Hindawi
Scientific Programming
Volume 2022, Article ID 7494108, 11 pages
https://doi.org/10.1155/2022/7494108

Research Article

Multiprojector Interreflection Compensation Using a Deep Convolution Network

Xiaofeng Li,[1,2] Zhihui Hu,[1,2] Haiyan Gu,[2] Ting Yang,[1,2] Qinghua Lei,[1,2] Hu Chen,[1,2] Peng Cheng,[2,3] and Zhisheng You[1,2]

[1]*School of Computer Science, Sichuan University, Chengdu 610065, China*
[2]*Sichuan University National Key Laboratory of Fundamental Science on Synthetic Vision, Chengdu 610065, China*
[3]*School of Aeronautics and Astronautics, Sichuan University, Chengdu 610065, China*

Correspondence should be addressed to Haiyan Gu; 2019226049045@stu.scu.edu.cn

Received 7 July 2021; Revised 23 September 2021; Accepted 30 September 2021; Published 23 February 2022

Academic Editor: Jianping Gou

The aim of multiprojector interreflection compensation is to modify input images to remove complex physical stray-light effects (interreflection) from a multiprojector immersive system. This is an important but often ignored problem, which can lead to degradation of a projection image. Traditional methods usually address this problem by computing a matrix inversion. These traditional methods often ignore issue of the clarity of the generated images. In this paper, we describe a method for learning the inversion using a deep convolutional neural network (CNN), named Superresolution Compensation Net (SRCN). SRCN consists of four convolution layers to learn interactions of global light, six convolution layers, and two transposed convolution layers to extract multilevel features and generate compensation images. We also used a subpixel convolution layer to increase the resolution. To make compensation images more consistent with human visual perception, we used a perceptual loss, which compares the differences between feature maps on the VGG16 network. We implemented an immersive projector-camera display prototype (Pro-Cam) and calculated the quality index of the compensation images and the projection results. Our method achieved better results than previous methods in both objective evaluations and subjective visual perception.

1. Introduction

Multiprojector systems are used in virtual reality (VR) systems, exhibitions, and tower simulators. For these applications, a good projection effect is crucial, since it can bring people an immersive visual experience, producing a highly realistic effect. However, the imaging effect of multiprojector systems is usually not ideal due to multiple factors, such as generation of points with noise and interreflection. Approaches to the multiprojector noise problem are relatively mature, and some researchers [1, 2] have proposed techniques to solve the noise problem, but the interreflection problem is very common and easily ignored, and there is still considerable scope for the development of effective solutions. When interreflection is serious, such as when there are too many projectors, folding projection surfaces, or curved projection screens, the display image mixed by the light from the projector and the interference light of superimposed reflections leads to poor display image quality. The contrast of the projection display images is low, which disturbs user immersion and becomes an important factor hampering the popularization, application, and development of these systems. For example, this phenomenon has already led to the failure of an aerospace industry tower simulator to be put into practical teaching in a university. Internal reflection is a serious problem that prevents multiprojection systems from meeting application requirements and therefore needs to be solved. Methods for multiprojector reflection compensation can help multiprojector systems produce an optimal immersive visual experience. Conventional methods require considerable optical knowledge, and they are often ineffective in dealing with the interreflection problem in large, complex systems. Therefore, we aimed to minimize the heavy reliance on optical knowledge and explore the use of deep learning methods to solve the interreflection problem in sophisticated environments.

Generally, a light transport matrix (LTM) is used to describe the multiplication of the projected incident light and the reflection from the immersive scene [3–8], and the interreflection compensation is regarded as a matrix inversion problem. Inverting the LTM, we can calculate the compensated images, so when they are reprojected, the interreflection will be eliminated. The acquisition of an LTM, however, is a laborious process. The available methods require the devices to be radiometrically calibrated and carefully set up. The LTM, which is determined by the resolution of the projector-camera system (Pro-Cam), is very large, so it is very difficult to calculate the inversion. To obtain an LTM suitable for matrix inversion, methods are used to downsize or simplify the matrix. The performance of these methods is limited, so they are not practical for use in a huge immersive environment.

In the image processing field, the inverse problem is common. Assuming that an observed image y represents the output of model T, and x is the input of T, then given output y, calculating the input x is the inverse problem [9]. In recent years, convolutional neural networks (CNNs) have become a popular way to solve the inverse problem [10–15] in problems such as dehazing, style transfer, and image superresolution reconstruction. They have shown outstanding performance on large databases of images.

Interreflection compensation is also an inverse problem. However, in the last decade, few researchers have solved the problem using CNNs. We realized that this was a possible approach to reduce the undesired interreflection without LTM inversion. Compared with traditional methods, CNN does not require a large amount of knowledge about optics, such as the radiometric precalibration of the Pro-Cam. The method generates a visually optimal compensation image even with slight misalignment.

In this study, we innovatively used a deep learning-based image processing method to solve the problem of interreflection in complex immersive projection systems, built a projector-camera display prototype (Figure 1), and developed a novel neural network, named Superresolution Compensation Net (SRCN), for multiprojector interreflection compensation, to improve the projection performance, as shown in Figure 2. First, we used a geometric correction subnet to autocorrect the sampling images captured by the camera. Second, by connecting with several convolution layers, SRCN could be trained to perform matrix inversion to modify the input images. We used a superresolution layer (hereinafter referred to as SR layer) as proposed by Habe et al. [3] to double the resolution of the output images and used a loss calculated from the differences in the outputs of the max-pooling layer of the VGG network to improve the perceptual quality [16], Finally, we evaluated the proposed network model and evaluated its performance compared to that of conventional methods.

