*Article*

# A Closed-Loop Network for Single Infrared Remote Sensing Image Super-Resolution in Real World

Haopeng Zhang [1,*] , Cong Zhang [2], Fengying Xie [1,3,4] and Zhiguo Jiang [1,3,4]

1    Department of Aerospace Information Engineering, School of Astronautics, Beihang University, Beijing 102206, China
2    AVIC DIGITAL, Beijing 100028, China
3    Beijing Key Laboratory of Digital Media, Beijing 102206, China
4    Key Laboratory of Spacecraft Design Optimization and Dynamic Simulation Technologies, Ministry of Education, Beijing 102206, China
*    Correspondence: zhanghaopeng@buaa.edu.cn; Tel.: +86-10-6171-6978

**Abstract:** Single image super-resolution (SISR) is to reconstruct a high-resolution (HR) image from a corresponding low-resolution (LR) input. It is an effective way to solve the problem that infrared remote sensing images are usually suffering low resolution due to hardware limitations. Most previous learning-based SISR methods just use synthetic HR-LR image pairs (obtained by bicubic kernels) to learn the mapping from LR images to HR images. However, the underlying degradation in the real world is often different from the synthetic method, i.e., the real LR images are obtained through a more complex degradation kernel, which leads to the adaptation problem and poor SR performance. To handle this problem, we propose a novel closed-loop framework that can not only make full use of the learning ability of the channel attention module but also introduce the information of real images as much as possible through a closed-loop structure. Our network includes two independent generative networks for down-sampling and super-resolution, respectively, and they are connected to each other to get more information from real images. We make a comprehensive analysis of the training data, resolution level and imaging spectrum to validate the performance of our network for infrared remote sensing image super-resolution. Experiments on real infrared remote sensing images show that our method achieves superior performance in various training strategies of supervised learning, weakly supervised learning and unsupervised learning. Especially, our peak signal-to-noise ratio (PSNR) is 0.9 dB better than the second-best unsupervised super-resolution model on PROBA-V dataset.

**Keywords:** super-resolution; closed-loop structure; channel attention; infrared remote sensing image; deep learning

## 1. Introduction

Infrared remote sensing has strong anti-interference ability and penetration ability for cloud cover compared with visible remote sensing. Even in a dark environment, it can obtain available images through infrared radiation radiated by objects. Thus, infrared remote sensing technology is widely used owing to these advantages. However, due to hardware limitations in the processing design of infrared detection elements, the spatial resolution of infrared remote sensing images is usually low, which has a great impact on the interpretation tasks that rely on high-quality images. Updating infrared remote sensing hardware is difficult, costly and time-consuming. Therefore, it becomes an ideal solution to improve the spatial resolution of infrared remote sensing images through super-resolution reconstruction algorithms [1]. Super-resolution (SR) is to recover the natural and sharp detailed high-resolution (HR) counterparts from low-resolution (LR) images. It can not only improve the visual quality of the image itself, but also help to improve the performance of image classification [2,3], target detection, target tracking [4,5] and

other tasks [6–9] that are limited by the spatial resolution. Hence, the super-resolution algorithm can effectively make up for the defects of infrared remote sensing imaging, and encourage it to play a greater role in application [10]. The SR algorithms can be divided into multiple images super-resolution (MISR) [11,12] and single image super-resolution (SISR) [13] according to the number of LR images used. MISR needs multiple images of the same scene, which is not practicable in the field of infrared remote sensing. SISR requires only a single image of the target scene to recover the corresponding HR image, which has wide applicability and practical value [14,15]. Therefore, in this paper, we focus on SISR for infrared remote sensing.

The research of SR algorithms starts from the methods based on the frequency domain [16] and gradually develops into the spatial domains [17,18]. Traditional SR algorithms in spatial domains can be divided into methods based on interpolation [17,18] and reconstruction [19]. Interpolation-based algorithms, such as nearest, bilinear, bicubic [18], etc., do not essentially generate missing high-frequency information, but simply weight the adjacent pixels. Therefore, the blur and aliasing effect in the reconstructed images are usually obvious. The reconstruction-based methods introduce prior information and establish a reasonable model to obtain the final HR images through iterative calculation. The representative methods are iterative back-projection (IBP) [20], maximum a posterior (MAP) [21], and projections onto convex sets (POCS) [22]. The reconstruction quality is better than interpolation-based methods, but these reconstruction-based methods have high computational complexity, slow convergence speed, and poor practicability. The methods based on sparse representation [23,24] get significant improvement compared with interpolation-based and reconstruction-based methods, but their modeling capability and expression ability are limited. Thus, recently, SR methods based on sparse representation gradually have been replaced by deep learning based methods [25–27].

Most deep-learning-based SR methods rely on paired training data to learn the mapping between LR images and HR images. Since the real HR-LR image pairs are difficult to obtain, they usually use HR images and their corresponding bicubic-degraded LR counterparts. However, the real-world data does not necessarily have the same distribution as the LR image obtained by a specific degradation method. Therefore, even if the SR method can obtain good results on artificially synthesized LR images, it is very challenging to learn a good model for real-world applications [28]. To alleviate this problem caused by degradation in the real world, we propose a more practical framework, i.e., closed-loop network. It contains not only a network for super-resolution, but also a network for downsampling. These two networks have independent functions, facilitating the use of excellent super-resolution network structure. Most importantly, connecting these two networks can make full use of the information of real LR images during the training process to learn the real causes of degradation, so as to improve the performance of the SR network on real LR data. We compare and analyze the network training modes with full supervision, weak supervision and non-supervision, respectively. Furthermore, according to the characteristics of infrared remote sensing images with fixed resolution and multi-band imaging, the resolution adaptability and band adaptability are explored and evaluated for the proposed method. To summarize, the main contributions of this paper can be highlighted as follows.

- We proposed a novel closed-loop framework to better super-resolve infrared remote sensing LR images in real world. Since our closed-loop network can work with LR input only, i.e., in unsupervised manner, our method has better practical utility.
- We provided reasonable training suggestions about how to train our network in the ways of supervised learning, weakly supervised learning and unsupervised learning.
- We experimentally studied the resolution level adaptability and spectrum adaptability of our proposed infrared remote sensing image SR method. Such analyses are especially needed for infrared remote sensing, but usually neglected in natural image super-resolution.

The remainder of this paper is organized as follows. Section 2 presents an overview of learning-based SISR methods and their limitations. Section 3 describes details of our closed-

loop framework. Sections 4 and 5 show experimental results to validate the proposed approach and explore the resolution and band adaptability of infrared remote image super-resolution. Conclusions are given in Section 6.

## 2. Related Works

Generally, learning-based SISR algorithms can be categorized into three different groups, i.e., supervised learning, unsupervised learning and weakly supervised learning. Thus, we summarize related previous works from these three aspects in this section.

### 2.1. Supervised Learning

Supervised learning methods aim at learning the relationship between the LR domain and HR domain with an external training set containing HR images and corresponding LR images. In the past few years, many supervised learning-based methods have succeeded in SR tasks, including sparse coding [23,24], neighborhood embedding [29] and mapping functions [30,31].

The supervised learning-based methods get better performance than reconstruction-based methods with the help of training sets. However, the training data is also the reason for the shortcomings of this kind of method. Because real HR-LR image pairs are extremely difficult to obtain. In general, the remote sensing satellites will not set two different resolutions for the same load. Thus, LR images are generally captured by down-sampling from HR images with a degradation kernel, which is inconsistent with the degradation of real LR data. As a result, although the SR model based on supervised learning can achieve good performance on the down-sampled data, it can not meet expectations when applied in real LR images [32].

### 2.2. Unsupervised Learning

Unsupervised learning methods only rely on LR images to perform super-resolution. In a single image, the repeated appearance of small image patches has been proved to be a very strong characteristic of natural images [33]. Natural images follow the rule of perspective projection, where the near scene is large and the distant scene is small. It has laid the foundation for many unsupervised image enhancement methods, including super-resolution [34]. The early unsupervised learning methods are mainly based on the above characteristic, also known as self-similarity. Huang et al. [35] proposed a compositional model which can simultaneously handle perspective distortion and affine transformation to expand the internal patch space. Shocher et al. [36] exploited the internal recurrence of information inside the test image, and then trained a light convolutional neural network (CNN) to reconstruct the corresponding HR image. These algorithms work well for images with multi-scale similar structures. However, due to the far imaging distance in satellite remote sensing, the influence of the depth of field on imaging is negligible. Remote sensing images are mostly like the results of parallel projection. Compared with natural images, remote sensing images have poor self-similarity, so such methods may not achieve good performance.

Besides, there are other unsupervised methods in the remote sensing community. Haut et al. [37] proposed a deep generative network that used random noise to generate HR images. In [37], HR images were updated by minimizing mean square error (MSE) loss of the down-sampled version of the output HR image and the real LR image. Sheikholeslami et al. [38] took the LR image directly as input and simplified the network structure to further improve efficiency. Both methods down-sampled the output image to solve the problem of lacking ground-truth HR image. However, the fixed down-sampling function used in these methods may also cause the problem of inconsistent degenerate kernels, similar to supervised learning methods.

