

Research Article

Research on Image Super-Resolution Reconstruction Technology Based on Unsupervised Learning

Shuo Han¹, Bo Mo¹, Jie Zhao¹, Bolin Pan², and Yiqi Wang³

¹School of Aerospace Engineering, Beijing Institute of Technology, Beijing 100081, China

²Southwest Institute of Technical Physics, Chengdu 610046, China

³Shandong Institute of Aerospace Electronics Technology, Yantai 264043, China

Correspondence should be addressed to Shuo Han; 3220215010@bit.edu.cn

Received 31 May 2023; Revised 31 October 2023; Accepted 1 November 2023; Published 21 November 2023

Academic Editor: Giovanni Palmerini

Copyright © 2023 Shuo Han et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Affected by the movement of drones, missiles, and other aircraft platforms and the limitation of the accuracy of image sensors, the obtained images have low-resolution and serious loss of image details. Aiming at these problems, this paper studies the image super-resolution reconstruction technology. Firstly, a natural image degradation model based on a generative adversarial network is designed to learn the degradation relationship between image blocks within the image; then, an unsupervised learning residual network is designed based on the idea of image self-similarity to complete image super-resolution reconstruction. The experimental results show that the unsupervised super-resolution reconstruction algorithm is equivalent to the mainstream supervised learning algorithm under ideal conditions. Compared to mainstream algorithms, this algorithm has significantly improved its various indicators in real-world environments under nonideal conditions.

1. Introduction

Deploying deep learning-based target detection algorithms on infrared-guided weapon platforms can effectively improve guidance accuracy. However, deep learning-based object detection algorithms require high-quality dataset to train the network model. Usually affected by the movement of the aircraft platform and the accuracy limit of the infrared image sensor, the quality of the infrared image obtained is poor, which is manifested with low resolution, lots of noise, and artifacts. Based on this, this paper studies the super-resolution reconstruction technology of infrared images and visible light images.

Image super-resolution reconstruction technology is a technology that migrates images from low-resolution domains to high-resolution domains. It enhances image details and restores part of image information while improving image resolution. Traditional image super-resolution algorithms need to pair high-resolution (HR) images and low-resolution (LR) images and realize the conversion from LR images to HR images by learning the

mapping relationship between them. Since it is difficult to collect real HR and LR image pairs, the existing HR-LR image dataset construction method is mainly to establish a degradation model, and the corresponding LR image is obtained from the HR image through the degradation model. Currently, the performance of super-resolution algorithms in nonideal dataset largely depends on the accuracy of the degradation model.

In order to solve the above problems, this paper designs an image super-resolution algorithm based on unsupervised learning. Compared with the existing image super-resolution algorithm, this algorithm has the following advantages: Firstly, an image degradation model based on the generation countermeasure network is constructed. The network establishes a degradation model by learning the degradation relationship between image blocks and truly simulates the process of image degradation. Secondly, based on the idea of image self-similarity, a super division model of unsupervised learning is designed, which can realize the super-resolution reconstruction of input images without matching HR and LR datasets.

2. Literature Review

Image super-resolution reconstruction technology has important research significance and application value in both civilian and military fields [1]. Traditional methods are interpolation-based super-resolution algorithms, such as nearest neighbor interpolation and bicubic interpolation [2], which have simple principle and fast operation speed, but the image quality after super-resolution is poor, there are serious fuzzy phenomena, and the indicators after subjective and objective evaluations are relatively general. Later, modeling-based super-resolution algorithms were developed, such as the maximum a posteriori estimation method [3] and the iterative back-projection algorithm [4], which have better reconstruction effects than interpolation, but require a large amount of prior knowledge, high cost, and low operational efficiency.

In recent years, there has been significant progress in the application of deep learning in the field of super-resolution reconstruction, solving many problems in traditional algorithms. In 2014, the SRCNN algorithm proposed by Dong et al. applied the convolutional neural network structure to the super-resolution reconstruction task, and the obtained image quality was greatly improved [5]. With the superposition of the network depth, the parameters increase a lot, which may lead to the disappearance of the gradient, and the network is difficult to converge. Kim et al. introduced the residual network into the super-resolution algorithm, which alleviated the problems of gradient disappearance and gradient explosion while increasing the network depth [6]. He et al. introduced a dense model on the basis of the residual network and connected each convolutional layer in the network to the subsequent convolutional layer, so that each layer can learn the features of the previous layer, effectively improving the quality of the reconstructed image [7]. In 2018, Zhang et al. proposed the RCAN algorithm, introduced the attention mechanism, and enhanced the ability of the network to learn features by learning the importance of different channels in the network and assigning corresponding weights to them [8].

The proposal of adversarial networks has pushed the research on image super-resolution reconstruction to a new level, making the high-frequency details of reconstructed images more accurate, edge blur control better, and texture details more refined. Early researchers focused their research on reducing the minimum mean square error of the reconstructed image, resulting in excellent objective indicators of the reconstructed image, but the image is too smooth and lacks high-frequency details. In 2017, Ledig et al. first applied adversarial networks to image super-resolution reconstruction projects and proposed the SRGAN algorithm [9], which added L_1 loss to the basis of adversarial loss to make the reconstructed image visually more natural. In 2018, Wang et al. proposed the ESRGAN algorithm, which incorporated dense block connections into deep residual networks, used relative discriminators, and removed the batch normalization layer from SRGAN, and used the deep network to enhance the quality of reconstructed images, effectively improving the artifact problem [10].

The current image super-resolution reconstruction algorithm is usually based on supervised learning, which requires a training set composed of high-resolution (HR) and low-resolution (LR) images during training. However, it is difficult to collect images of different resolutions in the same scene, so usually performing predefined degradation on high-resolution images to generate low-resolution images leads to poor generalization of the trained network. In 2018, Shocher et al. proposed the ZSSR algorithm, which only relies on the internal information of a single image to complete network training and has achieved remarkable results in processing image super-resolution reconstruction under nonideal conditions [11]. In 2022, Karwowska and Wierzbicki conducted an extensive review of current deep learning-based image super-resolution methods, especially pointing out that generative adversarial networks show great potential in super-resolution satellite image reconstruction [12].