Our main contributions are as follows:

(1) We removed multiprojector interreflection using a learning process, greatly improving multiprojector system imaging and simplifying the process of obtaining an LTM and calculating its inverse

(2) We utilized SR compensation to further improve the definition of compensated images

(3) We used a perceptual loss with coefficients in addition to pixelwise loss [17], so that the compensated images are more invariant to changes in pixel space [18, 19]

(4) We created a dataset in our Pro-Cam environment and made the dataset public

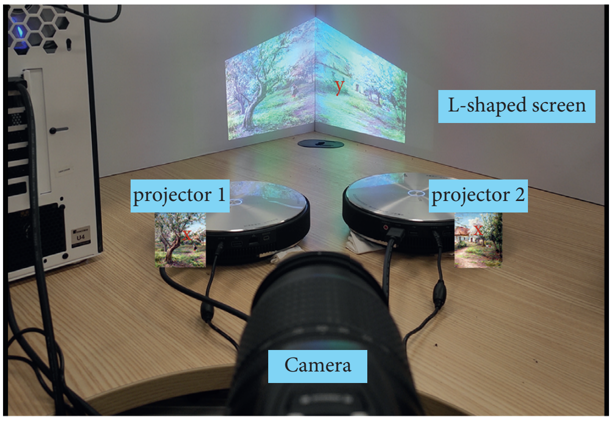


Figure 1: Our projector-camera (Pro-Cam) display prototype.

The rest of this paper is organized as follows. Section 2 reviews and discusses the relevant research. In Section 3, the multiprojector interreflection model-based deep learning is described. The experimental results are compared with other methods, and a comparison of the self-control method is introduced in Section 4, and Section 5 presents the conclusions.

2. Related Work

Two research fields—interreflection compensation and convolutional neural networks—are closely related to our proposed method. We introduce the related fields and discuss the development of our approach in this section.

2.1. Interreflection Compensation. Some techniques have been proposed to remove interreflection by modifying the uncompensated projection images. The method was initially presented by Bimber et al. [4], who divided the uncompensated projection image into small patches, and based on the Jacobi iteration, each patch was compensated offline to eliminate the scattered light. Similarly, Bai et al. [7] compute the compensation iteratively. These iterative methods do not require direct LTM inversion so it is effective to solve only a single image compensation. However, for multiple images, the compensated projector images need to be separately and iteratively computed for each input.

Other methods have been proposed to remove interreflection by precorrecting input images. This is a popular research method for the reduction of undesired lights on various types of surfaces. However, such approaches usually

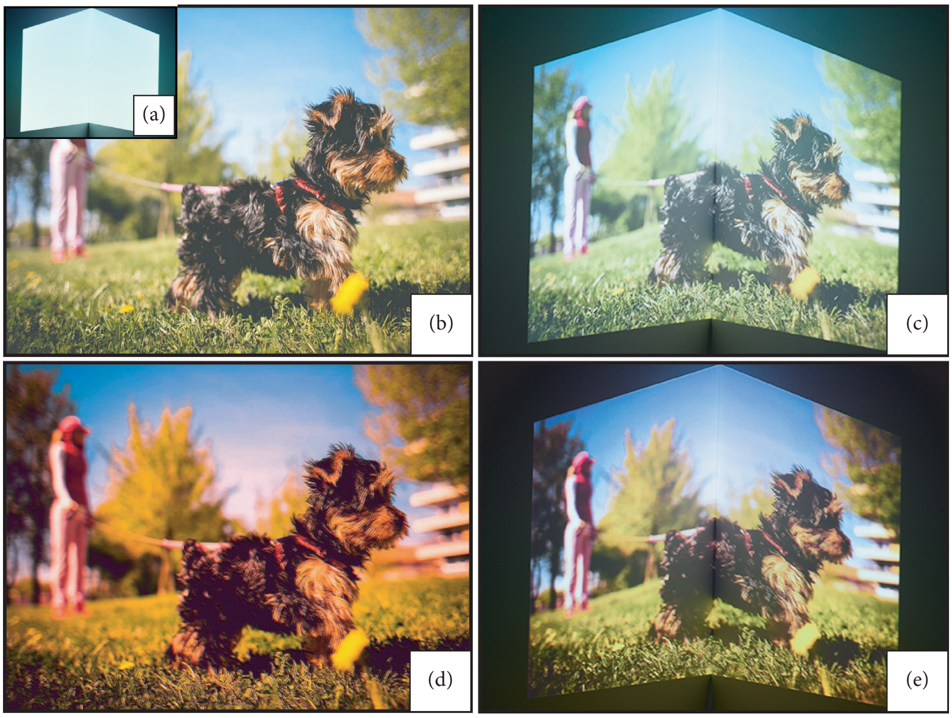


Figure 2: Interreflection compensation for an L-shaped projecting screen using SRCN. (a) White projection surface $\hat{s}$. (b) Input image (x). (c) The uncompensated projection result $y(x)$ taken by the camera when b is projected on screen. (d) Compensation image x^* using our SRCN. (e) The compensation projection result $y(x^*)$ taken by the camera when d is projected on screen. Comparing (c) and (e), we can conclude (e) is closer to (b) with higher quality and fewer interreflections.

need to find the inverse of the LTM. Compared with iterative computation, this approach precomputes the matrix inversion only once, and then any desired images can be compensated using matrix-vector multiplication. This approach was taken by Habe et al., Wetzstein et al., Ding et al., and Grundhöfer et al. A difficulty with this approach is the enormous size of the LTM, which is decided by the resolution of the Pro-Cam system. For example, a camera may have 3072 × 2304 pixels and the projector 1024 × 768. Then, the size of the LTM is 3072 × 2304 by 1024 × 768. To make the LTM inversion feasible, one approach is to downsize the matrix. Habe et al. [3] simplify the LTM by downsizing it to 64 × 64 pixels, Wetzstein and Bimber [6] describe a customized clustering scheme to approximate the inverse of the LTM. They traverse all the pixels in the projection image and find a pixel of the highest luminance contribution as the center of each patch in turn, and add some adjacent pixels into the patch. These clusters were computed for many small patches. Another approach is to simplify the matrix without changing the size of the LTM. Ding et al. [5] took pairs of white images in their Pro-Cam to acquire the LTM. Thereafter, they used the LTM to construct a matrix to approximate inversion. Recently, Grundhöfer and Iwai [8] proposed an interpolation based on TPS used to calculate an accurate color transformation.

2.2. Convolutional Neural Networks. Recently, CNNs have attracted considerable attention. They imitate the visual perception mechanism of biology and are widely used to solve the inverse problem in image processing. They can reach a stable effect, and there are no additional feature engineering requirements for the data [20, 21]. We review the network architecture and loss function in this part.

