Compare Deep Learning and Image processing methods For removing shadows from images

1st Puranjay Datta
19D070048
19D070048@iitb.ac.in

2nd Raja Kumar
190110070
190110070@iitb.ac.in

***Abstract*—The Deep Learning Model is multi-tasking which jointly learns both shadow detection and shadow removal. The framework is based on a novel Stacked Conditional Generative Adversarial Network (ST-CGAN), composed of two stacked CGANs, each with a generator and a discriminator. Essentially, a shadow image is given as input into the first generator, which provide a shadow detection mask. That shadow image, merged with its predicted mask, goes through the second generator to finally get back its shadow-free image. Moreover, the two corresponding discriminators will model the higher-level relationships and global scene characteristics for the detected shadow region and reconstruction via removing shadows, respectively.**
The image processing method rectifies illumination using diffusion model and the traditional binarization using K-means clustering.For the documents the traditional image processing methods are superior but the time complexity is a issue where as for natural images the Neural net gives better results.

I. Introduction

Deep-Learning-Both shadow detection and shadow removal reveal their particular advantages for scene understanding. The accurate recognition of shadow area (i.e., shadow detection) supplies good clues about the light sources, illumination conditions, object shapes, and geometry information. Meanwhile, removing shadows (i.e., shadow removal) in images is of great interest for the downstream computer vision tasks, such as efficient object detection and tracking.
Detection only: Their main objective is to search out the shadow regions, in a form of an image mask that set apart shadow and non-shadow areas.
Removal only: Produces the illumination attenuation effects on the whole image, which is also denoted as a shadow matte, to find again the image with shadows removed naturally.
Two stages for removal. Many of the shadow removal methods generally include two divided steps: shadow localization and shadow-free reconstruction by taking advantage of the intermediate results in the awareness of shadow regions.

Image processing Methods-With the increase in usage of mobiles for taking selfies and pictures,the need for increasing the robustness of image quality is the need of the hour.The online semester brought with its own challenge of online exam where the hand-clicked documents had shadows and illumination problems.Therefore such shadow removal techniques will greatly help in improving the image quality and provide better visual appearance.Figures show three examples of illumination distorted digitized documents captured in different lighting conditions using smartphone's cameras.

Fig. 1. Poor illumination

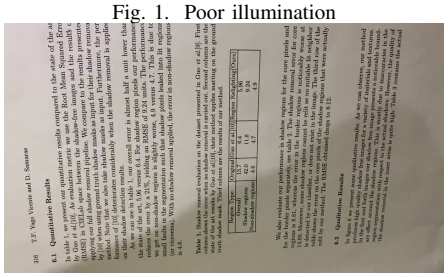


Fig. 2. Dark shadow

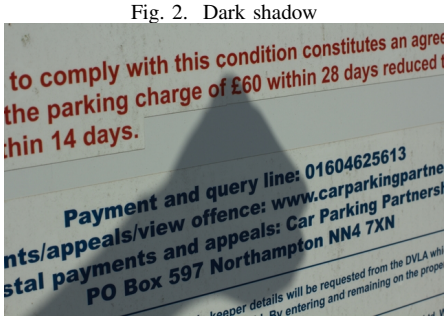


Fig. 3. Natural Image

II. Background and prior work

A. Deep-Learning

Besides, most previous methods, including shadow detection and removal, are heavily based on local region classifications or low-level feature representations, failing to reason about the global scene semantic structure and illumination

conditions. Consequently, a most recent study in shadow detection introduced a Conditional Generative Adversarial Network (CGAN), which proved to be effective for global consistency.

B. Image Processing Method

Classical binary image classification-One of the basic building blocks is the thresholding technique where each pixel is categorized into black or white pixel.The authors proposed a locally adaptive binarization technique based method which computes a threshold using local information.But this did not help much in digitized documents involving shadow as both the text and shadow will become black and thus their separation will become difficult.

Background Shading Estimation Based Approaches-In background shading estimation based methods, a digitized document is assumed to have two separate layers. The background layer–the shading layer which contains illumination distortions, and the foreground layer–the layer which contains the text and images.

Mask and Interpolation-A mask is created to cover the text in the documents using a canny edge detector followed by using morphological closing operation.This works well with documents of similar font sizes but has poor results with varying font size documents in poor illumination.Another method is to use the binarized image as a mask. However, the generated masks often fail to cover the photo regions on the digitized documents accurately.

