Document Image Shadow Removal Guided by Color-Aware Background

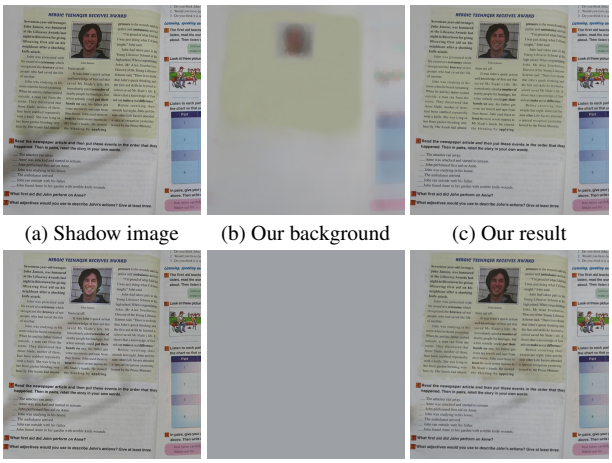
Ling Zhang[1], Yinghao He[1], Qing Zhang [2], Zheng Liu[3], Xiaolong Zhang[1], Chunxia Xiao[4*]
[1]School of Computer Science and Technology, Wuhan University of Science and Technology
[2]School of Computer Science and Engineering, Sun Yat-Sen University
[3]School of Computer Science, China University of Geosciences (Wuhan)
[4]School of Computer Science, Wuhan University
zhling@wust.edu.cn, hyhwust@163.com, zhangqing.whu.cs@gmail.com
liu.zheng.jojo@gmail.com, xiaolong.zhang@wust.edu.cn, cxxiao@whu.edu.cn

Abstract

Existing works on document image shadow removal mostly depend on learning and leveraging a constant background (the color of the paper) from the image. However, the constant background is less representative and frequently ignores other background colors, such as the printed colors, resulting in distorted results. In this paper, we present a color-aware background extraction network (CBENet) for extracting a spatially varying background image that accurately depicts the background colors of the document. Furthermore, we propose a background-guided document images shadow removal network (BGShadowNet) using the predicted spatially varying background as auxiliary information, which consists of two stages. At Stage I, a background-constrained decoder is designed to promote a coarse result. Then, the coarse result is refined with a background-based attention module (BAModule) to maintain a consistent appearance and a detail improvement module (DEModule) to enhance the texture details at Stage II. Experiments on two benchmark datasets qualitatively and quantitatively validate the superiority of the proposed approach over state-of-the-arts.

1. Introduction

Documents, such as textbooks, newspapers, leaflets, and receipts, are available daily, often saved as electronic documents for digital document archives or online message transfer. Since the wide use and convenience of mobile phones, people currently prefer to use mobile phones for digital document copying. However, the captured document images become highly susceptible to shadows when the light sources are blocked. The low brightness in shadow regions reduces the quality and readability of the document image, resulting in illegible content and unpleasant user experience [2, 10, 12, 19, 28, 29]. Thus, shadow removal for document images is a required image processing task in various vision applications [3, 4, 33, 37, 46, 48].



(a) Shadow image (b) Our background (c) Our result
(d) Result of [2] (e) Background of [24] (f) Result of [24]

Figure 1. Document image shadow removal. With the assumption of constant background, results of [2] and [24] may cause color distortion or shadow remnant. By using a spatially varying background, our method can produce more desirable result.

Although natural image shadow removal has made substantial progress [16, 20, 22, 40, 44, 45], these approaches generally perform poorly on document pictures due to their drastically different features from natural images. Natural image, for example, emphasizes background content (shadow-free image) without a shadow layer [6, 7, 13, 27, 41], whereas document image emphasizes text content [2, 19, 30]. Without considering the particular properties of the document image, traditional approaches to natural image generally yield incorrect result when they are applied to document image, as well as learning-based methods (see Figure 9(c-f)) due to the less attention to the content structures.

*Corresponding author: Chunxia Xiao (cxxiao@whu.edu.cn).

Several approaches on document image shadow removal are currently available, which dig into the specific characteristics of the document image [3, 21, 30, 49]. However, these approaches may cause color distortion or shadow remnant for image with complex backgrounds, as illustrated in Figure 1(d). Recently, Lin et al. [24] estimate a constant background for the image and propose BEDSR-Net for document image shadow removal. The constant background is the color of the paper (see Figure 1(e)), which ignores some other background colors, such as the printed colors. The constant background may provide inaccurate information for the shadow removal task, resulting in unsatisfactory results (see Figure 1(f)). To address this problem, we propose a color-aware background extraction network (CBENet) to extract a spatially varying background, which can preserve various background colors of the original image (see Figure 1(b)). The spatially varying background can provide more useful color information, which contributes to image shadow removal, as shown in Figure 1(c).

With the background image, we propose a background-guided shadow removal network (BGShadowNet) for document image that exploits the background image as auxiliary information. Figure 2 presents the framework of the proposed BGShadowNet, which removes shadows in a two-stage process. First, we introduce a background-constrained decoder to combine background features with image features, which can help to promote the realistic coarse shadow-removal result. Then, we refine the coarse result with a background-based attention module (BAModule) and a detail enhancement module (DEModule). In particular, BAModule is designed to eliminate the illumination and color inconsistency in the image by using the attention mechanism. Inspired by image histogram equalization, DEModule aims to enhance the detail features at low-level scales.

Due to the lack of large-scale real document image datasets, we construct a new dataset comprised of real document images to facilitate the performance of document image shadow removal. Experiments on extensive document images and evaluations on two datasets verify that our BGShadowNet outperforms existing approaches.

In summary, our main contributions are three-fold:

- We present a color-aware background extraction network (CBENet) for estimating a spatially varying background image that guides the shadow removal of document image.
- We propose a framework named BGShadowNet for document image shadow removal, which takes full advantage of the background image and is able to robustly produce high-quality shadow removal results.
- We design a background-based attention module (BAModule) to maintain a consistent appearance and a detail enhancement module (DEModule) to enhance texture details.