### 2.3. Weakly Supervised Learning

CycleGAN (Cycle-consistent Generative Adversarial Network) [39] is a pioneering work of unpaired training mode for image-to-image translation. Inspired by it, Yuan et al. [40] firstly proposed CinCGAN (Cycle-in-Cycle Generative Adversarial Network) for image super-resolution. It contains two cycle GAN structures, and discriminate the domain of generated images with a discriminator instead of matching HR constraints. Bulat et al. [41] learned the HR-to-LR degradation model from unpaired images, and then used HR image and the LR image generated by the degradation model to train the SR model. Wang et al. [42,43] proposed Cycle-CNN, using two cycles in different directions to make full use of the information of real high and low-resolution data, making it more suitable for SR task of real remote sensing images. Guo et al. [44] proposed a dual regression network using paired synthetic data and additional unpaired LR data for training. The above methods are obviously not fully supervised, since no matched HR-LR pairs are used for supervised training. However, they are not strictly unsupervised either, because there are HR images in the target domain used in training. In this paper, we prefer to regard these methods using unpaired training data as weakly supervised learning SR methods. These weakly supervised learning methods aim to make the SR algorithm more practical on real LR images by using unpaired HR images at target resolution. They have used various methods to solve the impact of the artificial down-sampling kernel of supervised learning, but real LR and HR images are both needed, still facing data acquisition difficulty in remote sensing applications.

Although the above-mentioned learning-based methods have their own limitations, single-image super-resolution is indeed an ill-posed problem and hard to solve perfectly. Therefore, inspired by supervised learning and unpaired training mode, we try to propose a flexible framework that can generally adapt the training mode in the case of supervised learning, unsupervised learning and weakly supervised learning. Such an SISR model can make full use of the mapping constraints of supervised learning, and can also introduce the characteristics of real images to reduce the negative impact caused by the difference between the synthetic degradation kernel and the real degradation kernel, so as to get better SR results for real LR images.

### 3. Method

#### 3.1. Framework

Our framework contains two networks, one for super-resolution (denoted as $S$) and one for down-sampling (denoted as $D$). Because of the opposite function, connecting these two networks can form a closed-loop and perform self-supervision without matching ground truth. As shown in Figure 1, there are two possible basic types of connections. The **forward loop** refers to the way that the image passes through the super-resolution network $S$ first and then the down-sampling network $D$. Starting from the LR image $\mathcal{LR}$, the super-resolution network can obtain its HR version $\mathcal{HR}$ and the down-sampling network takes the HR output as input to get the low-resolution image $\mathcal{LR}'$ of the same size as the original input $\mathcal{LR}$ in the end of the loop. The input and final output can be constrained by loss functions. The **backward loop** refers to the opposite process. The HR image $\mathcal{HR}$ is used as the input of down-sampling network first, then the super-resolution network performs SR on the obtained LR image $\mathcal{LR}$. As a result, we get a high-resolution image $\mathcal{HR}'$ which is of the same size as the original input $\mathcal{HR}$. The **double loop** is the combination of forward loop and backward loop. It is worth noting that the LR and HR image in each loop is relative concepts. The super-resolution network $S$ learns how to reconstruct during training and helps us get final results during testing. The closed-loop structure adjusts $S$ more suitable to reconstruct LR images in real world by taking the real images as self-supervision input. Our experimental results in Section 4.1 also demonstrate the effectiveness of the closed-loop structure in learning the true causes of degradation.
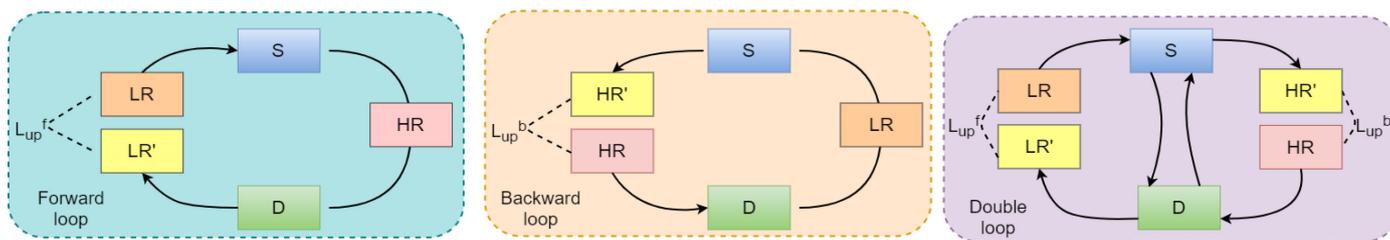
**Figure 1.** The cycle structure of our closed-loop framework. Our framework consists of two generative networks for super-resolution (*S*) and down-sampling (*D*), respectively. The left subfigure starting with the LR image is denoted as a forward loop, the center one starting with the HR image is denoted as a backward loop, and the right one is a double loop. $L_{up}^{f}$ and $L_{up}^{b}$ are loss functions denoted in Equations (1) and (2) separately.

We design the structure of the down-sampling network *D* according to Cycle-CNN [42]. As shown in Figure 2, it first downscales the feature maps using an average-pooling layer, and the resblock in *D* is much simpler, including only two convolution layers and a rectified linear unit (ReLU) [45] layer, which is determined by the difficulty of the task.
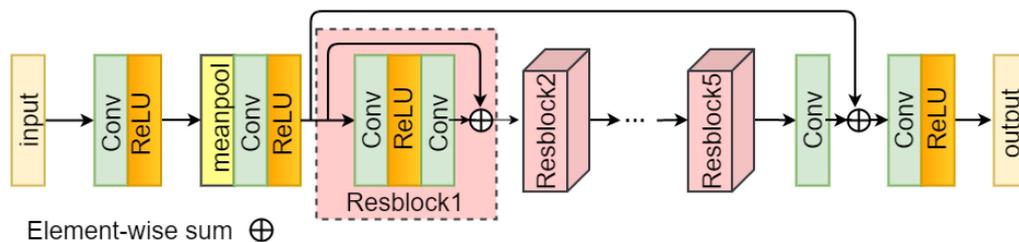


**Figure 2.** The down-sampling network of our closed-loop framework.

The structure of super-resolution network *S* is similar to RCAN (Residual Channel Attention Networks) [46], and its function is opposite to *D*. We adjust the number of modules in the network according to our SR task. The overview of super-resolution network *S* is shown in Figure 3. Specially, we use five residual modules as the main body, and each residual module contains five stacked blocks with a convolution layer at the end. Each block contains two convolution layers, a ReLU layer and channel attention (CA) module [47]. Multiple levels of skip connections are all over the network. Sub-pixel layer [48] is selected to up-sample feature maps to a larger size. Finally, the output image is obtained through a convolution layer. It should be noticed that the CA module used in this paper is to further tap the learning ability of convolutional neural networks. Its overall structure is shown in the lower right corner of Figure 3. The global pooling layer simplifies each channel of the previously extracted feature map to a single value, i.e., the average value of each channel. Then the number of channels is first reduced by a convolution layer and then restored to the original number by another convolution layer. The sigmoid activation function maps the value of each channel to the range of $[0, 1]$ as the weight of each channel feature map. The original feature map multiplies its weight to get the final feature map with channel attention. Through such a simple process, the channel attention module further enhances the nonlinear mapping ability of the convolutional neural network.
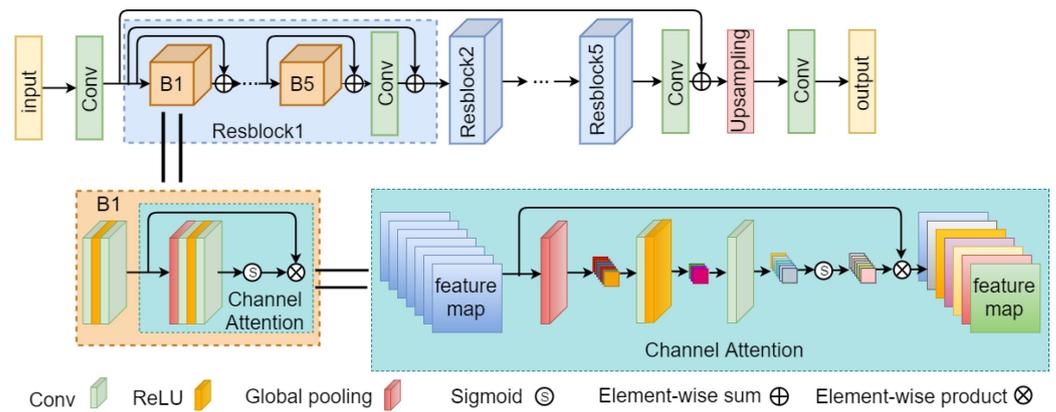
**Figure 3.** The super-resolution network of our closed-loop framework.

### 3.2. Loss Functions

A network is connected to the task through the loss function. As suggested by [47], we take mean square error (MSE) as the loss function. In our closed-loop framework in Figure 1, there are two ways to connect the SR network and down-sampling network, and we can choose one or both. Therefore, the self-supervised loss without matching ground truth has the following three forms, i.e.,

$$L_{up}^{f} = \frac{1}{N} \sum_{i=1}^{N} (\|D(S(\mathcal{LR})) - \mathcal{LR}\|_2) \tag{1}$$

$$L_{up}^{b} = \frac{1}{N} \sum_{i=1}^{N} (\|S(D(\mathcal{HR})) - \mathcal{HR}\|_2) \tag{2}$$

$$L_{up}^{d} = \lambda_1 L_{up}^{f} + \lambda_2 L_{up}^{b} \tag{3}$$

where $N$ is the batch-size, $up$ indicates the unpaired loss, and $f$, $b$ and $d$ represent forward loop, backward loop and double loop, respectively. $\lambda_1$ and $\lambda_2$ are the linear combination weights of $L_{up}^{f}$ and $L_{up}^{b}$ in $L_{up}^{d}$.