III. Data and Methodology

A. Deep-Learning

Datasets-Trained the model on the ISTD dataset. The Image Shadow Triplets dataset (ISTD) is a dataset for shadow understanding that contains 1870 image triplets of shadow image, shadow mask, and shadow-free image. Tested on the SBU shadow dataset which have 5 thousand images containing shadows from a wide variety of scenes and photo types. Annotations in the form of shadow binary masks are provided along with the actual images. The shadow label annotations for the 4K images in the training set are the result of applying our proposed label recovery method to reduce label noise. Whereas, the testing images were carefully annotated manually to produce precise shadow masks.

Methodology-Generative Adversarial Networks (GANs) is a framework for estimating generative models via an adversarial process, in which we simultaneously train two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G. The training procedure for G is to maximize the probability of D making a mistake. This framework corresponds to a minimax two-player game. Conditional Generative Adversarial Networks (CGANs) extend GANs by introducing additional observed information, named conditioning variable, to generator G and discriminator D. Our ST-CGAN consists of two Conditional GANs in which the second one is stacked upon the first. For the first CGAN of ST-CGAN in Figure 2, the generator G1 and discriminator D1 are conditioned on the input RGB shadow image x. G1 is trained to output the corresponding shadow mask G1(z, x), where z is the randomly sampled noise vector. It is basically a two-player zero-sum game. The first player is a team consisting of two generators (G1, G2). The second player is a team containing two discriminators (D1, D2). To defeat the second player, the first team members are encouraged to produce outputs that are close to their corresponding ground truths. The code is based on PyTorch. We train ST-CGAN with the Adam solver and an alternating gradient update scheme is applied. Specifically, we first adopt a gradient ascent step to update D1, D2 with G1, G2 fixed. We then apply a gradient descent step to update G1, G2 with D1, D2 fixed. We initialize all the weights of ST-CGAN by sampling from a zero-mean normal distribution with standard deviation 0.2. During training, augmentations are adopted by cropping (image size 286 $\rightarrow$ 256) and flipping (horizontally) operations. A practical setting for , where 1 = 5, 2 = 0.1, 3 = 0.1, is used. The Binary Cross Entropy (BCE) loss is assigned for the objective of image mask regression and L1 loss is utilized for the shadow-free image reconstruction respectively.

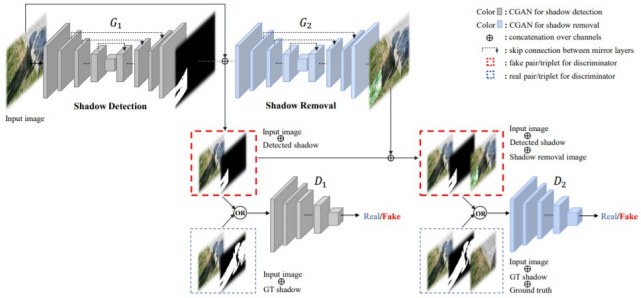


Figure 2. The architecture of the proposed ST-CGAN. It consists of two stacked CGANs: one for shadow detection and another for shadow removal, which are marked in different colors. The intermediate outputs are concatenated together as the subsequent components' input.

B. Image Processing Method

Datasets-Tried on images from SBU-shadow datasets available on Kaggle which had colored objects and natural images and Document shadow removal dataset available on github which had documents.

Method1-Three types of pixels present in a document namely shadow,text,global background image which can be separated using threshold by K-means clustering.The lowest value is assigned to text as its dark and highest value is assigned as global background for every channel(RGB) and a binarized image is created with text and equalized background.

Method2-Water filling algorithm Modelling by diffusion equation- We use two different methods to simulate this water-filling task: incrementally filling of catchment basins method, and flood-and-effuse method.
Let h be the altitude of the topographic surface, $w(x,t)$ be the water level on the topographic surface x at a point of time t.

$$w(x,t) \geq 0, \forall x \in D_I$$

$$w(x,t) = 0, \forall x \in \partial D_I$$

Equation (1) limits water level to non-negative value since water either be stored or flow out. Equation (2) describes the drop of water at the image boundary so that only the catchment basins at the interior region of the image can be filled.

$$\frac{\partial w(x,t)}{\partial t} = -\nabla J_w(x,t)$$

where J_w is the flux of the diffusing water. Equation (3) states that a change in water level in any part of the structure is due to inflow and outflow of water into and out of that part of the structure.

