

Shadow Removal via Shadow Image Decomposition

Name – Aryaman Patel, Venkatesh Ashok Desai

----

1. Abstract:

In this project, we delve into a deep learning approach for the challenging task of shadow removal. Inspired by the principles governing the formation of shadows, our method leverages a linear illumination transformation to accurately model the effects of shadows in images. This approach allows us to represent a shadow image as a combination of the shadow-free image, shadow parameters, and a matte layer. To realize this vision, we employ two distinct deep networks, SP-Net and M-Net, tasked with predicting the shadow parameters and shadow matte, respectively. The resulting system equips us with the capability to effectively eliminate shadow effects from images, significantly enhancing their visual quality. Our extensive evaluation and experimentation center on the ISTD dataset, renowned as one of the most challenging benchmarks for shadow removal.

----

2. Introduction and prior work:

This project is inspired by the paper Hieu Le, Dimitris Samaras, "Shadow Removal via Shadow Image Decomposition", ICCV19, 2019. The paper introduces an innovative approach to shadow removal that combines both shadow illumination modelling and deep learning techniques. Departing from earlier methods for shadow removal, they propose the use of a simplified physical illumination model to establish a connection between shadow pixels and their corresponding shadow-free counterparts.

The illumination model comprises a linear transformation characterized by scaling factors and additive constants, unique to each color channel, across the entire umbra region of the shadow. These scaling factors and additive constants serve as the model's parameters, as illustrated in the below Figure 1. The illumination model plays a pivotal role in the approach, enabling them to eliminate shadows from images when they accurately estimate these model parameters.

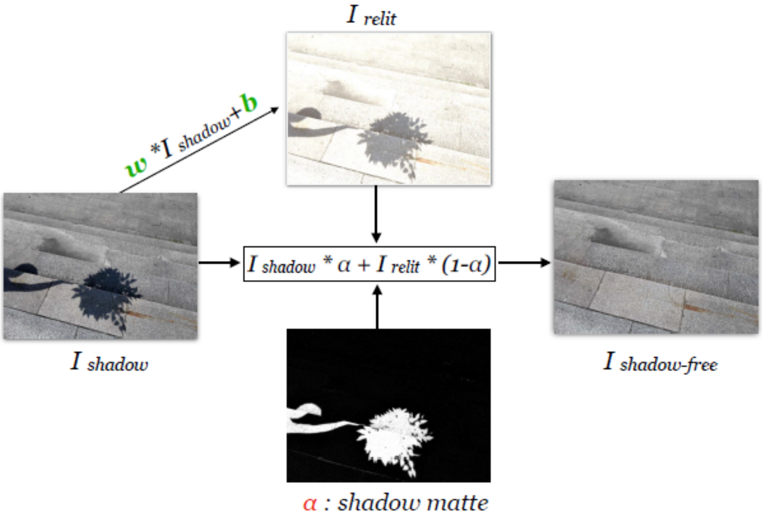

Fig 1 Shadow Removal via Shadow Decomposition

To achieve this, they introduce a deep network called SP-Net, trained to estimate the parameters of the shadow model. Through training, SP-Net learns a mapping function that predicts illumination model parameters from input shadow images. Additionally, they employ a shadow matting technique to address the penumbra area of shadows. They integrate the illumination model into an image decomposition framework, where the shadow-free image is expressed as a combination of the shadow image, the shadow model parameters, and a shadow density matte. This decomposition framework allows to reconstruct the shadow-free image.

The shadow parameters (w; b) define the transformation from shadowed pixels to illuminated pixels, while the shadow matte represents a per-pixel linear combination of the relit image and the shadow image, ultimately resulting in the shadow-free image. In contrast to previous approaches that often require user assistance or solving optimization systems to obtain shadow mattes, they propose training a second network, M-Net, to accurately predict shadow mattes.

