Shadow Removal and Intrinsic Images

CS7180 Image Enhancement Project

Hrigved Suryawanshi, Haard Shah, Matthew Ray
10/19/2024

Abstract

There were two parts to our approach to this project. The first to leverage past work inspired by Yun-Hsuan Lin to implement a model that removes shadows from handwritten documents. The second was to build on work by Chris Careaga and Yagiz Aksoy to identify shadows from intrinsic image decomposition for shadow removal.

Most shadow removal from the text that we encountered was structured from textbooks and did not take into account nuances from handwritten text. As we worked through the initial project of removing shadows from handwritten text documents, we ran into significant challenges getting code to run and switched gears to focus on shadow identification from the formation components of intrinsic images.

In their work with colorful diffuse intrinsic image diffusion, Chris and Yagiz apply a colorful shading method that makes up for the shortcomings of greyscale. This was interesting to us as a color scale may highlight shadow nuances that are not as prominent in greyscale, creating a better image for shadow detection and subsequent removal. To test this, the team implemented shadow detection on the formation components of the intrinsic images.

After overcoming numerous obstacles, we finally succeeded in getting the code to run for our first approach. We will discuss both approaches in detail within this paper.

Introduction and prior work

Inspired by several works on intrinsic image decomposition, shadow identification, and shadow removal, the team wanted to incorporate a method for separating and identifying shadows within the intrinsic image formation components to determine if one component performed better at identifying shadows for shadow removal within the image. Patterns of shadows across images can be arbitrary, varied, and often highly complex. This is especially true for wild images or those captured in an uncontrolled environment, such as images taken outdoors with many objects casting shadows.

In the paper, ‘Colorful Diffuse Intrinsic Image Decomposition in the Wild,’ Chris and Yagiz introduce a method to generate decompositions under the intrinsic residual model for in-the-wild photographs using color shading. This proved to eliminate some shortcomings of accurately modeling colorful lighting effects in greyscale. To achieve this, they start from a decomposition that uses the intrinsic diffuse model and gradually remove the single-color shading and the Lambertian world assumptions to estimate the diffuse albedo and the colorful diffuse shading at high resolutions. With this in mind, the team wanted to test the impact of shadow detection on the different formation components of intrinsic images.

Code from the Computational Photography Lab's repo, Intrinsic, was leveraged for this project. Their repo was cloned, and the model was tested and updated to incorporate shadow identification on the formation component images.

For shadow removal in documents, we were inspired by BEDSR-Net, a shadow removal network, and its application to remove shadows from a document image. BEDSR-Net consists of two sub-networks: BE-Net for estimating the global background color of the document and SR-Net for eliminating shadows. Given most of the sample images we found across different methods, we wanted to see how this model would perform on handwritten text that may vary in size, shape, and color.

Code from Yun-Hsuan et al. was leveraged as a starting point for implementation and testing against custom images of text documents and later applied to a dataset of handwritten records to test and compare the performance of the model on the two different datasets.

Additional References.

- "Colorful Diffuse Intrinsic Image Decomposition in the Wild" - Chris Careaga, Yagiz Aksoy, 2024
- "Image Thresholding" - OpenCV, 2024
- "BEDSR-Net: A Deep Shadow Removal Network from a Single Document Image" - Yun-Hsuan Lin, Wen-Chin Chen, Yung-Yu Chuang, 2020
- https://github.com/CV-Reimplementation/BEDSR-Net-Reimplementation/blob/main/README.md

Method

For shadow identification from intrinsic decomposition formation components, code from the Computational Photography Lab's repo, Intrinsic, was leveraged. The model from Chris and Yagiz leverages an input image to create a shading/albedo pair generated with a simplified grayscale. They then extend the image formation model to include colorful shading and estimate the shading color using a chroma network used as input in the second step, where they estimate the high-resolution diffuse albedo. In the final step, they remove the Lambertian-world assumption and estimate a colorful diffuse shading component and a non-diffuse residual layer.

