applied sciences

MDPI

Article

Image Shadow Removal Using End-To-End Deep Convolutional Neural Networks

Hui Fan [1,2], Meng Han [1,2] and Jinjiang Li [1,2,*]

[1] School of Computer Science and Technology, Shandong Technology and Business University, Yantai 264005, China; fanlinw@263.net

[2] Co-Innovation Center of Shandong Colleges and Universities: Future Intelligent Computing, Yantai 264005, China; hmsdibt@163.com

[*] Correspondence: lijinjiang@gmail.com

Received: 9 January 2019; Accepted: 5 March 2019; Published: 11 March 2019

Abstract: Image degradation caused by shadows is likely to cause technological issues in image segmentation and target recognition. In view of the existing shadow removal methods, there are problems such as small and trivial shadow processing, the scarcity of end-to-end automatic methods, the neglecting of light, and high-level semantic information such as materials. An end-to-end deep convolutional neural network is proposed to further improve the image shadow removal effect. The network mainly consists of two network models, an encoder–decoder network and a small refinement network. The former predicts the alpha shadow scale factor, and the latter refines to obtain sharper edge information. In addition, a new image database (remove shadow database, RSDB) is constructed; and qualitative and quantitative evaluations are made on databases such as UIUC, UCF and newly-created databases (RSDB) with various real images. Using the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) for quantitative analysis, the algorithm has a big improvement on the PSNR and the SSIM as opposed to other methods. In terms of qualitative comparisons, the network shadow has a clearer and shadow-free image that is consistent with the original image color and texture, and the detail processing effect is much better. The experimental results show that the proposed algorithm is superior to other algorithms, and it is more robust in subjective vision and objective quantization.

Keywords: end-to-end; encoder–decoder; convolutional neural network; shadow removal

1. Introduction

Today's world is an era of rapid information development. With the continuous upgrading of multimedia technology and electronic products, human beings are exposed to more multimedia information such as text, images, and audio in their daily life. Most of this multimedia information is obtained through people's visual systems. We define the multimedia information obtained through this channel as an image. However, when the image is acquired, it is susceptible to various conditions that generally degrade the quality of image. Shadows are one of them. Shadows are a phenomenon of quality degradation caused by imaging conditions. These conditions lead to missing or interfered information reflected by the target; therefore, the interpretation accuracy of the image is sharply reduced, and it affects various quantitative analyses and applications of the image.

Shadow detection and removal [1–4] is one of the most basic and challenging problems in computer graphics and computer vision. The removal of shadow images are important pre-processing stages in computer vision and image enhancement. The existence of shadow not only affects the visual interpretation of the image, but also affects the analysis of the image and the subsequent processing results. For instance, a darker area (caused by shadows) introduces incorrect segments in image segmentation [5,6]; radiation changes (caused by shadows) reduce the performance of the target recognition [7] system; and the presence of shadows reduces the performance of the target tracking [8] system. Therefore, it is necessary to perform shadow detection and analysis on the image in order to reduce or eliminate the influence of the image shadow, it also increases the visual authenticity and physical authenticity of the image through editing and processing.

Shadows are generated by different illumination conditions, and the shadow image I_s is represented by the multiple of the shadow-free image I_{sf} and the shadow ratio α (pixel level $\otimes$) according to [9–11], as shown in (1).

$$I_s = \alpha \otimes I_{sf}. \tag{1}$$

Shadows can be divided into two categories according to different causes [12]: one type is self-shadow, which is produced by the occluded object itself not being illuminated by the light source; the other type is projection (cast shadow), which is caused by an object blocking the light source while producing shadows on the surface of other object. The projection is further divided into an umbra zone and penumbra zone, wherein the umbra zone is formed by completely direct ray-blocking, and the penumbra zone is partially blocked by light.

Given a single shadow image, shadow removal aims to generate a high-quality, shadow-free image with the original shadow image area restored to the shadow-free image in texture, color, and other features. Existing methods of removing shadow regions typically involve two steps: shadow detection and shadow removal. Firstly, shadow detection is used to locate the shadow area [13,14] or the user manually marks the shadow area [9,15,16], then the model is constructed to rebuild both and to remove shadows.

However, shadow detection is an extremely challenging task. Traditional physics-based methods can only be applied to high-quality images, while statistical learning-based methods rely on features that users have to manually label [15,17]. With the development of neural networks, convolutional neural networks (CNNs) [14,18,19] are used to learn the features of shadow detection. CNNs overcome the shortcomings of traditional methods that require high-quality images and manual annotation features. However, they still limited to small network architectures, owing to the shortage of training data.

Similarly, even if the shadow area is known, it is still a challenge to remove it. At the same time, it can be clearly understood that the effect of shadow detection will seriously affect the result of shadow removal. If the effect of shadow detection is not very good, it is impossible to obtain a high-quality, shadow-free picture in the subsequent removal process.

Inspired by the classical adaptive algorithm, Calvo-Zaragoza and Gallego [20] used the convolutional automatic encoder for image binarization, which is an end-to-end network structure and shows good performance in image binarization processing. Image binarization can be formulated as a two-class classification task at the pixel level, which is similar to the shadow removal task. At the same time, one of the biggest drawbacks of the current image shadow removal is the lack of the end-to-end channel. Inspired by this, the purpose of this paper is to design an end-to-end and fully automated network structure to remove image shadows. Thereby, through the ex-trained network structure, the input of a single shadow image can be transformed to a high-quality, shadow-free image, which provides a good basis for subsequent target detection and target tracking.

Different from traditional methods, this paper applies the most prominent features of the deep encoding–decoding model framework of end-to-end and convolutional neural network (CNN) design. Inputting a single shadow image, a shadow mask is obtained. It can describe the global structure of the scene and the high-level semantic information of shadow characteristics. Using the relationship between the shadow ratio and shadow image, Equation (1) realizes an end-to-end fully automated algorithm to restore the shadow-free image. At the same time, this paper designs a small network structure for refinement in order to better handle local information, thus predicting a more accurate alpha mask and sharper edge information.

2. Related Work

In recent years, many scholars have analyzed the characteristics of shadows, established shadow generation models, and proposed a number of algorithms for shadow detection and shadow removal. Analyzing the research results comprehensively, it is not difficult to find that the shadow removal algorithm mostly follows these principles: detecting the shadow or marking it manually, then creating models to remove the shadow.

Gong and Cosker [15] proposed a statistical-based, interactive shadow removal method for shadows in extreme conditions with different shadow softness, irregular shadow shapes, and existing dark textures. First, using the dynamic learning method, two users roughly marked the shadow and non-shadow areas in the image, meaning the shadow area was highlighted and the non-shadow area was resisted, so the shadow area could be extracted easily. On this basis, the model was constructed for shadow removal. Gong and Cosker used a rough hand-marked form to detect shadows while sacrificing finer shadows and a full range of simpler user input autonomy. Gryka [16] and Yu [21] et al. also proposed to manually mark the shadows in an interactive manner to achieve the purpose of shadow removal. Different from Gong and Cosker, Yu [21] proposed a color line-based artificial interactive shadow removal algorithm, which required users to provide shaded areas, shadow-free areas, and areas with similar textures changing significantly in brightness. In the work of Gryka [16], they used the unsupervised regression algorithm to remove the shadows through inputting the shaded area as processed by user, and mainly focused on the soft shadows in the real scene.