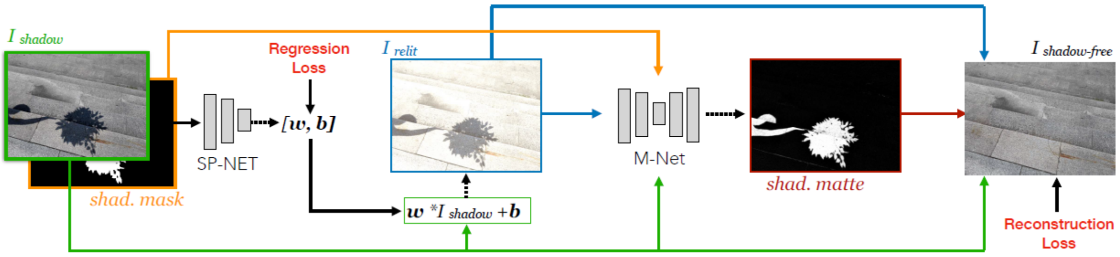

Fig 2 Shadow Removal Framework

We evaluate our proposed SP-Net and M-Net on the ISTD dataset, which stands as the most extensive and challenging dataset for shadow removal. The ISTD dataset encompasses image triplets—comprising shadow images, shadow masks, and shadow-free images—captured in diverse scenes. The training subset encompasses 1870 image triplets from 135 scenes, while the testing subset comprises 540 triplets from 45 scenes. However, it's important to note that the testing set of the ISTD dataset requires adjustments due to color inconsistencies between the shadow images and their corresponding shadow-free counterparts. This well-documented issue, mentioned in the original paper, arises from the images being captured at different times of the day, resulting in slight variations in environmental lighting conditions. In order to mitigate this color inconsistency, they use linear regression to transform the pixel values in the nonshadow area of each shadow-free image to map into their counterpart values in the shadow image. They use a linear regression for each color-channel, similar to the method for relighting the shadow pixels. This simple transformation transfers the color tone and brightness of the shadow image to its shadow-free counterpart.

Notably, the paper claims that the model surpasses state-of-the-art methods, delivering a remarkable 40% reduction in root mean square error (RMSE) for shadow areas. To improvise the model the authors, increase the training dataset by generating synthetic images. These images are generated by estimating shadow parameters and shadow mattes from an image, and then reintroduce the shadows into the shadow-free image using adjusted shadow parameters. By manipulating these parameters, the shadow effects can be precisely controlled. This allows to generate additional shadow images, which can serve as augmented training data. The paper claims that training the system on the ISTD dataset, augmented with these synthesized images, results in a 6% reduction in RMSE on shadowed areas compared to training solely on the original ISTD dataset.

----

3 .<u>Methods</u>

As stated above, SP-Net and M-Net are the two neural networks trained to obtain shadow-free image. Introducing SP-Net, a deep neural network designed for the task of estimating the parameters of the shadow model. During its training process, SP-Net acquires the capability to learn a mapping function that predicts illumination model parameters based on input shadow images. The shadow parameters, denoted as (w; b), play a crucial role in transforming shadowed pixels into illuminated ones, while the shadow matte represents a per-pixel linear combination of the relit image and the shadow image, ultimately resulting in the production of the shadow-free image. M-Net, is specifically designed to make accurate predictions of shadow mattes. This innovation eliminates the need for user intervention or complex optimization processes often required by previous approaches.

SP-Net

In the paper, they have used a ResNet architecture but we used a VGG model. It's input is a 4-channel tensor derived from the shadow image and the shadow mask. The network consists of 13 convolutional layers and three fully connected layers. All of the convolutional layers have a filter size of 3x3 and a stride of 1. Max pooling layers are used after every two convolutional layers to reduce the spatial dimensions of the feature maps. Regression loss is used to train the model using this neural-network.

SP-Net produces 6 feature maps as its output which are the parameters 'w' and 'b' to compute Image_relit. For a 256x256 input image, the output feature map is 1x1 in size. Importantly, SP-Net is fully-convolutional, making it adaptable to images of varying sizes.