In these inverse problems, dehazing provides the main trends in most papers. One important perspective on these dehazing results is that the CNN is learning a mapping between a hazy image and a clear image [12, 13]. To improve the image quality and increase the resolution, Ledig et al. [22] propose a four-time superresolution (SR) architecture constructed by connecting two two-time superresolution trained subpixel convolution layers. In [11], the authors propose an encoder-decoder structure named UNet, which keeps the size of feature maps unchanged. In the encoder, it can extract features faster by increasing the number of feature channels. The decoder gradually recovers the edge information of the images. Then, Zhang et al. [18] designed a UNet-like backbone network named CompenNet to remove the projected texture background. Both input and output images were 256 × 256 × 3. In CompenNet, the researchers innovatively added an encoder, which is the same as the encoder of the backbone network, to learn the global light. Each layer of the two encoders is connected by elementwise adding. Later, they proposed CompenNet++, which concatenates a geometric correction subnet with CompenNet, to realize both geometric correction and projection texture removal.

The choice of the loss function generally defaults to pixelwise approaches such as l_1 loss and l_2 loss (MSE). However, some limitations are apparent when using these pixelwise loss functions in image processing. For example, although the test index is high, the image quality is not necessarily good, because human visual perception is not taken into account [18], so a structural similarity index (SSIM) was proposed in [17]. SSIM uses luminance, contrast, and structure to evaluate images, to more closely match the human visual system (HVS). Generally, the results using SSIM are more detailed than those using pixelwise loss. In [19] the authors investigated different loss functions and found that using different losses in combination can obtain better results than using only one loss. Recently, a perceptual loss was proposed in [16], achieved by comparing the loss in a feature map. This approach was extended in SRGAN [22] to enhance the visual quality.

Our method addresses the problem of removing interreflection by precorrecting input images. Inspired by the research into the inverse problem using CNNs, we propose a novel neural network for multiprojector interreflection compensation, instead of computing the LTM inversion. The network can learn the mapping between input images and compensated images, which means that it can learn complex spectral interactions and generate a modified input image. To improve the visual quality, we used SSIM and perceptual loss as well as pixelwise loss.

3. Proposed Method

3.1. Problem Formulation. The purpose of interreflection compensation is to map the input image onto a compensated image. When the image is projected again, the interreflection is reduced or even eliminated. Our research focused on finding an LTM inversion to realize the mapping between the input image and compensated image.

Assuming x is an input image, f_p is the optical transfer function of two projectors, s and f_s are the surface reflectance property and surface bidirectional reflectance distribution function (BRDF), respectively, E is the global illumination, and f_c is the camera's composite capturing function, then we can formulate the camera-captured image y as

$$y = f_c\left(f_s\left(f_p(x), E, s\right)\right). \tag{1}$$

For simplicity, we can regard f_p, f_s, f_c as T, which is actually interreflection light transport mapping between the projected image and camera-captured image. Thus, equation (1) can be reformulated as

$$y = T(x, E, s). \tag{2}$$

However, the global illumination E and surface reflectance s are hard to measure without additional spectral devices. Because the multiprojection display prototype is fixed, we can use a camera-captured surface image, $\hat{s}$, to approximate global interactions:

$$\hat{s} = T(x_0, E, s), \tag{3}$$

where x_0 is a pure white image whose grayscale is 255. Thus, we can substitute E and s with $\hat{s}$ in equation (2):

$$y = T(x, \hat{s}). \tag{4}$$

Interreflection compensation aims to find a compensated image x^* so that the camera-captured result is the same as the original input image x (ground truth):

$$x = T\left(x^*, \hat{s}\right). \tag{5}$$

Multiprojector interreflection compensation can be formulated as follows:

$$x^* = T^{-1}(x, \hat{s}), \tag{6}$$

where T^{-1} is the T inversion. We used a deep neural network to model it.

3.2. Deep Learning-Based Formulation. From equation (4) we can get

$$x = T^{-1}(y, \hat{s}). \tag{7}$$

Because $\hat{s}$ is known, we can use sampled image pairs (x, y) to learn T^{-1}. We model T^{-1} using a deep convolutional neural network named SRCN and denote it as $T(\theta)^{-1}$, where θ is the learning parameter.

$$y^* = T(\theta)^{-1}(y, \hat{s}), \tag{8}$$

where y^* is the compensated image of camera-captured image y. Because the projection system is not a plane, y is out of shape, as shown in Figure 1. Generally, y requires manual geometric correction. In this paper, we add a geometric correction subnet G to realize the process automatically. We designed G inspired by [23], which uses a cascaded coarse-to-fine structure to generate a sampling grid Ω, and camera-captured images y can be corrected geometrically using a single bilinear interpolation $\oplus$.

We train $T(\theta)^{-1}$ with N sets of image pairs $\{(x_i, y_i)\}_{i=1}^{N}$. We want y^* to be as close to ground truth x as possible. So, using a loss function L, SRCN can converge by learning as follows:

$$\theta = \arg\min \sum_{i=1}^{N'} L\left(y_i^*, x_i\right). \tag{9}$$

3.3. Network Architecture. The architecture of SRCN is shown in Figure 3. We used a UNet-like [9] backbone network with several convolution layers to extract features. $\hat{s}$

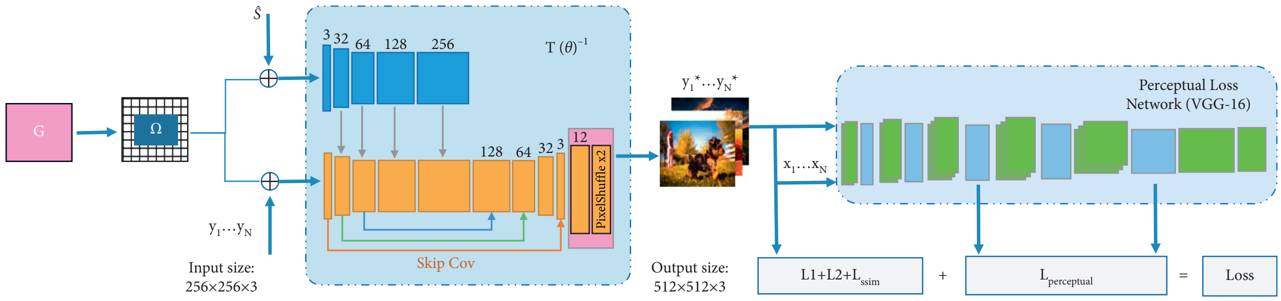


Figure 3: The architecture of SRCN. Operator ⊕ denotes bilinear interpolation. {3, 32, 64, 128, 256} is the number of channels in each layer.

and y are fed separately to the same four convolutional layers. Then, the multilevel feature maps generated by the convolution of each layer are combined using elementwise addition, which allows learning the complex interreflection information of the immersive environment. To keep the size invariant for the feature maps, the first two convolutional strides are set to 2, the last two convolutional strides are set to 1, and the number of convolution kernels is {32, 64, 128, 256}. Each is followed by a rectified linear unit (ReLU). In addition, we use three skip convolution connections [24] to enrich the representation ability of the network. Then we use two convolutional layers with stride 1, padding 1, and two transposed convolutional layers with stride 2, no-padding to gradually reduce the channel of the feature maps. We ultimately use the SR layer [25] to increase the resolution of the input images, which is then followed by ParametricReLU [26] as the activation function.

