---------------------------------------------------------------------------------------------------------

**Shadow Removal**

Name – Aryaman Patel, Venkatesh Ashok Desai

---------------------------------------------------------------------------------------------------------

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.

Furthermore, we introduce an augmented ISTD dataset generated through an image decomposition system, which modifies shadow parameters to produce synthetic shadow images. By training our model on this augmented dataset, we push the boundaries even further, resulting in a lowered RMSE of 7.4 for the shadow areas.

-----------

SP-Net, without matting, surpasses the state-of-the-art method by 29% in terms of RMSE (Root Mean Square Error) on shadowed areas, reducing it from 13.3 to 9.5 RMSE. When both SP-Net and M-Net are utilized, our system further enhances overall results by an additional 17%, resulting in a RMSE of 7.9.

**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 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.

A diagram of a shadow

Description automatically generated

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.

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 of 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.

-----------

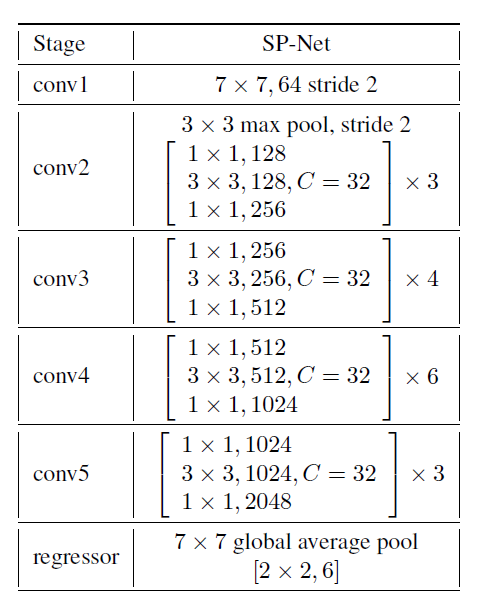
**Methods**

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**

The SP-Net is inspired by the ResNeXt model and is outlined below. It comprises 6 blocks, and its input is a 4-channel tensor derived from the shadow image and the shadow mask. Each stage (excluding "conv1" and "regressor") is composed of repeated building blocks. Convolutional layers use a stride of 1x1, and each is followed by Batch Normalization and a Rectified Linear Unit. The number of convolutional layers is denoted in the format "1x1, 128," which signifies 128 layers with a 1x1 kernel size, and "C=32" indicates the use of grouped convolutions with 32 groups.

SP-Net produces 6 feature maps as its output. 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.



**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.

We provide a concise summary and visual representation of the architecture of our M-Net. This diagram also illustrates the dimensions of the intermediate feature maps when handling a 256x256 input tensor. Each box in the 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.

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.

A diagram of a number of objects

Description automatically generated with medium confidence

-----------

Results: Show the result of your work on multiple appropriate images, showing both the strengths and weaknesses of the method. Include a short description with each result image indicating anything special about it.

-----------

Reflection and acknowledgements: reflect on what you learned and acknowledge any assistance or resources you received or used during the project.

-----------