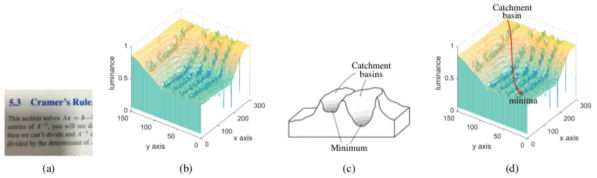


Fig. 2. A sample digitized image and its topographic surface visualization. (a) Example digitized document, (b) topographic surface representation of (a). (c) An illustration of catchment basins and corresponding minimum. (d) An example of catchment basin and the corresponding minima on a digitized document representing (c).

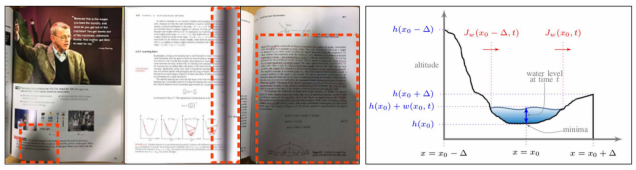


Fig. 3. (left) Examples of shaded region, marked in orange dashed rectangles, touching the boundary of the images. (right) An one dimensional cartoon topographic model of a catchment basin for illustrating the diffusion equation for document illumination correction.

Incremental filling of Catchment Basins-Let $G(x_0,t) = h(x_0) + w(x_0,t)$ and $J_w(x_0,t) \propto G(x_0,t) - G(x_0+\Delta,t)$,Similarly for 2D case $G(x,y,t+\Delta) = \eta\{G(x+\Delta,y,t)+G(x-\Delta,y,t)+G(x,y+\Delta,t)+G(x,y-\Delta,t)-4G(x,t,t)\}+w(x,y,t)$ where η decides the speed of the process.A small η might lead to slow convergence and a large η may lead to divergence.

Flood and Effuse-In this method, the diffusion equation can be decomposed into two independent processes: water effusion, and flood. For the effusion process, we consider the dynamics of water flow on the topographic surface without any external water supply.

$w_\phi(x_0,t) \propto min\{G(x_0+\Delta,t)-G(x_0,t),0\}+min\{G(x_0-\Delta,t)-G(x_0,t),0\}$ The effusive term $w_\phi(x0,t)$ has non-positive value and represents the amount of water to be effused at each location.
The flooding process $w_\Psi(x_0,t)$ is modelled by immersing the entire topography to the same altitude.

$w_\Psi(x_0,t) = (h_{max} - G(x,y,t))e^{-t}$ and $w(x,y,t) = w_\phi(x,y,t) + w_\Psi(x,y,t)$

Given the background layer I_b and foreground layer I_f , the final photometrically correct image, I_r, can be computed as follows.

$$I_r(x,y) = \lim_{t->\infty} \frac{I(x,y).l}{G(x,y,t)}, l \in [0,1]$$

IV. Experiments and Results

A. Image Processing K-means clustering binarization method

Fig. 4. input1

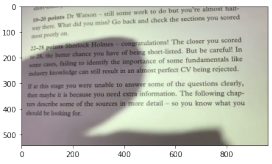


Fig. 5. output1

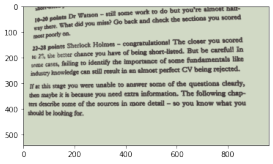


Fig. 6. input2

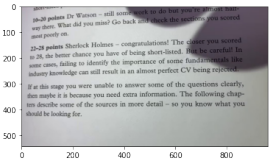


Fig. 7. output2

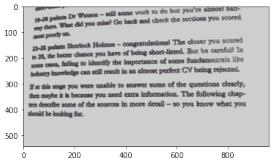


Fig. 8. input3

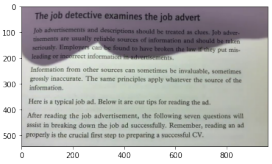
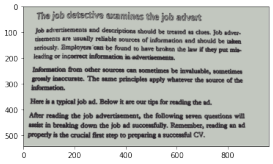


Fig. 9. output3



The K-means clustering algorithm(Figure 10,11,12) clearly shows that there are 3 different clusters which is shown by plotting histogram for R,G,B channel respectively.The lowest intensity cluster belongs to the text and the highest intensity cluster belongs to global background as it is light in color and the medium intensity belongs to the shadow region.