2. Related Work

2.1. Shadow removal for natural image

Traditional methods for natural image shadow removal usually focus on studying the different physical properties of shadows [1, 9, 25]. Finlayson et al. [8, 9] reconstructed shadow removal images based on gradient consistency. However, these methods can incur obvious shadow boundary artifacts due to the change of illumination. Shor et al. [34] defined an affine relationship between the shadow and non-shadow areas. Xiao et al. [34, 42, 43] and Zhang et al. [50] removed shadows by transforming illumination from non-shadow regions to shadow regions. However, these methods depended on the reference non-shadow areas, and often resulted in inconsistent illumination when the reference areas are unfortunate.

Numerous learning-based methods have been proposed for natural image shadow removal [6,11,15,16,18,23,26,27, 31, 35]. For example, Deshadow-Net [31] extracted multi-context features to predict shadow matte layers for shadow removal. Wang et al. [38] employed stacked conditional GANs for joint shadow detection and removal. Zhang et al. [47] explored residual and illumination with GANs for shadow removal. ARGAN [7] proposed a attentive recurrent generative adversarial network for shadow detection and removal. Liu et al. [27] utilized shadow generation for weakly-supervised shadow removal. More recently, Chen et al. [5] transferred the contextual information from non-shadow regions to shadow regions in the embedded feature spaces.

Although these methods are effective for natural images, they do not generalize well to document image shadow removal due to the difference characteristics between natural images and document images.

2.2. Shadow removal for document image

Most existing document shadow removal algorithms [2,3,19,21,30] use heuristics algorithms to dig into specific features of the document image. Bako et al. [2] removed shadows using the estimated shadow map. This method leaves slight trace at the boundary under strong shadows. Oliveira et al. [30] used natural neighborhood interpolation to estimate shadow image. Jung et al. [19] explored a water-filling method for correcting the illumination of document image by converting the input image to a topographic surface. This method can achieve good performance in weak or moderate level of shadows, but tends to produce color degraded results for scenes with heavy shadows.

Recently, Lin et al. [24] proposed a BEDSR-Net for document image shadow removal by estimating a constant background. It is the first deep network specifically designed for document image shadow removal, which takes advantage of specific properties of document images. Due

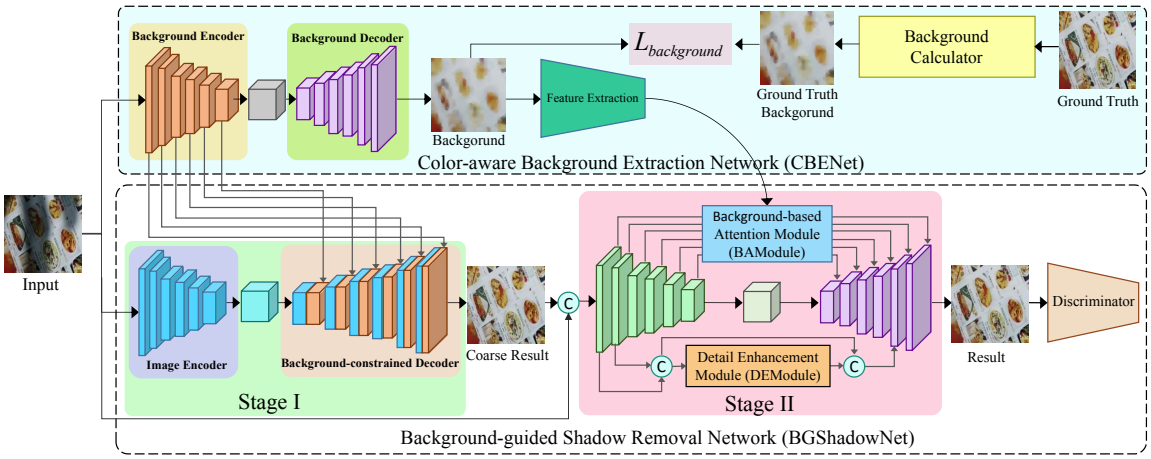


Figure 2. The framework of the proposed CBENet and BGShadowNet. Spatially varying background estimated by CBENet is used to help BGShadowNet produce a high-quality shadow removal result. With a background-constrained decoder, BGShadowNet first predicts a coarse shadow-removal result. Then, the coarse result is refined with a BAModule and a DEModule.

to ignoring some other background colors in the image, this method may introduce artifacts, such as shadow boundaries or unremoved shadows.

3. Document Dataset Construction

Although there are several document shadow removal datasets, such as Bako [2], Kligler [21], Jung [19], RDSRD [24], SDSRD [24], they mostly have some limitations. Bako, Kligler, Jung, and RDSRD are small-scale evaluation datasets that are unsuitable for training a deep model. While SDSRD is a large-scale dataset, it is a synthetic dataset.

Figure 3. Illustration of several captured shadow and shadow-free image pairs in RDD. The first row shows shadow images, and the second row demonstrates the corresponding shadow-free images.

Imaging is a physical generation process by the interaction between light and material. Light environments in the real world usually contains multiple different lights, which are difficult to simulate in synthetic environments accurately. The statistical features of the synthetic and real images are often different. To facilitate the performance of document image shadow removal, we construct a new real document dataset, named RDD, for shadow removal.

Concretely, we use documents as the background scene to construct our dataset, such as papers, books, publicity pamphlets, etc. We first capture a shadow image with illumination blocked by an object. Then, we obtain the corresponding shadow-free image by removing the occluder. Our RDD collects 4916 pairs of shadow and shadow-free images, divided into two groups, 4371 for training and 545 for testing. Figure 3 presents some shadow and shadow-free image pairs in our RDD. To the best of our knowledge, RDD is the first large-scale real document dataset for shadow removal.

4. Methodology

We first introduce a color-aware background extraction network (CBENet) to estimate a spatially varying background for shadow image. Then, we propose a background-guided shadow removal network (BGShadowNet) for document image, which uses the estimated background as auxiliary information to facilitate shadow removal task. Figure 2 presents the framework of our CBENet and BGShadowNet.