The repo was cloned, and the model was tested against sample images to validate. We then leveraged an outside dataset to test performance against the provided samples. In testing against an outside dataset, we found the model performed well. However, there were instances where the test dataset did not perform as well as the sample data, particularly in bright light.

After validating the model, the team reviewed several ways to incorporate shadow detection into the pipeline on each of the formation components of the image. Different thresholding methods were tested to segment the shadow. Otsu's thresholding, adaptive thresholding, and K-means clustering were compared to detect shadows from the image. While Otsu's thresholding performed well in certain types of images, K-means with 2 clusters was used because of its generally better performance in various images. Since the Albedo image represents the color properties and reflectance, it did not include shadowing, and we found that the residual model was often too dark for shadow detection. Therefore, we focused on the colorful diffuse images and the greyscale images.

Images were selected for complexity in shadow content and run through the model for intrinsic image decomposition. The output diffuse and greyscale images were then run through K-means detection to determine the success of shadow identification between the two resulting components.

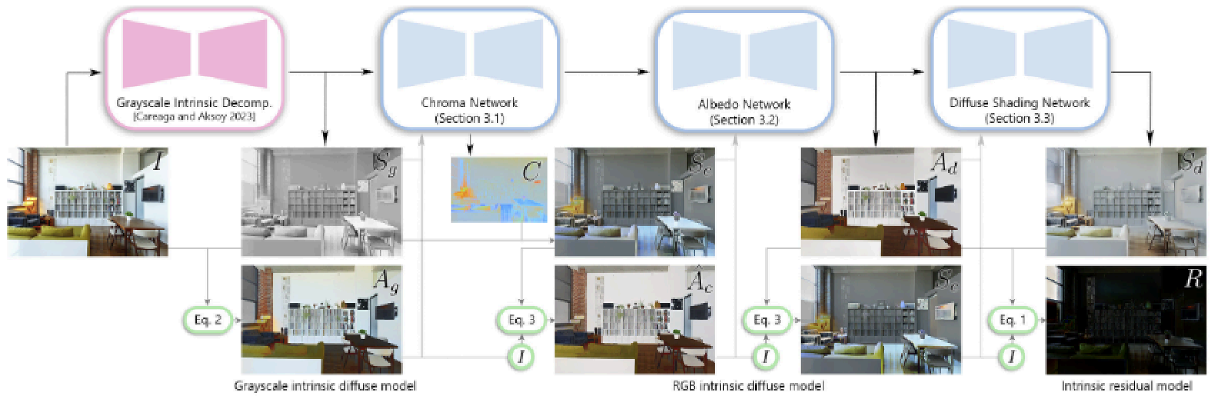
With shadow removal on text documents, code from a BEDSR-Net repo was leveraged. This model consists of two sub-networks: BE-Net for estimating the global background color of the document and SR-Net for removing shadows. Given the input shadow image, BE-Net predicts the background color. As a side product, it generates an attention map depicting how likely each pixel is to belong to the shadow-free background. With the help of the attention map, the model removes the typical requirement of ground-truth shadow masks for training. Along with the input shadow image, the estimated background color and the attention map are fed into the SR-Net to determine the shadow-free version of the input shadow image.

To start, the team tested several text documents as the validation set. This validation set successfully removed shadows from documents containing text as outlined in the example data. After testing and validating the model, the team implemented a new data set containing both document tests and handwritten images to test the model's success and identify any improvement areas.

Model Architecture:

Colorful Diffuse Intrinsic Image Decomposition

The colorful diffuse intrinsic image decomposition pipeline starts with an input image and a shading/albedo pair generated within the simplified grayscale intrinsic diffuse model generated via an off-the-shelf method. They first extend the image formation model to include colorful shading, and estimate the shading color using our chroma network. This color information is used as input in the second step where we estimate the high-resolution diffuse albedo. In the final step, we remove the Lambertian-world assumption and estimate a colorful diffuse shading component and a non-diffuse residual layer. A single variable is estimated at each step, and other intrinsic components are computed using the corresponding intrinsic image formation model with increasing representative power.



