 × 4368 and the size of the low resolution image is 2748 × 1092. After eliminating the invalid area, the darkest 1.5% of pixels, and the brightest 1.5% of pixels in FY4ASRcolor, the average percentage of valid pixels is 64.85% and the number of valid pixels is 5.14 billions. In contrast, the number of pixels in the DIV2K dataset is 6.69 billions.

It has to be noted that the low-resolution images in FY4ASRgray and FY4ASRcolor are not downsampled from the high-resolution images, but are acquired by separately mounted sensors of the same type. In addition to the spatial difference, there are minor differences between these two types of images, which we call the sensor difference. Possible causes of the sensor difference include various spectral response curves, solar altitude angles, preprocessing methods, and so on, which makes the difference stochastic and scene dependent [16]. After downsampling the high-resolution images and comparing them with the low-resolution images, the average absolute errors are 1.619 for FY4ASRgray and 29.63 for FY4ASRcolor. To obtain the optimal super-resolution performance, it is necessary for algorithms to model the sensor difference.

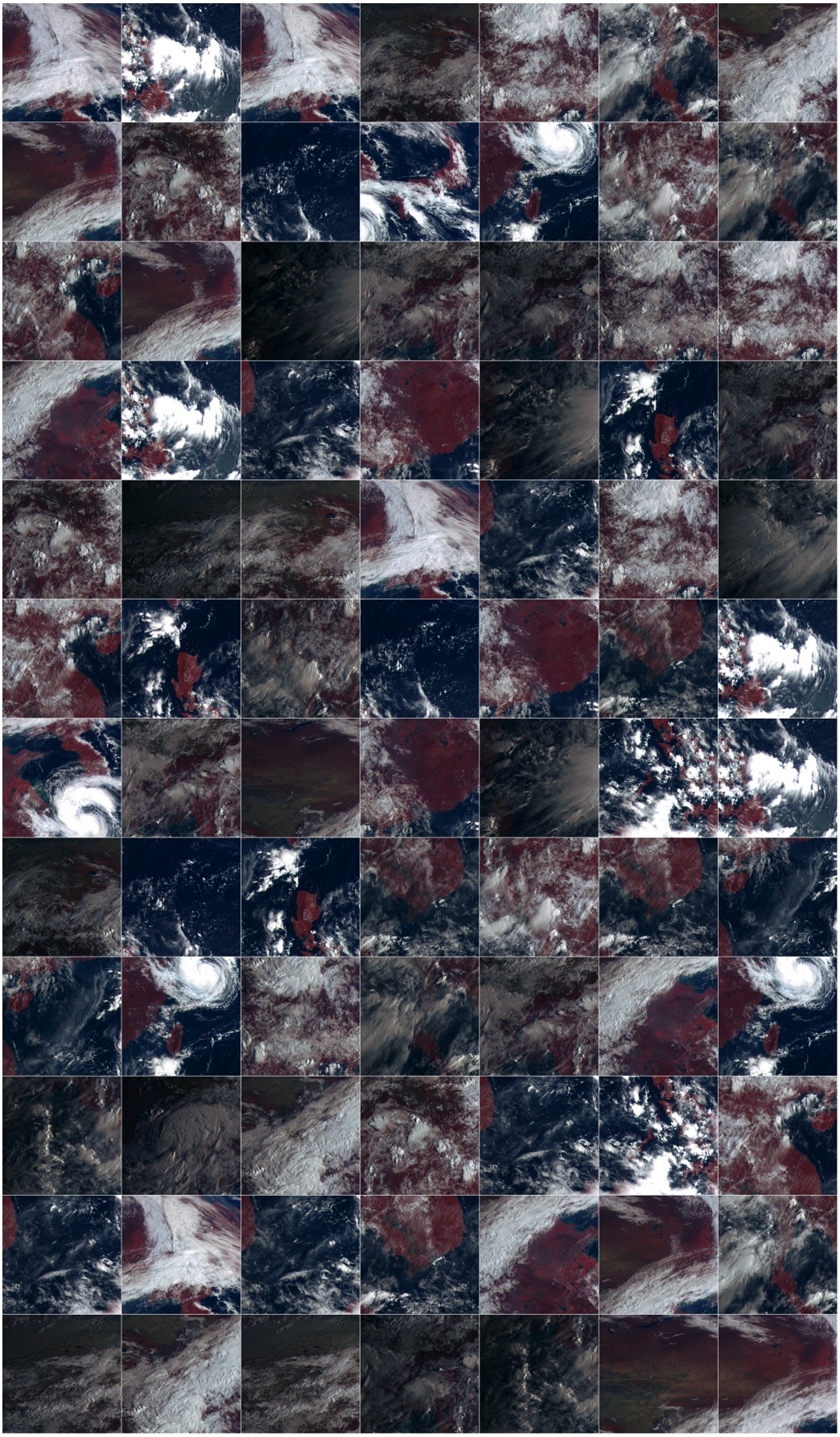


Figure 2. Image patches from the FY4ASRcolor dataset.

3. Experimental Scheme

The FY4A dataset file can be downloaded from the URL github.com/isstncu/fy4a, accessed on 25 October 2022, which takes up 39.1 GB of disk size. To test the usability of the new datasets, state-of-the-art super-resolution algorithms are gathered for contest. The technique of shift learning is used to shorten the training time and approach to the optimal parameters. The reconstruction results will be evaluated with objective metrics. The super-resolution algorithms, evaluation metrics, and the additional 16-bit processing are described in this section.

3.1. Methods for Validation

All the super-resolution algorithms for test were proposed in recent five years with deep neural networks modelling the upsampling process, as will be introduced in Figures 3–11. In these figures, (a, b) denotes that the channel numbers are a for input and b for output, conv3 denotes the 3×3 convolution, conv1 denotes the 1×1 convolution, FC denotes the fully connected layer, ReLU denotes the rectified linear unit for activation, and Mult denotes the multiplication operation.

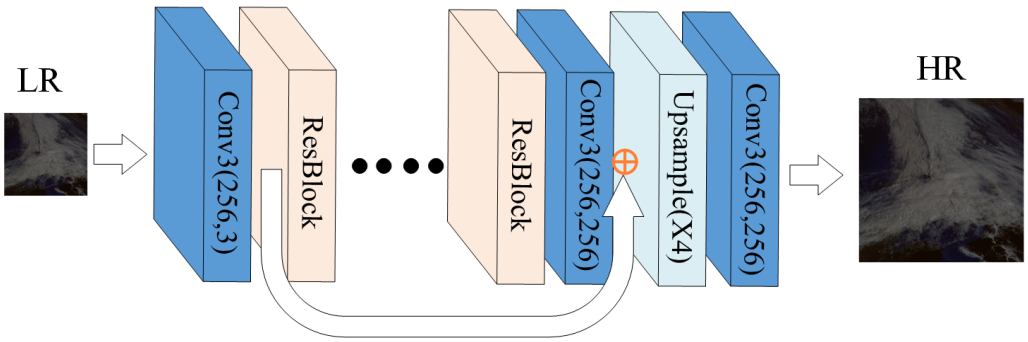


Figure 3. Structure of enhanced deep residual networks (EDSR).