These loss functions can make the input and output of the closed-loop tend to be consistent, but there is not an effective way to ensure the performance of the SR network. Therefore, we need an additional loss function to constrain the SR network. In order to ensure that the HR images generated by the SR network maintain a high degree of consistency in content with the input LR images, HR-LR image pairs are still required. In general, LR images paired with HR images are obtained by down-sampling HR images using bicubic kernels, and we denote them as $\mathcal{LRBI}$. Then, the loss of paired data is as

$$L_p = \frac{1}{N} \sum_{i=1}^{N} (\|S(\mathcal{LRBI}) - \mathcal{HR}\|_2) \tag{4}$$

In this way, the paired loss $L_p$ makes the SR network more stable, and the unpaired losses $L_{up}^{f}$ and $L_{up}^{b}$ make use of real data and the closed-loop structure in order to make the network more adaptable to real data.

### 3.3. Datasets for Study

We use two datasets in our research, i.e., PROBA-V dataset and Landsat-8 dataset.

#### 3.3.1. PROBA-V Dataset

The full name of PROBA-V satellite [49] is Project for On-Board Autonomy-Vegetation, which is mainly used for vegetable monitoring. There are multiple payloads on the satellite that can be used to obtain images of different resolution levels. The data used in our paper

comes from the PROBA-V Super-Resolution competition [50] (https://kelvins.esa.int/proba-v-super-resolution/data/ accessed on 20 November 2019). We can download the near-infrared (NIR) dataset with 566 different scenes. Each data-point consists of one 100 m resolution image and several 300 m resolution images from the same scene. The 300 m resolution data is delivered as 128 × 128 gray-scale images, while the 100 m resolution data is delivered as 384 × 384 gray-scale images. Different from the original task of multi-image super-resolution using PROBA-V data, we reorganize the dataset for our single image super-resolution task by selecting the one with the best quality from the 300 m data as the LR image in each scene according to the attached quality map. The total 566 scenes in the PROBA-V dataset are split into 500 scenes for training, 6 scenes for validating and 60 scenes for testing. The specific information of the PROBA-V dataset is shown in Table 1. It is worth noting that the HR and LR images in PROBA-V dataset are real and matched.

**Table 1.** The information of PROBA-V dataset used in this paper.

| Dataset | Spectral Range (μm) | Training Set | Validation Set | Testing Set | Image Size | Resolution (m) |
|---------|---------------------|--------------|----------------|-------------|------------|----------------|
| PROBA-V | 0.772–0.902 | 500 | 6 | 60 | HR:384 × 384 | HR:100 |
|         |             |     |   |    | LR:128 × 128 | LR:300 |

### 3.3.2. Landsat-8 Dataset

The Landsat-8 satellite carries two sensors, namely the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS). Landsat-8 can provide multiple infrared spectrum images, which is convenient for us to study the influence of different imaging bands on the performance of super-resolution. In addition, since the resolution of the Landsat-8 infrared image is 30m, which is different from the resolution of PROBA-V, the impact of the resolution of training data can also be explored. To solve the problem of resolution adaptability, we collect 566 near infrared images from Landsat-8 website [51] (https://earthexplorer.usgs.gov/ accessed on 20 November 2019), and divide the dataset according to the same ratio as PROBA-V. As for the issue of spectrum adaptability, we choose, in total, five bands to construct the dataset separately, including the near-infrared band (B5), short-wave infrared band 1 (B6), short-wave infrared band 2 (B7), thermal infrared band 1 (B10) and thermal infrared band 2 (B11). For each band, we collect eight remote sensing scenes, crop them into 2005 images in the size of 384 × 384, and split 2000 images for training and 5 for validating. Considering the practicability of remote sensing image super-resolution, we directly choose two complete infrared remote sensing scenes of each band as a testing set instead of cropped data. Detailed information on the Landsat-8 dataset used in our paper is shown in Table 2.

**Table 2.** The information of Landsat-8 dataset used in this paper.

| Dataset | Spectral Range (μm) | Training Set | Validation Set | Testing Set | Image Size | Resolution (m) |
|---------|---------------------|--------------|----------------|-------------|------------|----------------|
| Landsat-8-NIR | 0.845–0.885 | 500 | 6 | 60 | HR:384 × 384 | HR:30 |
|         |             |     |   |    | LR:128 × 128 | LR:90 |
| B5 | 0.845–0.885 | 2000 | 5 | 2 scenes | HR:384 × 384 | HR:30 |
| B6 | 1.560–1.660 | 2000 | 5 | 2 scenes | HR:384 × 384 | HR:30 |
| B7 | 2.100–2.300 | 2000 | 5 | 2 scenes | HR:384 × 384 | HR:30 |
| B10 | 10.60–11.19 | 2000 | 5 | 2 scenes | HR:384 × 384 | HR:30 |
| B11 | 11.50–12.51 | 2000 | 5 | 2 scenes | HR:384 × 384 | HR:30 |

*3.4. Training Mode*

Corresponding to the three different closed-loops in Figure 1 and the three unpaired losses in Section 3.2, there could be three training modes. The proposed training process of these three modes is summarized in Algorithm 1. For clear description in a unified form, we use $\mathbf{x}_p$ and $\mathbf{y}_p$ to represent the paired LR data and HR data, respectively. $\mathbf{x}_{up}$ is the input of the forward loop and $\mathbf{y}_{up}$ is the input of the backward loop. Our goal is to get a well-trained super-resolution network $S$ and a down-sampling network $D$. $\theta_S$ and $\theta_D$ represent the parameters of these two networks, respectively. Different training modes are selected by setting the weight parameters of each loss function. In practice, we can select the training mode first to reduce unnecessary operations.

---

**Algorithm 1** Pseudo-codes of our framework in three training modes.

---

**Require:** Infrared remote sensing training images $(\mathbf{x}_{pi}, \mathbf{y}_{pi})$, $\mathbf{x}_{upi}$, $\mathbf{y}_{upi}$, $i = 1, ..., N$,
**Goal** The well-trained $S$ and $D$
 1: Initialize $\theta_S$, $\theta_D$
 2: **repeat**
 3: 　　　$\mathbf{x}_{pi} \to S \to \mathbf{y}'_{pi}$
 4: 　　　$L_p = MSE(\mathbf{y}'_{pi}, \mathbf{y}_{pi})$
 5: 　　　$\mathbf{x}_{upi} \to S \to D \to \mathbf{x}'_{upi}$
 6: 　　　$L^f_{up} = MSE(\mathbf{x}'_{upi}, \mathbf{x}_{upi})$
 7: 　　　$\mathbf{y}_{upi} \to D \to S \to \mathbf{y}'_{upi}$
 8: 　　　$L^b_{up} = MSE(\mathbf{y}'_{upi}, \mathbf{y}_{upi})$
 9: 　　　$L^d_{up} = \lambda_1 L^f_{up} + \lambda_2 L^b_{up}$
10: 　　　$L_{total} = \omega_1 L_p + \omega_2 L^d_{up}$
11: 　　　Choose training mode:
12: 　　　$--$forward loop: $\lambda_1 = 1, \lambda_2 = 0, \omega_1 = 1, \omega_2 = 0.1$
13: 　　　$--$backward loop: $\lambda_1 = 0, \lambda_2 = 1, \omega_1 = 1, \omega_2 = 0.1$
14: 　　　$--$double loop: $\lambda_1 = 1, \lambda_2 = 1, \omega_1 = 1, \omega_2 = 2$
15: 　　　ADAM-optimizer($L_{total}$,variable$= (\theta_S, \theta_D)$)
16: **until** Reach maximum iteration of minibatch updating
17:
18: **function** TEST($\mathcal{LR}$)
19: 　　　$\mathcal{LR} \to S \to \mathcal{HR}$
20: 　　　**return** $HR$
21: **end function**

---

Supervised, weakly supervised and unsupervised data situations could happen in all three different training modes in Algorithm 1. Table 3 summarizes these possible combinations. The supervised data includes paired HR-LR images donated as $\mathcal{HR}_p$ and $\mathcal{LR}_p$. Weakly supervised data includes unpaired HR and LR images donated as $\mathcal{HR}_{up}$ and $\mathcal{LR}_{up}$, where $\mathcal{HR}_{up}$ are matched with their down-sampling version $\mathcal{LRBI}$. The unsupervised case only contains LR images donated as $\mathcal{LR}$, which are matched with down-sampled lower-resolution images $\mathcal{LRBI}_{\mathcal{LR}}$.

To detailed show how to select the training mode in various data situations, we give an example experimental setting on real LR images of the PROBA-V dataset. As shown in Table 4, the training data corresponding to supervised learning, weakly supervised learning, and unsupervised learning are real paired HR-LR images, unpaired HR-LR images, and LR images, respectively. Training details are described as follows.

