A Deep Learning Method for Document Shadow Removal with Sobel Prior under Mask Supervision

Jiarui Li

College of Computer Science and Technology
Qingdao University, Qingdao, China
Ljr20231130@gmail.com

Abstract

When digitizing documents using conventional equipment, shadows often appear, posing significant challenges to the visual quality and readability of the digital copies. Given that the removal of document shadows typically involves complex image processing and computational tasks, which require substantial computational resources and time, the cost can become prohibitive, limiting the practicality and efficiency of shadow removal algorithms. This research aims to address the critical task of designing a model capable of achieving superior shadow removal effects. We propose a deep learning model for document shadow removal that harnesses Sobel text prior and ground truth masks as supervision. This prior knowledge encapsulates regular information regarding document structure and shadow formation, thereby enhancing its ability to utilize edge information for shadow removal. Additionally, the integration of prior knowledge and supervised learning can help the model learn more quickly, reducing the amount of information the model needs to process and improving its efficiency.

Introduction

Document shadows significantly affect the clarity and readability of digital copies, thereby compromising the reading experience. Consequently, the removal of document shadows is crucial to address this issue.

Currently, a variety of methods exist to tackle this problem, including traditional techniques and deep learning approaches. Among traditional methods, several techniques have been employed to confront the challenge of shadow removal. These include edge detection techniques (Fredembach and Finlayson 2005), various filters (Yang, Tan, and Ahuja 2012)(Parisotto et al. 2019)(Dare 2005), and morphology-based approaches (Xu, Landabaso, and Pardàs 2005)(Nair, Kosal Ram, and Sundararaman 2019), among others. Each of these methods plays a significant role in mitigating the impact of shadows on digital imagery.

However, traditional methods are often sensitive to noise and distortion, leading to issues such as blurred edges and artifacts. Additionally, these methods heavily rely on manual parameter adjustments, making them less suitable for handling diverse and complex situations. In light of these challenges, and in line with advancements in deep learning, deep learning methods have increasingly come to prominence, such as BEDSR-Net(Lin, Chen, and Chuang 2020), Waterfilling(Jung, Hasan, and Kim 2018), FSENet(Li et al. 2023), and others. Nevertheless, these methods typically have a large information burden, which increases the computational cost of the models.

In this paper, we propose an efficient deep learning method for document shadow removal. By utilizing a model with Sobel text prior to detect text edges and employing the supervision of ground truth masks for shadow removal, we aim to achieve a more efficient algorithm.

Methods

MaskNet for Mask Prediction

In our network, we propose MaskNet to predict masks.

We utilize F_{out} as the input to the MaskNet. F_{out} represents a 4-channel feature map, with the gate block dividing it into a 3-channel latent and a 1-channel latent. The 1-channel latent output is processed through a sigmoid activation function, resulting in a monochromatic, pixel-wise local mask forecast that spans a range of values from 0 to 1. It is important to mention that this monochromatic mask prediction at the pixel level is guided by the ground truth mask. In the final stage of the gate mechanism, the 3-channel latent is combined with the estimated mask to produce the residual image. Ultimately, the residual images across various scales are fused together to construct the final predicted mask images.

We denote each level's predicted mask as F_m and each level's feature map as F. Both F_m and $F - F_m$ are utilized as inputs for the FusionNet to enhance feature fusion.

FusionNet for Feature Fusion

The FusionNet facilitates feature fusion using two inputs: the original feature map F and the feature map with the mask $F - F_m$.

In the framework of F, the Kernel Generation Units (KGUs) synthesize separate weight tensors, denoted as f_g, which span areas of 3×3, 7×7, and 11×11 pixels respectively. The architecture of KGUs, influenced by densely connected architectures, incorporates four densely linked layers that bolster the propagation and representation of features, thereby facilitating the recycling of features and augmenting