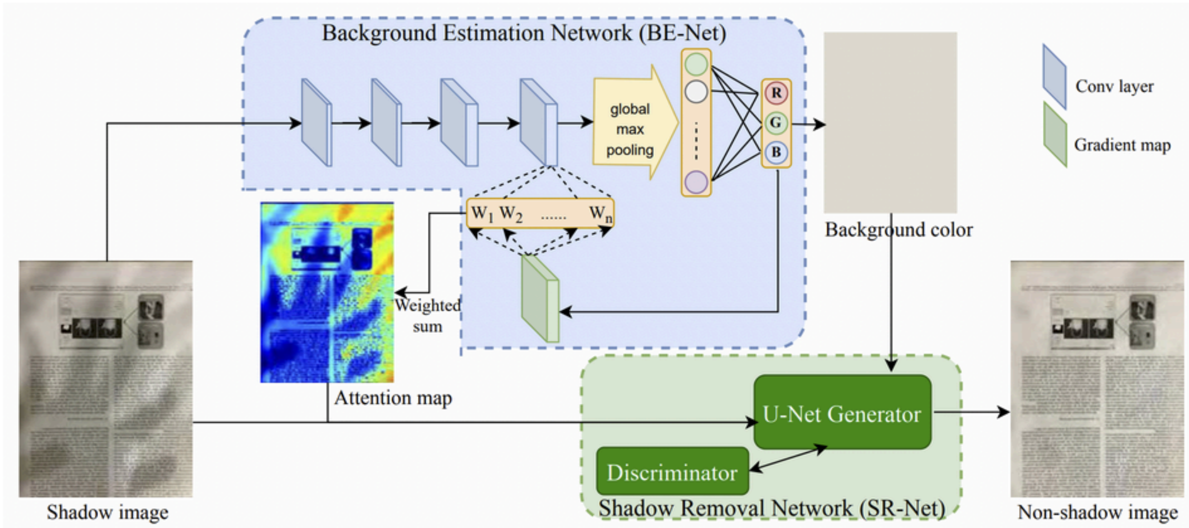
Shadow Detection with Thresholding

The shadow detection with thresholding model takes the formation components of the colorful diffuse intrinsic image decomposition model, converts the image to grayscale, increases kernel size for blurring, flattens the image for clustering, then applies a KMeans clustering, separating the image into 2 clusters, shadow and non-shadow. The image is then reshaped and scaled.

```python
# Function to display images side by side
def show_images_side_by_side(images, titles, cmap='gray'):
    plt.figure(figsize=(15, 5))
    for i in range(len(images)):
        plt.subplot(1, len(images), i + 1)
        plt.imshow(images[i], cmap=cmap)
        plt.title(titles[i])
        plt.axis('off')
    plt.show()

# Convert to grayscale
gray = cv2.cvtColor(diffuse_shading, cv2.COLOR_BGR2GRAY)
#gray = cv2.cvtColor(residual, cv2.COLOR_BGR2GRAY)

# Increase kernel size for more blurring (e.g., (15, 15))
blurred = cv2.GaussianBlur(gray, (15, 15), 0)

# Flatten the blurred image to a 1D array for clustering
pixel_values = blurred.reshape((-1, 1))
pixel_values = np.float32(pixel_values)

# Define criteria and apply KMeans clustering
# KMeans will separate the image into 2 clusters (shadow and non-shadow)
kmeans = KMeans(n_clusters=2, random_state=42)
kmeans.fit(pixel_values)

# Reshape the clustered labels back into the image format
segmented_image = kmeans.labels_.reshape(blurred.shape)

# Since KMeans clusters may assign labels differently, we will assign white to shadows and black to non-shadows
# We assume the darker regions (with lower intensity) are shadows, so we reverse the labels if necessary
if np.mean(blurred[segmented_image == 0]) > np.mean(blurred[segmented_image == 1]):
    segmented_image = np.where(segmented_image == 0, 1, 0)

# Scale the segmented image to 0-255 for visualization
segmented_image = segmented_image * 255

# Plot the original grayscale and the K-Means clustered image side by side
show_images_side_by_side(
    [gray, diffuse_shading, segmented_image],
    ['Grayscale Image', 'diffuse shading', 'Shadow Segmentation']
)
```