Our multiprojector interreflection compensation overall architecture consists of three main steps. (1) We first split a plain white image $x0$ and N sampling images $x_1, x_2. \ldots . .x_N$ into two parts and project them using two projectors. With a camera, we can capture $y_1, y_2. \ldots . .y_N$. Then, we resize the images to 256 × 256 and preprocess them by gamma correction. (2) All of the camera-captured images are input to the geometric correction subnet G and then enter the deep convolution layer to output the compensation images $y_1{}^*$, $y_2{}^*. \ldots . .y_N{}^*$. Because of the superresolution mechanism in SRCN, the resolution of the output is doubled. Finally, using our four-loss functions, we can train SRCN to converge. (3) With the converged SRCN, we can input the desired image x and obtain a compensated image x^*. If x^* is projected, we find that the result is the same as the ground truth x, as Figure 2 shows.

3.4. Loss Function. The loss function in a neural network compares the difference between the predicted value and the true value. In SRCN, we use the loss function below to jointly optimize the color fidelity (pixelwise l_1 and l_2), structural similarity (SSIM), and perceived similarity (perceptual loss):

$$L = l_1 + l_2 + l_{\text{ssim}} + \lambda l_{\text{perceptual}}, \tag{10}$$

where λ is a coefficient to balance perceptual loss and other loss functions. In our experiment, we set λ to 0.02 to balance perceptual loss and other loss functions. All of these loss functions are explained in detail in Sections 3.4.1, 3.4.2, and 3.4.3.

3.4.1. Pixelwise Loss. The output of the network is compared with the ground truth pixel by pixel. l_1 is most simple:

$$l_1 = \frac{1}{N} \sum_{i=1}^{N} \left| y_i^* - x_i \right|. \tag{11}$$

l_2 is also called MSE loss and can be computed by

$$l_2 = \frac{1}{N} \sum_{i=1}^{N} \left\| y_i^* - x_i \right\|^2. \tag{12}$$

3.4.2. SSIM Loss. This approach compares the structure similarity of y_i^* and x_i in three dimensions: luminance $\ell(y_i^*, x_i)$, contrast $c(y_i^*, x_i)$, and structure $s(y_i^*, x_i)^2$

$$\begin{aligned} \text{SSIM}_i &= \ell(y_i^*, x_i) \times c(y_i^*, x_i) \times s(y_i^*, x_i) \\ &= \frac{\left(2\mu_{y_i^*}\mu_{x_i} + C_1\right)\left(2\sigma_{y_i^* x_i} + C_2\right)}{\left(\mu_{y_i^*}^2 + \mu_{x_i}^2 + C_1\right)\left(\sigma_{y_i^*}^2 + \sigma_{x_i}^2 + C_2\right)}, \end{aligned} \tag{13}$$

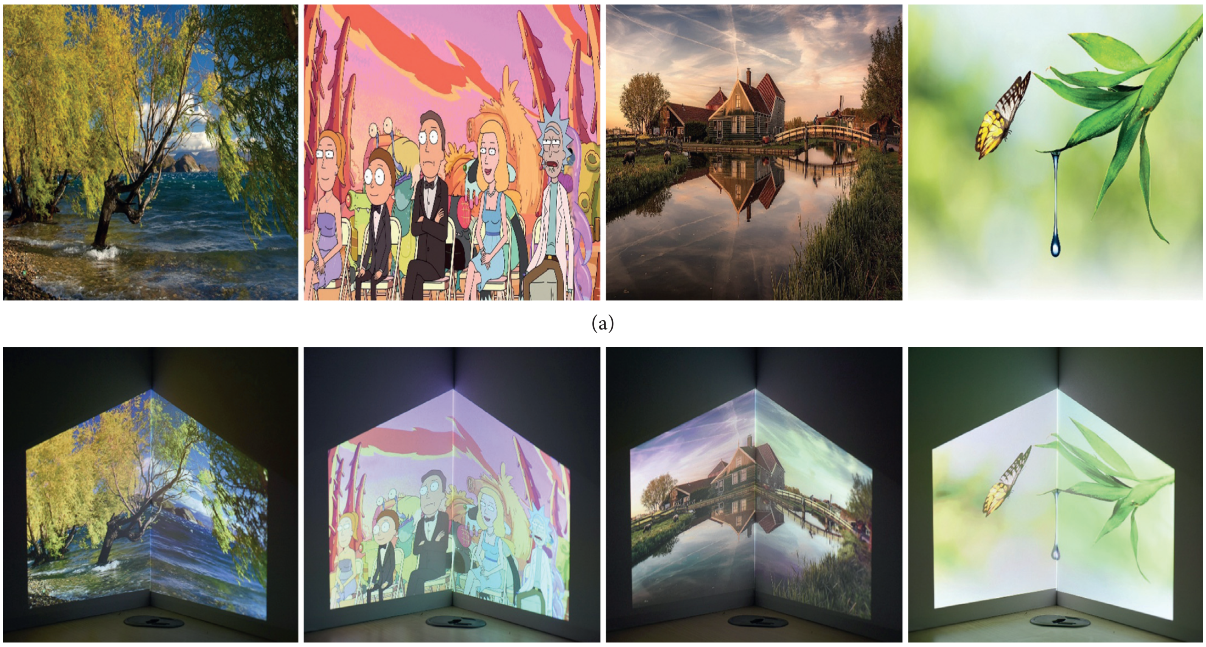
where μ and σ are the means and standard deviations and C_1 and C_2 both are invariant constants. The SSIM loss can be computed by

$$l_{\text{ssim}} = \frac{1}{N} \sum_{i=1}^{N} \left(1 - \text{SSIM}_i\right). \tag{14}$$