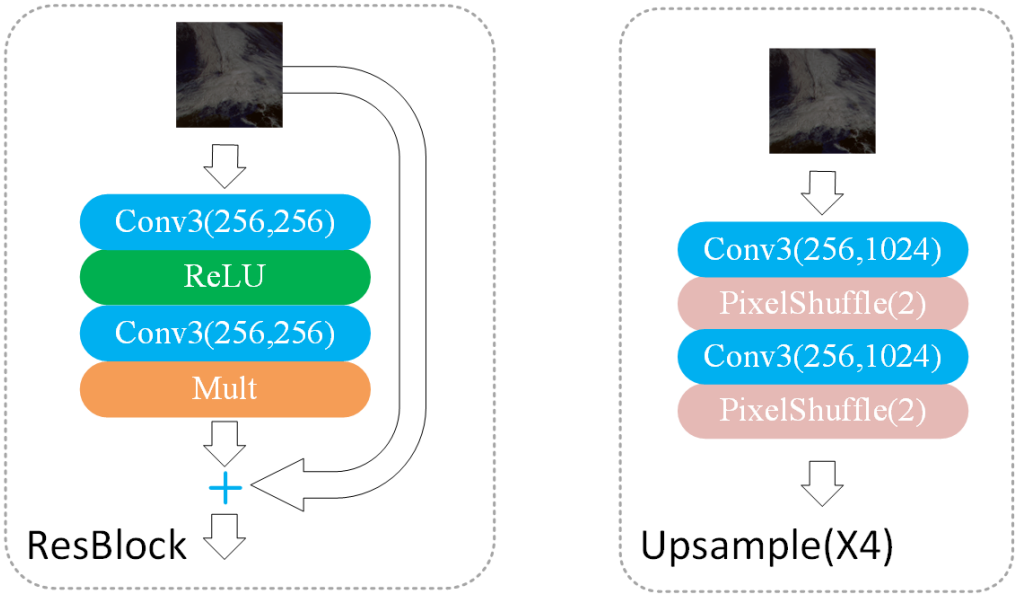


Figure 4. Blocks in EDSR.

Ref. [21] won the NTIRE2017 super-resolution challenge with the proposed enhanced deep residual networks (EDSR) by removing unnecessary modules in conventional residual networks. The structure and creative blocks of EDSR are presented in Figures 3 and 4.

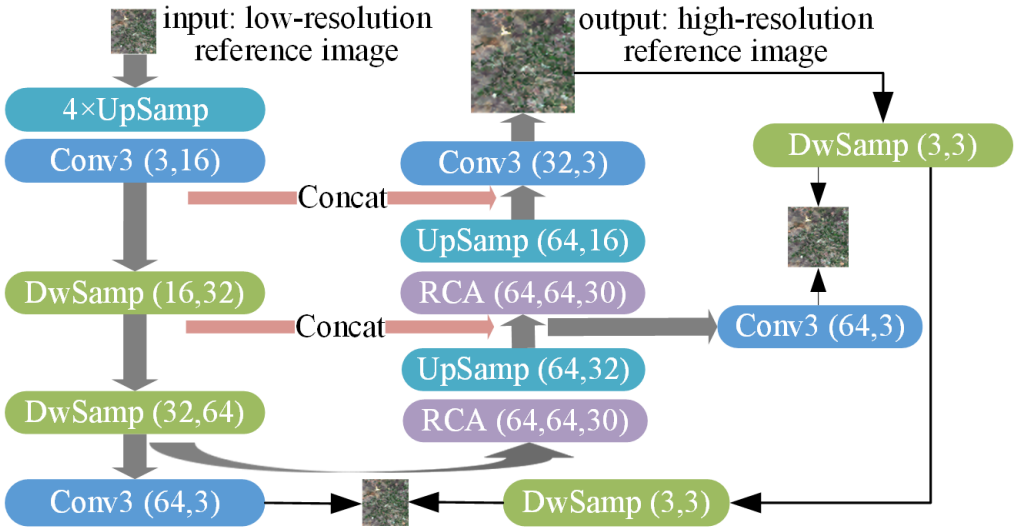


Figure 5. Structure of dual regression networks (DRN).

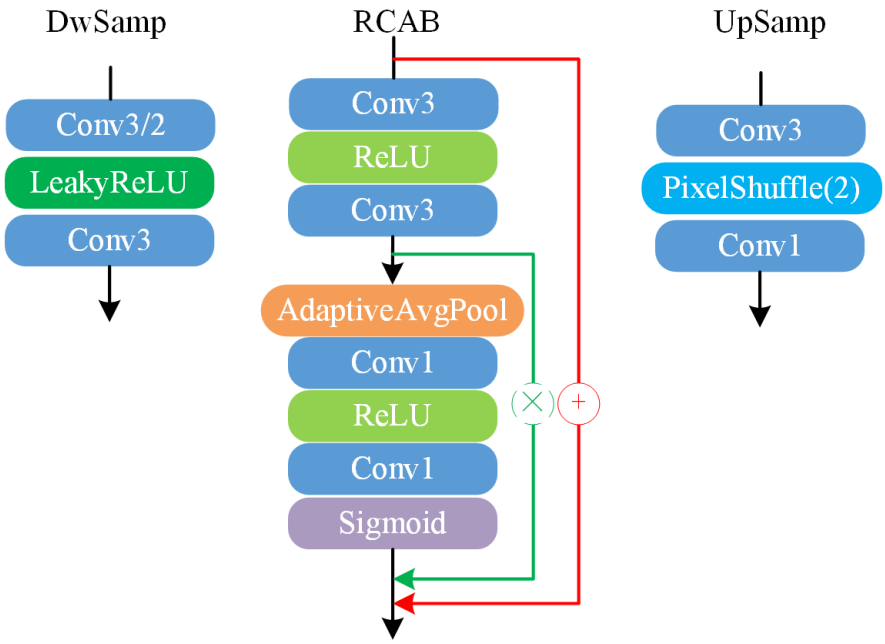


Figure 6. Blocks in DRN.

Ref. [22] proposed the dual regression networks (DRN) for paired and unpaired super-resolution tasks. In DRN, an additional constraint is introduced on the low-resolution data to reduce the span space of possible solutions. A closed loop is designed with a feedback mapping to estimate the down-sampling kernel targeting the original low-resolution images. The structure and blocks of DRN are presented in Figures 5 and 6.

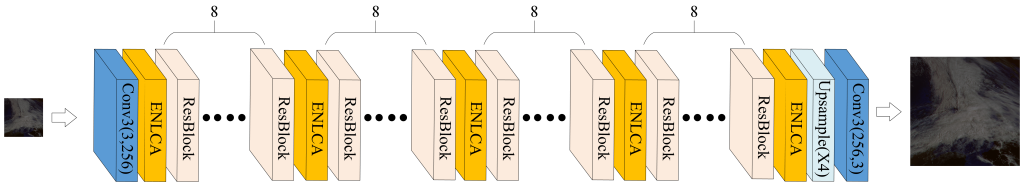


Figure 7. Structure of the efficient non-local contrastive attention (ENLCA) network.

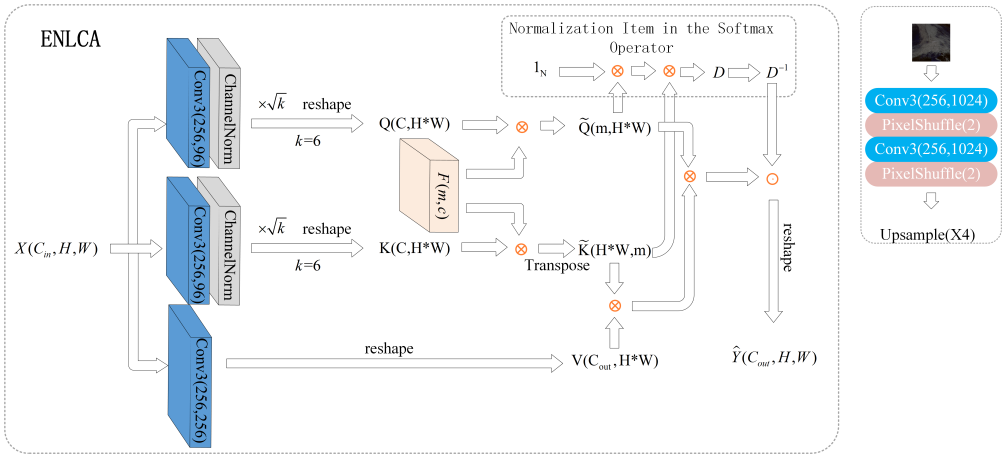


Figure 8. Blocks in ENLCA.

Ref. [23] propose a new model with an efficient non-local contrastive attention (ENLCA) module to perform long-range visual modeling. The ENLCA module the leverage more relevant non-local features. The network structure is illustrated in Figure 7, and the critical ENCLA module is in Figure 8.