Finlayson [22] proposed a shadow removal algorithm based on Poisson's equation, which had a good effect on a considerable part of the image. However, there was no consideration of the influence of ambient illumination and material changes, thus resulting in poor texture restoration in shadow areas. Generally, the brightness contrast between the shadow area and the non-shadow area in the image are relatively large while the features such as texture and color are similar. Therefore, the shadow area and the non-shadow area have similar gradient fields. Liu et al. [23] used this feature to propose an image shadow removal algorithm based on gradient domain, which solved some shortcomings of Poisson equation, but it had a poor performance on discontinuous or small shadow regions. Xiao et al. [24] proposed a shadow removal algorithm based on Subregion Matching Illumination Transfer. This method takes into account the different reflectivity characteristics of different materials, which makes the processing of shadow regions in complex scenes better.

On the one hand, Zheng [25] proposed a projection shadow detection and elimination algorithm based on the combination of texture features and luminance chrominance chroma (YUV) color space. Firstly, a moving object was detected using a pixel-based adaptive segmenter (PBAS) algorithm that resisted shadows. In addition, a portion of the projected shadow candidate area was obtained by texture detection. Then, another projected shadow candidate area was obtained by shadow detection that was based on the luminance chrominance chroma (YUV) color space. Finally, the two portions of the projected shadow candidate regions were filtered and merged by the shadow features. In the work of Murali [26], they proposed a simple method for detecting and removing shadows from a single red green blue (RGB) image. The shadow detection method was selected based on the average value of the RGB images in the A and B planes of the CIELab color model (LAB) equivalent image, and the shadow removal method was based on the recognition of the amount of light irradiated on the surface. The brightness of the shaded areas in the image was increased, and then the color of the surface portion was corrected to match the bright portion of the surface. Tfy et al. [27] trained a kernel least-squares support vector machine (LSSVM) to separate shadow and non-shadow regions. It was embedded it into the Markov random field (MRF) framework and pairwise contextual cues were added to further enhance the performance of the region classifier so the detection of shadow area was realized. A shadow removal method based on area re-lighting is raised based on this.

On the other hand, for the umbra and penumbra regions in projection, researchers have proposed an image shadow removal algorithm based on convolutional neural network structure (CNN) [14,18] or intensity domain [9,11] in deep learning. Both Khan [14] and Shen [18] used convolutional neural networks (CNN) for learning. The former used multiple CNN structures for fusion applications and conditional random field (CRF) for image shadow region predictions, while the latter applied structured label information to the classification process. The two methods were different in the process of extracting and removing shadows. The former used the Bayesian framework, while the latter applied the least-squares method of optimization to perform the final recovery optimization. Arbel et al. [9] used the Markov random field to determine the penumbra region, and then created a smoothing model in the shadow region to construct a shadow mask. Wu [11] extracted the shadow mask by considering features such as color shift, texture gradation, and shadow smoothness. Meanwhile, it reserved the texture of the shadow area during the extraction process.

Analyzing and summarizing the above methods, we find that the current method of shadow removal can either restore the texture of the shadow area effectively, or neglect the influence of environment and material of the object. Most methods are interactive rather than fully automated, which greatly reduces the efficiency of use. Therefore, the purpose of this paper is to propose an end-to-end, fully automatic network structure to remove image shadow.

With the development of artificial intelligence in recent years, some researchers have established fully automatic models for image shadow removal [28–31]. A shadow removal algorithm based on generative adversarial networks (GANs) is proposed in the work [30], which uses two GANs for joint learning in order to detect and remove shadows. Although this allowed fully automatic shadow detection using GANs, it relied on the shadow detection result of the previous generation against the network in the shadow removal stage, and it did not perform well when dealing with small shadow areas and shadow edges. In the work of Yang [28], a three-dimensional inherent image restoration method based on bilateral filtering and a two-dimensional intrinsic image is proposed. Two-dimensional original images are derived from the original image, and then the shadow restoration method of the three-dimensional original image is proposed by deducing two-dimensional original images and bilateral filtering. Finally, an automatic shadow removal method for a single image is realized. However, since the reconstruction of the image may result in changes in the non-shaded area, a high-quality image consistent with the shadow-free image generally cannot obtained.

Therefore, this paper aims to explore the end-to-end deep convolutional neural network for image shadow removal. An encoding–decoding neural network is used, and a small neural network with a two-layer neural network structure is used to train learning. Then, the original image (shadow image) is inputted into the trained network structure to realize fully automatic shadow removal. Finally, a high-quality, shadow-free image is obtained.

3. End-To-End Convolutional Neural Network

This article aims to solve the shadow removal problem using an end-to-end deep convolutional network structure, which is named RSnet. It aims to learn a mapping function between the shadow image and its shadow matte. The RSnet model mainly includes two parts: an encoding–decoding stage and an elaboration stage. Specifically, we used a network structure of a deep convolutional encoder–decoder that took an image as the input, and it took into account penalties for the shadow mask factor (α) that predicted loss and penalties for new composition loss. At the same time, a refined, small network structure was introduced in order to consider multi-context scenarios as well as process local information and edge information, etc. The related information of the two networks, the training of RSnet model, and how to get the shadow-free image will be described in detail in the following chapters.

3.1. Encoding–Decoding Phase

The deep encoding–decoding network structure is a very common model framework in deep learning. The encoder and decoder parts can be any text, voice, image, or video data. The model can use CNN, recurrent neural network (RNN), long short-term memory (LSTM), gated recurrent unit (GRU), and so on. One of the most notable features of the encoder–decoder framework is that it is an end-to-end learning algorithm. It has been widely used in various fields and has succeeded in many other computer vision tasks, such as image segmentation [32], object edge detection [33], and image restoration [34]. The encoding–decoding network structure in this paper consisted of two parts, one was the encoder network for extracting features and the other was the decoder network for reconstructing images. Therefore, the encoder network inputted a shadow image, and the output was a related feature; the decoder network took the output of the encoder network as an input and finally outputted a shadow mask through the deconvolution layer and the unpooling layer.

Encoder network: The encoder network was based on the VGG16 network [35], which was originally designed for object recognition. Recent studies have shown that CNNs train a large amount of data on image classification tasks and can be well generalized in databases and tasks, such as depth prediction [36,37] and semantic segmentation [38]. Therefore, this paper used the pre-trained convolution layer of the VGG16 model [35] and transferred its feature representations to shadow masks by fine-tuning, for the function of task-prediction.

As is known, the VGG16 network consists of thirteen 3 × 3 convolutional layers (five convolutional blocks) and three fully connected layers, as well as five maximum pooling and subsampling layers. These five convolutional layers and spatial aggregations greatly increased the acceptance of the network, so that the characteristics of the global context and semantic scenes could be extracted. However, these five largest pooling layers introduced a 32-pixel stride in the network, so the final prediction graph was quite rough. Therefore, in order to extract more detailed features and obtain more intensive predictions, the step sizes of the first, third, and fifth layer pooling layers in the VGG16 model were modified to 1. In addition to this modification, the sub-sampling layer was abandoned, and all fully connected layers in the VGG16 network were replaced with a 1 × 1 convolutional layer [38]. These 1 × 1 convolutional layers enabled the network to operate in a fully convolutional manner. Specifically, the encoder network had a total of 14 convolutional layers and five largest pooling layers.

Decoder network: In the decoder stage, this paper used the deconvolution network proposed by Zeiler [39], which was mirror-symmetric with the convolutional network. The deconvolution network had an unpooling layer and deconvolution layer. With the successful application of deconvolution in neural network visualization, it had been adopted in more and more work, such as scene segmentation, model generation, etc. Deconvolution is also known as transposed convolution, and it can be clearly seen that the forward propagation process of the convolutional layer is the reverse propagation process of the deconvolution layer, and the reverse propagation process of the convolutional layer is the forward propagation process of the deconvolution layer. The unpooling layer captures instance image space information. Therefore, it can effectively reconstruct the detailed structure of an object with finer resolution.