3.4.3. Perceptual Loss. Perceptual loss can bring the high-level information, content, and global structure closer by comparing the features of a generated image with that of the real image. It uses a pretrained 16-layer VGG network V_k [27] to obtain the feature map of y_i^* and x_i, where V_k indicates the feature map obtained by the k-th max-pooling layer. Then, it compares the difference between $V_k(y_i^*)$ and $V_k(x_i)$:

$$l_{\text{perceptual}} = \frac{1}{N} \sum_{i=1}^{N} \left\| V_k(y_i^*) - V_k(x_i) \right\|^2. \tag{15}$$

Rather than encouraging the pixels of the generated image y_i^* to exactly match the pixels of the input image x_i,



(a)

(b)

Figure 4: Samples from our Pro-Cam dataset. (a): the input images (ground truth). (b): the corresponding camera-captured images preprocessed using gamma correction.

we instead encourage them to have similar feature representations as computed by the VGG network. Using the total losses above, we achieved comparable performance with better reconstructed fine details and edges.

3.5. Training Details. We trained our deep convolutional architecture SRCN for 8 epochs on NVIDIA GEFORCE 1060TI GPUs with a batch size of 8, using 4,200 images for the training set and 72 images for the validation set. Backpropagating the derivative of the loss throughout the network, the network's parameters, θ, such as weight and bias, were updated via the Adam optimizer [28] with the following specifications: we set the fixed l2 penalty factor to 10^{-4}. We started with a learning rate of 0.001. The size of input images was 256×256. As mentioned in equation (10), we set λ to 0.02 to balance perceptual loss and other loss functions. We used the third and fifth max-pooling layers within the VGG16 network to compute the perceptual loss, so $k = 3, 5$ in equation (15). We also provided a pretrained model to make the method more practical, using 5,000 pairs of sampling images. θ was initialized by loading the saved weights. During the test time, we used SRCN without geometric correction subnet G to compensate 52 1920×1080 colorful images.

4. Experiments

In this work, we created a dataset using our Pro-Cam. We conducted the experiment under a multiprojector immersive environment and compared the results with different values of gamma correction in the data preprocessing phase. We compared the experimental results from SRCN with those from other methods, from objective and subjective aspects.

4.1. Projection Display Prototype and Dataset. Our immersive multiprojector-camera system consists of a Nikon DX VR camera with a resolution of 2992×2000 and two JMGO G7 projectors with a projection resolution of 1920×1080. We set the distance between the camera and the two projectors to 300 mm. The L-shaped screen was located approximately 550 mm in front of the projectors. The camera's white balance mode, shutter speed, ISO, and focus were set to Auto, 1/90, 200, and $f = 5.6$, respectively. To simulate a real immersive projection system, we captured the pictures in the dark to exclude the influence of global lighting.

To ensure the dataset was as diverse as possible, we projected 5,000 1920×1080 colorful images crawled from several free picture websites and obtained $N = 5000$ camera-captured images automatically by setting the camera mode to interval shooting. Then, we resized the images to 256×256 and preprocessed them using gamma correction (Figure 4).

4.2. Comparison of Different Methods

4.2.1. Objective Evaluations. We used objective evaluations to compare the quality of the compensation images. We compared our SRCN model with other methods, including CompenNet [13], CompenNet++ [29], and TPS [8].

To compare the compensation effect of different kinds of images, we divided the test images into four groups. Each group had 13 images. In the first group, the images were bright, and some areas were even overexposed. In the second group, the images were dark. In the third group, the images were multicolored. In the last group, the images were solid color. We named them Group_Bright,