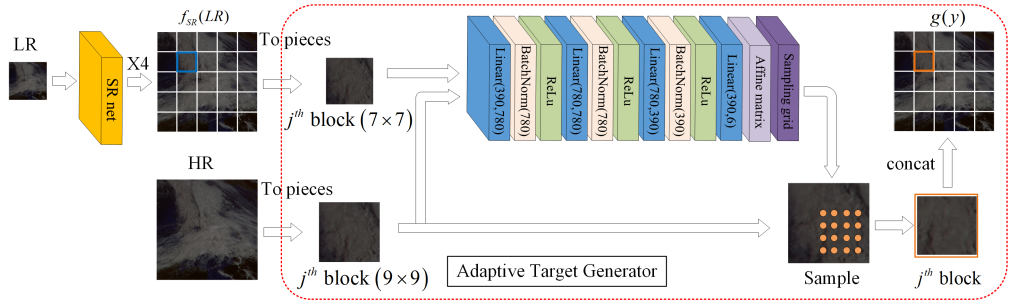


Figure 9. Structure of adaptive target generator (AdaTarget).

The concept of adaptive target is introduced in [24] by generating from the original ground truth target with a transformation to match the output of the super-resolution network. To deal with the ill-posed nature of super-resolution, the adaptive target provides the flexibility of accepting a variety of valid solutions. Their adaptive target generator (AdaTarget) is an improvement to existing super-resolution networks. In [24], ESRGAN is advised as the baseline super-resolution network (SR net in Figure 9). In the PIRM2018-SR challenge, AdaTarget outweighed all the other super-resolution algorithms. The structure of AdaTarget is presented in Figure 9.

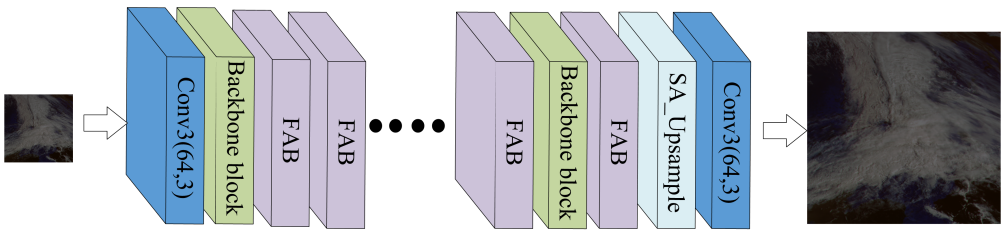


Figure 10. Structure of scale-arbitrary super-resolution (ArbRCAN).

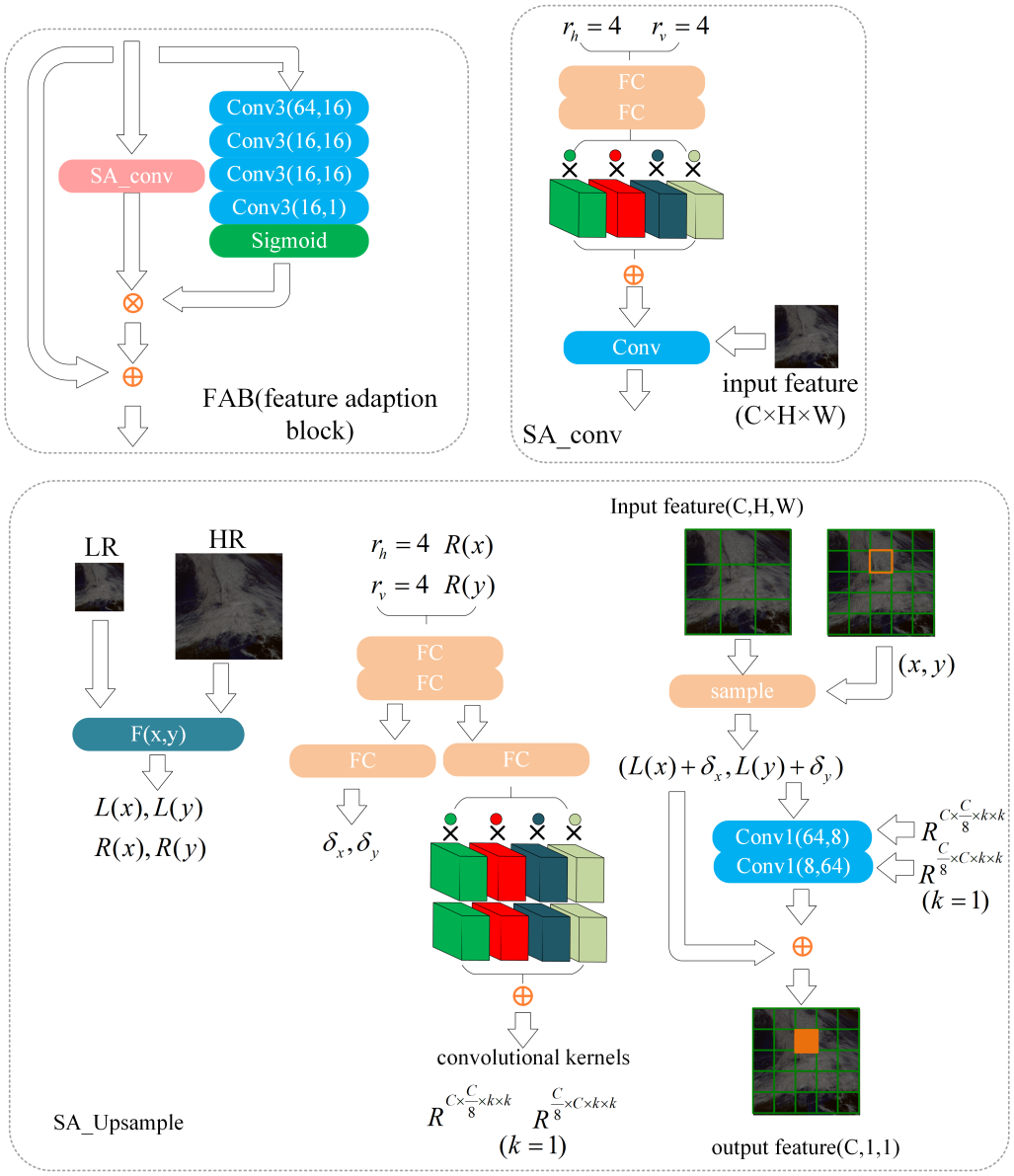


Figure 11. Blocks in ArbRCAN.

Ref. [25] proposed the scale-arbitrary super-resolution (ArbRCAN) network in the form of plug-in modules to enable existing super-resolution networks for scale-arbitrary super-resolution with a single model. ArbRCAN can be easily adapted to scale-specific networks with small additional computational and memory cost, which is very help for remote sensing images as sources may differ a lot. The structure and blocks of ArbRCAN are presented in Figures 10 and 11, where the backbone network can be existing super-resolution network where EDSR was used in [25].

3.2. Training and 16-bit Preprocessing

Two sets of model parameters are prepared for each dataset. All the methods were designed for natural images, and they have been trained with natural datasets, such as DIV2K or Flickr1024. Intuitively, we hope to know the performance of the model trained on natural image datasets and used for remote sensing images. On the other hand, an improvement in reconstruction accuracy is expected by training the model on a more matched dataset. The performance differences correspond to the migration ability of the models. Therefore, two experiments are designed using models either pre-trained on natural images or trained on proposed datasets. To accelerate the model training, the initial

parameters for FY4ASRgray and FY4ASRcolor training are derived from the results of pre-training on natural images.