3. Model Establishment

3.1. Construction of Degradation Model Based on Generative Adversarial Network. Most super-resolution algorithms usually assume that the low-resolution image is obtained by downsampling the high-resolution image in a fixed way, such as bicubic downsampling. However, the degradation process of low-resolution images under real conditions is different from others. When the set downscaling kernel deviates from the real downscaling kernel, the performance of the super-resolution reconstruction model will drop significantly, resulting in poor algorithm performance under nonideal conditions. The degradation model of real low-resolution images will be affected by sensor optics and jitter, and the use of incorrect degradation models has the greatest impact on algorithm performance [13, 14]. Therefore, establishing an accurate degradation model is of great significance for improving the performance of super-resolution algorithms.

3.1.1. Overall Network Design. The natural degradation model we designed is based on the generative adversarial network and uses the cross scale cooccurrence property [15] to realize the training of the degradation kernel of a single image and restore the degradation process of the image to the greatest extent. The cross scale cooccurrence property can be understood as the consistency of overall pixel distribution between individual downsampled images of different sizes in natural scenes, although the sizes of image blocks (patches) and downsampled image blocks may be different. At the same time, since the information entropy inside a single image is the lowest, that is, it contains the most information, the patch set inside a single image is sufficient for network training. The overall architecture of the network is shown in Figure 1. The image Real to be reconstructed is input, the degraded output image Fake is generated by the generator, and the patches of the same size in Real and Fake are paired as real samples and fake samples. Finally, the two crops are sent to the discriminator for discrimination.

The degradation network we designed can establish corresponding degradation models for each image by

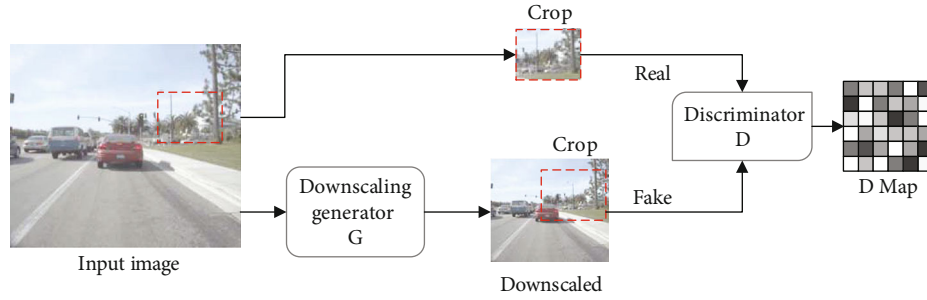


FIGURE 1: Overall structure of degenerate networks.

learning the degradation methods of different image blocks within a single image. The goal of this network is to generate a downsampled image such that the pixels in each image patch match the real LR image as much as possible. First, the input image is downsampled to generate a downsampled image, and then the downsampled image and the input image are cropped to obtain real image blocks and samples for generating downsampled image blocks, and then, both are sent to the discriminator Discriminate to optimize the generator. Finally, the downsampled image generated by the generator is consistent with the input image in distribution. At this time, the confrontation between the discriminator and the generator also reaches a balance, and the generator G can be regarded as the degradation model of the input image.

3.1.2. Generative Network Design. The basic assumption of the super-resolution model is that the low-resolution input image is the result of n -fold downscaling of the HR image. Here, we use generative adversarial networks to learn the degradation relationship of images to establish a more accurate image degradation model. Some influencing factors such as Gaussian noise and Poisson's noise can be ignored and do not affect the overall effect of the experiment. In most cases, the conversion process from HR images to LR images is described as the convolution of HR images and degradation kernels, as shown in

$$I_{LR} = (I_{HR} * k_n) \downarrow_n. \quad (1)$$

In the equation, I_{LR} is a low-resolution image, I_{HR} is a high-resolution image, k_n is a degenerated kernel, and n is a downscaling multiple. It can be approximated that low-resolution images are mainly generated by linear transformation of high-resolution images, so the generation network is designed as a deep and completely linear network, which does not contain batch normalization layers and nonlinear activation functions. Afterwards, the relative loss function and regularization constraint are added to the overall network to avoid the tendency of degenerate kernel towards error and improve the anti-interference performance of the designed degenerate model. In addition, if the generator has no linear display constraints, it may cause unconstrained changes in the overall pixel distribution of the image block, making its pixel distribution deviate from the high-resolution image [16], so it is necessary to use a linear network as the generator. The generative network designed

in this paper is shown in Figure 2, which is a completely linear structure and adopts a deep network structure to enhance the network's feature extraction ability and optimal solution solving ability.

The input to the generated network is a single-channel image. When the input is infrared image, it enters the generation network directly as a single-channel image; when the input is visible light image, it cannot be directly input into the generation network as a three-channel image, so it needs to be merged into a single-channel image by using the merge function. Before entering the network, the image is randomly cropped, and the $2n \times 2n$ image blocks are cut out and sent to the network, and the output of the network is $n \times n$ downsampled image blocks. When $n = 32$, the parameters of each layer of the network are shown in Table 1.

In Table 1, c , h , and w represent the number of channels, height, and width of the input and output features, respectively, while k , n , s , and p are the convolution kernel size, number of convolution kernels, step, and padding, respectively. It can be seen from the table that the first convolutional layer is a downsampling layer, and the feature size is reduced from 64×64 to 32×32 ; the second to fifth convolutional layers are equal-scale convolution, and the feature size remains unchanged; the last layer of the network reduces the number of channels to 1 and outputs 32×32 image blocks.

3.1.3. Discriminative Network Design. The goal of the discriminative network is to learn the distribution of individual patches of the input image and to discriminate between real patches belonging to this distribution and fake patches generated by the generative network. The real samples input by the discriminative network are cropped from real LR images, and the fake samples come from low-resolution images generated by the generator. The discriminator designed in this paper is a fully convolutional network, as shown in Figure 3.