However, in order to speed up the training of the network, the network structure of the VGG16 was rejected. A small network structure was defined to sample the features acquired in the encoding network to obtain the desired output. Specifically, the decoder network was defined as eight convolutional layers, five deconvolution layers, and the last layer was the prediction layer of the shadow mask factor α.

In order to prevent over-fitting and achieve local optimum values, dropout was applied after each convolutional layer of the encoding–decoding network, so that the over-fitting problem could be effectively mitigated and achieve the regularization effect to some extent. This article no longer used the activation function rectified linear unit mentioned in the VGG16 network, ReLU [40]. Instead, the parameter rectification linear unit (PReLU) was used throughout the encoding–decoding phase, which was an improved ReLU proposed by the He Kaiming group [41] and featured non-saturation. PReLU introduced an additional parameter to avoid the over-fitting problems. The definition of PReLU is as follows:

$$p(x_i) = \begin{cases} x_i & x_i \geq 0 \\ \lambda x_i & x_i < 0 \end{cases}, \tag{2}$$

where x_i is the input to the activation function p at channel i, and λ is the learning parameter. In particular, when λ is a fixed non-zero smaller number, it is equivalent to LeakyReLU (a special version of the Rectified Linear Unit (ReLU)); when the parameter λ is 0, PReLU is equivalent to ReLU.

3.2. Refinement Phase

While the first part of the RSnet network prediction worked well, the results could sometimes be too smooth as a result of the encoder–decoder structure. Therefore, we designed a multi-context mechanism for local detail correction and the previous stage of the encoding–decoding network. Thereby, the results of the first phase were further improved so that the prediction result α of the entire network structure was more precise, and the edge processing effect was more refined.

Network structure: The refinement phase was the second phase of the RSnet network, which aimed to improve the prediction result α obtained in the first phase. Therefore, the input of this stage connected the image patch with the obtained α predicted value of the first stage and produced a 4-channel input, and the output was the true shadow scale factor α in the image. To ensure convergence speed and granularity, the network was designed as a small, fully convolutional network with three convolutional layers. Each one had a nonlinear ReLU layer without a down-sampling layer in the first two convolutional layers, as the purpose of this article was to preserve the very subtle structures in the first phase. The specific configuration is shown in Figure 1.

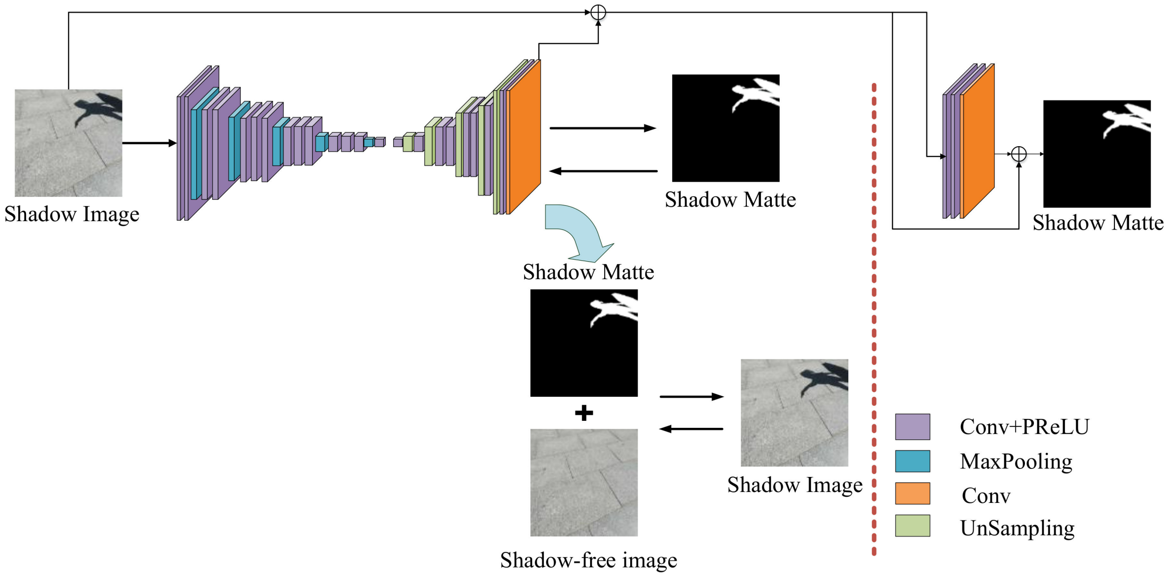

Figure 1. The architecture of the RSnet. RSnet consists of two cooperative sub-networks. The left side of the dotted line shows the encoding–decoding network, and the right side shows the refinement network.

In the refinement phase, instead of making large-scale changes to the alpha mask, we refined and sharpened solely the alpha values.

3.3. Training

The relationship between the shadow image I_s and its shadow mask α is given by Equation (1). In training, the actual shadow mask α was calculated according to a given pair of shadows and shadow-free images. Then, to establish the relationship between the shadow image I_s and the shadow mask α, the mapping function was as follows:

$$\alpha = F(I_s, \beta). \tag{3}$$

Among them, β was the learning parameter of the convolution network. During training, a linear fusion of two loss functions was used to adaptively train the RSnet model.

α prediction loss: The first loss was called α prediction loss, which was the difference between the absolute ground actual α value and the predicted α' value of each pixel. The mean square error (MSE) is used as a loss function to approximate it:

$$L(\beta) = \frac{1}{N}\sum_{i}^{N}\left\|F(I_s^i, \beta) - \alpha_i\right\|, \tag{4}$$

where N is the total number of training samples in the batch, $F(I_s^i, \beta)$ is the shadow mask of the *ith* image pair output by the prediction layer through the training network, and α_i is the true shadow mask of the *ith* image pair.

Composition loss: The second loss was called the composition loss, which was the discrepancy between the true shadow image RGB color and the predicted shadow image RGB color. The predicted shadow image RGB color was synthesized by the shadow-free image and the shadow layer factor α' outputted in the prediction layer. The loss function approximation is as follows:

$$L(c) = \sqrt{(c' - c)^2 + \varepsilon^2}, \tag{5}$$

where c represents the RGB channel of the real shadow image, c' represents the RGB channel of the predicted shadow image synthesized by the predicted shadow mask factor α', ε represents a small value, and 10^{-5} is taken in this training. The derivative $\frac{\partial L}{\partial c'}$ is easy to calculate, its formula is defined as follows:

$$\frac{\partial L}{\partial c'} = \frac{c' - c}{\sqrt{(c' - c)^2 + \varepsilon^2}}. \tag{6}$$

Overall loss: It is a linear fusion of two losses, namely:

$$L_{overall} = \lambda L(\beta) + (1 - \lambda)L(c) \quad \lambda \in [0, 1]. \tag{7}$$

Here, λ was set to 0.5.

Training method: Although the performance increased significantly with the increase of network depth, it was an extraordinary task to train a deeper network due to the problem of vanishing gradient and exploding gradient. In this paper, the following three methods were used to achieve fast convergence and avoid over-fitting:

(1) Staged training. Since the RSnet network structure was a combination of two networks, each network was trained separately in the training process. When the network reached a certain precision, the two were combined and trained together, and finally the combined optimization of the two was realized.

(2) Multi-level training. According to the size of the shadow scale factor, different levels of images were set for training, for instance, training hard shadow image databases before soft shadow image databases, and combining the two to form a database for training in the last step. At the same time, considering the difference in pixel size of the user input image, images of different pixel sizes can also be divided into multiple levels for training. All of those methods are more robust in this paper, and help the network to learn context and semantics information better, thus, multi-level training results were finally achieved.