- **Supervised learning:** The real HR data and LR data in the PROBA-V dataset are represented by $\mathcal{HR}100$ and $\mathcal{LR}300$, respectively. In order to prove the superiority of our framework, we also separately train the super-resolution network $S$ for comparison

in addition to three different training modes. When the super-resolution network is trained separately, the paired loss $L_p$ in Algorithm 1 is used as the total loss.

- **Weakly supervised learning:** The real unpaired HR-LR data can be obtained by shuffling the order of PROBA-V data. In order to distinguish from the data for supervised learning, the unpaired HR and LR data are represented by $\mathcal{HR}100_{up}$ and $\mathcal{LR}100_{up}$, respectively. For the purpose of calculating paired loss, we need to down-sample the 100m HR data to obtain the matching LR data $\mathcal{LRBI}300$.
- **Unsupervised learning:** In most cases, the super-resolution task is carried out with only LR images. We use the real LR data from the PROBA-V dataset and denote it as $\mathcal{LR}300$ and down-sample it to get paired data $\mathcal{LRBI}900$.

**Table 3.** The training data of different training modes at different data situation. (— means not used.)

| Data Situation | Mode | $x_p$ | $y_p$ | $x_{up}$ | $y_{up}$ |
|---|---|---|---|---|---|
| Supervised | Forward loop | $\mathcal{LR}_p$ | $\mathcal{HR}_p$ | $\mathcal{LR}_p$ | — |
| | Backward loop | $\mathcal{LR}_p$ | $\mathcal{HR}_p$ | — | $\mathcal{LR}_p$ |
| | Double loop | $\mathcal{LR}_p$ | $\mathcal{HR}_p$ | $\mathcal{LR}_p$ | $\mathcal{HR}_p$ |
| | S | $\mathcal{LR}_p$ | $\mathcal{HR}_p$ | — | — |
| Weak supervised | Forward loop | $\mathcal{LRBI}$ | $\mathcal{HR}_{up}$ | $\mathcal{LR}_{up}$ | — |
| | Backward loop | $\mathcal{LRBI}$ | $\mathcal{HR}_{up}$ | — | $\mathcal{HR}_{up}$ |
| | Double loop | $\mathcal{LRBI}$ | $\mathcal{HR}_{up}$ | $\mathcal{LR}_{up}$ | $\mathcal{HR}_{up}$ |
| | S | $\mathcal{LRBI}$ | $\mathcal{HR}_{up}$ | — | — |
| Unsupervised | Forward loop | $\mathcal{LRBI}_{\mathcal{LR}}$ | $\mathcal{LR}$ | $\mathcal{LR}$ | — |
| | Backward loop | $\mathcal{LRBI}_{\mathcal{LR}}$ | $\mathcal{LR}$ | — | $\mathcal{LR}$ |
| | Double loop | $\mathcal{LRBI}_{\mathcal{LR}}$ | $\mathcal{LR}$ | $\mathcal{LRBI}_{\mathcal{LR}}$ | $\mathcal{LR}$ |
| | S | $\mathcal{LRBI}_{\mathcal{LR}}$ | $\mathcal{LR}$ | — | — |

**Table 4.** Correspondence table of training data in different training modes on the PROBA-V dataset. (— means not used.)

| Data Situation | Mode | $x_p$ | $y_p$ | $x_{up}$ | $y_{up}$ |
|---|---|---|---|---|---|
| Real paired HR-LR data | Forward loop | $\mathcal{LR}300$ | $\mathcal{HR}100$ | $\mathcal{LR}300$ | — |
| | Backward loop | $\mathcal{LR}300$ | $\mathcal{HR}100$ | — | $\mathcal{HR}100$ |
| | Double loop | $\mathcal{LR}300$ | $\mathcal{HR}100$ | $\mathcal{LR}300$ | $\mathcal{HR}100$ |
| | S | $\mathcal{LR}300$ | $\mathcal{HR}100$ | — | — |
| Real unpaired HR-LR data | Forward loop | $\mathcal{LRBI}300$ | $\mathcal{HR}100_{up}$ | $\mathcal{LR}300_{up}$ | — |
| | Backward loop | $\mathcal{LRBI}300$ | $\mathcal{HR}100_{up}$ | — | $\mathcal{HR}100_{up}$ |
| | Double loop | $\mathcal{LRBI}300$ | $\mathcal{HR}100_{up}$ | $\mathcal{LR}300_{up}$ | $\mathcal{HR}100_{up}$ |
| | S | $\mathcal{LRBI}300$ | $\mathcal{HR}100_{up}$ | — | — |
| Real LR data | Forward loop | $\mathcal{LRBI}900$ | $\mathcal{LR}300$ | $\mathcal{LR}300$ | — |
| | Backward loop | $\mathcal{LRBI}900$ | $\mathcal{LR}300$ | — | $\mathcal{LR}300$ |
| | Double loop | $\mathcal{LRBI}900$ | $\mathcal{LR}300$ | $\mathcal{LRBI}900$ | $\mathcal{LR}300$ |
| | S | $\mathcal{LRBI}900$ | $\mathcal{LR}300$ | — | — |

## 4. Experimental Results

### 4.1. Training Mode Selection

We first explore the training mode under different data situations on real LR images of the PROBA-V dataset as shown in Table 4. We use peak signal-to-noise ratio (PSNR) and

structural similarity index (SSIM) as evaluation indicators like many other related studies. We train 12 SR models and test them using real LR images in the testing set of the PROBA-V dataset. Experimental results are shown in Table 5.

**Table 5.** Results of different training modes under different data situations on the PROBA-V dataset. The best results are marked in red.

| Data Situation | S<br>PSNR(dB)/SSIM | Forward Loop<br>PSNR(dB)/SSIM | Backward Loop<br>PSNR(dB)/SSIM | Double Loop<br>PSNR(dB)/SSIM |
|---|---|---|---|---|
| Real paired HR-LR dataset | 40.0677/0.9671 | 40.3701/0.9675 | 40.3701/0.9675 | 40.0874/0.9647 |
| Real unpaired HR-LR dataset | 39.1032/0.9635 | 39.4057/0.9621 | 39.9832/0.9645 | 39.4217/0.9631 |
| Real LR dataset | 39.1266/0.9635 | 40.1415/0.9645 | 39.6151/0.9645 | 39.9268/0.9637 |

First of all, regardless of the data situation, simply training a super-resolution network $S$ cannot achieve the best performance. It is due to the fact that $S$ only uses the information of paired data. When the dataset does not contain real HR-LR image pairs, it introduces too much influence of the down-sampling kernel of the synthetic data on SR model.

When using real paired HR and LR data to train the SR model, the forward loop and backward loop simultaneously achieve the best results, and the double loop does not seem to be so effective. When the dataset contains real unpaired HR-LR data, the overall performance is reduced, but the advantage of the backward loop is comparatively obvious among the three training modes. It indicates that when the data is not matched, the data of the target resolution has a greater effect. The double loop mode does not superimpose the advantages of a forward loop and backward loop, and increases the computation complexity of the model. This indicates that the double loop mode may be too complicated for the SR task with the data used in this paper.

It should be noticed that when we perform unsupervised learning, the input of the forward loop is $\mathcal{LR}$, not the $\mathcal{LRBI}900$ with relatively lower resolution, because the original intention of the closed-loop structure is to introduce the information of real data. The input of the backward loop is also $\mathcal{LR}300$, because there is no higher-resolution data in our dataset. The forward loop and the backward loop achieve the same SSIM performance on the testing set, while the forward loop achieves better PSNR performance. The reason may be the training process of the forward loop is more consistent with the testing process.

### 4.2. Comparison with Learning Based SR Methods

We compare with some state-of-the-art SR methods on the PROBA-V dataset, including supervised learning, weakly supervised learning and unsupervised learning. The results are shown in Table 6. In the case of supervised learning, our closed-loop framework performs PSNR of 40.3701dB, while the SR network $S$ only obtains 40.0677dB, which has approved the effectiveness of our closed-loop structure. Since the PROBA-V dataset contains real paired HR-LR data, EDSR [52] and SRFBN [53] also achieve good results in supervised learning experiments. As for the cases of weakly supervised learning and unsupervised learning, our method achieves the best PSNR and the second-best SSIM. The visual comparison of each unsupervised method is shown in Figure 4. It can be shown in Table 6 that UGSR [37] and EUSR [38] do not always get satisfactory results for infrared remote sensing images with relatively low pixel values and low contrast. We separately show some results of UGSR and EUSR in Figure 5. UGSR reconstructs the HR image from random noise while EUSR does this from LR image directly. EUSR performs relatively better and sometimes could get good results such as subfigure (h) of Figure 5.