Five convolutional layers are used in the discriminant network. The parameters of each convolutional layer are shown in Table 2, which does not contain any pooling layer. After each convolutional layer, the IN layer and the ReLU layer are connected, and spectral normalization is performed. The activation function in the last layer is replaced by the sigmoid function. The convolution kernel size of the first layer of convolution is 7×7 , and the convolution kernel sizes of the second to sixth layers are all 1×1 . The input of the discriminative network is 32×32 image blocks. Through the calculation and processing of the network, the final

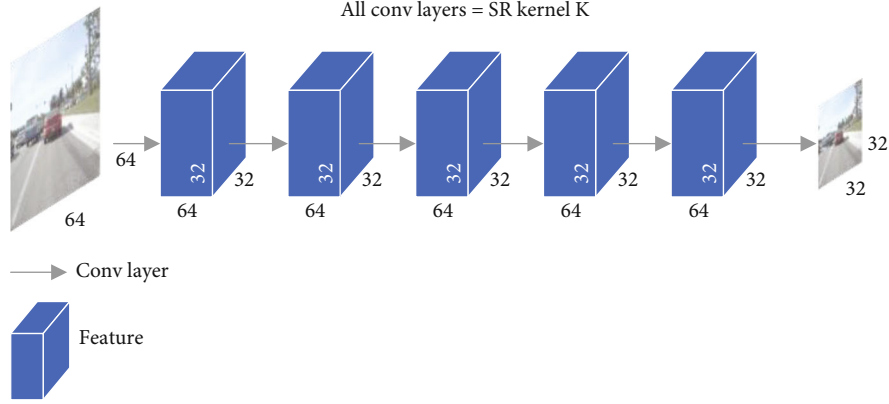


FIGURE 2: Linear generative network structure.

TABLE 1: Generate network parameters.

Network layer	Input feature (c, h, w)	Convolutional layer parameters (k, n, s, p)	Output feature (c, h, w)
Conv1	$(1, 64, 64)$	$(4, 64, 2, 1)$	$(64, 32, 32)$
Conv2	$(64, 32, 32)$	$(3, 64, 1, 1)$	$(64, 32, 32)$
Conv3	$(64, 32, 32)$	$(3, 64, 1, 1)$	$(64, 32, 32)$
Conv4	$(64, 32, 32)$	$(1, 64, 1, 0)$	$(64, 32, 32)$
Conv5	$(64, 32, 32)$	$(1, 64, 1, 0)$	$(64, 32, 32)$
Conv6	$(64, 32, 32)$	$(1, 1, 1, 0)$	$(1, 32, 32)$

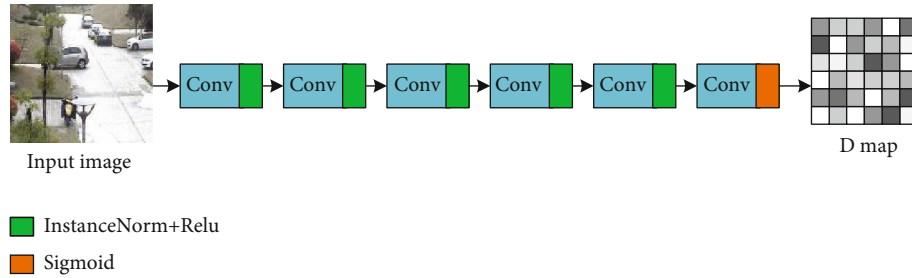


FIGURE 3: Schematic diagram of discriminant structure.

TABLE 2: Discriminant network parameter.

Network layer	Input feature (c, h, w)	Convolutional layer parameters (k, n, s, p)	Output feature (c, h, w)
Conv1	$(1, 32, 32)$	$(7, 64, 1, 0)$	$(64, 26, 26)$
Conv2	$(64, 26, 26)$	$(1, 64, 1, 0)$	$(64, 26, 26)$
Conv3	$(64, 26, 26)$	$(1, 64, 1, 0)$	$(64, 26, 26)$
Conv4	$(64, 26, 26)$	$(1, 64, 1, 0)$	$(64, 26, 26)$
Conv5	$(64, 26, 26)$	$(1, 64, 1, 0)$	$(64, 26, 26)$
Conv6	$(64, 26, 26)$	$(1, 1, 1, 0)$	$(1, 26, 26)$

output is a discriminative matrix of size 26×26 . After multiple trainings, we calculated that the best discrimination effect is achieved in this size.

In Table 2, c , h , and w represent the number of channels, height, and width of the input and output features, respectively, while k , n , s , and p are the convolution kernel size, number of convolution kernels, step, and padding, respectively. From the above table analysis, the discriminant network only reduces the feature size in the first layer, while

expanding the number of channels to 64; use 1 for the second to fifth floors $\times 1$ convolutional kernel to extract features; the number of contraction channels in the last layer is 1, and the output is limited between 0 and 1 using the sigmoid function to output the final discrimination matrix.

3.1.4. Degenerate Kernel Constraint and Loss Function Design. At any iteration during network training, the generator G performs an estimated downscaling operation on a

specific LR input image whose degenerated kernel is obtained from the generator's implicit training weights. The final output required by the network is a degraded kernel that can be directly applied to the image rather than a downscaling network, because the degraded kernel can be directly applied to any super-resolution algorithm, and a priori operation can be performed on the degraded kernel such as artificially constrained degraded kernel, making it more in line with the natural degradation process of the image. The degenerate kernel is extracted from the trained generator, and its specific operation is to input the matrix whose element values are all 1 into the generating network for calculation.

In addition to the usual adversarial loss, the loss function of the generator adds regularization restrictions to limit the possible error tendency of the degenerate kernel, as shown in the following equations:

$$l_{\text{sum_to_1}} = \left| 1 - \sum_{i,j} k_{i,j} \right|, \quad (2)$$

$$l_{\text{boundaries}} = \sum_{i,j} |k_{i,j} \bullet m_{i,j}|, \quad (3)$$

$$l_{\text{sparse}} = \sum_{i,j} |k_{i,j}|^{1/2}, \quad (4)$$

$$l_{\text{center}} = \left\| (x_0, y_0) - \frac{\sum_{i,j} k_{i,j} \cdot (i, j)}{\sum_{i,j} k_{i,j}} \right\|, \quad (5)$$

where $k_{i,j}$ is the value of each point in the degenerate kernel, m is a constant weight matrix whose weight increases exponentially with the distance from the center of k , (x_0, y_0) represents the center index, $l_{\text{sum_to_1}}$ ensures that the overall sum of the degenerate kernel is as close to 1 as possible, $l_{\text{boundaries}}$ punishes nonzero values close to the boundary to ensure that the value of the degenerate kernel boundary is small, l_{sparse} guarantees the sparsity of the coefficients to prevent excessive smoothing of the degenerate kernel causing the image generated by degeneration to be too blurry, and l_{center} ensures that the pixel value in the center of the degeneration kernel is as large as possible. In the first round of training, the output of the generator network will be trained with the bicubic downsampled image as a constraint at the same time, as the initialization of subsequent training.