To test the FY4ASRcolor dataset with the natural image pre-trained models, the 16-bit quantization has to be dealt with. The pixel values of natural images usually range from 0 to 255. However, the FY4ASRcolor dataset is 16-bit quantized with the maximum value 4095, which cannot be tested directly using the pre-trained parameters. Due to outlier pixels, a linearly stretched remote sensing image is much too dark to maintain structural information. By approaching the style of training data, we propose to non-linearly stretch the pixel values in FY4ASRcolor with saturate thresholds. In the transformation, the values of the darkest 1.5% pixels are set to 0, and the values of the brightest 1.5% pixels are set to 255. The values of the remaining image pixels are linearly stretched to [0, 255] and recorded as floating point numbers. Stretched images can be put into the network for training and reconstruction. The reconstructed results should be linearly stretched back to [0, 4095] using the original thresholds defining the darkest and brightest 1.5% pixels. The forward stretch and backward stretch are performed band by band. Our tests show that a threshold of 1.5% allows the contrast of the image to be enhanced significantly without noticeable loss of radiometric fidelity. The full processing steps are demonstrated in Figure 12.

For the FY4ASRgray dataset, after eliminating the surrounding invalid areas, the high-resolution image size was cropped to 7704 × 4276. By slicing the 1000 M resolution image into 1024 × 1024 blocks, a total of 4662 image blocks were obtained, each with a 256 × 256 low-resolution image block corresponding to it. Among the extracted image blocks, 4608 were used for training, 36 for validation, and 18 for testing.

As for the FY4ASRcolor dataset, the same block extraction strategy as FY4ASRgray was used after the 1.5% saturate stretch. Finally, we obtained 3057 pairs of image blocks. The size of high-resolution image blocks is 1024 × 1024, and the corresponding low-resolution size is 256 × 256. Among the extracted image blocks, 2507 are used for training, 270 for validation, and 280 for testing.

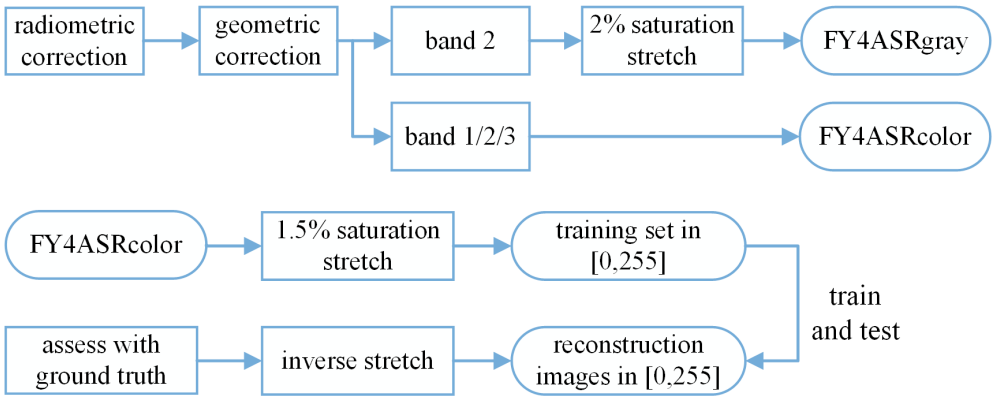


Figure 12. Processing steps for the two datasets.

3.3. Metrics

Referenced metrics are used to assess the performance of the reconstructed images. Peak signal-to-noise ratio (PSNR), Root Mean Square Error (RMSE), and Correlated Coefficient (Corr) measure the radiometric discrepancy. Structural Similarity (SSIM) measures the structural similarity. Spectral angle mapper (SAM), relative average spectral error (RASE) [26], relative dimensionless global error in synthesis (ERGAS) [27], and Q4 [28] measure the color consistency. The ideal results are 1 for SSIM, Corr, and Q4 while 0 for SAM, ERGAS, and RASE. The formulas for these metrics are listed below.

Two different PSNRs are calculated to deal with the various word lengths. For a FY4ASRgray reconstruction result, PSNR is calculated with $\mathrm{PSNR} = 20\log(255/\mathrm{RMSE})$. For FY4ASRcolor, PSNR is calculated with $\mathrm{PSNR} = 20\log(4095/\mathrm{RMSE})$.

In addition to the full reference metrics, no-reference approaches [29] are also introduced to assess image quality, including the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [30], Naturalness Image Quality Evaluator (NIQE) [31], and the Perception-based Image Quality Evaluator (PIQE) [32]. Since these methods do not work properly on 16-bit data, they are only used on the FY4ASRgray test to calculate the no-reference image quality scores.

4. Experimental Results

With the super-resolution methods and metrics, the reconstruction results are evaluated. The scores are listed in this section for digital comparison. Some reconstruction details are also demonstrated for visual comparison.

4.1. Visual Comparison

Four image patches are selected from the test set for the presentation of the results, which are presented in Figures 13–16. A 2% saturation stretch is used to convert the 16-bit result images to the range [0, 255]. In particular, each group of images was stretched using the same thresholds as the ground truth images used. By comparing the results, it is observed that the differences between them are small. The spectral fidelity is good for all algorithms. However, when the structural information is concerned, a large loss of detail can be observed. By carefully comparing the different algorithms, it is inferred that EDSR has the best color consistency, as well as the blurring details, and AdaTarget shows a slight advantage in detail.

4.2. Digital Comparison

PSNR, RMSE, and Corr measure radiometric accuracy, SSIM measures structural similarity, and BRISQUE, NIQE, and PIQE reflect the degree of acceptance by the human eye. The evaluation results are listed in Tables 1–6, where the bolded numbers highlight the best scores across algorithms.

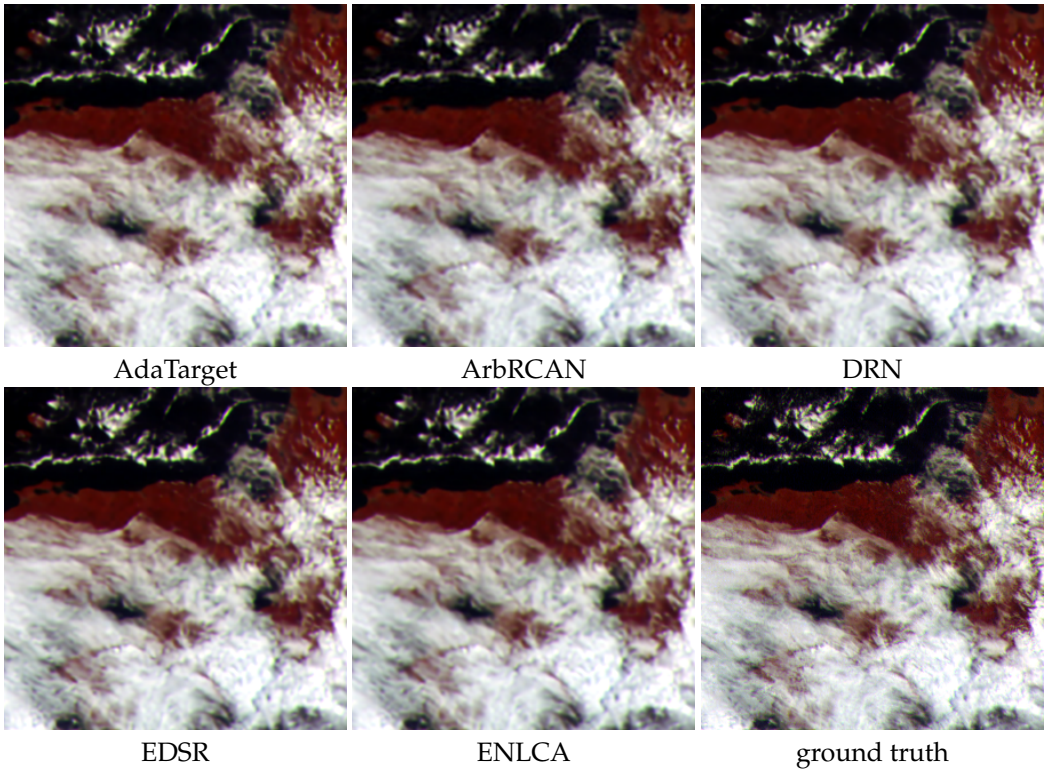


Figure 13. Local manifestation of the super-resolution results on the FY4ASRcolor dataset: Patch 1.