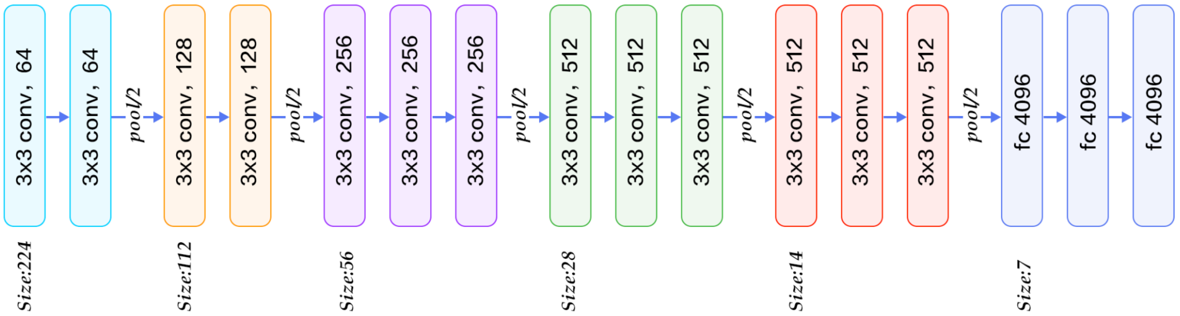

Fig 3. SP-Net Architecture

M-Net

The M-Net is structured on the foundation of the U-Net architecture, comprising four skip-connection modules, an input layer, and an output layer. Each skip-connection module consists of two branches: a down-branch and an up-branch. The down-branch comprises a sequence of operations, starting with a Leaky-ReLU activation function (with α = 0.2), followed by a convolutional layer using a (4x4) kernel, a (2x2) stride, and (1x1) padding, and finally, a Batch Normalization layer.

The up-branch, mirroring the down-branch, includes a Leaky-ReLU activation, a deconvolutional layer, and a Batch Normalization layer, all with identical parameter configurations as those in the down-branch. Importantly, each up-branch takes its input not only from the preceding layer but also from the corresponding down-branch, ensuring a comprehensive flow of information throughout the network.

Each box in the below diagram corresponds to an output from either a down-branch (depicted in yellow) or an up-branch (depicted in gray). These boxes represent multi-channel feature maps, and their respective sizes are indicated within the box as white text. Additionally, three values are associated with each branch, representing the input channel count, the output channel count, and the stride, respectively. Reconstruction loss is used during training of this neural-network.

Notably, the diagram employs dotted arrows to signify copy operations. The initial layer is a convolutional layer that takes as input a tensor of size Cin x h x w, where Cin is 7, signifying a stack of the shadow image, the relit image, and the shadow mask. Finally, the last layer is a deconvolutional layer that produces a shadow matte with dimensions of 3 x 256 x 256.