BEDSR-Net Shadow Removal on Documents

BEDSR-Net. It consists of two sub-networks: BE-Net for estimating the global background color of the document and SR-Net for removing shadows. Given the input shadow image, BE-Net predicts the background color. As a side product, it generates an attention map, which depicts how likely each pixel belongs to the shadow-free background. With the help of the attention map, the model removes the typical requirement of ground-truth shadow masks for training. Along with the input shadow image, the estimated background color and the attention map are fed into the SR-Net for determining the shadow-free version of the input shadow image.



Results

Shadow segmentation using thresholding:

Comparing different methods of thresholding and shadow detection, we found that KMeans with 2 clusters performed best at detecting shadows in grayscale images. Because of its performance this method was chosen for shadow detection in intrinsic image formation components.

Otu’s Thresholding

Adaptive Thresholding

KMeans

Separation of Albedo and Diffuse shading, Shadow segmentation on diffuse shading:

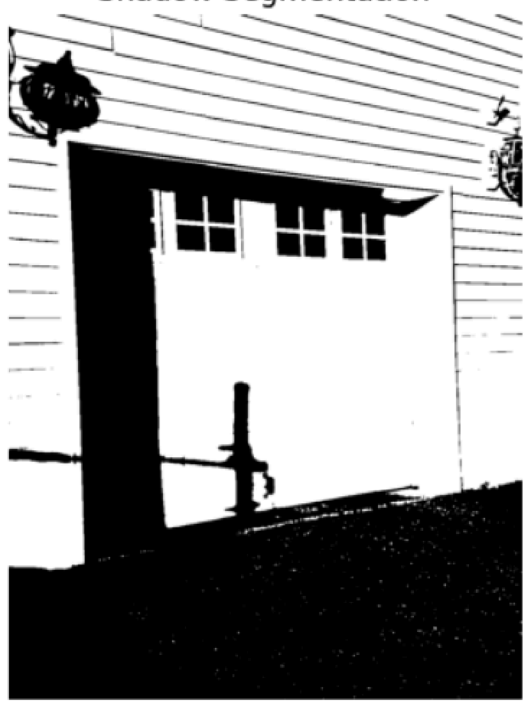
<u>IMG:1</u>

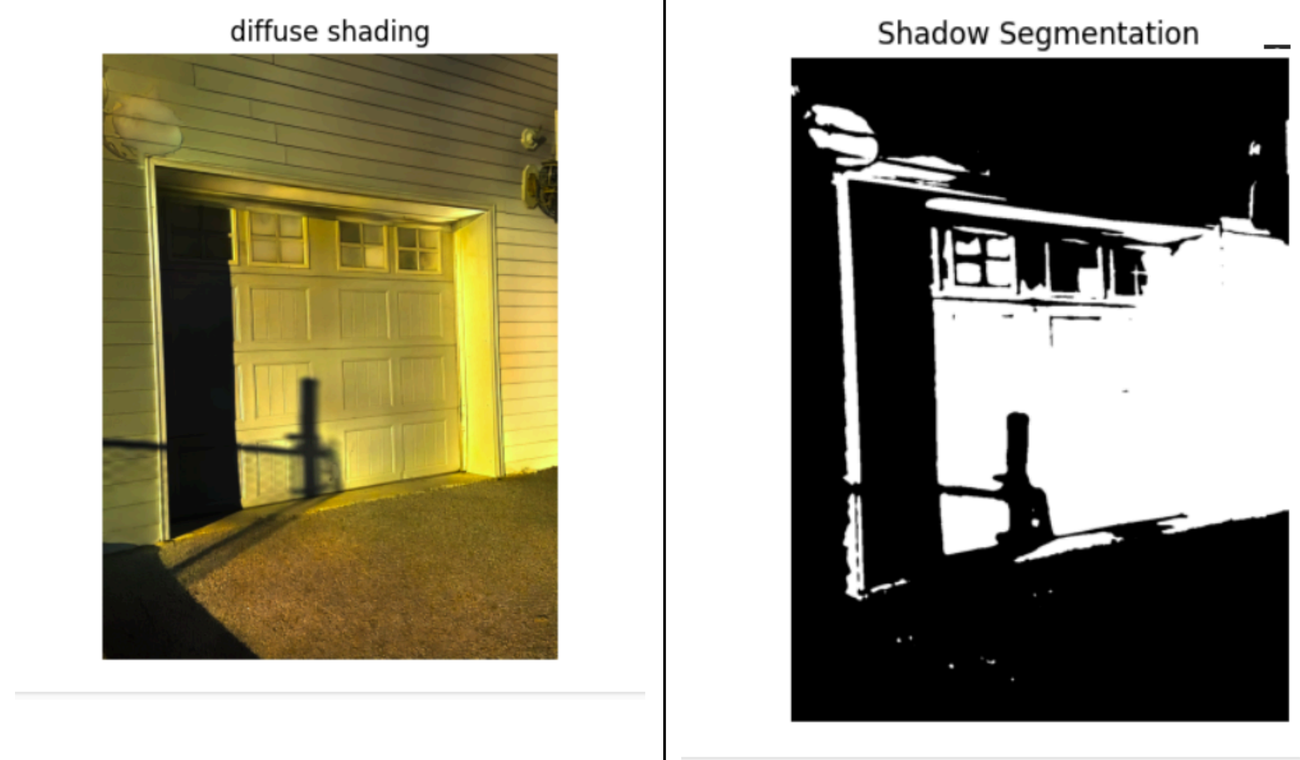
Comparing shadow segmentation in the original grayscale image and the diffuse shading image, we saw better performance on the diffuse shading image in certain areas. More specifically, when there was finer detail present in the overhang of the garage door and the window sections of the garage door. The original grayscale image tends to blanket shadow across darker sections as seen in the ground and the windows of the garage door. The one area that the grayscale shading performs well at is with consistent shading as seen with the siding.