The total network loss parameter is shown in

$$G^* = \arg \min_G \max_D \{E_x[\log D(x)] + E_x[\log (1 - D(G(x)))] + R\}. \quad (6)$$

In the equation,

$$R = \alpha l_{\text{sum_to_1}} + \beta l_{\text{boundaries}} + \gamma l_{\text{sparse}} + \delta l_{\text{center}}, \quad (7)$$

where α , β , γ , and δ are the weight parameters, which are calculated through training. Here, we take the values 0.5, 0.5, 5, and 1, respectively, after training the network.

3.2. Image Super-Resolution Algorithm Based on Unsupervised Learning. In recent years, the development of deep learning has greatly improved the performance of super-resolution reconstruction (SR) algorithms. However, due to the limitation of supervised learning, these SR algorithms are restricted to be used in a specific training set, which requires the existence of both high-resolution images and matching low-resolution images in the training set. It is very difficult to collect paired high- and low-resolution images in reality, and it is more difficult to collect infrared dataset. Therefore, the effect of super-resolution reconstruction of infrared images using traditional super-resolution algorithms with supervised learning is not ideal. Therefore, we propose an unsupervised super-resolution algorithm based on the theory of image self-similarity, which realizes the training of the super-resolution network using the training set of a single image.

3.2.1. Algorithm Overall Design. Shocher et al. [11] argue that supervised training overly focuses on the differences between large amounts of data, making it impossible to recover this image-specific information. Compared with the external entropy of the regional blocks in the general natural image collection, the information entropy of the internal region blocks of a single image is much smaller. According to information theory, parts with low entropy are richer in information [17], so internal image statistics tend to have stronger predictive power than external information obtained from general image collections. Based on this idea, this paper designs and builds a single-image super-resolution model, using the internal recursion of a single-image information, while training and testing a residual network, whose training set is extracted from the image to be super-resolution itself. Among them, the principle structure of supervised learning and unsupervised learning super-resolution algorithms is shown in Figure 4.

The super-resolution reconstruction algorithm based on supervised learning requires a training set composed of a large number of paired HR and LR images. The LR images are input into the network, and the training network outputs the corresponding HR images. The image super-resolution algorithm we designed based on unsupervised learning requires only one image for the training set. Using the similarity of the image itself, the dataset is constructed from the image itself, and then, the self-supervised network is trained, and finally, the converged network is applied to the original low-resolution image to complete the super-resolution reconstruction. The overall process of the algorithm is as follows:

Step 1. Perform data expansion and image enhancement on the input single image. Data expansion is mainly because the image obtained by the aircraft platform in a nonstationary state may be an incomplete image, so it needs to be filled into a complete image. Image enhancement mainly includes contrast adjustment, probability flipping, and other preprocessing. Then, the image is divided into multiple images with smaller sizes as HR samples and fed into the degradation model built in the previous section to obtain LR samples, thereby matching specific HR-LR image pairs. The network can be trained on these image pairs randomly.

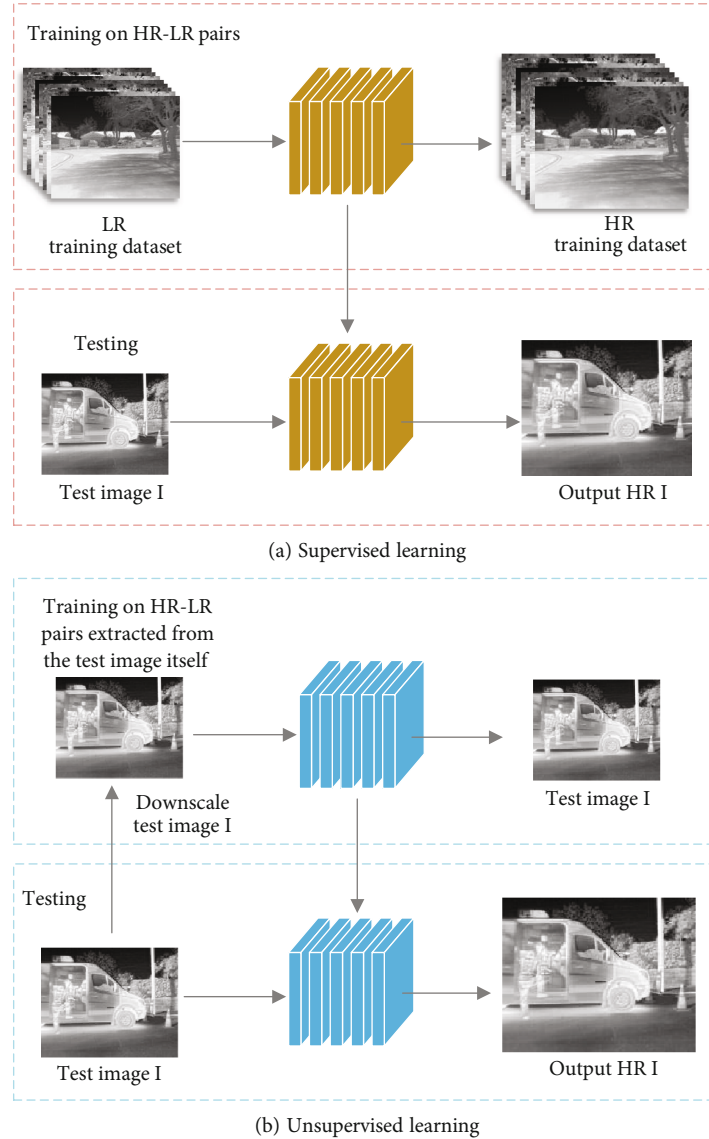


FIGURE 4: Schematic diagram of super-resolution algorithm.