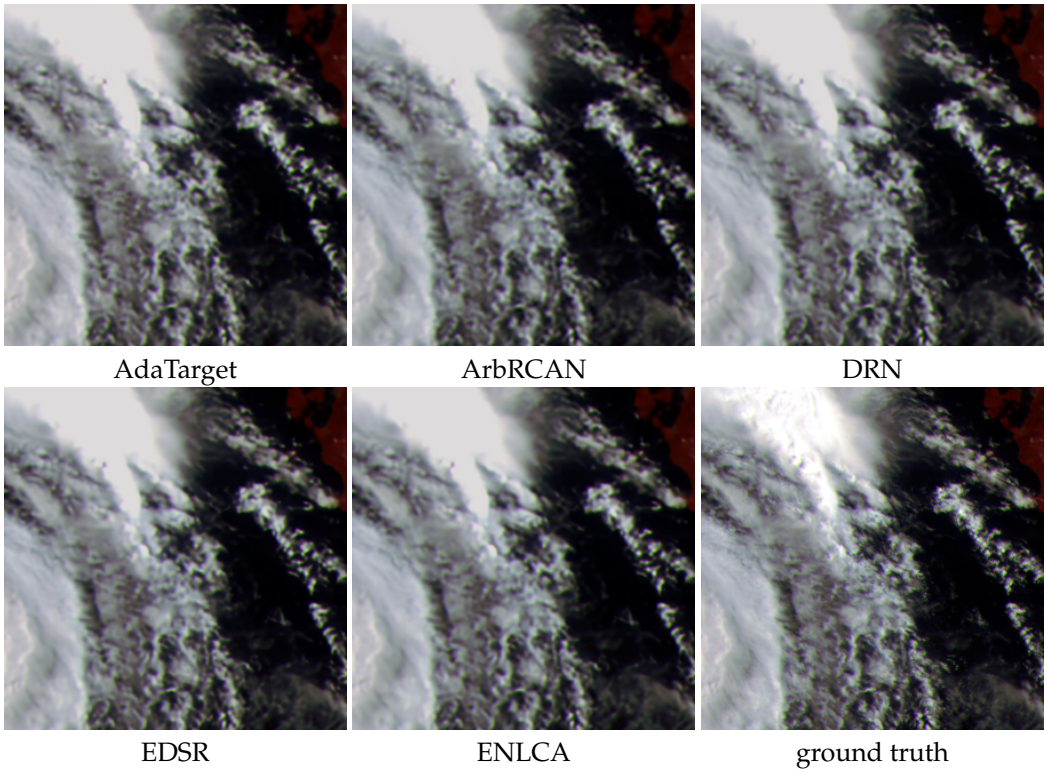


Figure 14. Local manifestation of the super-resolution results on the FY4ASRcolor dataset: Patch 2.

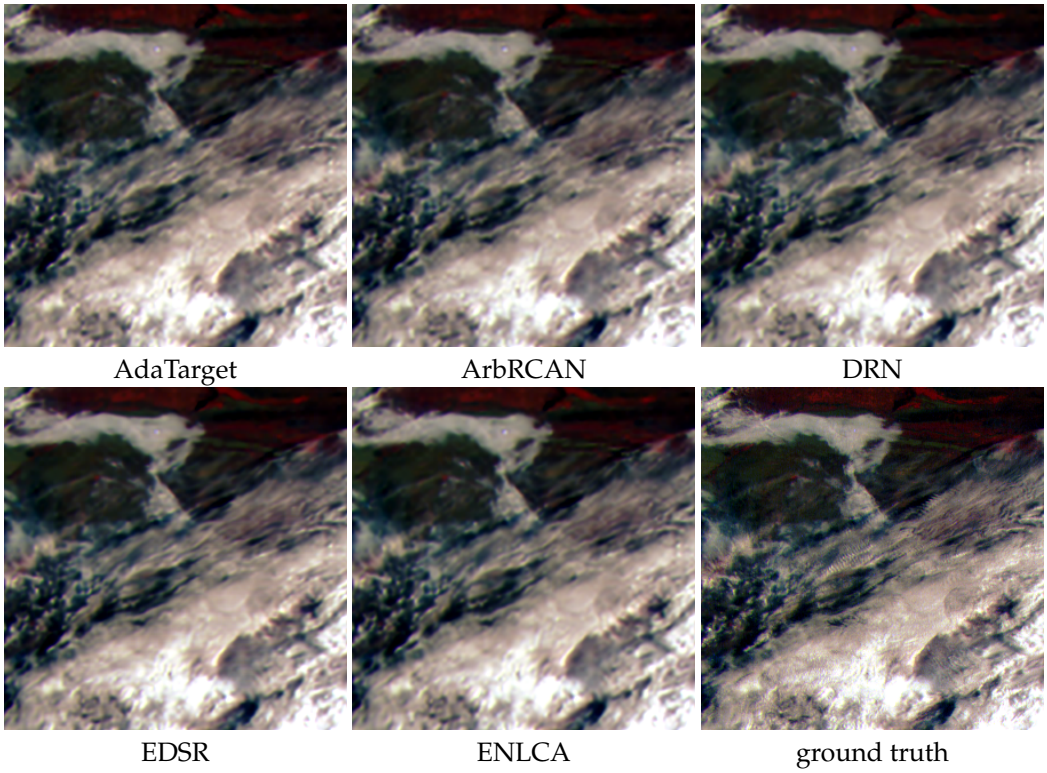


Figure 15. Local manifestation of the super-resolution results on the FY4ASRcolor dataset: Patch 3.

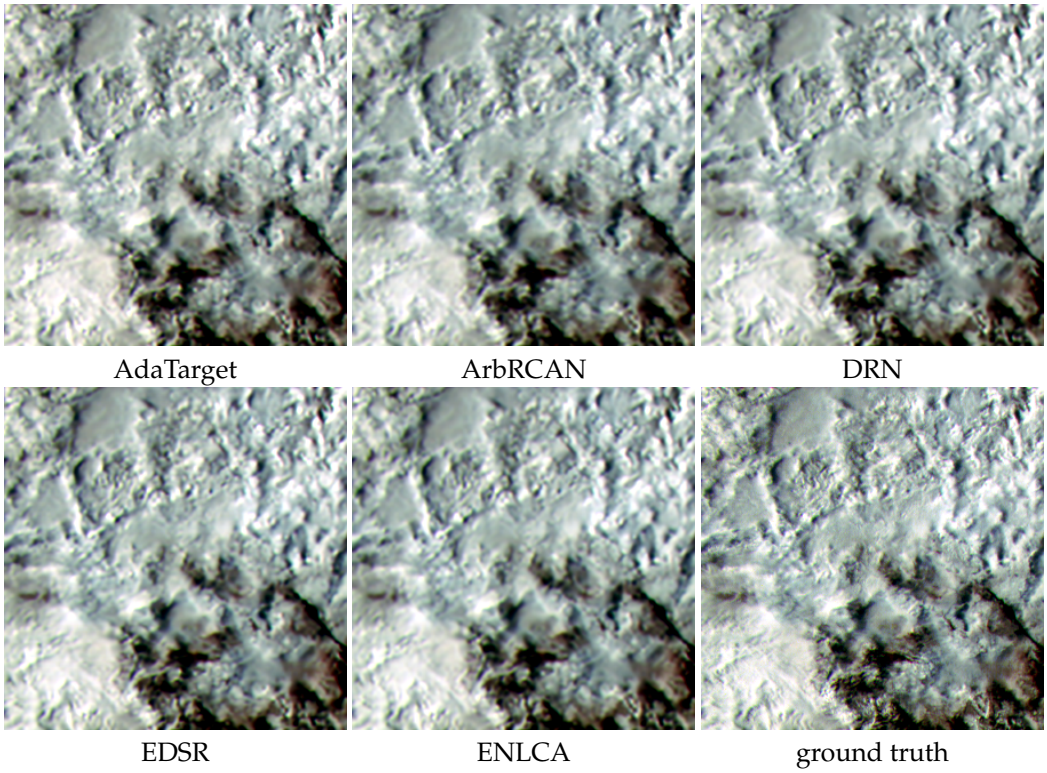


Figure 16. Local manifestation of the super-resolution results on the FY4ASRcolor dataset: Patch 4.