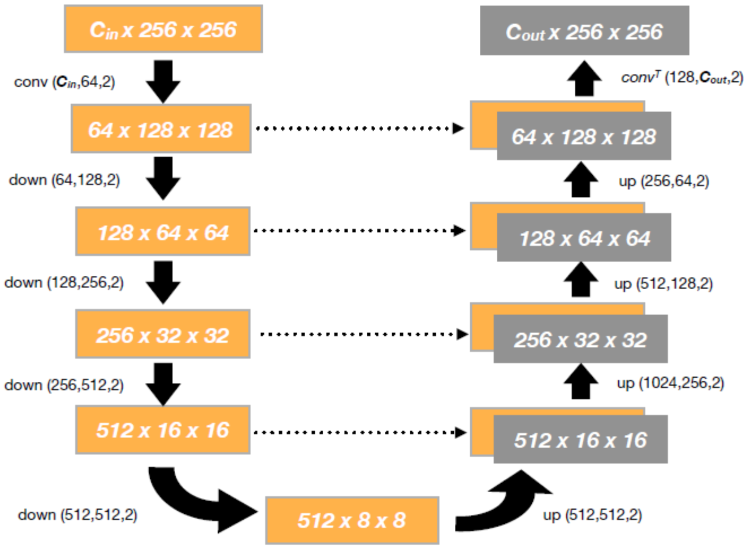

Fig 4 M-Net Architecture

---

4. Results

We conducted training on the model using the ISTD dataset, which comprises 102 images. The training dataset includes Shadow Images, Shadow Masks, Ground Truth Images (images devoid of shadows), and shadow parameters. The SP-Net is trained employing the VGG16 network architecture, while the M-Net is trained using the U-Net network architecture. The model underwent 200 epochs of training, with a batch size set to 16.

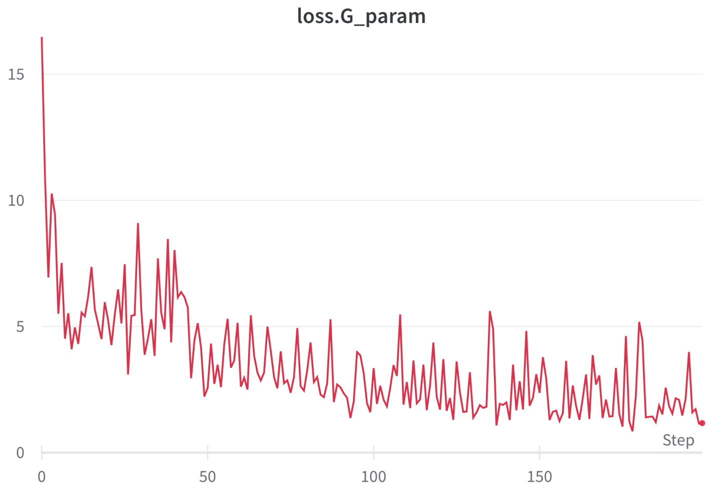

Fig 5 Regression Loss

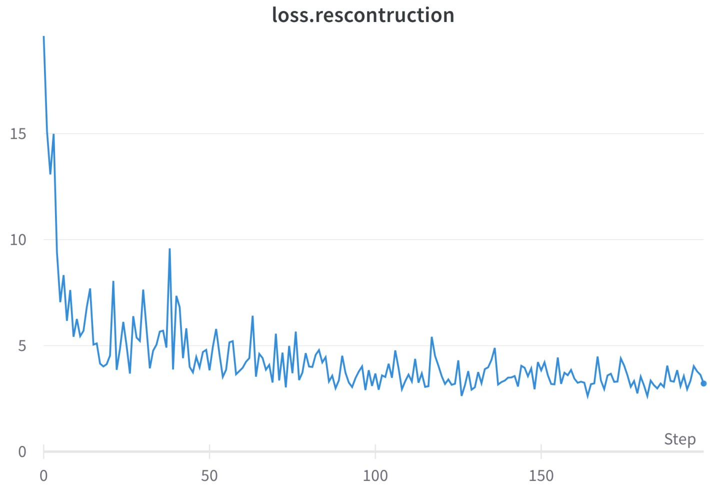

Fig 6 Reconstruction Loss

Subsequently, the model underwent testing on our designated test dataset, resulting in a root mean square error (RMSE) of 4.4. This RMSE value is in line with the performance of Gong et al.'s method, which achieved an RMSE of 4.2. Worth noting is that Gong et al.'s method involves user interaction to define shadow and non-shadow regions, thus yielding minimal error in non-shadow areas.