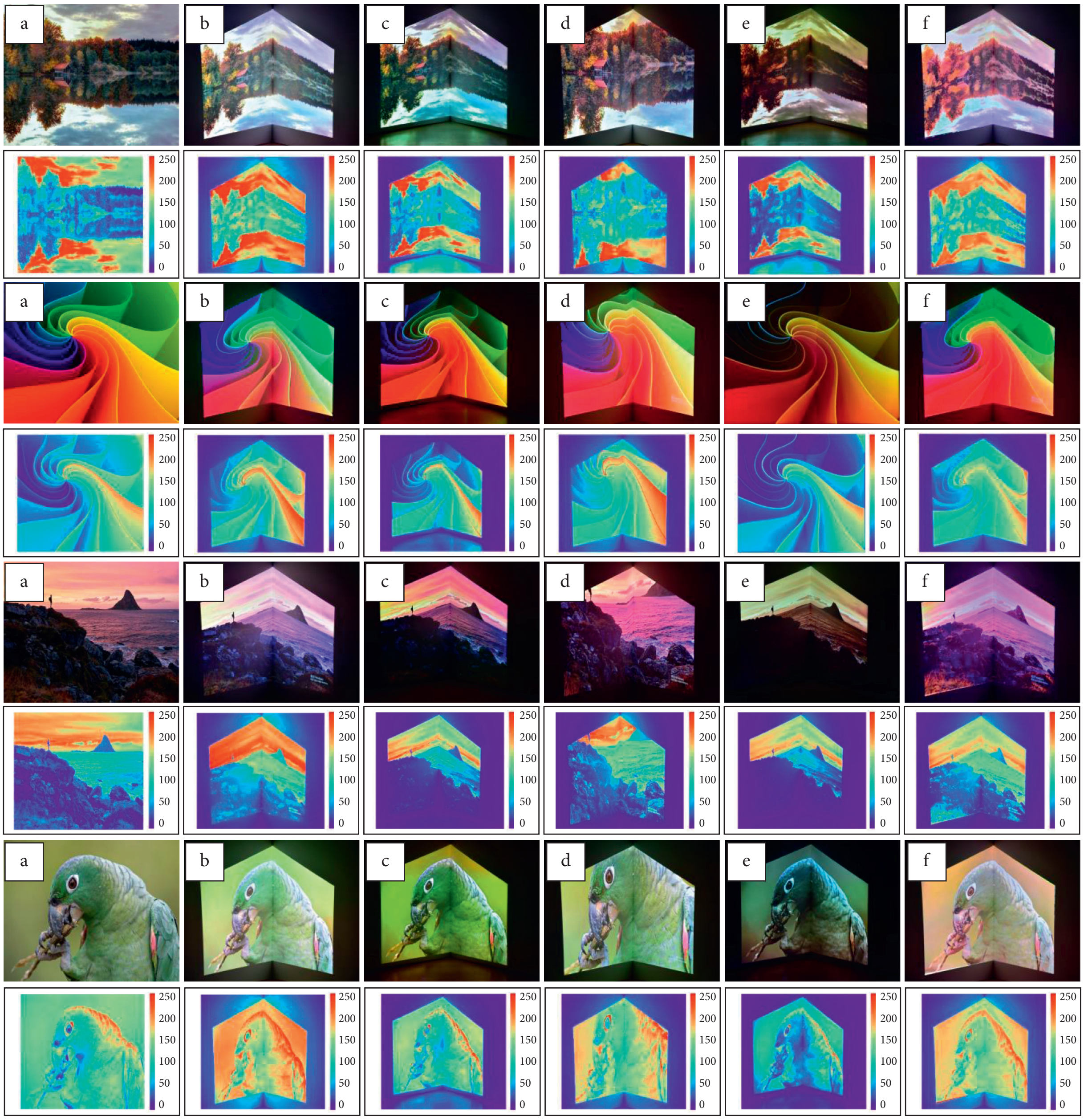


Figure 5: Comparison of projection results with (b) uncompensated, (c) SRCN, (d) CompenNet++, (e) CompenNet, and (f) TPS. (a) is the ground truth image (GT). In the heat maps, red and blue colors indicate that the image has higher and lower intensity, respectively.

Group_Dark, Group_Multicolor, and Group_Solidcolor, respectively. The projection results are shown in Figure 5. We compared four groups of projection results on PSRN, SSIM, and RSME. Table 1 shows specific evaluation values for four group images. We can conclude that SRCN generally produced better compensation image quality.

4.2.2. Subjective Evaluations

(1) Compensation Image. We used subjective evaluations to compare the clarity of the compensation images. We utilized the SR mechanism to increase the resolution of the input images. We found that the compensation image was clearer than those of other methods. As shown in Figure 6, we observed that using SRCN yielded better texture detail. The cat's eyes are clearer in our images than in those produced by CompenNet and TPS. CompenNet++ takes geometric correction into consideration, so the compensation result is a little distorted. When it is projected, the image will be distorted back. In this work, we were solely concerned with reducing interreflection. The resolution of the other three methods depends on the training dataset. If we want to obtain a 1920 × 1080 compensation image, we must use thousands of 1920 × 1080 images to train the model. This approach requires more computer memory and a longer training time than ours.

Table 1: Objective evaluations of different compensation algorithms.

Model		SSIM↑	PSRN↑	RSME↓
Group_Bright	SRCN	**0.54**	**16.85**	**0.14**
	CompenNet	0.51	15.37	0.17
	CompenNet++	0.39	11.52	0.27
	TPS	0.46	12.37	0.24
Group_dark	SRCN	**0.39**	**16.79**	**0.15**
	CompenNet	0.34	14.31	0.20
	CompenNet++	0.23	11.49	0.28
	TPS	0.31	12.74	0.23
Group_Multicolor	SRCN	**0.49**	**15.52**	**0.17**
	CompenNet	0.46	13.67	0.21
	CompenNet++	0.31	9.92	0.32
	TPS	0.41	11.73	0.26
Group_Solidcolor	SRCN	**0.49**	**15.81**	**0.16**
	CompenNet	0.48	14.72	0.19
	CompenNet++	0.34	11.16	0.28
	TPS	0.46	13.14	0.22
Average	SRCN	**0.48**	**16.24**	**0.16**
	CompenNet	0.45	14.52	0.19
	CompenNet++	0.32	11.02	0.29
	TPS	0.41	12.50	0.24

The best results are in bold. "Average" was obtained by averaging the metric scores of four groups of test images.

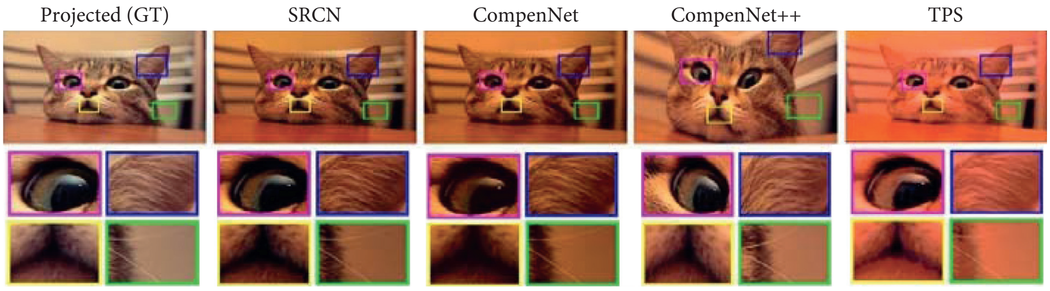


Figure 6: Comparison of clarity of compensation images produced using SRCN, CompenNet, CompenNet++, and TPS.

Table 2: Comparison of MOS of projection results with SRCN, CompenNet, CompenNet++, and TPS.

Methods	SRCN	CompenNet	CompenNet++	TPS
MOS	3.3	2.4	2.6	3.1

(2) Projection Results. Our multiprojector display prototype was designed for an immersive viewing experience, so subjective assessment with respect to the human visual system is very important. We invited 25 raters to evaluate the quality of the compensation images in a mean opinion score (MOS) test and asked them to choose the scores (1 to 5) for different projection images. A score of 1 indicated that the quality of the projection image was poor, and 5 indicated excellent quality. All raters had normal visual power and color vision. We used the ground truth as the reference image. The MOS of each compensated method was used as the final subjective evaluation index, as shown in Table 2.