Table 1. Evaluation on FY4ASRgray with Pre-trained Parameters.

	PSNR	RMSE	SSIM	Corr	BRISQUE	B	PIQE
AdaTarget	30.889	7.359	0.923	0.961	**48.662**	5.175	76.970
ArbRCAN	31.887	6.658	0.931	0.964	51.222	5.330	82.094
DRN	**32.098**	**6.481**	**0.936**	**0.971**	50.181	**5.104**	**76.836**
EDSR	31.924	6.618	0.934	0.966	51.880	5.575	86.715
ENLCA	31.734	6.734	0.933	0.967	50.007	5.526	87.142

Table 1 shows the evaluation results of the FY4ASRgray dataset, where all the parameters of the algorithms were pre-trained on the DIV2K dataset. Although DIV2K is designed for natural images, the experimental results give very high reconstruction quality. The best radiometric accuracy comes from the DRN algorithm proposed in 2021, but EDSR and ENLCA are more acceptable by human vision. The good performance is partially explained by the non-linear enhancement of the 16-bit satellite data to make it close to the structural features of natural images.

Table 2. Evaluation on FY4ASRgray with re-trained parameters.

	PSNR	RMSE	SSIM	Corr	BRISQUE	NIQE	PIQE
AdaTarget	**32.936**	**5.845**	**0.949**	**0.974**	55.277	5.321	**69.261**
ArbRCAN	32.020	6.525	0.934	0.969	**52.239**	5.695	85.274
DRN	32.678	6.036	0.945	0.973	56.544	**5.305**	73.763
EDSR	32.707	6.008	0.947	**0.974**	57.697	6.064	87.603
ENLCA	32.907	5.874	**0.949**	**0.974**	58.092	6.112	85.989

Table 2 shows the evaluation results of training using the FY4ASRgray dataset. Compared to the results of the DIV2K pre-training in Table 1, the reconstruction scores from the

re-training parameters are all slightly improved. Interestingly, the evaluation results for all algorithms are close. ArbRCAN falls slightly behind in terms of radiometric accuracy, while AdaTarget is the least recognized by human vision. All PSNRs are between 32 dB and 33 dB. It can be inferred from the small differences that these algorithms may have coincidentally reached their performance limits.

Table 3. Radiometric and structural evaluation on FY4ASRcolor with pre-trained parameters.

	PSNR	SSIM	RMSE			Correlated Coefficient		
			Band1	Band2	Band3	Band1	Band2	Band3
AdaTarget	**29.223**	**0.618**	**116.264**	**158.832**	**166.606**	**0.913**	**0.897**	**0.903**
ArbRCAN	28.717	0.608	123.249	170.134	175.077	0.912	0.893	0.901
DRN	28.922	0.599	119.516	166.740	171.436	0.908	0.886	0.897
EDSR	28.542	0.595	125.390	175.339	178.229	0.909	0.887	0.898
ENLCA	28.073	0.593	131.300	183.972	185.968	**0.913**	0.891	0.901

Table 4. Spectral evaluation on FY4ASRcolor with pre-trained parameters.

	SAM	ERGAS	RASE	Q4
AdaTarget	0.079	**0.261**	**0.253**	**0.903**
ArbRCAN	**0.078**	0.263	0.255	0.901
DRN	0.079	0.271	0.263	0.895
EDSR	**0.078**	0.269	0.260	0.897
ENLCA	**0.078**	0.269	0.260	0.900

Table 3 presents the radiometric and structural evaluation results on the FY4ASRcolor dataset, where all parameters are from the DIV2K pre-training. It can be seen that the reconstruction quality of the parameters using DIV2K pre-training is very low. The highest performance comes from AdaTarget with a PSNR of only 29.223 and a structural similarity of only 0.618. The normalized correlation coefficient around 0.9 is also not good enough. The spectral evaluation in Table 4 shows very close spectral similarities for all algorithm results, but all have significant spectral distortions.

Table 5. Radiometric and structural evaluation on FY4ASRcolor with re-trained parameters.

	PSNR	SSIM	RMSE			Correlated Coefficient		
			Band1	Band2	Band3	Band1	Band2	Band3
AdaTarget	37.617	0.964	51.624	64.531	61.548	0.986	0.987	0.990
ArbRCAN	37.183	0.963	53.435	68.251	64.617	0.985	0.986	0.989
DRN	37.835	0.968	50.906	63.390	59.562	0.986	**0.988**	**0.991**
EDSR	**38.095**	**0.971**	**49.805**	**61.710**	**57.742**	**0.987**	**0.988**	**0.991**
ENLCA	37.790	0.968	51.004	63.693	59.905	0.986	0.987	0.990

Table 6. Spectral evaluation on FY4ASRcolor with re-trained parameters.

	SAM	ERGAS	RASE	Q4
AdaTarget	0.045	0.094	0.092	0.987
ArbRCAN	0.047	0.099	0.097	0.986
DRN	0.044	0.092	0.090	**0.988**
EDSR	**0.043**	**0.089**	**0.088**	**0.988**
ENLCA	0.044	0.092	0.090	**0.988**

The results of training and testing using the FY4ASRcolor dataset are shown in Figures 5 and 6. By comparing them with pre-trained results in Figures 3 and 4, it can

be concluded that the quality of the reconstruction after targeted training is substantially improved. Both radiometric, structural, and spectral errors are very small and difficult to be detected subjectively by the human eye. EDSR achieves the highest performance in all metrics. Although the results of ArbRCAN are not as good as the other algorithms, the performance difference is small.

Different from natural images, an important purpose of remote sensing image reconstruction is for quantitative remote sensing, such as inversion of vegetation indices, water indices, forest indices, and so on. It requires that the radiometric error must be within a certain range. To evaluate the possibility of super-resolution for quantitative remote sensing, we use the widely used relative radiometric index. For an image, it is calculated as the RMSE divided by the mean value of the true image. Then, 5 and 15% are two commonly used thresholds in ground pre-processing systems for satellite data. For multispectral images, the relative radiometric error should not exceed 5%. When the data are used for quantitative analysis, the qualified error should not exceed 15%. Corresponding to the two strategies of pre-training and re-training, we counted the correctness of the reconstructed pixels that meet the requirements of these two thresholds, respectively, and the results are presented in Table 7, where the bolded numbers present the highest percentage of correctness for various thresholds.

Table 7. Correctness ratio on FY4ASRcolor (%).

	Pre-Trained		**Re-Trained with FY4ASRcolor**	
	<5% Allowed	**<15% Allowed**	**<5% Allowed**	**<15% Allowed**
AdaTarget	8.17%	27.08%	42.97%	77.98%
ArbRCAN	9.32%	25.87%	45.52%	77.25%
DRN	9.82%	28.36%	46.46%	79.69%
EDSR	10.01%	**30.59%**	**46.75%**	**81.13%**
ENLCA	**10.06%**	28.97%	46.60%	79.85%

The correct image element rates in Table 7 again reflect the need for re-training. Using a more stringent 5% criterion, half of the re-trained results meet the requirements for quantitative remote sensing applications, while only 10% of the pre-trained results do. This large gap can be explained by the large difference between natural and remotely sensed images. It also indirectly explains the necessity of constructing this dataset. In this test, the EDSR results achieved the highest score, while AdaTarget had the lowest score.