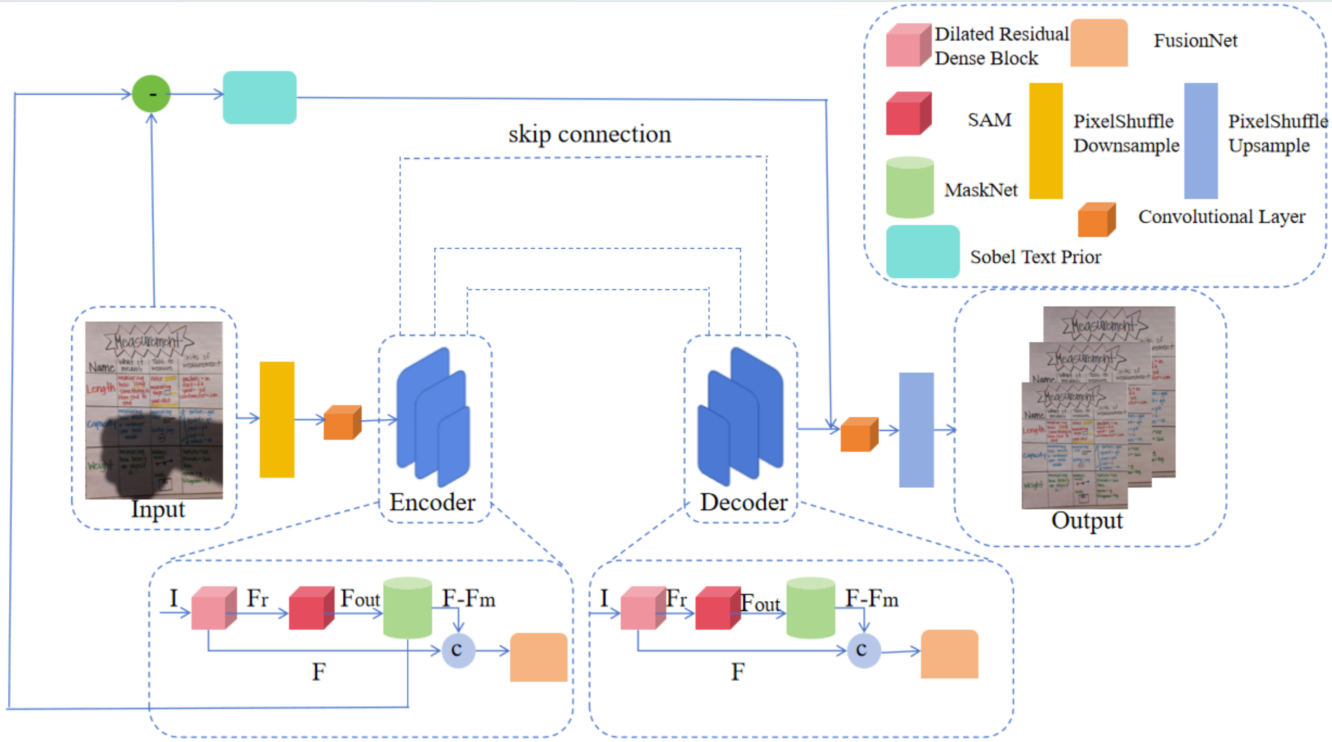

Figure 1: Illustration of our architecture, it includes Dilated Residual Dense Block, SAM, MaskNet, Sobel Text Prior, FusionNet, PixelShuffle Downsample, PixelShuffle Upsample, and Convolutional Layer.

the efficiency of parameters. Following this, Kernel Transformation Units (KTUs) are utilized to forge conventional convolutional kernels with various dilation levels, achieved through the reassembly of kernel tensors and the strategic insertion of zero values.

For $F - F_m$, after preliminary dimension reduction, the input is re-weighted and integrated into three parallel branches to obtain enhanced features. This channel-wise adjustment ensures independent operations for each channel.

This method effectively predicts masks and utilizes them for feature fusion, enhancing the performance of document shadow removal.

Experiments

Experimental Settings

Data and Evaluation. We evaluate the performance of different network structures at various resolutions using Jung's dataset and Kligler's dataset.

For evaluation metrics, we utilize three widely recognized measures: Root Mean Square Error (RMSE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity (SSIM).

Selected Baseline Methods. To evaluate the effectiveness of our proposed method, we compare it with several representative methods for document shadow removal, including traditional techniques such as edge detection, filtering, and morphology-based approaches, as well as state-of-the-art deep learning-based methods like Kligler(Kligler, Katz, and Tal 2018), AEFNet(Fu et al. 2021), BEDSR-Net(Lin, Chen, and Chuang 2020), DC-ShadowNet(Jin, Sharma, and Tan 2021), DHAN(Cun, Pun, and Shi 2020), LG-ShadowNet(Liu et al. 2021), Mask-ShadowGAN(Hu et al. 2019), SG-ShadowNet(Wan et al. 2022), ST-CGAN(Wang, Li, and Yang 2018), and ShadowFormer(Guo et al. 2023).