The output images of our model are as follows:

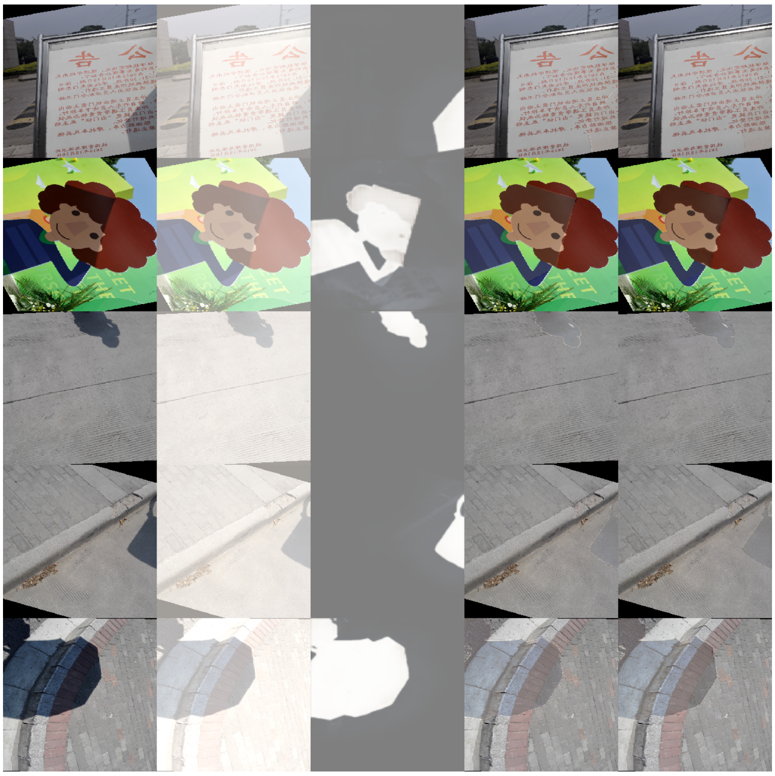

Fig 7. The above images are Input Image, Relit Image, Shadow Mask, only SP-Net output, SP-Net + MNet output (from left to right)

In the image above, we have a sequence of visual elements: the initial input image (the shadow image), the Relit Image generated using 'w' and 'b' parameters produced by the SP-Net, the Shadow Mask, and the SP-Net Output (which is the shadow-free image created using the Input Image, Relit Image, and Shadow Mask). The final image, as our ultimate result, is computed by combining the Input Image with the shadow matte (the output of the M-Net) and the Relit Image, following the formula below:

$$I^{\text{shadow-free}} = I^{\text{shadow}} \cdot \alpha + I^{\text{relit}} \cdot (1 - \alpha)$$

where I_shadow and I_shadow-free are the shadow and shadowfree image respectively, \alpha is the matting layer, and I_relit is the relit image.

It's evident that the SP-Net tends to excessively enhance the lighting in shadowed regions, but the shadow matte derived from M-Net effectively corrects these issues. This occurs because M-Net is specifically trained to seamlessly blend the relit and shadow images, resulting in a more accurate shadow-free image. On the whole, the model excels at estimating overall illumination changes without introducing color inconsistencies, blurriness, or random artifacts in the relit area.

However, it's important to note that the model isn't flawless; in some cases, it may either excessively brighten the shadowed areas or produce incorrect colorations, as demonstrated in the second example.

----

5 .Reflection and acknowledgements

Learnings:

1. Understanding the challenges of shadow removal and the limitations of traditional methods.

2. Learning about the concept of deep learning and how it can be used for shadow removal.

3. Understanding the architecture of the proposed model, including the SP-Net and M-Net, and how they work together to remove shadows.

4. Learning about the importance of data augmentation and how it can improve the performance of a model.

5. Gaining experience in implementing deep learning models using popular frameworks such as PyTorch or TensorFlow.

Assistance and Resources -

1. Paper - https://arxiv.org/abs/2012.13018
2. Official Github Repository - https://github.com/cvlab-stonybrook/SID
3. Official Research paper website https://www3.cs.stonybrook.edu/~cvl/projects/SID/index.html
4. https://chat.openai.com/
5. Northeastern Cluster - https://ood.discovery.neu.edu/