5. Discussion

The time of the images in the FY4ASRcolor dataset is on 16 September 2021 which starts at 00:30 and ends at 23:43. The acquisition duration of each image is 258 s. In terms of the time intervals of two adjacent images, most of them are 258 s, and the maximum value is 3084 s. The strong temporal continuity allows the data to be used for time-related studies. Therefore, two new experimental schemes were explored, namely sequence super-resolution and spatiotemporal fusion. The results of these studies are expected to uncover the feasibility of the new datasets for prediction of temporal correlation. An additional test is made to find the generalization ability. In these studies, only the FY4ASRcolor dataset is used as it better suits for the purpose of remote sensing.

5.1. Sequence Super-Resolution

We constructed a training set and a test set by considering the sequence images as a video, and performed a test of video super-resolution. To construct the training set, 40 various locations were selected. In total, 84 pairs of temporally consecutive patches were extracted from each location, which were divided into 12 groups in time order. Each group contains 7 pairs of temporally contiguous patches as a video clip for reconstruction. The patch sizes of each pair are 1024×1024 and 256×256, cropped from the 1 km and 4 km

images, respectively. After removing the groups with excessive darkness, 347 valid groups of video clips were finally obtained out of 480 groups of sequential images for training.

Similar to the training set, the test set was constructed to contain 10 groups of sequential patches at 10 various locations. These 10 locations are included in the 40 locations of the training set. Each group contains 10 pairs of temporally consecutive patches. The patch sizes of each pair are 1024×1024 and 256×256, cropped from the 1 km and 4 km images, respectively. The 100 images used in the test set are beyond the training set.

The Zooming Slow-Mo algorithm [33] is chosen to perform the sequence super-resolution, and its results are evaluated on the 100 test images. Only with the pre-trained model, the average PSNR is 28.2371 dB. With the pre-trained model as the initial value and the constructed dataset for re-training, the average PSNR is 29.2174 dB. When trained only with constructed training set where the pre-trained model is not used, the average PSNR is 30.1253 dB. By comparing these scores with that in Tables 3 and 5, it is concluded that the gap between our FY4ASRcolor dataset and commonly used video sequence datasets is huge. To reconstruct sequence remote sensing images, both the a priori image structure and the sequence change pattern have to be learned, such that matching datasets is becoming difficult.

5.2. Spatiotemporal Fusion

Spatiotemporal fusion is a solution to enhance the temporal resolution of high spatial resolution satellites by exploiting the complementarity of spatial and temporal resolutions between satellite images of different sources. Typical studies are carried out between MODIS and Landsat satellites, which have revisit periods of 1 and 16 days, respectively. A typical spatiotemporal fusion needs three reference images. Assuming that MODIS captures images at moments t_1 and t_2 while Landsat took an image only at moment t_1, spatiotemporal fusion algorithms try to predict the Landsat images at moment t_2 with the three known images.

The FY4ASRcolor dataset is ideal for conducting spatiotemporal fusion studies. Different from MODIS and Landsat, the two known images in the FY4ASRcolor dataset were taken at the exactly same time. They also have the same sensor response, which eliminates the fatal sensor discrepancy issue in fusing MODIS and Landsat. A similar work was carried out by us for the spatiotemporal-spectral fusion of the Gaofen-1 images [17], but it has only a 2-fold difference in spatial resolution. The use of the FY4ASR dataset for spatiotemporal fusion may provide new fundamental data support for this research topic.

We try to use two methods for spatiotemporal fusion, namely FSDAF and SSTSTF [16]. FSDAF is a classical algorithm, while SSTSTF is one of the latest algorithms based on neural networks. SSTSTF requires large amount of data for training, otherwise the performance is not as good as FSDAF. However, FSDAF fails in our test because it cannot give legible images. The changing sunshine intensities lead to a huge variation in the reflection of the features, which may exceed the temporal difference tolerance that FSDAF can reconstruct for surface reflectance. In contrast, SSTSTF can accomplish the reconstruction successfully.

For SSTSTF, paired images from 12 moments were used to construct the dataset. Each high-resolution image has a size of 2880×7680. Images from 9 moments were used for training, and they formed 8 groups. The test used images from 3 other moments, two of which were set as the prediction time. The reconstruction PSNRs are 32.9605 dB for 6:30 and 36.8904 dB for 11:30, respectively, after removing the dark areas from the reconstructed images as it may elevate PSNR values unfairly. The results show that the reconstruction quality of spatiotemporal fusion is weakly less than that of single-image super-resolution. Considering that the amount of training data is far smaller than that used for training of single-image super-resolution, spatiotemporal fusion algorithms need to be carefully designed to adapt to this new dataset.

5.3. Generalization of Trained Models across Datasets

In order to evaluate the generalization of the model trained based on the FY4ASRcolor dataset for remote sensing, it is planned to apply the trained models to other datasets. Unfortunately, the existing studies use images with ultra-high resolutions, which prevents us from finding matching application scenarios. Finally, the datasets used in [34] were tested. Two datasets were involved in the experiment, the 0.3 m UC Merced dataset and the 30 m to 0.2 m NWPU-RESISC45 dataset. The tested images are denseresident191 in the UC Merced dataset and railwaystation565 in the NWPU-RESISC45 dataset. The scores of PSNR evaluation are listed in Table 8 where the bolded numbers highlight the best scores across algorithms. The results from pre-training models are close to the values in [34]. However, re-training on FY4ASRcolor leads to a substantial decrease in reconstructing high-resolution remote sensing images. This convinces us again that the characteristics of our data are quite different from other datasets. This conclusion is easily concluded, as meteorological satellite images have to lose spatial resolution to ensure high temporal resolution. Knowledge of structural details cannot be learned from low-resolution images to reconstruct complex structures of high-resolution images. On the contrary, the temporal repetition and spectral features play much greater roles in the reconstruction process.

Table 8. PSNR Evaluation for generalization with DIV2K pre-training or FY4ASRcolor re-training.

		AdaTarget	ArbRCAN	DRN	EDSR	ENCLA
denseres-idential91	pre-train	26.0674	28.6735	**28.9303**	28.4478	24.4844
	re-train	21.8658	**23.1110**	23.0275	22.8833	21.8143
railways-tation565	pre-train	19.8586	**20.1817**	19.4924	19.2759	20.0070
	re-train	17.2072	16.7615	15.8631	16.1601	**17.2335**

6. Conclusions

Two meteorological satellite datasets are developed for super-resolution test which are named as FY4ASRgray and FY4ASRcolor, respectively. The images in datasets were captured the Chinese Fengyun-4A satellite. Images in FY4ASRgray are 8-bit quantized with single channel, while images in FY4ASRcolor are 16-bit quantized with three channels. All images in FY4ASRgray and FY4ASRcolor are paired with high temporal resolutions captured by true sensors, where the ground resolution of the high-resolution image has a ground resolution of 1 km and the low-resolution image has a ground resolution of 4 km. At the experimental stage, five state-of-the-art super-resolution algorithms are used with four types of evaluation criteria to assess the usability of the data. Overall, the radiometric and spectral fidelity of the data are good, and almost half of the pixels can meet the requirement of quantitative remote sensing. Visually comparison shows that the further improvement is needed for super-resolution algorithms to recover missing details. Additional experiments are made on FY4ASRcolor for sequence super-resolution, spatiotemporal fusion, and generalization test.

Author Contributions: Methodology, Z.C.; software, C.Z.; validation, C.Z.; formal analysis, J.W. (Jingbo Wei); investigation, Z.C.; resources, J.W. (Jingsong Wang); data curation, J.W. (Jingsong Wang); writing—original draft preparation, J.W. (Jingbo Wei); writing—review and editing, J.W. (Jingbo Wei); visualization, J.W. (Jingbo Wei). All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (No. 42267070 and 61861030), and the Jiangxi Provincial Institute of Water Sciences (No. 2021SKTR07 and 202224ZDKT11).

Data Availability Statement: The FY4ASRcolor and FY4ASRgray dataset files can be downloaded from https://github.com/isstncu/fy4a.