Step 2. Send the LR image to the network for super-resolution, generate an HR image, and compare it with the HR image in the dataset to calculate the loss.

Step 3. Backpropagation, the loss value obtained in step 2, is forwarded to update the parameters of the network.

Step 4. Repeat steps 2 and 3 until the loss approaches the lowest value and converges. At this time, the image generated by the network is closest to the HR image.

Step 5. Send the original image to the network to obtain a high-resolution image whose resolution is enlarged by n times.

3.2.2. Network Structure Design. The supervised convolutional network (CNN) trained on a large external set of diverse HR-LR images [18] must capture all possible diversity in the weights of its network, learning as much as possible from all HR-LR images in the training set. The mapping relationship of LR image pairs, therefore, the number of layers of the supervised learning network, is more, and the structure is more complex. In the unsupervised learning

designed in this paper, the diversity of the HR-LR relationship of a single image is much smaller, so a smaller and simpler dedicated network can be used for encoding. The designed network structure is shown in Figure 5.

The network module includes an upsampling layer and four residual blocks. The first layer is an upsampling layer, which upsamples the image by n times according to the scaling factor n and restores it to a high-resolution image. The second to fifth layers are the residual connection layers, in which the design structure of the residual block is shown in the Bottleneck structure at the bottom of Figure 5, in which the 1×1 convolution kernel of the first layer performs downsampling operation on the input features, reducing the number of channels of the matrix is a quarter, the second layer performs feature learning, and the 1×1 convolution of the third layer performs a dimension-boosting operation on the feature matrix, restoring the depth of the feature matrix to the number of channels of the input residual block. Through the design of the Bottleneck residual block, the

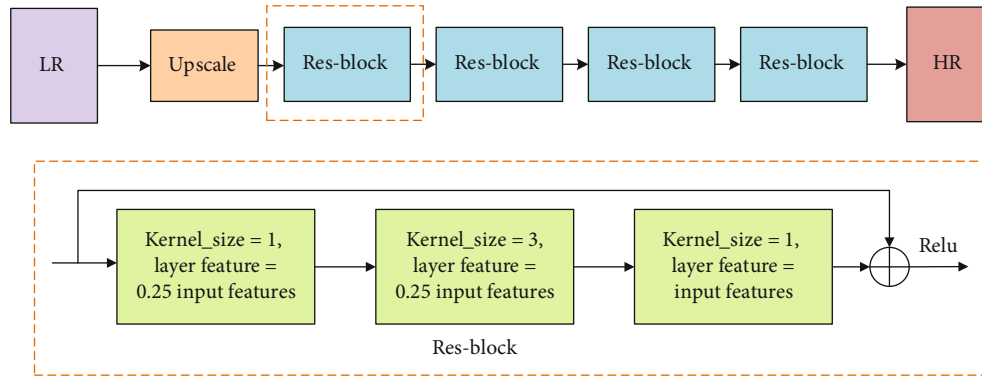


FIGURE 5: Schematic diagram of unsupervised learning network structure.

number of parameters can be effectively reduced. Compared with the residual block structure of Res-Block [19], the calculation amount of input features of the same dimension can be reduced by half when passing through the residual block. Reduce the probability of gradient disappearance while training.

The loss function of the network uses the *Smooth L₁* function. In the early stage of training, it can quickly converge to the center. In the later stage of training, the gradient value is small enough to smoothly converge to the optimal solution.

4. Verification and Analysis

4.1. Dataset Production and Network Training Strategy. In this paper, the FLIR thermal infrared imaging dataset and the DIV2K high-definition visible light dataset are selected as the experimental dataset for super-resolution reconstruction and test the effect of the algorithm. Among them, the FLIR thermal infrared imaging dataset has obvious characteristics and rich detailed information, so it is suitable for use as a dataset for unsupervised super-resolution algorithms; DIV2K is a classic super-resolution algorithm training set, which includes 1000 high-definition images with 2K resolution and high clarity, making it suitable for comparative experimental verification of the algorithm. An example of FLIR and DIV2K datasets is shown in Figure 6.

The algorithm in this paper is based on the network design of unsupervised learning, and the establishment of the degradation model and the super-resolution image reconstruction can be completed through a single image. Among them, the establishment of the training dataset is mainly for the comparison experiment with the supervised image super-resolution algorithm. We firstly perform 4 times bicubic downsampling on the FLIR dataset to obtain the infrared LR image under ideal conditions and form an HR-LR image pair with the original image; then, process the DIV2K dataset, in order to verify the effect of the proposed algorithm under nonideal conditions, the training set of visible light is set to the image pair composed of the original image and its 4 times bicubic downsampled image, and the test set is set to the image pair composed of the original image and the image after its composite degradation model.

The algorithm model training in this paper uses an RTX3060Ti GPU, CUDA version 11.4, and the program is



FIGURE 6: Example of FLIR and DIV2K datasets.

written based on the PyTorch deep learning framework. Use the Adam optimizer to optimize the training, and the parameters are set to $\beta_1 = 0.5$, $\beta_2 = 0.99$, and $\epsilon = 0.5$. The initial learning rate is set to 0.0002, which is reduced to 0.0001 after training for 150 epochs, and a total of 300 epochs are trained. When training on the FLIR dataset, the batch size is set to 2. When training on the DIV2K training set, the batch size is set to 1, because of its large image resolution and large memory usage.

4.2. Degradation Model Comparison Experiment Analysis. In this section, the degradation model is firstly analyzed by comparative experiments and tested on the infrared dataset FLIR and the visible light dataset DIV2K, respectively. Figure 7 shows the results of testing on infrared images, and Figure 7(a) shows the infrared images from the original FLIR dataset. Figure 7(b) shows the degradation results of the degradation model on the original image in this paper. The left side shows the degradation kernel calculated by the algorithm in this paper. The degradation kernel is used as a filter to process the original image and obtain the degradation image on the right side. Figure 7(c) shows the degraded effect of bicubic and Gaussian blurring on the original image. It can be seen from the figure that the algorithm in this paper can learn the degradation mode of the image and construct different degradation kernels according to different images. Bicubic downsampling only scales the image on the pixel, without adding blur, resulting in the loss of blur information. Gaussian blur applies the same blur kernel to all images, which easily leads to the problem of excessively high generated blur in Figure 7(c).