4.1. Color-aware Background Extraction Network

As document image focuses mostly on the text content, a common strategy [2, 24] to perform document image shadow removal is to utilize the background layer extracted from the image, that only contains the color information for the image without text content. These methods assume the document has a constant color background (the color of the paper), as shown in Figure 4(b). But discrepancy may exist between a constant color background and the image. For example, there may be other background colors due to color printing, as shown in Figure 4(a). The constant background will result in unsatisfactory results, as shown in Figure 4(c).

To address this problem, we propose a color-aware background extraction network (CBENet) to extract a spatially varying color background $\hat{B}$ for the document image, which preserves different background colors in the image, as shown in Figure 4(d). Compared with the constant background, our spatially varying background can provide more useful color information for the following shadow removal network. Note that our background is shadow-free, which can help BGShadowNet learn more shadow-free features,

(a) (b) (c) (d) (e) (f)

Figure 4. Extracted background images and the corresponding shadow removal images. (a) is input images. (b) and (c) are the background images and shadow removal results predicted by BEDSR-Net [24]. (d) and (e) are our spatially varying background images and shadow removal results. (f) is the corresponding ground-truth images.

(a) (b) (c)

Figure 5. Visualization of background: (a) shadow images, (b) local background images, and (c) final background images.

contributing to shadow removal while better avoiding illumination or color artifacts in the image (see Figure 4(e)).

To train our CBENet, we employ a background calculator with local-to-global strategy to construct the ground-truth background as the supervisor. Specifically, we first divide the ground-truth shadow-free image I_{gt} into 16×16 patches to obtain the local background $\overline{B}$. For each patch, we use the Gaussian Mixture Model (GMM) to cluster it into two clusters according to the pixel intensity of the image [2], corresponding to text content and background, respectively. With the observation that the background color of the document is usually brighter [2, 24], we consider the cluster with a higher intensity as the background cluster and apply the average color of the background cluster as the background of the patch. Since different patches have different background colors, the local background $\overline{B}$ usually has obvious patch boundaries, as shown in Figure 5(b). Thus, inspired by the guided filter for edge-preserving image smoothing [14], we employ a color-preserving smoothing operator to refine $\overline{B}$ and get a desired ground-truth background image B for the image, as shown in Figure 5(c).

The value of pixel i in B can be expressed as:

$$B_i = \sum_{j \in N(i)} W_{ij} \overline{B}_j, \tag{1}$$

where $N(i)$ is a local neighborhood of pixel i. W_{ij} is the filter kernel, which measures the color similarity between pixel i and pixel j. Due to the original image has the edge information, we use I^{gt} as the guidance image to calculate the filter kernel W_{ij}, that is,

$$W_{ij} = \frac{1}{|\omega|^2} \sum_{i,j} (1 + \frac{(I_i^{gt} - \mu_k)(I_j^{gt} - \mu_k)}{\sigma_k^2 + \epsilon}), \tag{2}$$

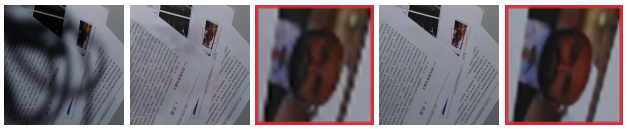
where μ_k and σ_k^2 are the mean and variance of I^{gt} in $N(i)$, $|\omega|$ is the number of pixels in $N(i)$, and ϵ is a regularization parameter preventing W_{ij} from being too large.

We adopt the U-Net structure to implement our CBENet. The U-Net first applies five Conv+BN+LReLu to extract features from the image. Then, it takes five deconvolutional layers with batch normalization and ReLU activation function to predict the background image. Skip connection is applied between convolutional layers and deconvolutional layers, increasing the number of channels in the network and preserving the context information of the front layer.

4.2. Background-guided Shadow Removal Network

As aforementioned, the background can provide some useful information to facilitate shadow removal. Thus, we propose a background-guided shadow removal network (BGShadowNet) exploiting the background image as supplementary information. As shown in Figure 2, our BGShadowNet includes two stages. At Stage I, besides an image encoder, a background-constrained decoder is introduced to produce a coarse shadow-removal result. At Stage II, to improve the coarse result and produce the final shadow-free image, a background-based attention module (BAModule) and a detail enhancement module (DEModule) are embedded into an encoder-decoder network. A discriminator is stacked at the end to distinguish whether the produced image is real or not. We choose DenseUnet [32] and Markovian discriminator [17] as our encoder-decoder structure and discriminator.

Background-constrained Decoder. To take full advantage of the features from background image, we replace the common decoder with a background-constrained decoder at Stage I. Concretely, features from the background encoder are integrated into the background-constrained decoder at each corresponding level. The integrated features can com-



(a) (b) (c) (d) (e)

Figure 6. Results of stage I and II. (a): Input. (b): Coarse result at stage I. (d): Result at stage II. (c)&(e): close-ups of (b) and (d).

plement the image features and help to produce a satisfactory shadow-removal result.

Background-based Attention Module. Generally, areas with a similar background should have a similar appearance (color and illumination) in an image. However, there might be illumination or color artifacts in the coarse shadow-removal result (see Figure 6(b)). To preserve the overall consistency of the image, we introduce a background-based attention module (BAModule). Using the learned background features and attention mechanism, the BAModule helps to eliminate the appearance inconsistency in the image (see Figure 6(d)).

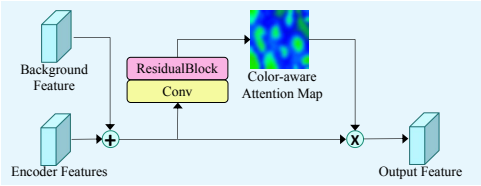


Figure 7. The network of our background-based attention module (BAModule).