**Table 6.** Comparison between our method and other advanced SR algorithms on PROBA-V dataset. The best results are marked in red.

| SR Type | Method | PSNR(dB) | SSIM |
|---------|--------|----------|------|
| Supervised learning | S | 40.0677 | 0.9671 |
| | Ours | 40.3701 | 0.9675 |
| | RDN [54] | 33.5486 | 0.9187 |
| | EDSR [52] | 40.5574 | 0.9675 |
| | SRFBN [53] | 40.6953 | 0.9679 |
| Weakly supervised learning | Ours | 39.9832 | 0.9645 |
| | Cycle-CNN [42] | 39.9691 | 0.9646 |
| | S | 39.1031 | 0.9635 |
| | RDN [54] | 33.8350 | 0.9527 |
| | EDSR [52] | 39.1036 | 0.9636 |
| | SRFBN [53] | 39.0796 | 0.9635 |
| Unsupervised learning | Ours | 40.1415 | 0.9645 |
| | IBP [20] | 39.1382 | 0.9643 |
| | BDB [33] | 37.3865 | 0.9348 |
| | FSR [55] | 39.2305 | 0.9674 |
| | GPR [56] | 37.5735 | 0.9631 |
| | UGSR [37] | 35.4371 | 0.9255 |
| | EUSR [38] | 23.6012 | 0.4987 |
| | ZSSR [36] | 38.1762 | 0.9597 |
| | MZSR [57] | 38.0026 | 0.8942 |
| | dSRVAE [58] | 38.3937 | 0.9486 |
| | DASR [59] | 39.7955 | 0.9637 |

### 4.3. Ablation Study of Attention Mechanism

To prove the effectiveness of our proposed closed-loop structure, we choose one SR model with channel attention (CA) module [47] as our S network and compare itself with closed-loop structure as shown in the supervised learning part of Table 6. In order to analyze the effect of different attention modules, we compare them with another attention module named convolutional block attention module (CBAM) [60]. Results in Table 7 show that CBAM has better performance than CA, and achieves the best SR performance compared with other supervised SR methods. It can be seen that a more powerful attention module can improve the SR performance, indicating that our closed-loop network has the potential to get better SR performance by updating the basic modules.

**Table 7.** Comparison of different attention modules in supervised manner on PROBA-V dataset. The best results are marked in red.

| Method | PSNR(dB) | SSIM |
|--------|----------|------|
| Ours (CA [47]) | 40.3701 | 0.9675 |
| Ours (CBAM [60]) | 41.1102 | 0.9689 |
| RDN [54] | 33.5486 | 0.9187 |
| EDSR [52] | 40.5574 | 0.9675 |
| SRFBN [53] | 40.6953 | 0.9679 |

(**a**) HR-histequal     (**b**) Ours     (**c**) BDB     (**d**) IBP

PSNR (dB)/SSIM     43.7637/0.9807     40.1598/0.9534     43.0150/0.9806

(**e**) HR     (**f**) FSR     (**g**) GPR     (**h**) ZSSR

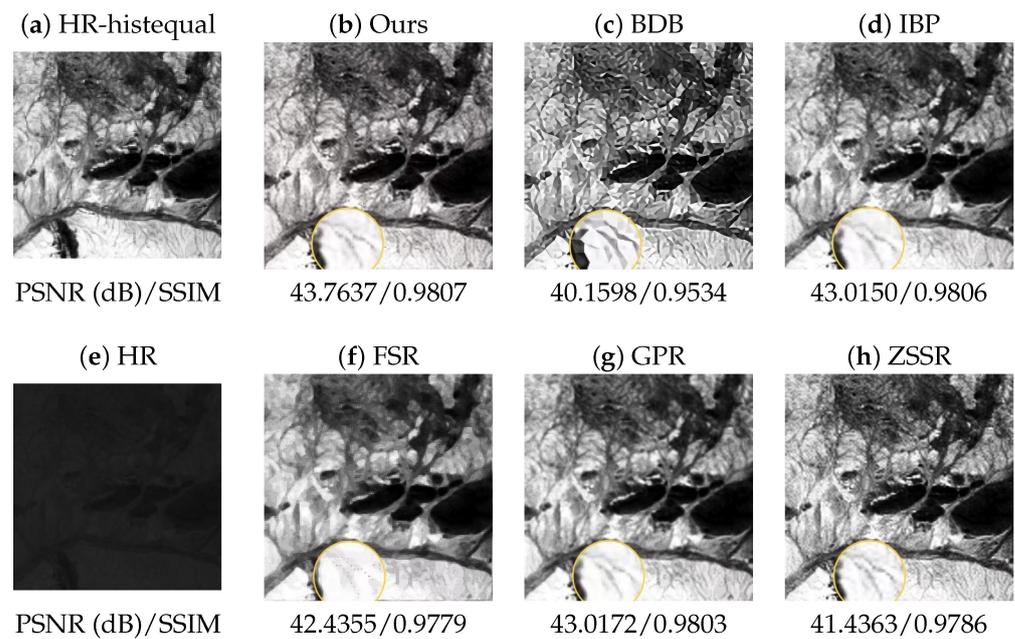PSNR (dB)/SSIM     42.4355/0.9779     43.0172/0.9803     41.4363/0.9786

**Figure 4.** Visual results of our method and other unsupervised methods (BDB [33], IBP [20], FSR [55], GPR [56], and ZSSR [36]) on the PROBA-V dataset. Since the pixel value of the infrared HR image is relatively low as shown in (**e**), histogram equalization (histequal) has been performed on other images in this figure for better visualization. Noting that infrared lights are not visible to human eyes, we show a single channel of infrared remote sensing image in gray-scale. The yellow circle in each image is the result of a local enlargement. In addition, we ignore the geographical locations in the figure since SR focuses on image quality improvement. Such visualization setting is also applied in other figures of this paper.

(**a**) HR     (**b**) HR-histequal     (**c**) UGSR     (**d**) EUSR

PSNR(dB)/SSIM     PSNR(dB)/SSIM     26.3796/0.7741     33.1514/0.9163

(**e**) HR     (**f**) HR-histequal     (**g**) UGSR     (**h**) EUSR

PSNR(dB)/SSIM     PSNR(dB)/SSIM     19.4671/0.0833     35.2808/0.9830
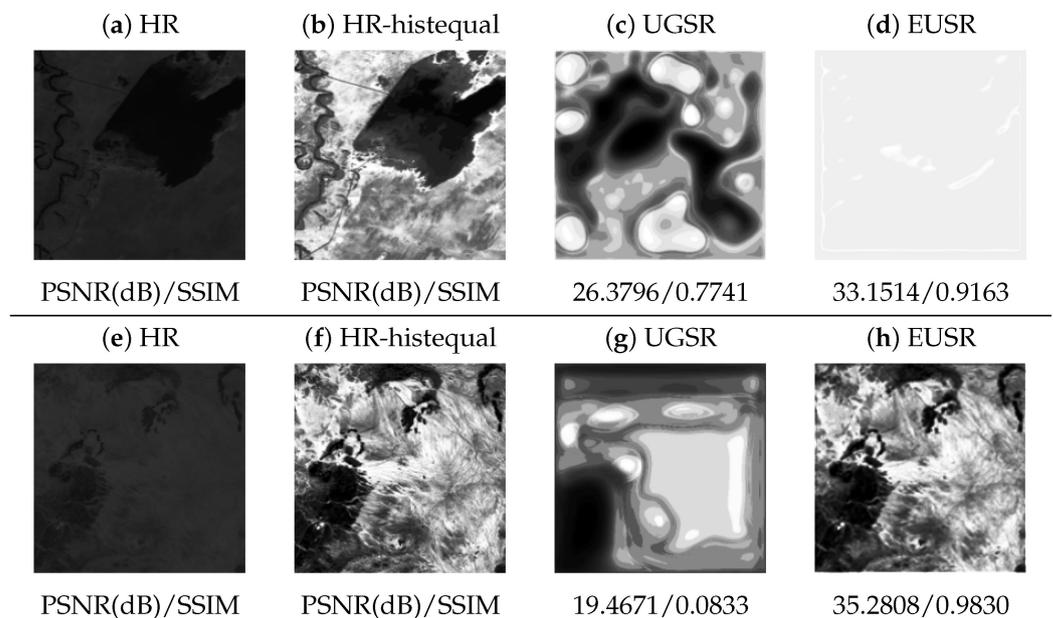
**Figure 5.** Visual results of UGSR [37] and EUSR [38] on the PROBA-V dataset. Since the pixel value of the infrared HR images is relatively low as shown in (**a**) and (**e**), histogram equalization (histequal) has been performed on other images in this figure for better visualization.

### 4.4. Resolution Adaptability Analysis

Since remote sensing images have clear resolution levels, it is valuable to study the SR performance of testing data with different resolution levels from the training data. Therefore, we conduct a cross-comparison experiment in this subsection by training different SR models of $\times 3$ scale and testing LR images with different spatial resolutions. Since Landsat-8-NIR does not contain real 90 m resolution data, the LR data used in supervised training is obtained by down-sampling from PROBA-V and Landsat-8 datasets (i.e., $\mathcal{HR}100$ and $\mathcal{HR}30$), denoted as $\mathcal{LRBI}300$ and $\mathcal{LRBI}90$. We adapt the backward loop mode and train two models using PROBA-V and Landsat-8-NIR data, respectively, in a supervised way. As for the unsupervised mode with only LR images, we use the forward loop mode for training. The lower-resolution data paired with $\mathcal{LRBI}300$ and $\mathcal{LRBI}90$ are further down-sampled and represented by $\mathcal{LRBI}900$ and $\mathcal{LRBI}270$, respectively. In order to keep the experimental condition as consistent as possible, we also use down-sampled LR data during the testing process of all trained models, namely $\mathcal{LRBI}300_{test}$ and $\mathcal{LRBI}90_{test}$, respectively. The experimental results are listed in Table 8. It can be seen that a $\times 3$ SR model trained for super-resolving the data from 90m to 30m does not perform well for the task of super-resolving the data from 300 m to 100 m, and vice versa. The situation is more obvious in unsupervised learning. However, the performance gap on $\mathcal{LRBI}300_{test}$ is relatively smaller than $\mathcal{LRBI}90_{test}$. This may indicate that data with higher resolution is more sensitive to the resolution of training data.