Next, experiments were conducted on the visible light dataset DIV2K, which provided both 2 times and 4 times downsampling degraded image dataset under the composite degradation model. In this section, the

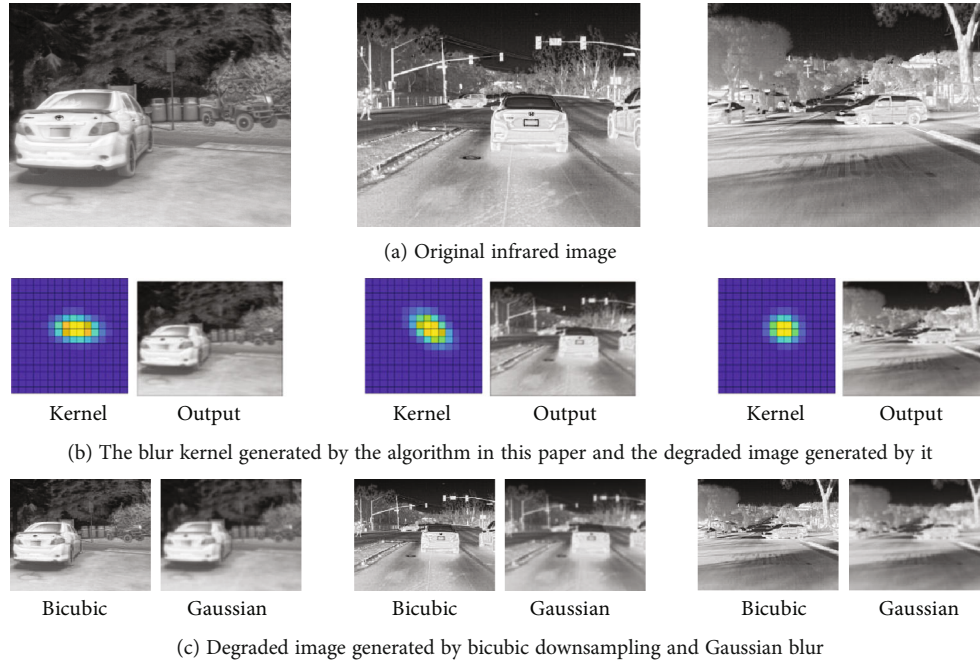


FIGURE 7: Infrared image degradation experiment results.



FIGURE 8: Visible light image degradation experiment results.

2 times downsampling images were first degraded and then compared with its 4 times downsampling images. The experimental results are shown in Figure 8.

It can be seen from Figure 8 that the degraded image in the DIV2K dataset is very similar to the image processed by the Gaussian blur kernel and the algorithm blur kernel proposed in this paper in terms of blur. It can also be seen from the degraded core extracted by the degradation network that the distribution of degraded pixels is very similar to the traditional isotropic Gaussian blur kernel, as shown in Figure 9.

Then, it is objectively evaluated and analyzed for PSNR and SSIM indicators. The experimental results are shown in Table 3.

It can be seen from the table that the image after Gaussian blur kernel processing has the highest score, because the LR dataset of DIV2K was produced using Gaussian blur like check images for degradation. The score of the algorithm in this paper is slightly lower than that of the Gaussian blur kernel, but compared with the bicubic method, the PSNR index is improved by 14.1%, and the SSIM index is improved

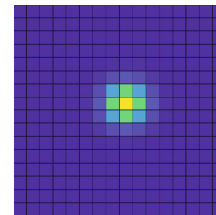


FIGURE 9: Isotropic Gaussian blur kernel.

TABLE 3: Comparison of objective evaluation indexes of different degradation models.

Index	Image degraded model		
	Bicubic	Gaussian	Ours
PSNR	22.282	25.935	25.440
SSIM	0.738	0.842	0.819

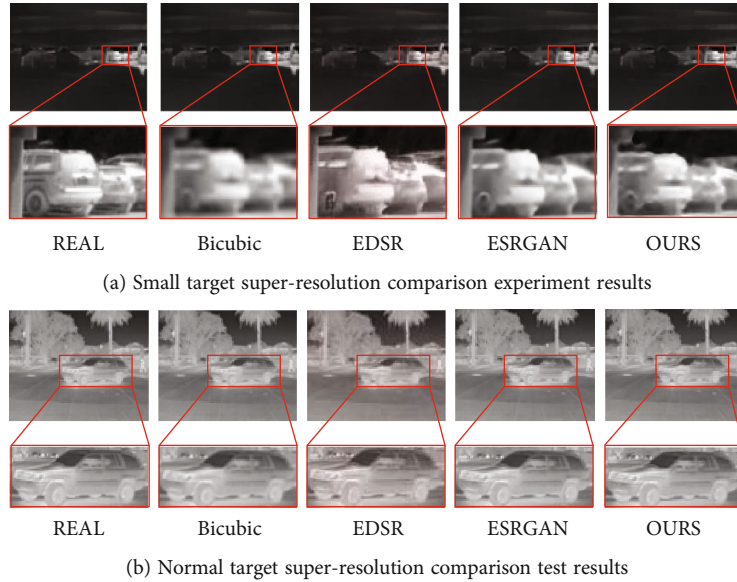


FIGURE 10: FLIR dataset super-resolution reconstruction results.

by 10.98%. The overall analysis of the experimental results can lead to the following conclusions:

- (1) The degradation network designed in this paper can effectively learn the degradation relationship of the image itself, thereby simulating the degradation process of the image, and has good performance on both infrared and visible light image datasets. Compared with the traditional bicubic algorithm, it can more realistically simulate the image degradation process
- (2) The degradation algorithm in this paper does not perform as well as the Gaussian blur kernel in the dataset of unknown degradation model. Because the LR image in DIV2K dataset is artificially degraded, Gaussian blur kernel is used as a part of the degradation model, resulting in Gaussian blur kernel performing better in this dataset. The experimental results show that the fuzzy kernel generated by the degradation algorithm in this paper is very close to the Gaussian blur kernel, which proves that the algorithm can better fit the image degradation process

4.3. Image Super-Resolution Reconstruction Experimental Results. We first trained different super-resolution reconstruction algorithms on the FLIR dataset under ideal conditions and then tested them on the test set and selected two representative results. The experimental results are shown in Figure 10. Among them, the first line in the picture (a) and picture (b) is the full-scale image after the infrared image super-resolution enhancement, and the image in the second line is the reconstruction detail display of the red frame part. It can be seen from Figure 10(a) that the algorithm in this paper can better restore the contour information of the vehicle and the details of the pedestrian on the right, and the overall sensory performance is slightly inferior to the ESRGAN [10] algorithm based on the generative

TABLE 4: Comparison of objective evaluation indexes of each algorithm in FLIR dataset.