Figure 7 illustrates the architecture of the proposed BAModule. We first fuse the encoder and background features to obtain the integrated features by channel-wise concatenation. Then, the integrated features are fed into an attention computing unit to generate a color-aware attention map. The attention computing unit consists of a convolutional layer, a Leaky ReLu activation function, batch normalization, and a ResidualBlock Layer. Finally, we fuse the color-aware attention map and the integrated features by element-wise multiplication to reconstruct the features, and then embed these features into the corresponding decoder level. The color-aware attention map can make the network adaptively focus on the regions with a similar background, promoting a consistent appearance in these regions.

Detail Enhancement Module. As multiple convolutions and downsampling operators in the network, partial detail information will be lost at high-level layers, resulting in detail-blurred results (see Figure 6(c)). Compared to the high-level features, low-level features in CNN layers usually contain more texture details. Thus, we introduce a detail enhancement module (DEModule) to restore the texture details of the coarse result by utilizing the low-level features of the network.

As we know, the statistical texture information of the image reflects the texture intensities to some extent. Therefore, our DEModule is inspired by image histogram equalization,

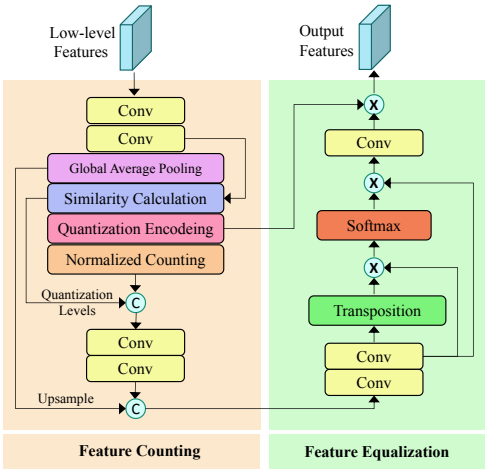


Figure 8. The pipeline of our detail enhancement module (DEM).

consisting of two parts. One is feature counting to obtain the statistical information for low-level features, and the other is feature equalization to enhance texture details. Figure 8 illustrates the pipeline of the proposed DEModule. Specifically, we fuse the features of the first two low-level layers from the encoder to get the concatenated low-level features F, which are fed into DEModule for statistical analysis.

(1) Feature Counting. The purpose of feature counting is to get the quantization encoding map and statistical feature. We first use two 2×2 convolution layers to produce a feature map M and perform global average pooling to obtain the global averaged features $\overline{M}$ for M. Next, we calculate the correlation between M and $\overline{M}$ by using cosine similarity, denoted as S.

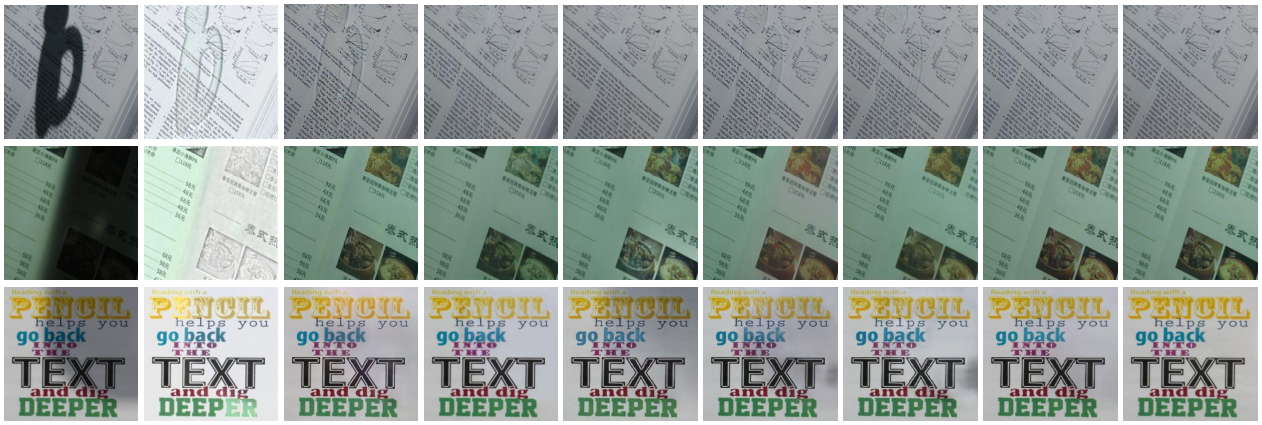
To effectively conduct quantification and statistics, we construct a set of quantization levels L, which divide the range of the minimum and maximum values of S into N equal parts. Then, the correlation matrix S can be quantized to a quantization encoding matrix E by using L:

$$E_{i,n} = \begin{cases} 1 - |L_n - S_i|, & if\ -\frac{0.5}{N} \le L_n - S_i \le \frac{0.5}{N} \\ 0, & else \end{cases}, \tag{3}$$

where $i \in [1, HW]$ and $n \in [1, N]$. H and W are the length and width of the image. L_n is the nth level of L, and S_i is the ith row of S. In our experiments, we set $N = 128$.

To avoid eliminating gradient information, we perform a normalization operation for matrix E. We integrate the normalized result and quantization levels L into a quantization counting map C, which reflects the relative statistics of the low-level input features. Due to the concatenation operation, the channel number of C is 2. Thus, we perform two 1×1 convolution operations for C to increase the channel number, followed by a concatenation operation with $\overline{M}$ to further get absolute statistical information H. H denotes the statistical features which play the role of the histogram.

(2) Feature Equalization. Feature equalization is used to enhance the texture details of low-level layers by recon-



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Figure 9. Visual comparison among state-of-the-art shadow removal methods: (a) input images, (b) Jung [19], (c) DSC [15], (d) Fu [11], (e) DHAN [6], (f) CANet [5], (g) BEDSR-Net [24], (h) our BGShadowNet, and (i) ground-truth images.

structing a new set of quantization levels. We first perform a 1×1 convolution operation for H to get G. Inspired by the attention mechanism, we perform matrix multiplication of G and its transposed matrix, followed by a softmax operation, to build a learned adjacent matrix X. Matrix X can be regarded as a similarity coefficient matrix. Then, we can reconstruct the new quantization levels L' with a matrix multiplication for X and G.