**Table 8.** Results of resolution adaptability (PSNR(dB)/SSIM). The best results under each data situation are marked in red.

| Training Data \ Testing Data | $\mathcal{LRBI}300_{test}$ | $\mathcal{LRBI}90_{test}$ |
|---|---|---|
| $\mathcal{HR}100/\mathcal{LRBI}300$ | 47.1727/0.9904 | 39.5389/0.9769 |
| $\mathcal{HR}30/\mathcal{LRBI}90$ | 47.1168/0.9904 | 39.9341/0.9793 |
| $\mathcal{LRBI}300/\mathcal{LRBI}900$ | 47.1564/0.9904 | 39.5677/0.9771 |
| $\mathcal{LRBI}90/\mathcal{LRBI}270$ | 47.0138/0.9902 | 39.8937/0.9791 |

### 4.5. Spectrum Adaptability Analysis

The spectrum is also an important factor affecting SR performance. In this subsection, we explore whether the SR model trained with a certain band is applicable to other infrared bands. The Landsat-8 dataset is used since it has multiple infrared spectrum bands with the same spatial resolution. The 30m data of Landsat-8 is regarded as HR data. LR data for training is down-sampled from HR data. As for supervised training, we train SR models with $\times 2$, $\times 3$ and $\times 4$ scales for bands B5, B6, B7, B10 and B11, respectively. Besides, we only train the unsupervised model on scale $\times 2$, to avoid large-scale down-sampling during the training process. Experimental results are shown in Tables 9 and 10.

We can see that although they are all infrared remote sensing data with the same resolution, the difference in the spectrum of the training data will also have a great impact on the performance, which is more obvious in the PSNR index. In most cases, the SR model achieves the best results when testing images have the same spectrum as the training data. Sometimes, an SR model trained by a similar spectrum performs better than the model trained by its corresponding spectrum, and the gap between them is usually small. Such gap can be clearly shown in Figure 6, where we use the results obtained by the model trained by the same spectrum as the performance benchmark for each band and show the difference of each model with its benchmark. It can be seen that for bands B5, B6, and B7, the models trained by B5, B6, and B7 usually perform well. The models trained by B10 and B11 always achieve better SRs of B10 and B11. B5 maintains the rule that the benchmark result is the best, and the training data cannot be replaced by other spectra. Figure 7 shows a partial image of B5 testing set. It can be seen that the model trained by

B10 and B11 has a poor recovery effect on the small lines. Compared with other bands, B10 and B11 have less volatility and better band adaptability. However, when the scale factor is 4, the band adaptability is extremely reduced, and the best performance can be obtained only when the data matches.

**Table 9.** Results of spectrum adaptability for supervised learning (PSNR(dB)/SSIM). The best results of every spectrum are marked in red.

| Scale | Test / Train | B5 | B6 | B7 | B10 | B11 |
|---|---|---|---|---|---|---|
| ×2 | B5 | 42.5729/0.9759 | 48.5109/0.9900 | 50.0885/0.9929 | 58.3780/0.9987 | 58.7213/0.9988 |
| | B6 | 42.2653/0.9744 | 48.5291/0.9900 | 50.1445/0.9929 | 58.3357/0.9987 | 58.7127/0.9988 |
| | B7 | 42.2117/0.9743 | 48.5502/0.9900 | 50.2106/0.9929 | 58.6805/0.9988 | 59.0529/0.9989 |
| | B10 | 38.7340/0.9443 | 45.3003/0.9804 | 47.0148/0.9864 | 59.6865/0.9990 | 59.9845/0.9991 |
| | B11 | 38.8092/0.9454 | 45.4116/0.9809 | 47.1310/0.9867 | 59.8778/0.9990 | 60.1900/0.9991 |
| ×3 | B5 | 38.5995/0.9430 | 44.9772/0.9790 | 46.6235/0.9852 | 55.7259/0.9983 | 56.0587/0.9984 |
| | B6 | 38.3891/0.9403 | 45.0956/0.9794 | 46.7962/0.9856 | 56.1163/0.9985 | 56.4618/0.9986 |
| | B7 | 38.3776/0.9402 | 45.0734/0.9793 | 46.7922/0.9856 | 56.2103/0.9986 | 56.5410/0.9987 |
| | B10 | 36.2983/0.9069 | 42.9761/0.9677 | 44.6739/0.9772 | 56.1572/0.9986 | 56.4653/0.9986 |
| | B11 | 36.2152/0.9058 | 42.8671/0.9670 | 44.5453/0.9766 | 55.5850/0.9984 | 55.9038/0.9985 |
| ×4 | B5 | 36.4659/0.9108 | 42.9500/0.9680 | 44.5885/0.9773 | 52.8712/0.9978 | 53.1948/0.9979 |
| | B6 | 36.3036/0.9075 | 42.9999/0.9682 | 44.6664/0.9775 | 53.6337/0.9978 | 53.9671/0.9980 |
| | B7 | 36.2905/0.9072 | 43.0121/0.9682 | 44.6960/0.9776 | 53.6969/0.9979 | 54.0170/0.9980 |
| | B10 | 35.0405/0.8796 | 41.6516/0.9575 | 43.3137/0.9696 | 54.8017/0.9981 | 54.0919/0.9982 |
| | B11 | 35.0061/0.8789 | 41.6026/0.9571 | 43.2639/0.9694 | 54.4611/0.9982 | 54.7579/0.9983 |

**Table 10.** Results of spectrum adaptability for unsupervised learning (PSNR(dB)/SSIM). The best results of every spectrum are marked in red.