Original

Grayscale Image

Shadow Segmentation





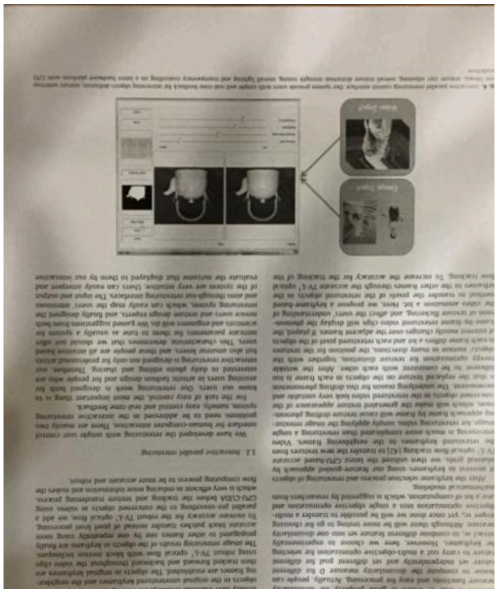
IMG:2

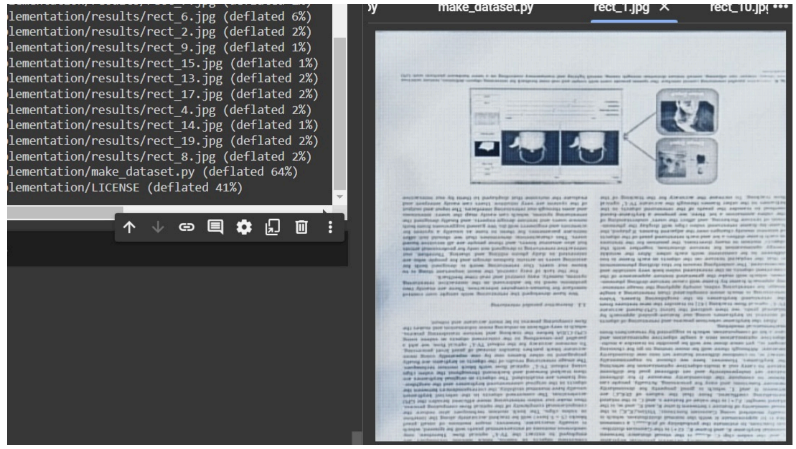
Shadow segmentation on the diffuse shagging image picked up some of the finer detail of the shadows within the image. However, as we saw with the garage door, it took the similar colored areas and blanked shadow detection.

<u>BEDSR-Net Shadow Removal</u>

Validation Image