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

References

1. Wei, J.; Huang, Y.; Lu, K.; Wang, L. Nonlocal Low-Rank-Based Compressed Sensing for Remote Sensing Image Reconstruction. *IEEE Geosci. Remote. Sens. Lett.* **2016**, *13*, 1557–1561. [CrossRef]
2. Agustsson, E.; Timofte, R. NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Honolulu, HI, USA, 21–26 July 2017.
3. Young, P.; Lai, A.; Hodosh, M.; Hockenmaier, J. From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions. *TACL* **2014**, *2*, 67–78. [CrossRef]
4. Xie, Y.; Feng, D.; Chen, H.; Liao, Z.; Zhu, J.; Li, C.; Wook Baik, S. An omni-scale global-local aware network for shadow extraction in remote sensing imagery. *ISPRS J. Photogramm. Remote. Sens.* **2022**, *193*, 29–44. [CrossRef]
5. Xie, Y.; Feng, D.; Chen, H.; Liu, Z.; Mao, W.; Zhu, J.; Hu, Y.; Baik, S.W. Damaged Building Detection From Post-Earthquake Remote Sensing Imagery Considering Heterogeneity Characteristics. *IEEE Trans. Geosci. Remote. Sens.* **2022**, *60*, 1–17. [CrossRef]
6. Xie, Y.; Feng, D.; Shen, X.; Liu, Y.; Zhu, J.; Hussain, T.; Baik, S.W. Clustering Feature Constraint Multiscale Attention Network for Shadow Extraction From Remote Sensing Images. *IEEE Trans. Geosci. Remote. Sens.* **2022**, *60*, 1–14. [CrossRef]
7. Lei, Z.; Zeng, Y.; Liu, P.; Su, X. Active deep learning for hyperspectral image classification with uncertainty learning. *IEEE Geosci. Remote. Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]
8. Wang, J.; Zheng, Z.; Ma, A.; Lu, X.; Zhong, Y. LoveDA: A Remote Sensing Land-Cover Dataset for Domain Adaptive Semantic Segmentation. In Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks, Virtual, 6–14 December 2021; Vanschoren, J., Yeung, S., Eds.; Volume 1.
9. Chen, B.; Feng, Q.; Niu, B.; Yan, F.; Gao, B.; Yang, J.; Gong, J.; Liu, J. Multi-modal fusion of satellite and street-view images for urban village classification based on a dual-branch deep neural network. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *109*, 102794. [CrossRef]
10. Sun, X.; Wang, P.; Yan, Z.; Xu, F.; Wang, R.; Diao, W.; Chen, J.; Li, J.; Feng, Y.; Xu, T.; et al. FAIR1M: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery. *ISPRS J. Photogramm. Remote. Sens.* **2022**, *184*, 116–130. [CrossRef]
11. Wang, Q.; Yuan, Z.; Du, Q.; Li, X. GETNET: A General End-to-End 2-D CNN Framework for Hyperspectral Image Change Detection. *IEEE Trans. Geosci. Remote. Sens.* **2019**, *57*, 3–13. [CrossRef]
12. Mohajerani, S.; Saeedi, P. Cloud-Net: An End-To-End Cloud Detection Algorithm for Landsat 8 Imagery. In Proceedings of the IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1029–1032.
13. Mohajerani, S.; Saeedi, P. Cloud and Cloud Shadow Segmentation for Remote Sensing Imagery Via Filtered Jaccard Loss Function and Parametric Augmentation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* **2021**, *14*, 4254–4266. [CrossRef]
14. Ji, S.; Dai, P.; Lu, M.; Zhang, Y. Simultaneous Cloud Detection and Removal From Bitemporal Remote Sensing Images Using Cascade Convolutional Neural Networks. *IEEE Trans. Geosci. Remote. Sens.* **2021**, *59*, 732–748. [CrossRef]
15. Gao, Y.; Guan, J.; Zhang, F.; Wang, X.; Long, Z. Attention-Unet-Based Near-Real-Time Precipitation Estimation from Fengyun-4A Satellite Imageries. *Remote. Sens.* **2022**, *14*, 2925. [CrossRef]
16. Ma, Y.; Wei, J.; Tang, W.; Tang, R. Explicit and stepwise models for spatiotemporal fusion of remote sensing images with deep neural networks. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *105*, 102611. [CrossRef]
17. Wei, J.; Yang, H.; Tang, W.; Li, Q. Spatiotemporal-Spectral Fusion for Gaofen-1 Satellite Images. *IEEE Geosci. Remote. Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
18. Liu, P.; Li, J.; Wang, L.; He, G. Remote Sensing Data Fusion With Generative Adversarial Networks: State-of-the-art methods and future research directions. *IEEE Geosci. Remote. Sens. Mag.* **2022**, *10*, 295–328. [CrossRef]
19. Zhu, S.; Ma, Z. Does AGRI of FY4A Have the Ability to Capture the Motions of Precipitation? *IEEE Geosci. Remote. Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
20. Zhang, B.; Ma, M.; Wang, M.; Hong, D.; Yu, L.; Wang, J.; Gong, P.; Huang, X. Enhanced resolution of FY4 remote sensing visible spectrum images utilizing super-resolution and transfer learning techniques. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* **2022**, *15*, 1–9. [CrossRef]
21. Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K.M. Enhanced Deep Residual Networks for Single Image Super-Resolution. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 1132–1140.
22. Guo, Y.; Chen, J.; Wang, J.; Chen, Q.; Cao, J.; Deng, Z.; Xu, Y.; Tan, M. Closed-Loop Matters: Dual Regression Networks for Single Image Super-Resolution. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 5406–5415.
23. Xia, B.; Hang, Y.; Tian, Y.; Yang, W.; Liao, Q.; Zhou, J. Efficient Non-local Contrastive Attention for Image Super-resolution. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 22 February–1 March 2022; Volume 36, pp. 2759–2767.
24. Jo, Y.; Wug Oh, S.; Vajda, P.; Joo Kim, S. Tackling the Ill-Posedness of Super-Resolution through Adaptive Target Generation. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 16231–16240.

25. Wang, L.; Wang, Y.; Lin, Z.; Yang, J.; An, W.; Guo, Y. Learning A Single Network for Scale-Arbitrary Super-Resolution. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 4781–4790.
26. Ranchin, T.; Wald, L. Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation. *Photogramm. Eng. Remote. Sens.* **2000**, *66*, 49–61.
27. Du, Q.; Younan, N.H.; King, R.; Shah, V.P. On the Performance Evaluation of Pan-Sharpening Techniques. *IEEE Geosci. Remote. Sens. Lett.* **2007**, *4*, 518–522. [CrossRef]
28. Alparone, L.; Baronti, S.; Garzelli, A.; Nencini, F. A Global Quality Measurement of Pan-Sharpened Multispectral Imagery. *IEEE Geosci. Remote. Sens. Lett.* **2004**, *1*, 313–317. [CrossRef]
29. Lyu, W.; Lu, W.; Ma, M. No-reference quality metric for contrast-distorted image based on gradient domain and HSV space. *J. Vis. Commun. Image Represent.* **2020**, *69*, 102797. [CrossRef]
30. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Trans. Image Process.* **2012**, *21*, 4695–4708. [CrossRef] [PubMed]
31. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a "Completely Blind" Image Quality Analyzer. *IEEE Signal Process. Lett.* **2013**, *20*, 209–212. [CrossRef]
32. Venkatanath, N.; Praneeth, D.; Bh, M.C.; Channappayya, S.S.; Medasani, S.S. Blind image quality evaluation using perception based features. In Proceedings of the 2015 Twenty First National Conference on Communications (NCC), Mumbai, India, 27 February–1 March 2015; pp. 1–6.
33. Xiang, X.; Tian, Y.; Zhang, Y.; Fu, Y.; Allebach, J.P.; Xu, C. Zooming Slow-Mo: Fast and Accurate One-Stage Space-Time Video Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 3370–3379.
34. Gu, J.; Sun, X.; Zhang, Y.; Fu, K.; Wang, L. Deep Residual Squeeze and Excitation Network for Remote Sensing Image Super-Resolution. *Remote. Sens.* **2019**, *11*, 1817. [CrossRef]