We computed the heat maps [30] of four projection images. These images represent Group_Bright, Group_Dark, Group_Multicolor, and Group_Solidcolor, respectively. The heat maps represent the luminance of projection images. The redder the color, the brighter the projection images, and the more the interreflection. As shown in Figure 5(b), when the image was projected directly onto the L-shaped screen without any processing, the projection image contained some scattered light and became lighter than the original image. CompenNet (Figure 5(e)) reduced the interreflection but also reduced the color. TPS (Figure 5(f)) exhibited color deviation problems when the images were projected. In CompenNet++, the authors thought, from the surface patches illuminated by the projector, the rest of the surface outside the projector FOV did not provide useful information for compensation, so they cropped the images to achieve better geometric correction. When the size of the compensation image is 256 × 256, it can approximate to the original image. However, in our situation, in which we compensate 1920 × 1080 images, the cropping seriously affects the results (Figure 5(d)). Our method (Figure 5(c)) can reduce interreflection most effectively, while maintaining the color information.

4.3. Effectiveness of Gamma Correction. In the immersive projection system, observation by the human eye is the most

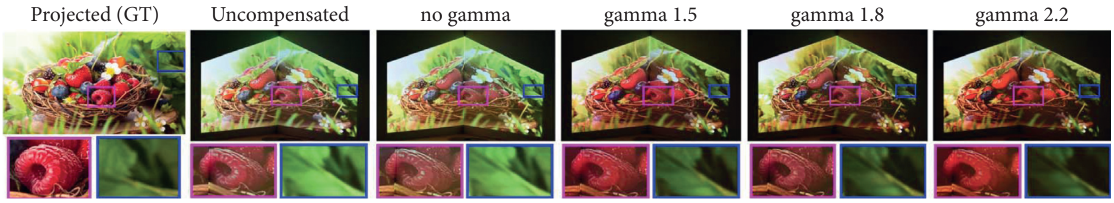


Figure 7: Comparison of projection results with different gamma corrections in the data preprocessing phase.

Table 3: Objective evaluations of compensation images between SRCN and SRCN w/o SR.

Model	SSIM↑	PSRN↑	RSME↓
SRCN	**0.80**	**19.69**	**0.11**
SRCN w/o SR	0.71	17.92	0.13

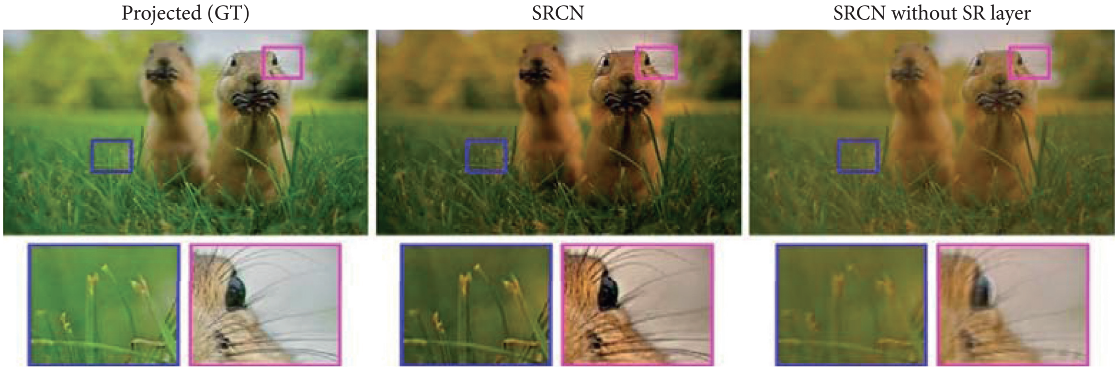


Figure 8: Subjective visual perception of compensation images between SRCN and SRCN w/o SR.

Table 4: Objective evaluations of projection results with different loss functions.

Loss	SSIM↑	PSRN↑	RSME↓
Uncompensated	0.49	12.90	0.23
l_1	**0.51**	15.58	0.17
l_2	**0.51**	15.19	0.18
$l_1 + l_2 + l_{\text{ssim}}$	0.48	14.29	0.20
$l_1 + l_2 + l_{\text{ssim}} + l_{\text{perceptual}}$	0.48	**16.24**	**0.16**

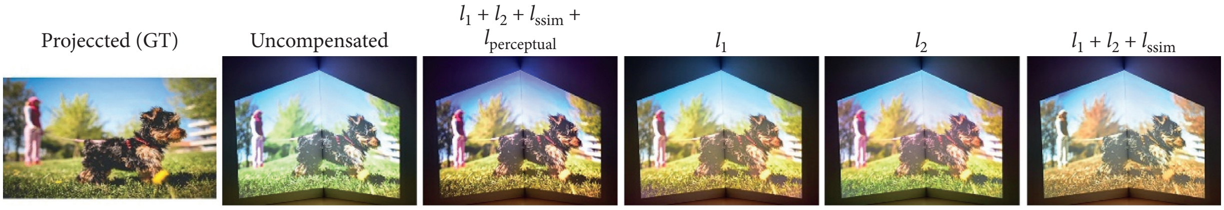

Projeccted (GT)
Uncompensated
$l_1 + l_2 + l_{\text{ssim}} + l_{\text{perceptual}}$
l_1
l_2
$l_1 + l_2 + l_{\text{ssim}}$

Figure 9: Comparison of projection results with different loss functions.

important index of image quality. However, the human eye's response to radiation is not a linear function but a curve that is similar to the gamma curve. Generally, our eyes have a greater dynamic range in the shadow than under high illumination, and we are more sensitive to low illumination and less sensitive to bright light. In our research, we took this situation into full consideration and performed a gamma correction on the camera-captured images.

As shown in Figure 7, we compared the different training data sets, with all conditions being the same, other than gamma correction. The values of the gamma correction were 1 (no gamma), 1.5, 1.8, and 2.2. The larger the value of the gamma correction, the darker the compensation image. If there was no gamma correction in the data preprocessing, the interreflection of the compensation image was only slightly reduced. Therefore, we set the value of the gamma correction to 1.5, which produced the closest results to the ground truth.

4.4. Effectiveness of the SR Layer. In order to investigate whether our learning-based formulation and the SR layer (subpixel convolutional layer) were necessary, we compared the results of the proposed SRCN with and without the SR layer (SRCN w/o SR). The results are shown in Table 3, and visual comparisons are shown in Figure 8. SRCN with an SR layer clearly yielded a better result.