Index	Algorithm			
	Bicubic	EDSR	ESRGAN	Ours
PSNR	29.395	28.867	30.508	29.763
SSIM	0.628	0.621	0.668	0.650

confrontation network. Compared to the traditional bicubic method and the deep network structured EDSR [20] algorithm, the restoration degree is higher, and the suppression of blur and noise is better, resulting in clearer generated images. From Figure 10(b), it can be seen that the algorithm in this paper effectively restores the detailed information of the vehicle, such as tires, rearview mirrors, and doors, while effectively suppressing the fuzzy effects present in the bicubic method; due to the rich detailed information of the vehicle in this test image, the algorithm in this paper is very close to the supervised super-resolution algorithms ESRGAN and EDSR in terms of restored visual clarity, but there are some jagged effects on the edges of the object.

Under ideal conditions, the algorithm in this paper achieves similar results to EDSR and ESRGAN algorithms in terms of visual effects. Then, the above-mentioned algorithms are analyzed objectively, and PSNR and SSIM are used as specific indicators to evaluate the reconstruction effect. The experimental results are shown in Table 4.

Combined with the experimental data in Figure 10 and the above table, it can be concluded that

- (1) the bicubic interpolation algorithm performs well in the objective indicators of the dataset, even surpassing the EDSR, because the HR-LR image pair of the dataset itself is obtained by bicubic downsampling, and then, the bicubic interpolation is performed to achieve super-resolution reconstruction; that is, the reconstruction model coincides with the regression

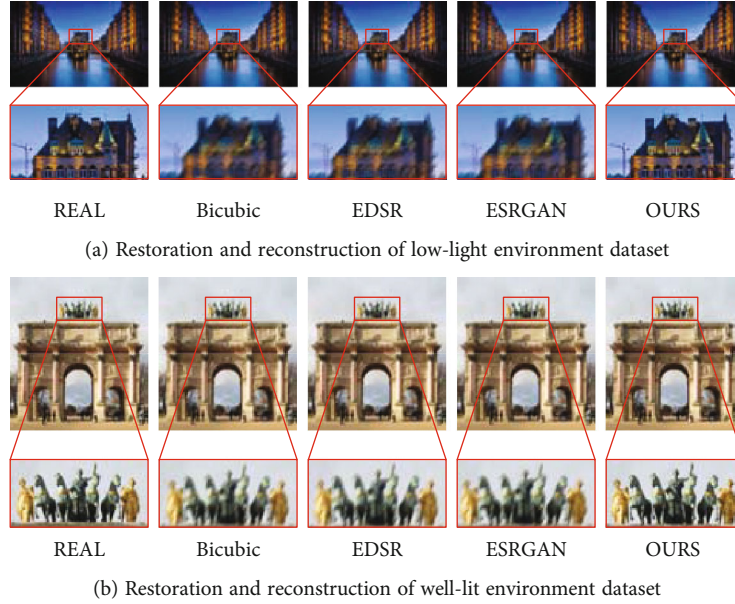


FIGURE 11: DIV2K dataset super-resolution reconstruction results.

model, resulting in the high objective indicators of the bicubic method in the dataset. From Figure 10, it can also be seen that the image reconstructed by the bicubic method has the worst visual clarity among the above algorithms, indicating that objective indicators cannot completely replace subjective visual perception in visual tasks

- (2) the PSNR and SSIM indicators of the algorithm in this paper are only slightly lower than ESRGAN and higher than EDSR and bicubic, indicating that the super-resolution reconstruction effect of the algorithm in this paper is better. It is worth noting that the algorithm in this paper is unsupervised learning and uses the internal information of an image for small sample training, while both EDSR and ESRGAN belong to supervised learning and require a large number of dataset for training. It can be concluded that the network in this paper is ideally similar to the supervised learning algorithm. At the same time, due to the characteristics of unsupervised learning, the network only needs to be trained on a small sample on a single image, which makes the network more generalizable and robust

Then, test on the nonideal DIV2K dataset, and use the traditional bicubic method downsampling LR dataset for supervised network training. In order to test the network reconstruction effect under real conditions, the test set uses the LR dataset processed by the composite degradation model provided by DIV2K. The test results are shown in Figure 11.

Figure 10 shows the enhancement effect of different algorithms on the DIV2K dataset under nonideal conditions, with scaling factors of 4. From Figure 11(a), it can be seen that when the degradation model is unknown, the perfor-

TABLE 5: Comparison of objective evaluation indexes of each algorithm in DIV2K dataset.

Index	Algorithm			
	Bicubic	EDSR	ESRGAN	Ours
PSNR	21.873	21.482	21.633	25.654
SSIM	0.594	0.562	0.582	0.776

mance of the bicubic algorithm decreases sharply, becoming the most blurry among all results. At the same time, the performance of the EDSR and ESRGAN trained in the ideal training set also decreases significantly, as it can recover some details of the window but cannot remove blurring and artifacts; thanks to the ability to construct a degradation network for a single image, the algorithm in this paper can learn the degradation model of a single image, perform super-resolution reconstruction for its specific degradation mode, and generate images that effectively reconstruct details such as windows, roofs, and tower cranes. From Figure 11(b), it can be seen that the details of characters, horses, and other details generated by the algorithm in this paper are significantly better than EDSR and ESRGAN, far exceeding bicubic, and the generated images almost completely filter out blurring effects. The experimental results show that the super-resolution image generated by our algorithm can reconstruct and restore the image defects such as blur, offset, and artifact and perform well in the nonideal dataset. The super-resolution reconstruction algorithms EDSR and ESRGAN based on supervised learning show excellent results in the ideal dataset, but the results are general when processing real low-resolution images.