(3) Data augmentation. Hundreds of individual scene images cannot achieve good results for the purpose of training a deeper network. In order to diversify the training data and expand the number of data pairs in the shadow database, the training of multiple scenes images was carried out, and the database of the same scene was expanded by adopting the methods of cutting and rotating. All this had realized the diversification of the database.

Obtain shadow-free image: The RSnet model aimed to learn a mapping function between the shadow image and its shadow mask. When the RSnet model was trained to obtain the mapping function between the two, the shadow mask corresponding to the shadow image could be obtained through the trained RSnet model, and the corresponding shadow-free image could also be obtained according to the Formula (1).

In the Formula (1), I_s is the shadow image, I_{sf} is the shadow-free image, α is the shadow mask, and $\otimes$ is the product of the pixel level.

To be specific, the shadow image was given by the user input, and the shadow mask was obtained by the trained RSnet model. The shadow-free image can be obtained by dividing them at the pixel level according to Formula (1).

4. Experiment

Given the complexity of the image composition, it was rather difficult to obtain accurate evaluations through visual observation alone. Therefore, this paper used RSV to evaluate RSnet on different databases. Specifically, during the experiment course, the UCF shadow database [42], the LRSS database [16], the UIUC shadow database [13], and the RSDB database (a new remove shadow database) were used for testing, respectively, through the RSnet network. The restored image and the shadow-free image were compared to ensure the diversity and enrichment of the data. At the same time, structural similarity (SSIM) and peak signal-to-noise ratio (PSNR) were used as evaluation indicators to evaluate the performance of RSnet.

4.1. Database

Shadow removal has important research significance. Researchers at home and abroad have been extensively studying this issue, especially in the past decade, but a database specifically constructed for this problem is very rare. Currently, the most widely used databases are presented by Guo [13] and Gryka [16], but they only contain 76 and 37 pairs of shadow/shadow-free image pairs. In order to better evaluate the shadow removal algorithm proposed in this paper, we constructed a new RSDB database, which contained 2685 shadow and shadow-free image pairs. The comparison with other shadow removal databases in recent years is shown in Table 1.

Table 1. Comparison table with other shadow databases.

Database	Amount	Content of Images	Type
LRSS [16]	37	Shadow/Shadow-free	pair
UIUC [13]	76	Shadow/Shadow-free	pair
UCF [42]	245	Shadow/Shadow mask	pair
RSDB (ours)	2685	Shadow/Shadow-free	pair

In order to better collect data and build a new shadow removal database, we used a tripod to fix the camera in the specified position and a wireless Bluetooth remote control unit to capture images during the shooting process to ensure that the shadow/shadow-free image was consistent with the background. Firstly, to set a fixed exposure compensation, focal length, and other parameters, the exposure compensation was set to 0, and the focal length was 4 mm; then, the shadow image projected by different objects was taken. Finally, the shadow cast by the object was removed, that is, the projection of the object was removed, and the background image of the corresponding shadow image was captured, that is, the image without shadow.

In addition, to diversify the database, the new RSDB database had the following characteristics:

Brightness: In order to capture shadows, databases with different softness, that is, shadow databases with different light intensities, included soft shadows and hard shadows. We collected shadow images at the same time in different weather conditions and different time periods in the same conditions. For example, we can take shadow pictures at the same time (including 11 noon) on cloudy and sunny days, or we can choose 7:00 in the morning, 12 in the afternoon, and 6 in the afternoon at different times of the day. The data was collected to obtain a soft shadow and hard shadow database.

Shape: In order to ensure the shape diversity of the shadow database, during the shooting process the shadows were projected by various shapes of objects such as school bags, umbrellas, bicycles, etc. In the meantime, the characteristics of subjective initiative are fully utilized, and the body was projected as a shadow. The object performed a multi-pose change and then projected it into a specified area, thus obtaining a shadow database with various shapes.

Scenario: Multiple scenes were another requirement for shadow removal databases. Given this, we collected separate shadow images for different scenes in the process of shadow-collecting, such as grass, roads, campus, and walls.

Reflection: When we were shooting a shadow database, we reflect the shadows of different objects to mirror the existence of shadows in real life more realistically. An example of an RSDB database is shown in Figure 2.

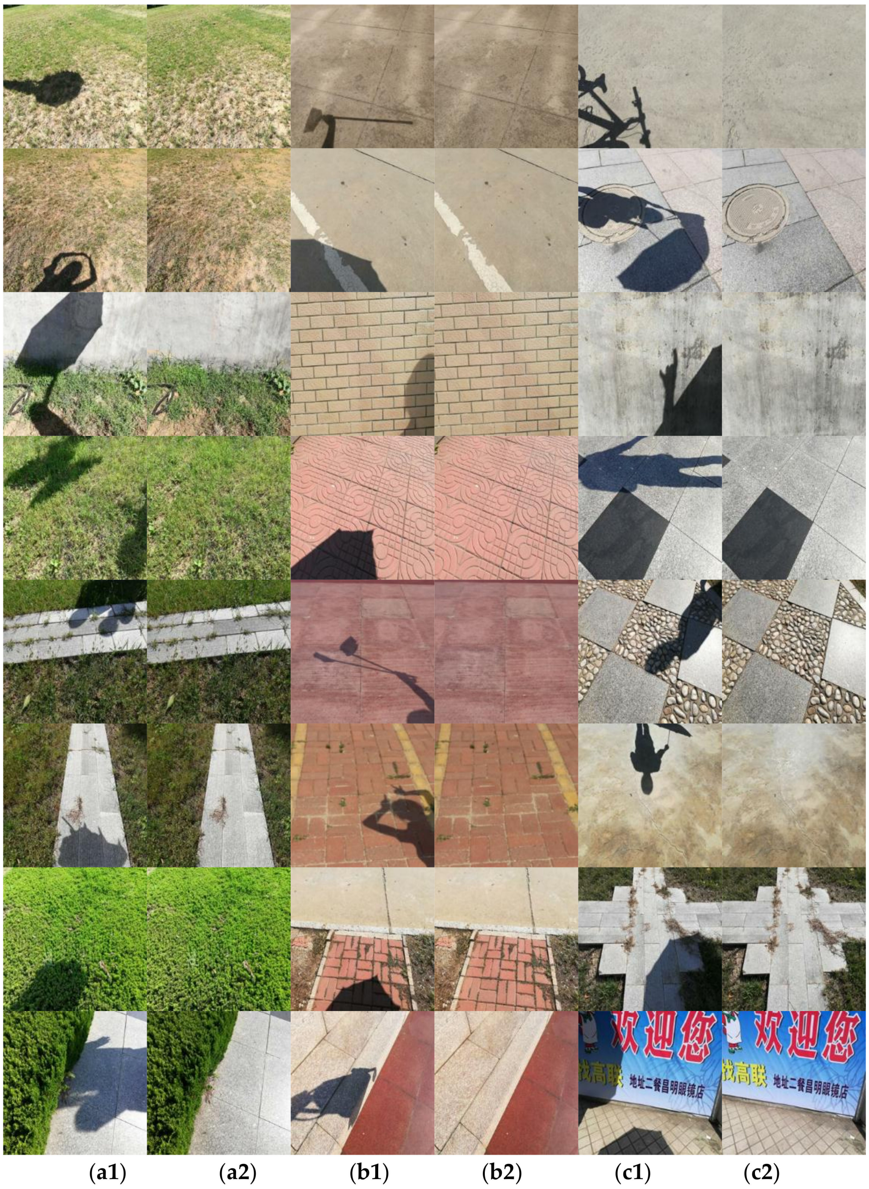

Figure 2. Illustration of several shadow and shadow-free image pairs in the RSDB database. The (**a1**, **b1**, **c1**) columns show the shadow images (inputs) in different scenes. The (**a2**, **b2**, **c2**) columns indicate true shadow-free images (targets) corresponding to the (**a1**, **b1**, **c1**) columns.