Based on the reconstructed quantization levels L', we conduct feature equalization for the original quantization encoding matrix E to enhance the detail features. The enhanced features R can be obtained using a matrix multiplication of quantization levels L' and matrix E. By using the enhanced texture details, the decoder can easily capture detail information.

4.3. Loss Function

Our loss functions for optimizing the proposed network contain four components: background reconstruction loss $\mathcal{L}_{background}$, appearance consistency loss $\mathcal{L}_{appearance}$, structure consistency loss $\mathcal{L}_{stucture}$ and adversarial loss $\mathcal{L}_{adv}$.

Background reconstruction loss is used to constrain the CBENet to obtain the desirable background image, which uses the ℓ_1 distance between $\hat{B}$ produced by CBENet and the ground-truth background image B , that is,

$$\mathcal{L}_{background} = ||B - \hat{B}||_1. \tag{4}$$

Appearance consistency loss evaluates the data loss between the predicted results and the ground-truth image, which is calculated in the ℓ_1 distance:

$$\begin{aligned}\mathcal{L}_{appearance} &= \lambda_1 \mathcal{L}_{coarse} + \lambda_2 \mathcal{L}_{final} \\ &= \lambda_1 ||I_{gt} - I_{coarse}||_1 + \lambda_2 ||I_{gt} - I_{free}||_1,\end{aligned} \tag{5}$$

where λ_1 and λ_2 are the weight parameters. I_{coarse} is the coarse result produced at Stage I, and I_{free} is the final shadow-removal result produced at Stage II.

Structure consistency loss aims to preserve image structure, which is calculated as,

$$\mathcal{L}_{stucture} = \lambda_3 ||\text{VGG}(I_{gt}) - \text{VGG}(I_{free})||_2^2, \tag{6}$$

where λ_3 is the weight parameter, and $\text{VGG}(\cdot)$ is the feature extractor from the pre-trained VGG19 model.

Adversarial loss is designed for the discriminator to judge whether the produced results are real or fake, which is described as:

$$\mathcal{L}_{adv} = \lambda_4 \mathbb{E}_{(I, I_{free}, I_{gt})}[log(D(I_{gt})) + log(1 - D(I))], \tag{7}$$

where D is the discriminator, and I is the shadow image.

5. Experiments

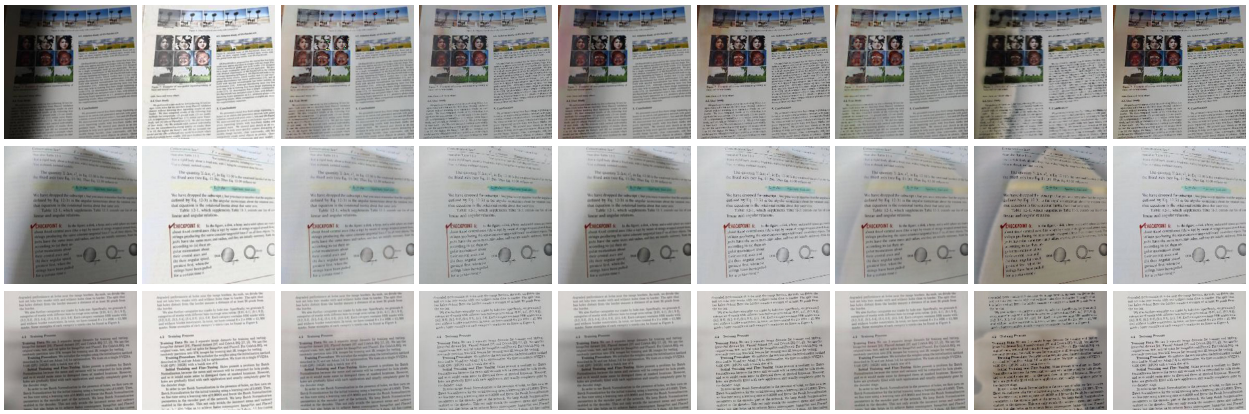
5.1. Implementation Detail

Our network is implemented in Pytorch. In our experiments, CBENet and BGShadowNet are trained separately. We first train CBENet for 200 epochs using the background ground-truth as the supervisor. Next, we fix the CBENet and train BGShadowNet for 200 epochs on a NVIDIA GeForce RTX2080Ti. We use Adam optimizer to optimize our generator and the discriminator with attenuation rate $betas = (0.5, 0.999)$. The initial learning rate is set to 0.0004. The weight parameters λ_1, λ_2, λ_3 and λ_4 are set to 1, 1, 0.05 and 0.01 in our experiments, respectively.

5.2. Comparison with State-of-the-arts

Datasets. Since SDSRD dataset [24] is not available, we use the proposed RDD dataset to train and evaluate our BGShadowNet. Apart from RDD, we also use Kligler's dataset [21] for evaluation.

Metrics. We utilize the root mean square error (RMSE) in LAB color space between the shadow removal result and the ground-truth shadow-free image to evaluate the shadow removal performance. In addition, we also report the PSNR



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Figure 10. Visual comparison among state-of-the-art shadow removal methods: (a) input images, (b) Jung [19], (c) DSC [15], (d) DHAN [6], (e) Fu [11], (f) BEDSR-Net [24], (g) CANet [5], (h) BMNet [51], and (i) our BGShadowNet.

Table 1. Quantitative comparisons of shadow removal on RDD and Kligler datasets in terms of RMSE, PSNR, and SSIM. All the learning-based methods are trained on RDD dataset. ↑ means the larger the better while ↓ means the smaller the better.