The performance of the above algorithms is analyzed by objective indicators, and PSNR and SSIM are used for evaluation. The experimental results are shown in Table 5.

After analysis, it can be concluded that

- (1) the performance of EDSR and ESRGAN models trained with ideal dataset is not ideal on nonideal dataset, because the models obtained from network training do not learn the parameter settings under nonideal conditions, resulting in a sharp decline in performance when applied to nonideal test set
- (2) due to its unsupervised nature, the algorithm proposed in this paper can theoretically adapt to images with any degradation mode in reality and reconstruct them. Among them, the performance under ideal conditions is basically consistent with the algorithm based on supervised learning; under nonideal conditions, the PSNR value increased by 19.4% and 18.6% compared to mainstream EDSR and ESRGAN algorithms, respectively, while the SSIM value increased by 38.07% and 25% compared to mainstream EDSR and ESRGAN algorithms, respectively

5. Conclusions

In this paper, image super-resolution enhancement technology based on unsupervised learning is studied. The construction of traditional image super-resolution reconstruction algorithm dataset needs to degrade high-resolution images to obtain low-resolution images, and the degradation model greatly affects the performance of the algorithm. Existing super-resolution reconstruction algorithms usually use bicubic downsampling to construct datasets, resulting in a sharp decline in performance of some algorithms that perform well during training when applied to nonideal datasets.

In response to the above issues, this paper first constructs a degradation model based on generative confrontation network and establishes a related degradation model by learning the degradation relationship between image blocks in the image, so as to simulate the real degradation process of the image to the greatest extent. Then, based on the principle of image self-similarity, an unsupervised learning residual network is designed, which realizes the super-resolution reconstruction of small sample training of a single image, which can be applied to any image.

The experimental results show that the performance of the proposed algorithm on the ideal test set is basically the same as that of the excellent mainstream super-resolution reconstruction algorithms such as EDSR and SRGAN, and the performance on the nonideal test set greatly surpasses these algorithms.

Data Availability

The dataset used to support the results of this research are open source, and some of them are selected for experimentation and verification in this paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work has been supported by the Laboratory of Flight Dynamics and Control, Beijing Institute of Technology.

References

- [1] X. Wei, F. Wang, J. Song, and S. Huang, "Medical image super-resolution reconstruction technology based on conditional generative adversarial network," *Basic & Clinical Pharmacology & Toxicology*, vol. 124, no. S3, 2019.
- [2] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 6, pp. 1153–1160, 1981.
- [3] J. Su, H. Fang, W. Bao, H. Sun, J. Gao, and L. Zhao, "A maximum a posteriori estimation based method for estimating pulse time delay," *Advances in Space Research*, vol. 69, no. 11, pp. 3966–3982, 2022.
- [4] C. Tian, J. Hu, X. Wu, and W. Wen, "Deep iterative residual back-projection networks for single-image super-resolution," *Journal of Electronic Imaging*, vol. 31, no. 2, 2022.
- [5] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Computer Vision – ECCV 2014. ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., vol. 8692 of Lecture Notes in Computer Science, pp. 184–199, Springer, Cham, 2014.
- [6] J. Kim, K. Lee, and K. M. Lee, "Accurate images super-resolution using very deep convolution networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1646–1654, Las Vegas, NV, USA, 2016.
- [7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, Las Vegas, NV, USA, 2016.
- [8] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, vol. 11211, pp. 286–301, Springer, Cham, Munich, Germany, 2018.
- [9] C. Ledig, L. Theis, F. Huszár 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 (CVPR)*, pp. 105–114, Honolulu, HI, USA, 2017.
- [10] X. Wang, K. Yu, S. Wu et al., "Esrgan: enhanced super-resolution generative adversarial networks," in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, vol. 11133, Springer, Cham, Munich, Germany, 2018.
- [11] A. Shocher, N. Cohen, and M. Irani, "'Zero-shot' super-resolution using deep internal learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3118–3126, Salt Lake City, UT, USA, 2018.
- [12] K. Karwowska and D. Wierzbicki, "Using super-resolution algorithms for small satellite imagery: a systematic review," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 3292–3312, 2022.
- [13] N. Efrat, D. Glasner, A. Apartsin, B. Nadler, and A. Levin, "Accurate blur models vs. image priors in single image super-resolution," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2832–2839, Sydney, NSW, Australia, 2013.

- [14] R. A. Hummel, B. Kimia, and S. W. Zucker, "Deblurring Gaussian blur," *Computer Vision, Graphics, and Image Processing*, vol. 38, no. 1, pp. 66–80, 1987.
- [15] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, 2002.
- [16] X. S. Tang, Y. Ding, and K. Hao, "A novel method based on line-segment visualizations for hyper-parameter optimization in deep networks," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 32, no. 3, pp. 1851002.1–1851002.15, 2018.
- [17] A. Adu-Bredu, Z. Zeng, N. Pusalkar, and O. C. Jenkins, "Elephants don't pack groceries: robot task planning for low entropy belief states," *IEEE Robotics and Automation Letters*, vol. 7, no. 1, pp. 25–32, 2022.
- [18] J. Zhu, S. An, W. Liu, and L. Li, "Improved algorithms for zero shot image super-resolution with parametric rectifiers," in *Advanced Data Mining and Applications. ADMA 2019*, J. Li, S. Wang, S. Qin, X. Li, and S. Wang, Eds., vol. 11888 of Lecture Notes in Computer Science, Springer, Cham, 2019.
- [19] Y. Wang, X. Guo, P. Liu, and B. Wei, "Up and down residual blocks for convolutional generative adversarial networks," *IEEE Access*, vol. 9, pp. 26051–26058, 2021.
- [20] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 1132–1140, Honolulu, HI, USA, 2017.