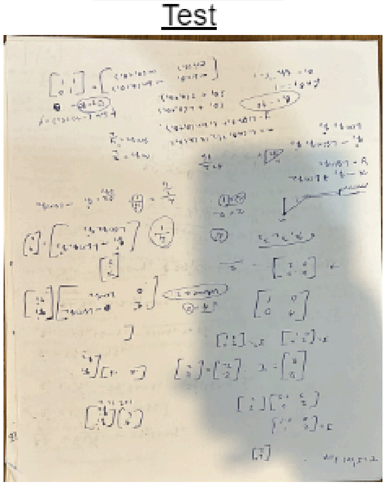
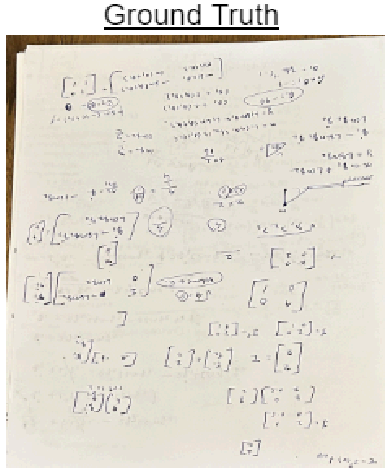
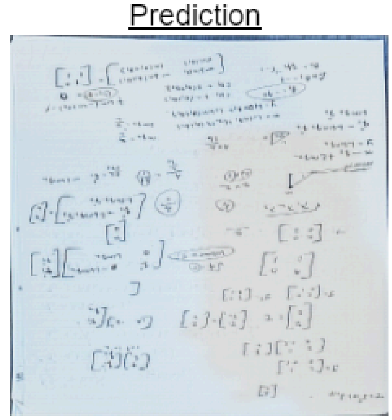
The validation image shows that the model was fairly successful at removing shadows from document images with text. However, we can still see some residual marks from the fingers on the left hand side.

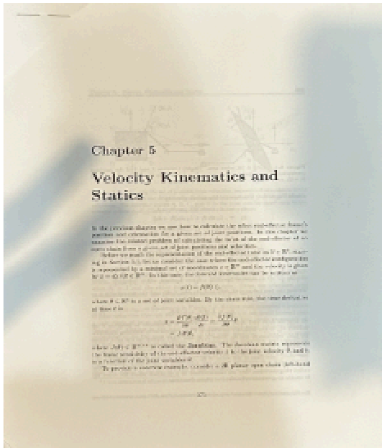
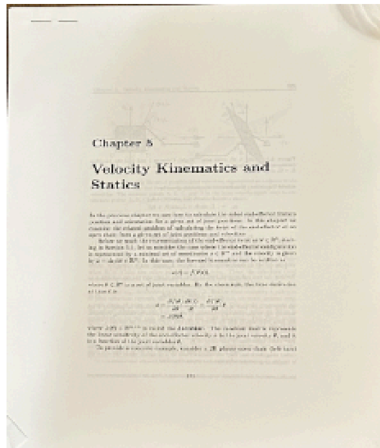
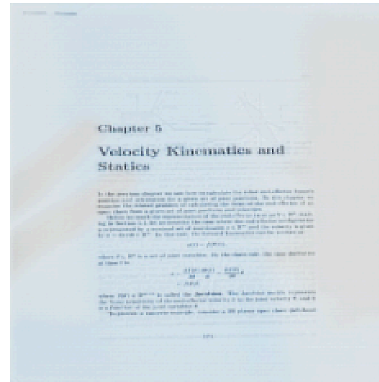