B. Image Processing Water filling algorithm

The text in Figure 14,16,18 written is blurred because of resizing as the time complexity=O(height.width.time) is very large.But the intensity of shadow is reduced but the hue is still a problem.But the intensity profile is greatly improved and very close to the ground truth.

As you can see the Histogram (Figure 23) in orange(the image after shadow correction) is very close to histogram in blue(original image) for the Y channel of YCR_CB.

C. Deep-Learning(Figure 24)

Component analysis of ST-CGAN on ISTD by using RMSE for removal-As the number of epoch increases the RMSE loss for both Generator and Discriminator decreases. As it was taking too much time to train this model on CPU

Fig. 10. Red-channel

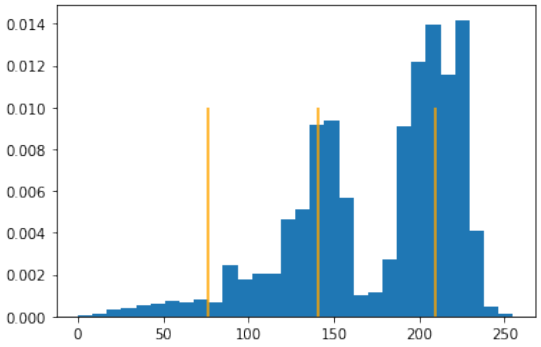


Fig. 11. Green-channel

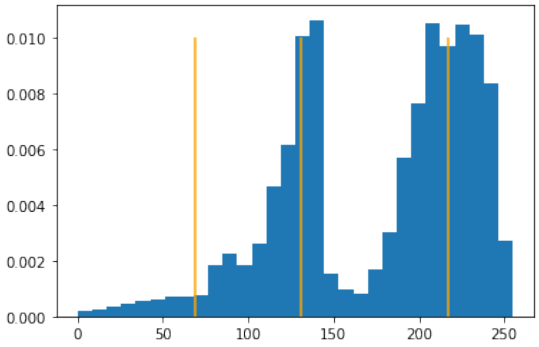


Fig. 12. Blue-channel

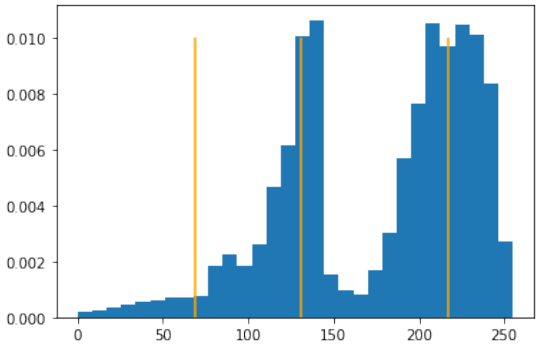


Fig. 13. input1

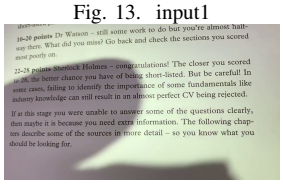


Fig. 14. output1

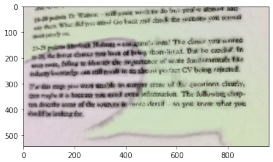


Fig. 15. input2

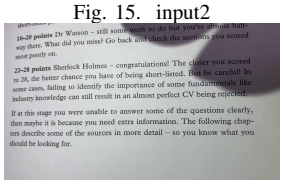


Fig. 16. output2

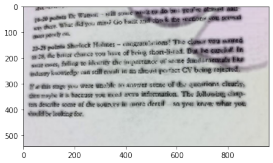


Fig. 17. input3

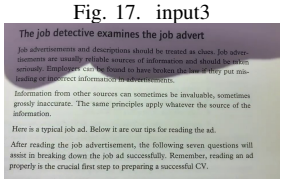


Fig. 18. output3

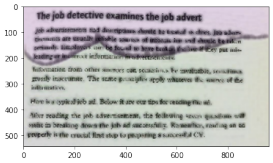


Fig. 19. input3

Fig. 20. output3

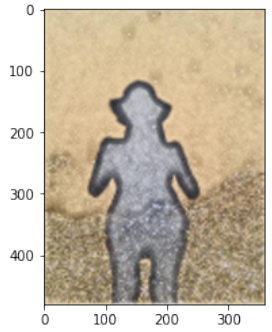


Fig. 21. input3

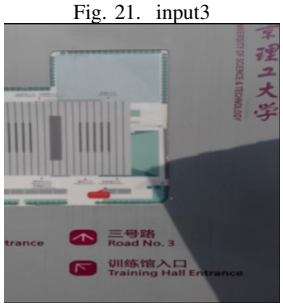
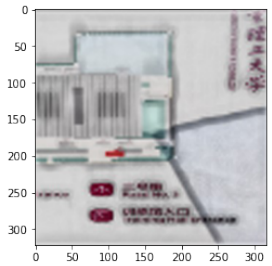


Fig. 22. output3



so we have included the graph for loss versus epoch for just 20 numbers of epochs.

Final Result-The rmse error for digitized documents is around 10.4 for image processing and around 63.6 for Neural nets.

V. Learning, Conclusions, and Future Work

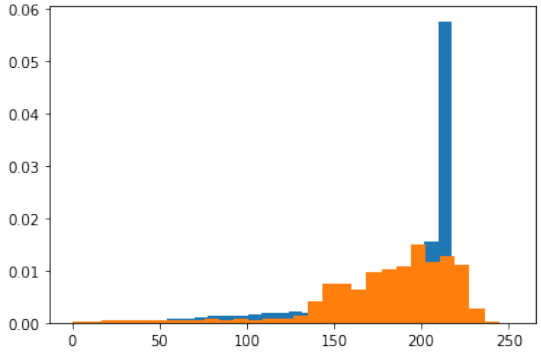
A. Deep-Learning

Read about GANs, C-GANs neural networks. Get to know how a single stacked Conditional GANs can handle both shadow removal and shadow detection. In the future, multiple tasks can be handled by a single neural network efficiently. The accuracy of test images from the ISTD dataset was good as we trained the model on the same dataset. But if we try on different datasets whose representation is low in the training dataset, the model fails to remove the shadow from images; however, it detects the shadow more efficiently.