4.2. Performance Comparison Analysis

In order to evaluate the performance of the RSnet network, this paper chose to use the structural similarity SSIM (structural similarity) and the peak signal-to-noise ratio (PSNR) as evaluation indicators to assess the model. On this basis, the RSnet algorithm proposed in this paper was compared with several popular algorithms in recent years.

SSIM is an index to measure the similarity between two images. It is widely used to evaluate the quality of image processing algorithms. The larger the SSIM value of two images, the higher the similarity between them and the better the corresponding shadow removal effect in the model is. Conversely, the smaller the value of SSIM, the lower the similarity between the two, and the worse the removal effect of the corresponding algorithm is. PSNR is the most common and widely used objective measurement method for evaluating image quality. It is similar to SSIM. When the PSNR value of two images is larger, it indicates that the corresponding shadow removal model is better. Otherwise, the smaller the PSNR value, the worse the removal effect of the corresponding algorithm is. (Notably, the PSNR score is not exactly the same as the visual quality seen by the human eye. It is possible that a higher PSNR seems to be worse than a lower PSNR. It is because the sensitivity of the human eye to the error is not absolute, and the perceived result is affected by many factors.) The definitions of SSIM and PSNR are as follows:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \tag{8}$$

$$PSNR = 10 \times \lg(\frac{L \times L}{MSE}) \text{ here } MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2. \tag{9}$$

It is worth noting that in order to better reflect the removal effect of the shadow image, we chose shadow/output and target/output as the quality measurement properties of the algorithm, where shadow/output represented the corresponding evaluation index value between the original shadow image and the shadow removal obtained by the algorithm. target/output represented the actual ground shadow-free image (target) and the evaluation index value corresponding to the shadow removal image obtained by the algorithm. In order to evaluate the performance of the algorithm more fairly, we only showed the shadow/output values of the UCF database in different algorithms, and the target/output value is not given since the UCF database did not provide a true ground-free shadow (target).

It can be clearly seen from Table 2 that the SSIM index values obtained in the RSnet model were superior to the SSIM values of other algorithms, regardless of shadow/output or target/output. In particular, as can be seen from the last line of Table 2, in the newly created RSDB database, when the SSIM values of other algorithms were compared with the SSIM values obtained by the RSnet model of the algorithm, the SSIM value of the former was lower, the shadow removal effect of the image was poorer, and the value of the RSnet model was much higher than other algorithms. Particularly, in the evaluation of the real ground shadow-free image and the shadow removal image obtained by the algorithm, it can be seen from Table 2 that it was 1.09 times the SSIM value obtained by the shadow removal algorithm proposed by Yu [21], and 1.47 times the SSIM value obtained by the model proposed by Yang [28].

Of course, the comparison on the database proposed in this paper was not standardized. Therefore, Table 2 also shows the SSIM values on other shaded databases such as LRSS. By analyzing the SSIM values on other databases in Table 2, it can be found that RSnet exhibited better robustness and better image shadow removal. The processing effect was much better, especially in the third line UIUC database, which demonstrated a desired shadow removal effect. The SSIM value between the original shadow image and the algorithm-derived shadow removal image was larger than the value under the LRSS database. The overall effect of the SSIM value was smoother and more stable than other databases. From the perspective of the algorithm, the algorithms proposed by Wang [30] and

Qu [31] were both based on deep learning, and both of them had better performances compared with the other five algorithms, indicating that the shadow removal effect of them was better than other algorithms. Even for UIUC data sets, the SSIM value between the shadow removal image (output) and the real ground shadow-free image (target) obtained by the model proposed by Wang [30] was the best, which was even better than the model proposed in this paper. It can be concluded that the model proposed by Yang [28] had the worst overall performance, since it was easy to change the non-shaded area in the original image during the shadow removal process, which resulted in poor similarity between the two. The image shadow removal model proposed by Yu [21] and Gong [15] was relatively better than other algorithms, but it was inferior to the RSnet model proposed in this paper. The RSnet model had more stability, and the image shadow processing effect was better.

Table 2. Quantitative results using structural similarity (SSIM) (the bigger the better), the maximum value is shown in bold.

Database	Description	Fin. [22]	Yang [28]	Guo [13]	Gong [15]	Yu [21]	Wang [30]	Qu [31]	Ours
LRSS [16]	Shadow/ Output	0.7558	0.8336	0.8507	0.8953	0.9580	0.9162	0.9587	**0.9772**
	Target/ Output	0.8976	0.8549	0.9074	0.9496	0.9633	0.9780	0.9679	**0.9814**
UIUC [13]	Shadow/ Output	0.7904	0.8549	0.9495	0.9423	0.9543	0.9676	0.9641	**0.9788**
	Target/ Output	0.8336	0.8889	0.9226	0.9576	0.9773	**0.9868**	0.9801	0.9848
UCF [42]	Shadow/ Output	0.8525	0.8651	0.8625	0.8931	0.9487	0.9121	0.9556	**0.9832**
	Target/ Output	–	–	–	–	–	–	–	–
RSDB (ours)	Shadow/ Output	0.8580	0.6904	0.7875	0.8609	0.8591	0.8953	0.9011	**0.9826**
	Target/ Output	0.8124	0.6725	0.7704	0.9122	0.9063	0.9344	0.9532	**0.9910**

The PSNR value in Table 3 showed a large difference between the shadow/output and target/output values when comparing the shadow/output (target/output) values of different algorithms under the same database or different databases of the same algorithm. We uncovered some valuable clues, which are needed to evaluate the algorithm. Comparing the PSNR values of different algorithms in the same shaded database in Table 3, it can be found that the value of the RSnet model was higher than other algorithms, especially in the RSDB database proposed in this paper, and the PSNR value was higher. The effect of the shadow removal model proposed by Yu [21] was similar to that of Gong [15]; the model PSNR values proposed by Yang [28] were relatively inferior, and so was the shadow removal performance. Qu [31] proposed an algorithm that had good performance, second only to the model proposed in this paper. As can be seen from the fourth line in Table 3, Qu's [31] proposed algorithm was superior to the model proposed in this paper, and the PSNR value of the model proposed by Wang [30] was slightly lower than that of the RSnet model. The three methods based on deep learning were all superior to the other five algorithms, which showed that the deep learning method achieved good results in the image shadow removal task. Comparing the PSNR values of the same algorithm under different databases, it can be clearly found that the proposed algorithm was different from other algorithms, and the PSNR values were higher under different databases. Therefore, the RSnet model was more stable, the image shadow removal effect was much better, and the algorithm was more robust. Comparing the PSNR values of the shadow removal model proposed by Guo [13] under a certain attribute and different databases, for example, the PSNR value on the shadow/output can be found, and the variation range of the PSNR was larger, indicating that the corresponding algorithm was not stable and the robustness was poor. At the same time, it can be concluded from Table 3 that the model proposed by Finlayson [22] had a smaller PSNR value under the target/output attribute of the RSDB database. Despite that it was higher than the model proposed by Yang [28], it was still 1.27 times lower than the value obtained by RSnet proposed in this paper.

Table 3. Quantitative results using peak signal-to-noise ratio (PSNR) (the bigger the better), the maximum value is shown in bold.