| Scale | Test / Train | B5 | B6 | B7 | B10 | B11 |
|---|---|---|---|---|---|---|
| ×2 | B5 | 42.4992/0.9760 | 48.3702/0.9899 | 49.9142/0.9927 | 55.0392/0.9985 | 55.3785/0.9986 |
| | B6 | 42.1859/0.9744 | 48.4360/0.9900 | 50.0237/0.9929 | 55.0141/0.9985 | 55.3801/0.9986 |
| | B7 | 42.1216/0.9741 | 48.4302/0.9900 | 50.0325/0.9929 | 54.9818/0.9985 | 55.3517/0.9986 |
| | B10 | 40.1325/0.9594 | 46.6713/0.9855 | 48.3643/0.9899 | 54.7714/0.9986 | 55.0933/0.9987 |
| | B11 | 40.1150/0.9593 | 46.6681/0.9854 | 48.3653/0.9863 | 55.1840/0.9988 | 55.5098/0.9989 |
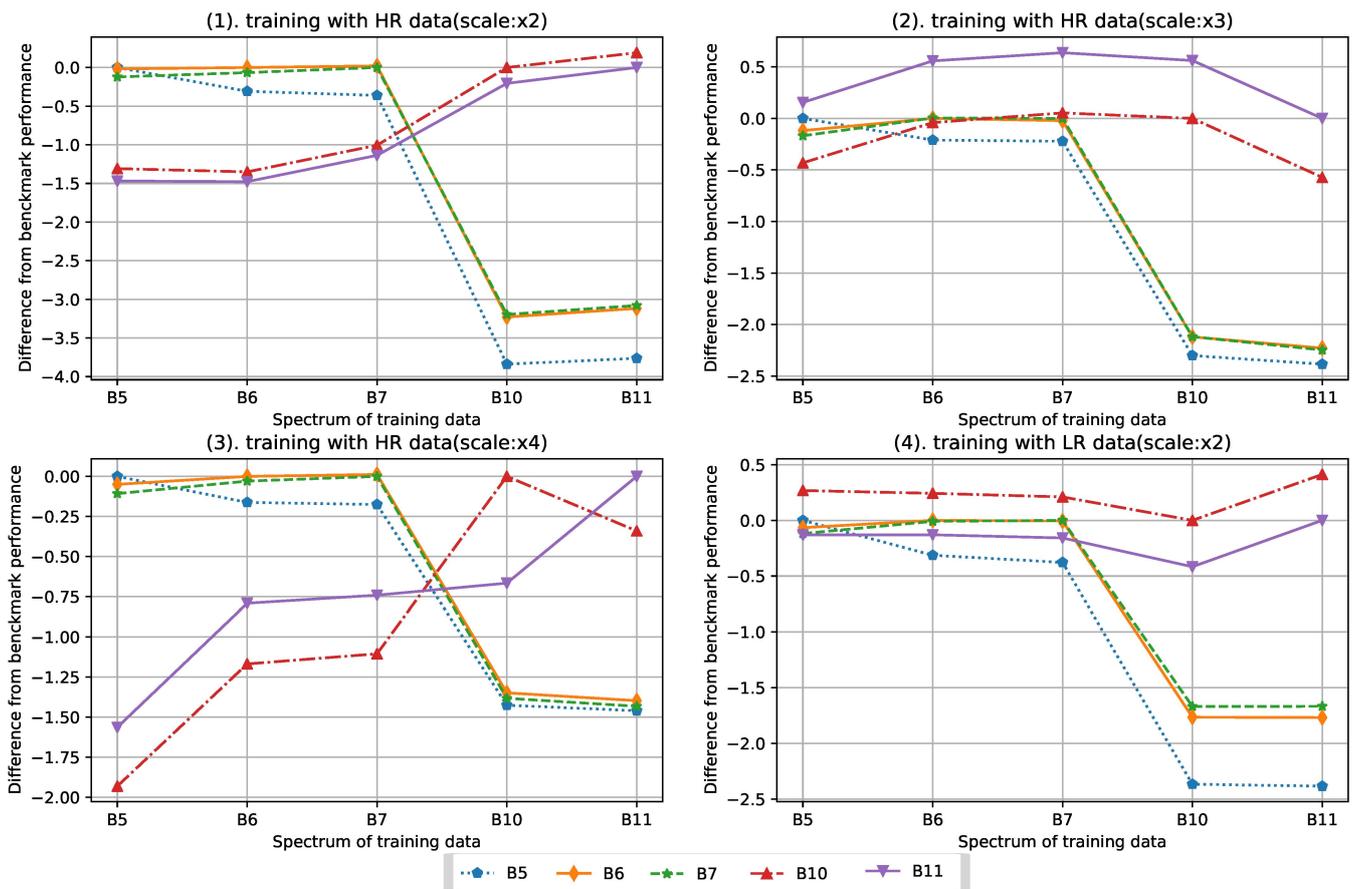
**Figure 6.** PSNR performance difference between each band and their benchmark result. The benchmark of each band refers to the result obtained by the model trained with the same band.
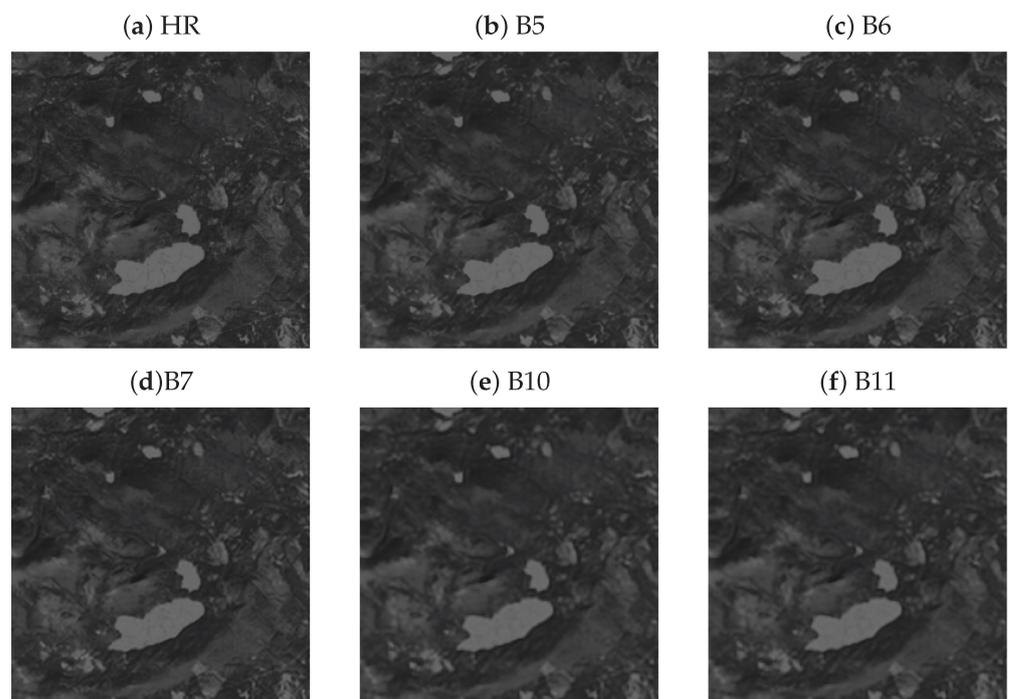


**Figure 7.** Visual results of B5 testing set. The titles of subfigures (**b**–**f**) represent the training set of SR models.

## 5. Discussions

### 5.1. Training Mode Selection

　　Our framework can derive a variety of training modes, and can be trained in supervised, weakly supervised or unsupervised ways. Therefore, there should be an optimal choice corresponding to different data situations. Based on the experimental analysis in Section 4.1, we can draw the following conclusion. First, when using real paired HR-LR images to train the SR model, a single loop is sufficient. Second, when there are HR images in the training set, the backward loop can achieve better SR performance on the real LR data. We even do not need to consider whether we have the unpaired real LR images, because they are not needed in the backward mode. The last and most important point is that when there is only LR training data, i.e., the SR model is trained in an unsupervised manner, the forward loop mode is the best choice.

### 5.2. Advantages and Limitations

　　Our proposed closed-loop framework is designed to super-resolve infrared remote-sensing LR images in the real world. It can be trained in the ways of supervised learning, weakly supervised learning and unsupervised learning. Especially, training with LR input only in an unsupervised manner gives our method better practical utility. However, given real paired HR-LR data, the supervised methods EDSR [52] and SRFBN [53] may perform better than ours. However, for more general situations when supervised methods cannot work, i.e., weakly supervised learning and unsupervised learning, our method has obvious advantages. It should be noticed that the way to train an unsupervised network by a single image may be unstable due to relatively low pixel values and low contrast of infrared remote sensing images, e.g., UGSR [37] and EUSR [38]. Such unstableness of SR results makes UGSR and EUSR face difficulties in the practice of infrared remote sensing image super-resolution. While owing to the design of the closed-loops, our framework can achieve stable and comprehensively optimal SR results.

### 5.3. Adaptability of Resolution and Spectrum

　　Remote sensing images and natural images are quite different. Due to the long imaging distance, the spatial resolution of each part of the remote sensing image tends to be consistent. However, the spatial resolution in different areas of natural images often differs due to the different distances between the imaging object and the lens, resulting in blocks of different resolutions on the same image. Therefore, the super-resolution reconstruction of natural images does not need to explain and study the specific resolution of the image, but only super-resolve the input image for a certain up-scaling factor. Since remote sensing images have clear resolution levels, it is valuable to study the SR performance of testing data with different resolution levels from the training data. In addition, the spectrum is also an important factor for infrared remote sensing. Typical satellites with infrared sensors like Landsat-8 usually have multiple bands of sensing spectrum. Such resolution adaptability and spectrum adaptability analyses are especially needed for infrared remote sensing, but usually neglected in natural image super-resolution. According to our experimental study in Sections 4.4 and 4.5, data with the higher spatial resolution is more sensitive to the resolution of training data, and SR network for the same up-scaling factor is still recommended to be re-trained with the data of targeting spatial resolution. Besides, as a summary of the experimental analysis of multiple infrared bands of Landsat-8 data, the short-wave infrared bands B6 and B7 are relatively similar in SR performance. We can replace or merge the training set to deal with the lack of training data. It is the same for B10 and B11.

## 6. Conclusions

　　In this paper, we have proposed a novel closed-loop framework for super-resolution of infrared remote sensing images. It contains two networks and can flexibly implement multiple closed-loop training modes. Experimental results demonstrate our framework can achieve good super-resolution performance in supervised, weakly supervised and

unsupervised situations, and effectively alleviate the impact of synthetic data. In view of the characteristics of infrared remote sensing images, we have conducted research on resolution adaptability and band adaptability, and done comprehensive analysis and discussion. The PSNR result of our method on the PROBA-V dataset is 0.9dB better than the second-best unsupervised super-resolution model. Our method shows outstanding advantages in super-resolution reconstruction of real remote sensing images, which means that our method has great application value in the tasks affected by the spatial resolution of infrared remote sensing data, such as object detection and recognition. Codes and data of this paper will be released to the public at https://github.com/haopzhang/CLN4SR (accessed on 31 January 2023) to accelerate the following related researches.