Methods	Venue/Year	RDD			Kligler		
		RMSE ↓	PSNR ↑	SSIM ↑	RMSE↓	PSNR ↑	SSIM ↑
ST-CGAN [39]	CVPR/2018	3.143	34.328	0.974	6.826	27.433	0.931
DSC [15]	PAMI/2020	6.357	28.151	0.914	7.705	25.615	0.898
DHAN [6]	AAAI/2020	2.467	36.337	0.978	6.610	27.707	0.937
Fu [11]	CVPR/2021	4.328	31.387	0.946	7.101	27.362	0.914
CANet [5]	ICCV/2021	5.561	28.951	0.918	7.855	25.625	0.899
SG-ShadowNet [36]	ECCV/2022	2.974	34.727	0.972	6.829	27.141	0.920
BMNet [51]	CVPR/2022	9.409	24.289	0.915	16.459	19.031	0.874
Bako [2]	ACCV/2016	14.648	20.741	0.894	9.058	24.777	0.895
Jung [19]	ACCV/2018	30.190	14.364	0.861	28.247	13.726	0.852
BEDSR-Net [24]	CVPR/2020	2.937	34.928	0.973	6.533	28.124	0.932
BGShadowNet	CVPR/2023	**2.219**	**37.585**	**0.983**	**5.377**	**29.176**	**0.948**

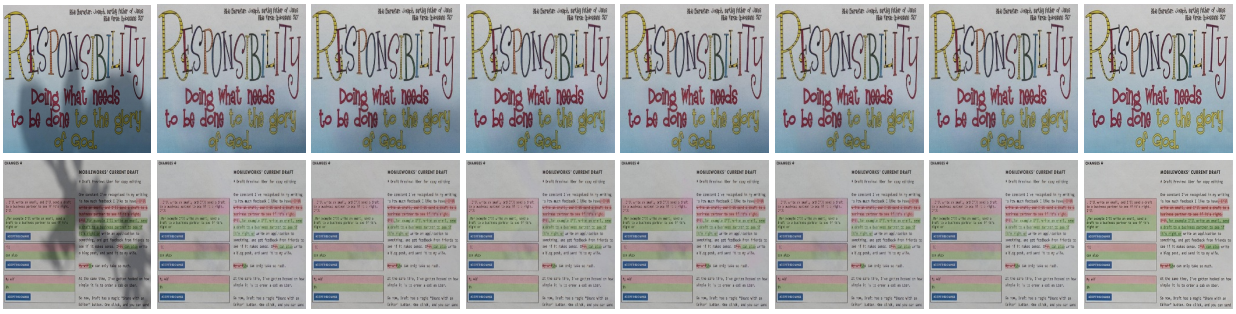
and SSIM in the RGB colour space to evaluate the performance of the proposed BGShadowNet.

To verify the effectiveness of our method, we compare our results with various state-of-the-art shadow removal methods including three document image shadow removal methods (Bako [2], Jung [19] and BEDSR-Net [24]) and six natural image shadow removal methods (ST-CGAN [39], DSC [15], DHAN [6], Fu [11], CANet [5], SG-ShadowNet [36] and BMNet [51]). To make fair comparison, we use RDD dataset to train all the learning-based methods on the same hardware. Table 1 concludes the comparison results. From the table, we can observe that, our method achieves the best values for all metrics among all the comparing methods, clearly demonstrating the effectiveness.

Figure 9 provides some visual shadow removal results to further demonstrate the superiority of our methods. It can be seen, DSC [15] fails in handling image with heavy shadows (see Figure 9(c)). The robustness of Fu [11] is limited, and their results may contain unremoved shadows, as shown in Figure 9(d). DHAN [6] and CANet [5] have the similar problem to Fu [11]. Ignoring the content characteristics of the documents, Jung [19] leads to obvious color and illumination distortion (see Figure 9(b)). With the constant background, results of BEDSR-Net [24] sometimes exhibit artifacts along the shadow boundaries (see Figure 9(g)). Comparatively, the proposed method effectively recovers illumination in shadow regions without artifacts, as shown in Figure 9(h), which is similar to the ground-truth image.

To further verify the robustness and generalization ability of the proposed method, Figure 10 presents some other shadow removal results for document images, containing some challenging cases, such as heavy shadows and inconsistent illumination in shadow regions. Apparently, results recovered by our method look more natural and have little artifacts.

User Study. We conduct an user study to evaluate the visual performance of our method and some state-of-the-art shadow removal methods. We prepare 100 sets of shadow removal images. Each set contains seven shadow removal results of our BGShadowNet, CANet, DSC, DHAN, Fu, BEDSR-Net, and Jung, respectively. We randomly select 100 volunteers. For each volunteer, we randomly provide them twenty image sets. The volunteers are required to select the best shadow-free image for each set. Counting all the results, we find that 20.32% of shadow-removal images generated by our BGShadowNet are chosen as the best shadow-free images, while 11.92%, 11.67%, 16.72%, 13.25%, 15.91%, and 10.21% of shadow removal results are chosen by CANet, DSC, DHAN, Fu, BEDSR-Net, and Jung, respectively.



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 11. Visual comparison for ablation study: (a) input images, (b) BASE_1, (c) BASE_2, (d) BGShadowNet_1, (e) BGShadowNet_2, (f) BGShadowNet_3, (g) BGShadowNet_4, and (h) our BGShadowNet.

Table 2. Quantitative results of ablation study on RDD and Kligler using RMSE, PSNR, and SSIM.

Methods	RDD			Kligler		
	RMSE ↓	PSNR ↑	SSIM ↑	RMSE ↓	PSNR ↑	SSIM ↑
BASE_1	2.942	34.821	0.938	6.253	28.267	0.944
BASE_2	2.897	35.976	0.945	5.811	28.895	0.947
BGShadowNet_1	2.603	36.052	0.980	5.805	28.371	0.944
BGShadowNet_2	2.583	36.135	0.981	5.731	29.035	0.947
BGShadowNet_3	2.433	36.681	0.982	5.538	29.180	0.947
BGShadowNet_4	2.344	37.049	0.982	5.633	28.840	0.948
BGShadowNet	**2.219**	**37.585**	**0.983**	**5.377**	**29.176**	**0.948**

5.3. Ablation Study

Ablation of BGShadowNet. To further evaluate the performance of each component applied in our BGShadowNet, we perform ablation experiments using six variants (with or without the specific component) as follows:
(1) *BASE*$_1$: one DenseUnet;
(2) *BASE*$_2$: two stacked DenseUnet;
(3) *BGShadowNet*$_1$: BGShadowNet without StageII;
(4) *BGShadowNet*$_2$: BGShadowNet without DEModule and BAModule;
(5) *BGShadowNet*$_3$: BGShadowNet without BAModule;
(6) *BGShadowNet*$_4$: BGShadowNet without DEModule.