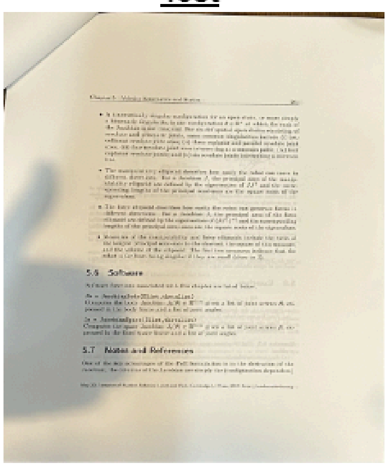
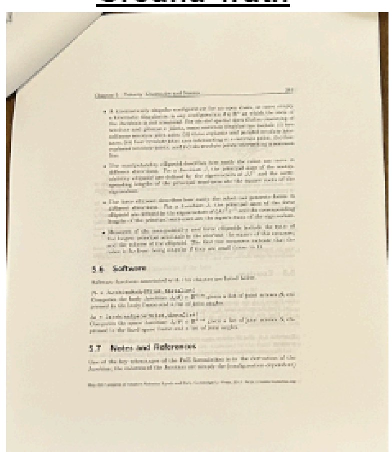
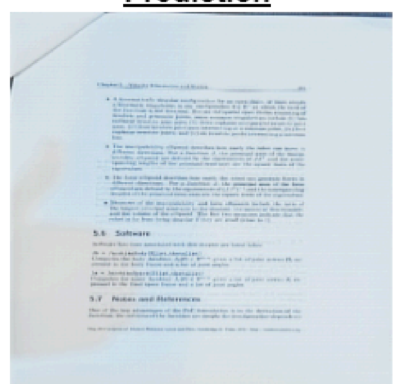


Test Images

In applying outside data of handwritten text and documents with folds included we see that performance was not as good as the original test document. What is interesting about the output is that while it removed the darkness of the shadow, it seemed to leave the original background color in place of where the shadow was, creating a color variance between the new background and the original. We suspect that this is happening when the background color is being estimated, the model is not seeing the shadowed area as part of the background.

Test	Ground Truth	Prediction
		

Test	Ground Truth	Prediction
 Chapter 5 Velocity Kinematics and Statics	 Chapter 5 Velocity Kinematics and Statics	 Chapter 5 Velocity Kinematics and Statics

Test	Ground Truth	Prediction
 5.6 Software 5.7 Notes and References	 5.6 Software 5.7 Notes and References	 5.6 Software 5.7 Notes and References

Acknowledgment and Reflection

We spent a fair amount of time on the project trying to get code to work properly to align with shadow removal in hand written text documents. That ended up eating up a fair amount of time before we switched gears to look at shadow detection from component formation of intrinsic images. While we were finally able to implement the model for shadow removal in text images, it was at the very end. Overall, we wanted to spend more time on shadow removal in hand-written text images.

Working with the colorful intrinsic image decomposition showed us the nuances in shading that using color shading can highlight additional shadows within an image. This is certainly different from traditional approaches that only assume white light sources. As we discussed how this

could be best applied, we thought about image editing software, augmented reality, and autonomous driving.

The idea to work with shadow removal in text images came from real work experience in automating document intake. While not the most glamorous area of focus, it is much needed in many industries. In implementing and running the BEDSR-Net we found that even though shadows were removed for the most part, residual shadows still remained. In the future, we would like to spend some additional time working with this problem.