Database	Description	Fin. [22]	Yang [28]	Guo [13]	Gong [15]	Yu [21]	Wang [30]	Qu [31]	Ours
LRSS [16]	Shadow/ Output	12.9899	15.5075	17.0192	18.8039	19.3035	20.3452	19.8817	**23.0102**
	Target/ Output	29.3137	29.6752	29.9656	30.0192	31.6604	31.1917	31.6946	**33.6863**
UIUC [13]	Shadow/ Output	17.6398	19.1518	22.6825	22.9292	21.2515	23.1865	23.9611	**24.4204**
	Target/ Output	29.6504	29.2484	32.2745	34.4854	34.0058	34.4204	35.6052	**35.9611**
UCF [42]	Shadow/ Output	16.6190	15.3573	19.5558	20.7869	20.7128	23.1614	**25.9801**	25.1759
	Target/ Output	–	–	–	–	–	–	–	–
RSDB (ours)	Shadow/ Output	16.8726	12.1715	19.1940	20.6937	22.3105	22.9801	23.7084	**25.7078**
	Target/ Output	28.5124	24.2104	29.5787	31.5248	33.1614	32.4272	32.6855	**36.1148**

4.3. Qualitative Analysis

In order to demonstrate the effect of RSnet network shadow removal properly, we selected part of the image in the database as the original image, and compared the processing results with the shadow removal effects of several other algorithms. The comparison diagram is shown in Figures 3–5.

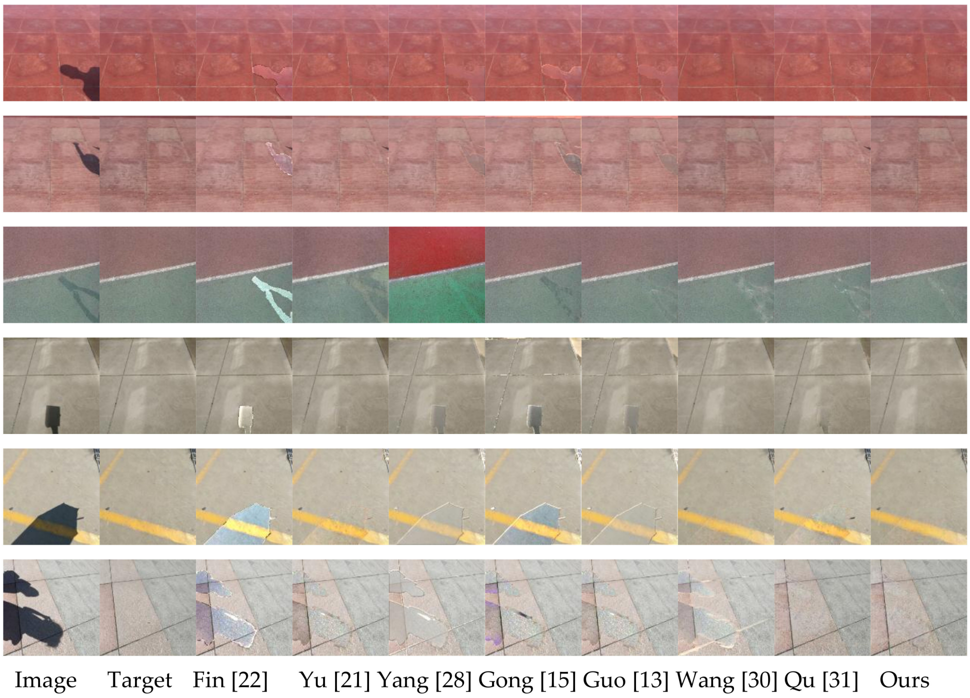

Figure 3. Schematic diagram of shadow removal in simple scenes.

Figures 3–5 show the image shadow removal effects of different algorithms under simple semantics, complex semantics, and special semantics. The first column is the original image (shadow image), the second column is the real shadow-free image, and the last eight columns are the shadow removal images obtained by eight different algorithms. The last column is the shadow removal image obtained by the RSnet model proposed in this paper. It is worth noting that the first few lines of Figures 3 and 4 show the image shadow removal effect when the projection object is smaller and more trivial, and the last few lines show the shadow removal effect when the projection object is larger. Comparing the processing effects of these two different projection objects, it can be clearly found that when the projection object was larger and the shadow area was more concentrated, that is, the shadow occupied a large proportion of the entire image, the overall removal effect of the image shadow was much better than that of the trivial projection object, and had a smaller shadow area.

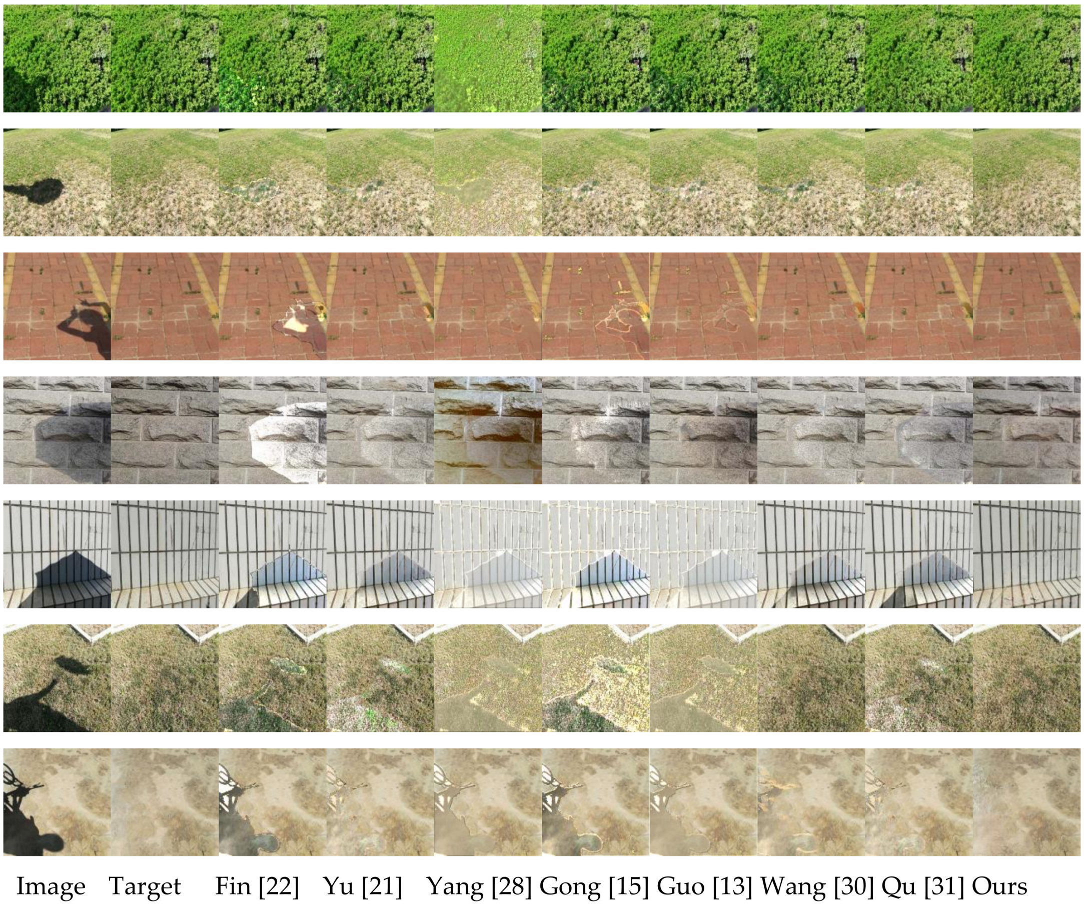

Figure 4. Schematic diagram of shadow removal in complex scenes.