We train the six variants and evaluate the results on RDD and Kligler's datasets. The results are summarized in Table 2. From the table, we can observe that the components embedded in the two stages can improve the shadow removal results. We also provide qualitative results in Figure 11, which shows that our BGShadowNet with all the components can produce more realistic results.

Ablation of Background Image. To verify the effectiveness of the spatially varying background, we rebuild BGShadowNet exploiting the constant background (denoted as BGShadowNet_b) estimated by BEDSR-Net [24], and reorganize BEDSR-Net employing our spatially varying background (denoted as BEDSR_b). Table 3 summarizes the comparison results. From the results, we can observe that, variant BEDSR_b using the spatially varying background can obtain better values than BEDSR-Net. Besides, our BGShadowNet can produce better results than that using the constant background. These results show that our spatially varying background contributes to superior results.

Table 3. Quantitative comparisons using different background images on RDD dataset.

Methods	RMSE ↓	PSNR ↑	SSIM ↑
BEDSR-Net	2.937	34.928	0.973
BEDSR_b	2.771	35.555	0.976
BGShadowNet_b	2.740	35.821	0.980
BGShadowNet	**2.219**	**37.585**	**0.983**

Limitation. Our BGShadowNet can effectively remove shadows in document images. However, when the images are corrupted by heavy noise, our shadow removal results may contain some residual noise, resulting in uneven illumination with the surrounding environment.

6. Conclusion

In this paper, we propose a CBENet to estimate a spatially varying background for the shadow image, which can facilitate the proposed BGShadowNet performs document shadow removal. Our BGShadowNet first predicts a coarse shadow-removal result using a background-constrained decoder. Then, we embed a BAModule and a DEModule into the encoder-decoder network to improve the coarse result and produce the final shadow-free result with a consistent appearance and detail-rich texture. Experiments comparing our BGShadowNet to state-of-the-art approaches demonstrate its superiority.

Acknowledgments

This work is partially supported by NSFC (No.61902286, No.61972299, No.U1803262, No.61972298) and CAAI-Huawei MindSpore Open Fund.

References

[1] E. Arbel and H. Hel-Or. Shadow removal using intensity surfaces and texture anchor points. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 33(6):1202–1216, 2011. 2

[2] S. Bako, S. Darabi, E. Shechtman, J. Wang, and P. Sen. Removing shadows from images of documents. In *ACCV*, pages 173–183, 2016. 1, 2, 3, 4, 7

[3] M. S. Brown and Y. C. Tsoi. Geometric and shading correction for images of printed materials using boundary. *IEEE Transactions on Image Processing*, 15(6):1544–1554, 2006. 1, 2

[4] Cen Chen, Kenli Li, Sin G Teo, Xiaofeng Zou, Keqin Li, and Zeng Zeng. Citywide traffic flow prediction based on multiple gated spatio-temporal convolutional neural networks. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 14(4):1–23, 2020. 1

[5] Zipei Chen, Chengjiang Long, Ling Zhang, and Chunxia Xiao. Canet: A context-aware network for shadow removal. In *ICCV*, pages 4743–4752, 2021. 2, 6, 7

[6] Xiaodong Cun, Chi-Man Pun, and Cheng Shi. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan. In *AAAI*, pages 10680–10687, 2020. 1, 2, 6, 7

[7] B. Ding, C. Long, L. Zhang, and C. Xiao. Argan: Attentive recurrent generative adversarial network for shadow detection and removal. In *ICCV*, pages 10213–10222, 2020. 1, 2

[8] G. D. Finlayson, M. S. Drew, and C. Lu. Entropy minimization for shadow removal. *International Journal of Computer Vision*, 85(1):35–57, 2009. 2

[9] G. D Finlayson, S. D Hordley, C. Lu, and M. S Drew. On the removal of shadows from images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):59–68, 2006. 2

[10] Gang Fu, Lian Duan, and Chunxia Xiao. A hybrid $l_2 - l_p$ variational model for single low-light image enhancement with bright channel prior. In *2019 IEEE International Conference on Image Processing (ICIP)*, 2019. 1

[11] Lan Fu, Changqing Zhou, Qing Guo, Felix Juefei-Xu, Hongkai Yu, Wei Feng, Yang Liu, and Song Wang. Auto-exposure fusion for single-image shadow removal. In *CVPR*, pages 10571–10580, 2021. 2, 6, 7

[12] Basilios Gatos, Ioannis Pratikakis, and Stavros J Perantonis. Adaptive degraded document image binarization. *Pattern Recognition*, 39(3):317–327, 2006. 1

[13] Maciej Gryka, Michael Terry, and Gabriel J. Brostow. Learning to remove soft shadows. *Acm Transactions on Graphics*, 34(5):1–15, 2015. 1

[14] Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6):1397–1409, 2013. 4

[15] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, Jing Qin, and Pheng-Ann Heng. Direction-aware spatial context features for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(11):2795–2808, 2020. 2, 6, 7

[16] X. Hu, Y. Jiang, C. W. Fu, and P. A. Heng. Mask-shadowgan: Learning to remove shadows from unpaired data. In *ICCV*, pages 2472–2481, 2019. 1, 2

[17] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *CVPR*, pages 1125–1134, 2017. 4

[18] Yeying Jin, Aashish Sharma, and Robby T Tan. Dc-shadownet: Single-image hard and soft shadow removal using unsupervised domain-classifier guided network. In *ICCV*, pages 5027–5036, 2021. 2

[19] Seungjun Jung, Muhammad Abul Hasan, and Changick Kim. Water-filling: An efficient algorithm for digitized document shadow removal. In *ACCV*, pages 398–414, 2018. 1, 2, 3, 6, 7

[20] S. H. Khan, M Bennamoun, F Sohel, and R Togneri. Automatic shadow detection and removal from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(3):431–446, 2016. 1