# References

1. Fernandezbeltran, R.; Latorrecarmona, P.; Pla, F. Single-frame super-resolution in remote sensing: a practical overview. *Int. J. Remote Sens.* **2017**, *38*, 314–354. [CrossRef]
2. Pham, M.; Aptoula, E.; Lefevre, S. Feature Profiles from Attribute Filtering for Classification of Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 249–256. [CrossRef]
3. Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of machine-learning classification in remote sensing: an applied review. *Int. J. Remote Sens.* **2018**, *39*, 2784–2817. [CrossRef]
4. Lin, H.; Shi, Z.; Zou, Z. Fully Convolutional Network With Task Partitioning for Inshore Ship Detection in Optical Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1665–1669. [CrossRef]
5. Wu, T.; Luo, J.; Fang, J.; Ma, J.; Song, X. Unsupervised Object-Based Change Detection via a Weibull Mixture Model-Based Binarization for High-Resolution Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 63–67. [CrossRef]
6. Bai, Y.; Zhang, Y.; Ding, M.; Ghanem, B. SOD-MTGAN: Small Object Detection via Multi-Task Generative Adversarial Network. In *Proceedings of the European Conference on Computer Vision*; Springer: Berlin, Germany, 2018; pp. 210–226.
7. Tatem, A.J.; Lewis, H.G.; Atkinson, P.M.; Nixon, M.S. Super-resolution target identification from remotely sensed images using a Hopfield neural network. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 781–796. [CrossRef]
8. Dai, D.; Wang, Y.; Chen, Y.; Van Gool, L. Is image super-resolution helpful for other vision tasks? In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Lake Placid, NY, USA, 7–10 March 2016; pp. 1–9. [CrossRef]
9. Zanotta, D.C.; Ferreira, M.P.; Zorte, M.; Shimabukuro, Y. A statistical approach for simultaneous segmentation and classification. In Proceedings of the IEEE Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, 13–18 July 2014; pp. 4899–4901. [CrossRef]
10. Hou, B.; Zhou, K.; Jiao, L. Adaptive Super-Resolution for Remote Sensing Images Based on Sparse Representation with Global Joint Dictionary Model. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2312–2327. [CrossRef]
11. Liao, R.; Tao, X.; Li, R.; Ma, Z.; Jia, J. Video Super-Resolution via Deep Draft-Ensemble Learning. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 531–539. [CrossRef]
12. Xu, J.; Liang, Y.; Liu, J.; Huang, Z. Multi-Frame Super-Resolution of Gaofen-4 Remote Sensing Images. *Sensors* **2017**, *17*, 2142. [CrossRef]
13. Yang, C.; Ma, C.; Yang, M. Single-Image Super-Resolution: A Benchmark. In Proceedings of the Europeon Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 372–386.
14. Yang, D.; Li, Z.; Xia, Y.; Chen, Z. Remote sensing image super-resolution: Challenges and approaches. In Proceedings of the IEEE International Conference on Digital Signal Processing, Singapore, 21–24 July 2015; pp. 196–200.

15. Pohl, C.; Genderen, J.V. Remote Sensing Image Fusion. In *Encyclopedia of Image Processing*; Laplante, P.A., Ed.; CRC Press: Boca Raton, FL, USA, 2018; pp. 627–640.

16. Harris, J.L. Diffraction and Resolving Power. *J. Opt. Soc. Am.* **1964**, *54*, 931–936. [CrossRef]

17. Turkowski, K. Filters for common resampling tasks. In *Graphics Gems*; Glassner, A.S., Ed.; Morgan Kaufmann: San Diego, CA, USA, 1990; pp. 147–165.

18. Keys, R. Cubic convolution interpolation for digital image processing. *IEEE Trans. Acoust. Speech Signal Process.* **1981**, *29*, 1153–1160. [CrossRef]

19. Nasrollahi, K.; Moeslund, T. Super-resolution: A comprehensive survey. *Mach. Vis. Appl.* **2014**, *25*, 1423–1468. [CrossRef]

20. Irani, M.; Peleg, S. Improving resolution by image registration. *Graph. Model. Image Process.* **1991**, *53*, 231–239. [CrossRef]

21. Schultz, R.R.; Stevenson, R.L. Extraction of high-resolution frames from video sequences. *IEEE Trans. Image Process.* **1996**, *5*, 996–1011. [CrossRef] [PubMed]

22. Stark, H.; Oskoui, P. High-resolution image recovery from image-plane arrays, using convex projections. *J. Opt. Soc. Am. A* **1989**, *6*, 1715–1726. [CrossRef] [PubMed]

23. Yang, J.; Wright, J.; Huang, T.S.; Ma, Y. Image super-resolution as sparse representation of raw image patches. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.

24. Zeyde, R.; Elad, M.; Protter, M. On single image scale-up using sparse-representations. In Proceedings of the Curves and Surfaces, Avignon, France, 24–30 June 2010; pp. 711–730.

25. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 295–307. [CrossRef]

26. Kim, J.; Lee, J.K.; Lee, K.M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.

27. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 105–114. [CrossRef]

28. Syrris, V.; Ferri, S.; Ehrlich, D.; Pesaresi, M. Image Enhancement and Feature Extraction Based on Low-Resolution Satellite Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 1986–1995. [CrossRef]

29. Timofte, R.; De, V.; Gool, L.V. Anchored Neighborhood Regression for Fast Example-Based Super-Resolution. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 1920–1927. [CrossRef]

30. Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a Deep Convolutional Network for Image Super-Resolution. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 184–199.

31. Polatkan, G.; Zhou, M.; Carin, L.; Blei, D.M.; Daubechies, I. A Bayesian Nonparametric Approach to Image Super-Resolution. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 346–358. [CrossRef]

32. Efrat, N.; Glasner, D.; Apartsin, A.; Nadler, B.; Levin, A. Accurate Blur Models vs. Image Priors in Single Image Super-resolution. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 2832–2839. [CrossRef]

33. Michaeli, T.; Irani, M. Blind Deblurring Using Internal Patch Recurrence. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 783–798.

34. Glasner, D.; Bagon, S.; Irani, M. Super-resolution from a single image. In Proceedings of the IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 349–356.

35. Huang, J.; Singh, A.; Ahuja, N. Single image super-resolution from transformed self-exemplars. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 5197–5206. [CrossRef]

36. Shocher, A.; Cohen, N.; Irani, M. "Zero-Shot" Super-Resolution Using Deep Internal Learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.

37. Haut, J.M.; Fernandezbeltran, R.; Paoletti, M.E.; Plaza, J.; Plaza, A.; Pla, F. A New Deep Generative Network for Unsupervised Remote Sensing Single-Image Super-Resolution. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6792–6810. [CrossRef]

38. Sheikholeslami, M.M.; Nadi, S.; Naeini, A.A.; Ghamisi, P. An Efficient Deep Unsupervised Superresolution Model for Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 1937–1945. [CrossRef]

39. Zhu, J.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2242–2251. [CrossRef]

40. Yuan, Y.; Liu, S.; Zhang, J.; Zhang, Y.; Dong, C.; Lin, L. Unsupervised Image Super-Resolution Using Cycle-in-Cycle Generative Adversarial Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 814–81409. [CrossRef]

41. Ferrari, V.; Hebert, M.; Sminchisescu, C.; Weiss, Y., Eds. *To Learn Image Super-Resolution, Use a GAN to Learn How to Do Image Degradation First*; Springer International Publishing: Cham, Switzerland, 2018.

42. Wang, P.; Zhang, H.; Zhou, F.; Jiang, Z. Unsupervised Remote Sensing Image Super-Resolution Using Cycle CNN. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 3117–3120. [CrossRef]

43. Zhang, H.; Wang, P.; Jiang, Z. Nonpairwise-Trained Cycle Convolutional Neural Network for Single Remote Sensing Image Super-Resolution. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 4250–4261. [CrossRef]

44. 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 IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5406–5415. [CrossRef]

45. Nair, V.; Hinton, G.E. Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the 27th International Conference on International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010; pp. 807–814.

46. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image Super-Resolution Using Very Deep Residual Channel Attention Networks. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 294–310.

47. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2011–2023. [CrossRef] [PubMed]

48. Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1874–1883.

49. Märtens, M.; Izzo, D.; Krzic, A.; Cox, D. Super-resolution of PROBA-V images using convolutional neural networks. *Astrodynamics* **2019**, *3*, 387–402. [CrossRef]

50. Kelvins - PROBA-V Super Resolution—Data. Available online: https://kelvins.esa.int/proba-v-super-resolution/data/ (accessed on 20 November 2019).

51. Earth Explorer. Available online: https://earthexplorer.usgs.gov/ (accessed on 20 November 2019).

52. Zhang, Y.; He, X.; Jing, M.; Fan, Y.; Zeng, X. Enhanced Recursive Residual Network for Single Image Super-Resolution. In Proceedings of the 2019 IEEE 13th International Conference on ASIC, Chongqing, China, 29 October–1 November 2019; pp. 1–4. [CrossRef]

53. Li, Z.; Yang, J.; Liu, Z.; Yang, X.; Jeon, G.; Wu, W. Feedback Network for Image Super-Resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3862–3871.

54. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual Dense Network for Image Super-Resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2472–2481.

55. Zhao, N.; Wei, Q.; Basarab, A.; Dobigeon, N.; Kouame, D.; Tourneret, J. Fast Single Image Super-Resolution Using a New Analytical Solution for $\ell_2 - \ell_2$ Problems. *IEEE Trans. Image Process.* **2016**, *25*, 3683–3697. [CrossRef] [PubMed]

56. He, H.; Siu, W. Single image super-resolution using Gaussian process regression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CL, USA, 20–25 June 2011; pp. 449–456.

57. Soh, J.W.; Cho, S.; Cho, N.I. Meta-Transfer Learning for Zero-Shot 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. 3513–3522. [CrossRef]

58. Liu, Z.S.; Siu, W.C.; Wang, L.W.; Li, C.T.; Cani, M.P.; Chan, Y.L. Unsupervised Real Image Super-Resolution via Generative Variational AutoEncoder. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; pp. 1788–1797. [CrossRef]

59. Wang, L.; Wang, Y.; Dong, X.; Xu, Q.; Yang, J.; An, W.; Guo, Y. Unsupervised Degradation Representation Learning for Blind Super-Resolution. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–25 June 2021; pp. 10576–10585. [CrossRef]

60. Woo, S.; Park, J.; Lee, J.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.