Analyzing the image shadow removal effect diagram in Figure 3, we found that Finlayson [22] and Yu [21] had relatively stable results for shadow images in simple scenes; the former was better than the latter. The algorithm proposed by Yu [21] had a better overall effect than the other four methods without deep learning, however, when comparing the shadow removal effect map obtained by the fully automatic RSnet model proposed in this paper, it can be seen that Rsnet's shadow-free image quality was higher, especially for the detail processing, which was far better than the interactive method proposed by Yu [21]. Observing the last three columns in Figures 3–5, it can be found that the depth learning-based methods proposed by Wang [30] and Qu [31] and the Rsnet algorithm proposed in this paper were better than the previous ones. It reflected that the overall effect of deep learning applied to image shadow removal was better. It can be clearly seen from the fifth column in Figures 3–5 that the bilateral filtering-based algorithm proposed by Yang [28] changed the non-shaded area. Of course, for the first and second rows of the fifth column in Figure 3, the image reconstruction effect was good, the reconstruction effect of the non-shadow region was good, and the non-shadow region was not significantly changed. However, the effect of shadow removal was not good enough, and the shadow area could still be observed more clearly from the shadow-free image after removal. At the same time, using these two images to compare the shadow-free image obtained by the Rsnet model proposed in this paper, it can be seen that the latter removal effect was better in both the non-shadow area and the shadow area, and the obtained shadow-free image quality was higher. Of course, for the shaded areas in the third and last lines, the algorithm proposed in this paper did not perform preferably on the removal of shadow areas, but still had advantage over other algorithms, especially for the detailed processing of edge information.

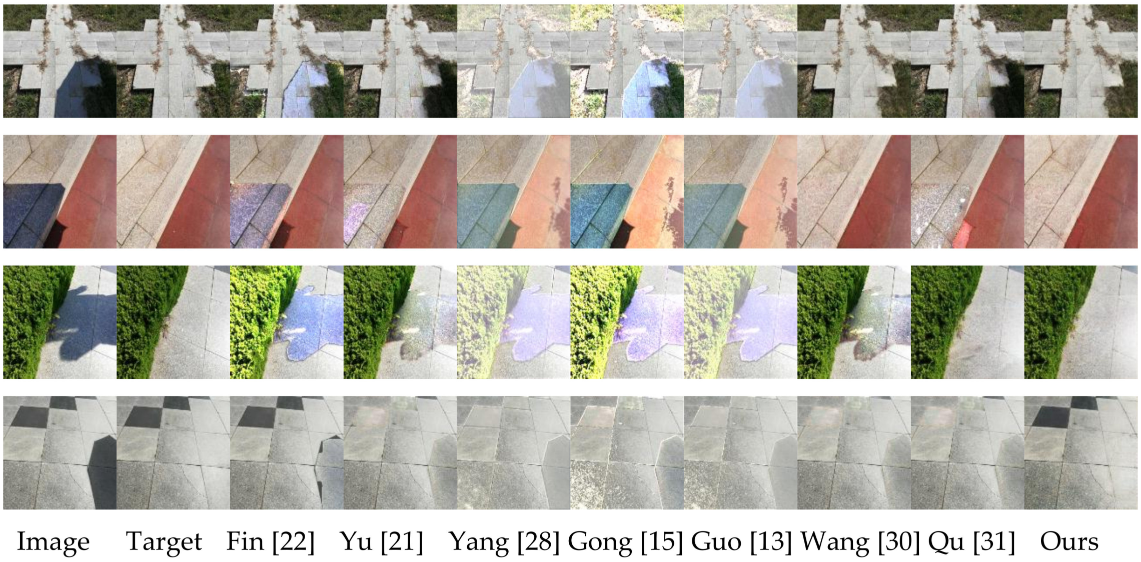

Figure 5. Schematic diagram of shadow removal in special scenes.

Figure 4 mainly shows the shadow removal effect of different algorithms under complex scene conditions. A comprehensive analysis of Figures 3 and 4 shows that some algorithms had better shadow removal effects in simple scenes, but the shadow removal effect in complex scenes was relatively poor. These algorithms of the overall analysis and comparison show that the quality of the shadow-free image obtained in the simple scene was higher than that of the shadow-free image obtained in the complex scene. Therefore, it can be explained that these algorithms had poor stability and could not adapt to shadow images with multiple semantic inputs at the same time. When the texture was complex and the shadows were small, these algorithms could not obtain high-quality, shadow-free images. Of course, the two methods based on deep learning were more stable than the other five algorithms. However, when comparing the RSnet model, it can be found that Wang [30] had a poor effect on small shadows, and Qu [31] had a poor shadow removal effect on complex scenes, such as grass. The Rsnet model had excellent shadow removal effects for both scenarios, and both had high quality, shadow-free images. Analysis of lines 4 and 5 in Figure 4 revealed that when shadows were projected onto a wall with complex textures, Yu [21], Gong [15], Guo [13], Yang [28], and Finlayson [22] did not get high-quality, shadow-free images, and the shadow removal effect was not ideal. The shadow-free image quality obtained by the Rsnet network in this case was relatively higher, and the shadow removal effect was better, even equivalent to the true shadow-free image in the second column.

The last line in Figure 4 shows the shadow removal effect of different algorithms when people and bicycles were used as projection objects. This was a common shadow image in our daily life, but the shadow removal methods of the other models were not satisfactory. In particular, the more complicated and fine shadow processing effects on the bicycle tire was extremely poor, and the corresponding shadow area was not substantially removed. The Rsnet model proposed in this paper obtained a high-quality, shadow-free image using a fully automated end-to-end convolutional neural network. Moreover, when the obtained shadow-free image was compared with the real shadow-free image, there was basically no difference between the two, and the shadow removal effect was excellent.

Figure 5 illustrates the effect of image shadow removal in some special cases. From the first three lines, it can be found that when the shadow area involved multiple semantic inputs, the other five shadow removal effects were extremely inferior. The model removal effect proposed in this paper was quite satisfactory. In particular, in the second row of Figure 5, which had stairs and involved two different colors, the Yang [28], Gong [15], and Guo [13] algorithms basically could not handle the shadow area, and they even changed the non-shadow area. From the last line in Figure 5, it can be found that when the real ground had a black texture, most of the shadow removal algorithms would mistake it for the shadow area to be removed, and the methods proposed by Wang [30] and Qu [31] presented the same situation. However, the RSnet model proposed in this paper could distinguish between shadow regions and non-shadow regions more accurately, and, thus, obtain higher quality, shadow-free images.

5. Summary and Outlook

Based on the analysis of the background, significance, and current research status of image shadow removal, this paper focuses on the shortcomings of existing image shadow removal algorithms. A fully automatic image shadow removal model based on an end-to-end deep convolutional neural network is proposed. This model uses a small-scale network to refine the processing that is based on the encode–decoder network to obtain finer semantic information, which enhances the quality of shadow-free images and better process the details. In the RSnet model, a more diverse training method is adopted in the training process, which performs the network training effect better and, thus, is applied to image shadow removal in a finer way. At the same time, a new image shadow removal RSDB database is provided, which offers more data types for future shadow removal. It can be seen from the comparison results that the image shading is removed by using this algorithm, the performance index is better, and the quality of the shadow-free image is higher, which has great application values in the later target recognition and target tracking stages. Of course, there are still some limitations with our method. In the following work, we would continue to optimize this method. The database size of the image shadow can be further expanded; the network expression ability can be improved, the convergence speed can be accelerated, and, thus, a more robust algorithm may be obtained.

Author Contributions: H.F. participated in the writing of the manuscript and assisted in the analysis of the results and in the performance of the experiments. M.H. did the setup of the experiments, the analysis of the results, and the writing of the manuscript. J.L. did the setup of the experiments, the definition, and development of the experiments.

Funding: This research was funded by National Natural Science Foundation of China (NSFC), grant number 61472227, 61772319, 61602277.