Conclusion

We introduce a novel deep learning approach that leverages Sobel text prior and ground truth mask supervision to achieve remarkable performance in the removal of shadows from document images.

For further study, I would like to optimize my model to have a better performance both on memory usage and performance. What's more, more datasets would be into consideration. Such as the SD7K dataset, ISTD dataset, and so on. This would prove our model's robustness. Related work is also included in my research, which can be a motivation for my new ideas.

If I could solve this problem completely in the future, it would not only advance the state-of-the-art(SOTA) in document shadow removal but also establish a benchmark for solutions capable of handling document imagery efficiently. Moving forward, our approach will open avenues for further research in developing more efficient and effective methods for document image processing, contributing to advancements in the field of computer vision and document analysis.

References

Cun, X.; Pun, C.-M.; and Shi, C. 2020. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 10680–10687.

Dare, P. M. 2005. Shadow analysis in high-resolution satellite imagery of urban areas. *Photogrammetric Engineering & Remote Sensing*, 71(2): 169–177.

Fredembach, C.; and Finlayson, G. 2005. Hamiltonian path-based shadow removal. In *Proceedings of the 16th British Machine Vision Conference (BMVC)*, volume 2, 502–511.

Fu, L.; Zhou, C.; Guo, Q.; Juefei-Xu, F.; Yu, H.; Feng, W.; Liu, Y.; and Wang, S. 2021. Auto-exposure fusion for single-image shadow removal. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10571–10580.

Guo, L.; Huang, S.; Liu, D.; Cheng, H.; and Wen, B. 2023. Shadowformer: Global context helps image shadow removal. *arXiv preprint arXiv:2302.01650*.

Hu, X.; Jiang, Y.; Fu, C.-W.; and Heng, P.-A. 2019. Mask-shadowgan: Learning to remove shadows from unpaired data. In *Proceedings of the IEEE/CVF international conference on computer vision*, 2472–2481.

Jin, Y.; Sharma, A.; and Tan, R. T. 2021. Dc-shadownet: Single-image hard and soft shadow removal using unsupervised domain-classifier guided network. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 5027–5036.

Jung, S.; Hasan, M. A.; and Kim, C. 2018. Water-filling: An efficient algorithm for digitized document shadow removal. In *Asian Conference on Computer Vision*, 398–414. Springer.

Kligler, N.; Katz, S.; and Tal, A. 2018. Document enhancement using visibility detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2374–2382.

Li, Z.; Chen, X.; Pun, C.-M.; and Cun, X. 2023. High-Resolution Document Shadow Removal via A Large-Scale Real-World Dataset and A Frequency-Aware Shadow Erasing Net. *arXiv preprint arXiv:2308.14221*.

Lin, Y.-H.; Chen, W.-C.; and Chuang, Y.-Y. 2020. Bedsrnet: A deep shadow removal network from a single document image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12905–12914.

Liu, Z.; Yin, H.; Mi, Y.; Pu, M.; and Wang, S. 2021. Shadow removal by a lightness-guided network with training on unpaired data. *IEEE Transactions on Image Processing*, 30: 1853–1865.

Nair, V.; Kosal Ram, P. G.; and Sundararaman, S. 2019. Shadow detection and removal from images using machine learning and morphological operations. *The Journal of Engineering*, 2019(1): 11–18.

Parisotto, S.; Calatroni, L.; Caliari, M.; Schönlieb, C.-B.; and Weickert, J. 2019. Anisotropic osmosis filtering for shadow removal in images. *Inverse Problems*, 35(5): 054001.

Wan, J.; Yin, H.; Wu, Z.; Wu, X.; Liu, Y.; and Wang, S. 2022. Style-guided shadow removal. In *European Conference on Computer Vision*, 361–378. Springer.

Wang, J.; Li, X.; and Yang, J. 2018. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1788–1797.

Xu, L.-Q.; Landabaso, J. L.; and Pardàs, M. 2005. Shadow removal with blob-based morphological reconstruction for error correction. In *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, volume 2, ii–729. IEEE.

Yang, Q.; Tan, K.-H.; and Ahuja, N. 2012. Shadow removal using bilateral filtering. *IEEE Transactions on Image processing*, 21(10): 4361–4368.