[21] N. Kligler, S. Katz, and A. Tal. Document enhancement using visibility detection. In *CVPR*, pages 2374–2382, 2018. 2, 3, 6

[22] Hieu Le and Dimitris Samaras. From shadow segmentation to shadow removal. In *ECCV*, pages 264–281, 2020. 1

[23] Hieu Le and Dimitris Samaras. Shadow removal via shadow image decomposition. In *ICCV*, pages 8578–8587, 2020. 2

[24] Y. H. Lin, W. C. Chen, and Y. Y. Chuang. Bedsr-net: A deep shadow removal network from a single document image. In *CVPR*, pages 12905–12914, 2020. 1, 2, 3, 4, 6, 7, 8

[25] Feng Liu and Michael Gleicher. Texture-consistent shadow removal. In *ECCV*, pages 437–450, 2008. 2

[26] Z. Liu, H. Yin, Y. Mi, M. Pu, and S. Wang. Shadow removal by a lightness-guided network with training on unpaired data. *IEEE Transactions on Image Processing*, 30. 2

[27] Z. Liu, H. Yin, X. Wu, Z. Wu, Y. Mi, and S. Wang. From shadow generation to shadow removal. In *CVPR*, page 4927C4936, 2021. 1, 2

[28] Gaofeng Meng, Shiming Xiang, Nanning Zheng, and Chunhong Pan. Nonparametric illumination correction for scanned document images via convex hulls. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7):1730–1743, 2012. 1

[29] Daniel Marques Oliveira and Rafael Dueire Lins. A new method for shading removal and binarization of documents acquired with portable digital cameras. In *Third International Workshop on Camera-Based Document Analysis and Recognition*, pages 3–10, 2009. 1

[30] D. M. Oliveira, R. D. Lins, and Gabriel De Frana Pereira E Silva. Shading removal of illustrated documents. In *ICIAR*, 2013. 1, 2

[31] L. Qu, J. Tian, S. He, Y. Tang, and Rwh Lau. Deshadownet: A multi-context embedding deep network for shadow removal. In *CVPR*, pages 4067–4075, 2017. 2

[32] N Bharath Raj and N Venkateswaran. Single image haze removal using a generative adversarial network. In *CVPR*, pages 37–42, 2018. 4

[33] Vatsal Shah and Vineet Gandhi. An iterative approach for shadow removal in document images. In *ICASSP*, pages 1892–1896, 2018. 1

[34] Yael Shor and Dani Lischinski. The shadow meets the mask: Pyramid-based shadow removal. 27(2):577–586, 2008. 2

[35] Oleksii Sidorov. Conditional gans for multi-illuminant color constancy: Revolution or yet another approach? In *CVPRW*, pages 1748–1758, 2019. 2

[36] Jin Wan, Hui Yin, Zhenyao Wu, Xinyi Wu, Yanting Liu, and Song Wang. Style-guided shadow removal. In *ECCV*, 2022. 7

[37] Bingshu Wang, Shuang Feng, and CL Philip Chen. Strong shadow removal of text document images based on background estimation and shading scale. In *ICCSS*, pages 738–742, 2020. 1

[38] J. Wang, X. Li, and J. Yang. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *CVPR*, pages 1788–1797, 2018. 2

[39] Jifeng Wang, Xiang Li, and Jian Yang. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *CVPR*, pages 1788–1797, 2018. 7

[40] Jinjiang Wei, Chengjiang Long, Hua Zou, and Chunxiad Xiao. Shadow inpainting and removal using generative adversarial networks with slice convolutions. *Computer graphics forum*, 38(7):381–392, 2019. 1

[41] Tai Pang Wu, Chi Keung Tang, Michael S. Brown, and Heung Yeung Shum. Natural shadow matting. *Acm Transactions on Graphics*, 26(2):8, 2007. 1

[42] Chunxia Xiao, Ruiyun She, Donglin Xiao, and Kwan Liu Ma. Fast shadow removal using adaptive multi-scale illumination transfer. *Computer Graphics Forum*, 32(8):207–218, 2013. 2

[43] Chunxia Xiao, Donglin Xiao, Ling Zhang, and Lin Chen. Efficient shadow removal using subregion matching illumination transfer. *Computer Graphics Forum*, 32(7):421–430, 2013. 2

[44] Yao Xiao, Efstratios Tsougenis, and Chikeung Tang. Shadow removal from single rgb-d images. In *CVPR*, pages 3011–3018, 2014. 1

[45] Qingxiong Yang, Kar Han Tan, and Narendra Ahuja. Shadow removal using bilateral filtering. *IEEE Transactions on Image Processing*, 21(10):4361–4368, 2012. 1

[46] Yibing Yang and Hong Yan. An adaptive logical method for binarization of degraded document images. *Pattern Recognition*, 33(5):787–807, 2000. 1

[47] Ling Zhang, Chengjiang Long, Xiaolong Zhang, and Chunxia Xiao. Ris-gan: Explore residual and illumination with generative adversarial networks for shadow removal. In *AAAI*, pages 12829–12836, 2020. 2

[48] Li Zhang, Andy M Yip, Michael S Brown, and Chew Lim Tan. A unified framework for document restoration using inpainting and shape-from-shading. *Pattern Recognition*, 42(11):2961–2978, 2009. 1

[49] Li Zhang, Andy M Yip, and Chew Lim Tan. Removing shading distortions in camera-based document images using inpainting and surface fitting with radial basis functions. In *ICDAR*, volume 2, pages 984–988, 2007. 2

[50] Ling Zhang, Qing Zhang, and Chunxia Xiao. Shadow remover: Image shadow removal based on illumination recovering optimization. *IEEE Transactions on Image Processing*, 24(11):4623–36, 2015. 2

[51] Yurui Zhu, Jie Huang, Xueyang Fu, Feng Zhao, Qibin Sun, and Zheng-Jun Zha. Bijective mapping network for shadow removal. In *CVPR*, pages 5627–5636, 2022. 7