4.5. Comparison of Different Loss Functions. We compared four different loss functions: $l_1, l_2, l_1 + l_2 + l_{\text{ssim}}$, and $l_1 + l_2 + l_{\text{ssim}} + l_{\text{perceptual}}$ loss. The objective and subjective comparisons are shown in Table 4 and Figure 9, respectively. We found that the quality of the compensation image and that of the projection image using l_1 and l_2 were almost the same. The use of $l_1 + l_2 + l_{\text{ssim}}$ produced the worst compensation and projection results, while $l_1 + l_2 + l_{\text{ssim}} + l_{\text{perceptual}}$ produced the best results. Finally, we used $l_1 + l_2 + l_{\text{ssim}} + l_{\text{perceptual}}$ as our SRCN loss function.

5. Conclusions

In this paper, in order to solve the problem of serious interreflection in multiprojection system imaging, we developed an SRCN model that reduces interreflection from multiprojector immersive systems. We performed experiments using our own data set to establish the validity of the approach. The technique achieved consistently better perceptual quality than previous methods. We first used a deep convolution network specialized for multiprojector interreflection compensation. We formulated a novel architecture by adding a geometric correction subnet. We used SR layers to improve the resolution of the compensated images. Other useful techniques, including gamma correction and perceptual loss, were employed to improve the image quality and restore more accurate realistic textures.

Data Availability

The authors have created a dataset in their Pro-Cam. The artificial dataset is publicly available and its download link is https://drive.google.com/drive/folders/1bdT6tqDW9blAfiKSPDWeqnrs_tG3d6hD?usp=sharing.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

Acknowledgments

This research was funded by the National Natural Science Foundation of China, Grant no. U183310179.

References

[1] S. Tang, L. Zhang, and Y. Zhang, "A large-scale multi-projector glass-free 3D display system," *International Society for Optics and Photonics*, vol. 9273, 2014.

[2] S. Xia, B. Chen, G. Wang et al., "Two classification methods based on a novel multiclass label noise filtering learning framework," *In IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2021.

[3] H. Habe, N. Saeki, and T. Matsuyama, "Inter-reflection compensation for immersive projection display," in *Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-2, Minneapolis, MN, USA, June 2007.

[4] O. Bimber, A. Grundhofer, T. Zeidler, D. Danch, and P. Kapakos, "Compensating indirect scattering for immersive and semi-immersive projection displays," in *Proceedings of the IEEE Virtual Reality Conference (VR 2006)*, pp. 151–158, Alexandria, VA, USA, March 2006.

[5] Y. Ding, J. Xiao, K.-H. Tan, and J. Yu, "Catadioptric Projectors," in *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2528–2535, Miami, FL, USA, June 2009..

[6] G. Wetzstein and O. Bimber, "Radiometric compensation through inverse light transport," in *Proceedings of the 15th Pacific Conference on Computer Graphics and Applications*, pp. 391–399, Maui, HI, USA, October 2007.

[7] J. Bai, M. Chandraker, T.-T. Ng, and R. Ramamoorthi, "A dual theory of inverse and forward light transport," in *Proceedings of the European Conference on Computer Vision*, pp. 294–307, Heraklion, Greece, September 2010.

[8] A. Grundhöfer and D. Iwai, "Robust, error-tolerant photometric projector compensation," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5086–5099, 2015.

[9] A. Lucas, M. Iliadis, R. Molina, and A. K. Katsaggelos, "Using deep neural networks for inverse problems in imaging," *In IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 20–36, 2018.

[10] X. Wang, Q. Tao, L. Wang, D. Li, and M. Zhang, "Deep convolutional architecture for natural image denoising," in *Proceedings of the 2015 International Conference on Wireless Communications & Signal Processing (WCSP)*, pp. 1–4, Nanjing, China, October 2015.

[11] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Munich, Germany, October 2015.

[12] M. Qin, F. Xie, W. Li, Z. Shi, and Z. Wang, "Dehazing for multispectral remote sensing images based on a convolutional neural network with the residual architecture," *In IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 5, pp. 1645–1655, 2018.

[13] A. Golts, D. Freedman, and M. Elad, "Unsupervised single image dehazing using dark channel prior loss," *In IEEE Transactions on Image Processing*, vol. 29, pp. 2692–2701, 2020.

[14] H. Wang, W. Wu, Y. Su, Y. Duan, and P. Wang, "Image super-resolution using a improved generative adversarial network," in *Proceedings of the 2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, pp. 312–315, Beijing, China, July 2019.

[15] B. Huang and H. Ling, "End-to-end projector photometric compensation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6810–6819, Long Beach, CA, USA, June 2019.

[16] J. Johnson, A. Alahi, and Li Fei-Fei, “Perceptual Losses for Real-Time Style Transfer and Super-resolution,” 2016, http://arxiv.org/abs/1603.08155.

[17] E. P. Simoncelli, H. R. Sheikh, A. C. Bovik, and Z. Wang, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[18] L. Zhang, L. Zhang, X. Mou, and D. Zhang, “A comprehensive evaluation of full reference image quality assessment algorithms,” in *Proceedings of the2012 19th IEEE International Conference on Image Processing*, pp. 1477–1480, Orlando, FL, USA, September 2012.

[19] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, “Loss functions for image restoration with neural networks,” *IEEE Transcation on Computer Imaging*, vol. 3, no. 1, pp. 47–57, 2016.

[20] J. Gu, Z. Wang, J. Kuen et al., “Recent advances in convolutional neural networks,” *Pattern Recognition*, vol. 77, pp. 354–377, 2018.

[21] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[22] 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*, pp. 105–114, Honolulu, HI, USA, July 2017.

[23] I. Rocco, R. Arandjelovic, and J. Sivic, “Convolutional neural network architecture for geometric matching,” *In IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 11, pp. 2553–2567, 2019.

[24] 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, June 2016.

[25] W. Shi, J. Caballero, F. Huszár et al., “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1874–1883, Las Vegas, NV, USA, June 2016.

[26] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: surpassing human-level performance on imagenet classification,” in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1026–1034, Santiago, Chile, December 2015.

[27] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” 2015, https://arxiv.org/abs/1409.1556.

[28] D. P. Kingma and J. Ba, “Adam: a method for stochastic optimization,” 2014.

[29] B. Huang and H. Ling, “CompenNet++: end-to-end full projector compensation,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 7165–7174, Seoul, Korea, November 2019.

[30] S. Takeda, D. Iwai, and K. Sato, “Inter-reflection compensation of cmmersive crojection cisplay by spatio-temporal screen reflectance modulation,” *In IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 4, pp. 1424–1431, 2016.