Acknowledgments: The authors would like to thank the associate editor and anonymous reviewers for their comments to improve the paper.

Conflicts of Interest: The authors declare that there is no conflict of interest regarding the publication of this paper.

References

1. Wang, J.M.; Chung, Y.C.; Chang, C.L.; Chen, S.W. Shadow detection and removal for traffic images. In Proceedings of the IEEE International Conference on Networking, Sensing and Control, Taipei, Taiwan, 21–23 March 2004.
2. Salvador, E.; Cavallaro, A.; Ebrahimi, T. Shadow identification and classification using invariant color models. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Salt Lake City, UT, USA, 7–11 May 2001.

3. Krishnan, A.; Jayadevan, P.; Panicker, J.V. Shadow removal from single image using color invariant method. In Proceedings of the 2017 International Conference on Communication and Signal Processing, Chennai, India, 6–8 April 2017.
4. Su, N.; Zhang, Y.; Tian, S.; Yan, Y.; Miao, X. Shadow detection and removal for occluded object information recovery in urban high-resolution panchromatic satellite images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 2568–2582. [CrossRef]
5. Ngo, T.T.; Collet, C.; Mazet, V. Automatic rectangular building detection from VHR aerial imagery using shadow and image segmentation. In Proceedings of the 2015 IEEE International Conference on Image Processing, Quebec City, QC, Canada, 27–30 September 2015; pp. 1483–1487.
6. Huang, S.; Huang, W.; Zhang, T. A New SAR image segmentation algorithm for the detection of target and shadow regions. *Sci. Rep.* **2016**, *6*, 38596. [CrossRef] [PubMed]
7. Yan, T.; Hu, S.; Su, X.; He, X. Moving object detection and shadow removal in video surveillance. In Proceedings of the 2016 International Conference on Software, Knowledge, Information Management and Applications, Chengdu, China, 15–17 December 2016; pp. 3–8.
8. Sanin, A.; Sanderson, C.; Lovell, B.C. Improved shadow removal for robust person tracking in surveillance scenarios. In Proceedings of the International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 141–144.
9. Arbel, E.; Helor, H. Shadow removal using intensity surfaces and texture anchor points. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 1202–1216. [CrossRef] [PubMed]
10. Liu, F.; Gleicher, M. Texture-Consistent Shadow Removal. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 437–450.
11. Wu, T.P.; Tang, C.K.; Brown, M.S.; Shum, H.Y. Natural shadow matting. *ACM Trans. Graph.* **2007**, *26*, 8. [CrossRef]
12. Das, S.; Aery, A. A review: Shadow detection and shadow, removal from images. *Int. J. Eng. Trends Technol.* **2013**, *4*, 1764–1767.
13. Guo, R.; Dai, Q.; Hoiem, D. Paired regions for shadow detection and removal. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2956–2967. [CrossRef] [PubMed]
14. Khan, S.H.; Bennamoun, M.; Sohel, F.; Togneri, R. Automatic shadow detection and removal from a single image. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 431–446. [CrossRef] [PubMed]
15. Gong, H.; Cosker, D. Interactive removal and ground truth for difficult shadow scenes. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **2016**, *33*, 1798–1811. [CrossRef] [PubMed]
16. Gryka, M.; Terry, M.; Brostow, G.J. Learning to remove soft shadows. *ACM Trans. Graph. (TOG)* **2015**, *34*, 153. [CrossRef]
17. Tian, J.; Qi, X.; Qu, L.; Tang, Y. New spectrum ratio properties and features for shadow detection. *Pattern Recognit.* **2016**, *51*, 85–96. [CrossRef]
18. Shen, L.; Chua, T.W.; Leman, K. Shadow optimization from structured deep edge detection. In Proceedings of the 28th IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 2067–2074.
19. Hosseinzadeh, S.; Shakeri, M.; Zhang, H. Fast shadow detection from a single image using a patched convolutional neural network. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018.
20. Calvo-Zaragoza, J.; Gallego, A.-J. A selectional auto-encoder approach for document image binarization. *Pattern Recognit.* **2019**, *86*, 37–47. [CrossRef]
21. Yu, X.; Li, G.; Ying, Z.; Guo, X. A new shadow removal method using color-lines. In Proceedings of the CAIP 2017: Computer Analysis of Images and patterns, Ystad, Sweden, 22–24 August 2017; pp. 307–309.
22. Finlayson, G.D.; Hordley, S.D.; Lu, C.; Drew, M.S. On the removal of shadows from images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 59–68. [CrossRef] [PubMed]
23. Levin, A.; Lischinski, D.; Weiss, Y. A closed-form solution to natural image matting. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *30*, 228–242. [CrossRef] [PubMed]
24. Xiao, C.; Xiao, D.; Zhang, L.; Chen, L. Efficient shadow removal using subregion matching illumination transfer. *Comput. Graph. Forum* **2013**, *32*, 421–430. [CrossRef]
25. Zheng, C.; Sun, Z.L.; Wang, N.; Bao, X.Y. Moving Cast Shadow Removal Based on Texture Feature and Color Space. In *International Symposium on Neural Networks*; Springer: Cham, Switzerland, 2018; pp. 611–618.

26. Murali, S.; Govindan, V.K. Shadow detection and removal from a single image using LAB color space. *Cybern. Inf. Technol.* **2013**, *13*, 95–103. [CrossRef]
27. Tfy, V.; Hoai, M.; Samaras, D. Leave-one-out kernel optimization for shadow detection and removal. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 682–695.
28. Yang, Q.; Tan, K.H.; Ahuja, N. Shadow removal using bilateral filtering. *IEEE Trans. Image Process. A Publ. Ieee Signal Process. Soc.* **2012**, *21*, 4361–4368. [CrossRef] [PubMed]
29. Barron, J.T.; Malik, J. Shape, Illumination, and reflectance from shading. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1670–1687. [CrossRef] [PubMed]
30. Wang, J.; Li, X.; Hui, L.; Yang, J. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
31. Qu, L.; Tian, J.; He, S.; Tang, Y.; Lau, R.W. DeshadowNet: A multi-context embedding deep network for shadow removal. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
32. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A deep convolutional encoder-decoder architecture for scene segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef] [PubMed]
33. Yang, J.; Price, B.; Cohen, S.; Lee, H.; Yang, M.H. Object contour detection with a fully convolutional encoder-decoder network. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 193–202.
34. Pathak, D.; Krahenbuhl, P.; Donahue, J.; Darrell, T.; Efros, A.A. Context encoders: Feature learning by inpainting. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2536–2544.
35. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *Comput. Sci.* **2014**.
36. Eigen, D.; Puhrsch, C.; Fergus, R. Depth map prediction from a single image using a multi-scale deep network. In Proceedings of the 2014 International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 2366–2374.
37. Gatys, L.A.; Ecker, A.S.; Bethge, M. Image style transfer using convolutional neural networks. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2414–2423.
38. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651.
39. Zeiler, M.D.; Krishnan, D.; Taylor, G.W.; Fergus, R. Deconvolutional networks. *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2010**, *238*, 2528–2535.
40. Hara, K.; Saito, D.; Shouno, H. Analysis of function of rectified linear unit used in deep learning. In Proceedings of the 2015 International Joint Conference on Neural Networks, Killarney, Ireland, 12–17 July 2015; pp. 1–8.
41. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-Level performance on ImageNet classification. In Proceedings of the 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 11–18 December 2015; pp. 1026–1034.
42. Zhu, J.; Samuel, K.G.G.; Masood, S.Z.; Tappen, M.F. Learning to recognize shadows in monochromatic natural images. In Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 223–230.