For example, this is the input image of the document, and after testing on the model, we see the efficient shadow detection part of the input image. Still, the shadow is not entirely erased from the image.

In the future, we can tune the model or try to build that kind of dataset that almost covers all the necessary features of shadow images to perform better on different domains of datasets efficiently.

Fig. 23. Y-channel histogram comparison



B. Image Processing Method

Read about various shadow removal techniques.The simple process of diffusion and effusion can be modelled as shadow and non shadow regions.The methods were simple yet quite effective. The proposed waterfilling algorithm is based on simulating the immersion process of a topographic surface using water.The method gives unsatisfactory results with specular light conditions where the original luminance value of a point in the foreground layer is significantly damaged due to overexposure, it is hard to reconstruct the point using the Lambertian surface model.
Future work-To compensate for Time complexity when we down sample and upsample the image using Bi-cubic interpolation some information is lost and For natural images the results are not very accurate and if we run it separately for R,G,B channels the Hue and shade is completely different.These are the some of the drawbacks which needs to be improved.

VI. Contributions

Puranjay Datta(19D070048)-Worked on the traditional image processing method of water filling algorithm.Tried and tested few other basic methods of kmeans clustering binarization and tried canny edge detector algorithm to find the shadow border and observe a pattern HSV parameters of shadow and non shadow regions and prepared the report and video for his part too.

Raja Kumar(190110070)-Worked on implementing Stacked Conditional Generative Adversarial Networks for Jointly Learning Shadow Detection and Shadow Removal. The model was trained and tested on the ISTD datasets and some custom input images to know the model's shortcomings and prepared the report and video for his part.

References

[1] https://arxiv.org/pdf/1904.09763v2.pdf
[2] https://github.com/BingshuCV/DocumentShadowRemoval
[3] https://www.kaggle.com/sayantandas30011998/sbu-shadow/version/1
[4] https://arxiv.org/pdf/1406.2661.pdf
[5] https://paperswithcode.com/dataset/istd
[6] https://arxiv.org/pdf/1712.02478.pdf
[7] https://www.arxiv-vanity.com/papers/1712.02478/

Fig. 24. Deep-Learning Results