ShaDocNet: Learning Spatial-Aware Tokens in Transformer for Document Shadow Removal

{Xuhang Chen[1], Xiaodong Cun[1], Chi-Man Pun[1] and Shuqiang Wang[2]
[1]University of Macau [2]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences

Objectives

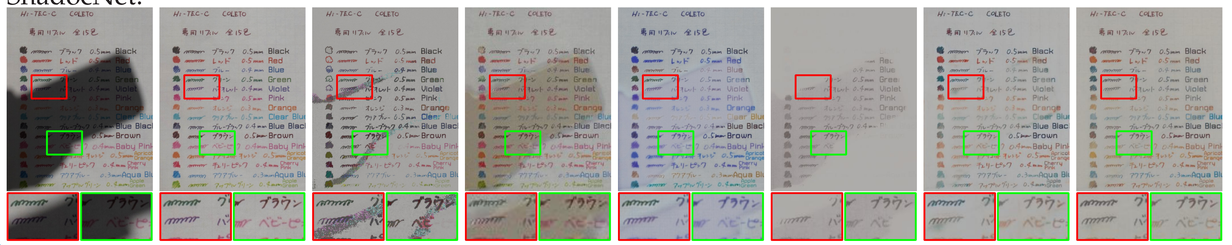
The main objective of this paper is to propose a new Transformer-based model called ShadocNet for document shadow removal. The paper aims to address the issue of shading errors in casual document photographs acquired in the wild due to occluders blocking light sources. The proposed model utilizes shadow context encoding and decoding in both shadow and shadow-free regions, as well as shadow detection and pixel-level enhancement in a coarse-to-fine process.

Introduction

Document shadow removal is a crucial computer vision task since casual document photographs often contain shading errors caused by occluders blocking the light sources, which can hinder the quality of the documents. Both traditional document shadow removal methods and deep learning methods have limitations, with traditional methods relying on heuristics and deep learning methods requiring a large dataset but not utilizing specific properties of document images.

Visual Result

Visual comparison of competing methods. From left to right: shadow image, ground truth, Wang *et al.*, Mask-ShadowNet, BEDSR-Net, ShadocNet w/o RefineNet, ShadocNet w/o Transformer encoder, ShadocNet.



Methodology

We propose the ShadocNet network to assess the document shadow removal method on ShaDocs. As shown in Figure 1, we build a multi-stage framework for document shadow removal that includes shadow detection, global color matching, and local pixel-by-pixel refining. Listed below are the specifics of each component and loss function.

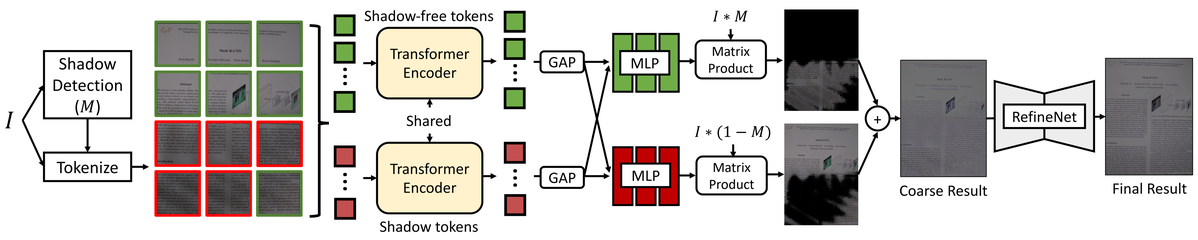


Figure 1: The network structure of the ShadocNet. It consists of three components: the Shadow Detection module, the Global Shadow Remapping for shadow color remapping, and the RefineNet for improving visual quality.

We utilize a retrained learning-based method for shadow detection on document images, which involves direction-aware spatial context modules to predict a score map at each layer, resulting in a shadow mask.

For shadow removal, we patch embed the shadow images and compute the self-attention on these domain-aware patch embeddings using the Transformer, resulting in a domain-aware region embedding for the shadow and shadow-free area. After obtaining the domain-aware tokens, we develop the MLP, a simple yet effective pixel mapping function to remap each region's original color pixels to their new value.

Inspired by Dual Hierarchical Aggregation Network [1], we also apply a multi-stage convolution and aggregation strategy for pixel-wise refinement.

Quantitative Result

We employ PSNR, SSIM and RMSE to evaluate visual quality as shown in Table 1.

Method	Jung PSNR	Jung SSIM	Jung RMSE	Kligler PSNR	Kligler SSIM	Kligler RMSE
Jung *et al.*	23.27	0.91	17.88	14.44	0.90	49.15
Wang *et al.*	16.86	0.86	40.50	22.97	0.88	19.35
Shah *et al.*	14.91	0.83	46.77	8.39	0.70	97.48
U-Net	22.47	0.86	20.44	21.68	0.83	21.58
BEDSR-Net	21.98	0.88	23.36	24.37	0.88	18.78
Ours	**24.60**	**0.91**	**15.30**	**26.20**	**0.94**	**13.48**

Table 1: Quantitative result

We also report the OCR performance on the recovered shadow-free images as shown in Table 2.

Method	Edit distance↓
Wang *et al.*	191.5
Shah *et al.*	187.8
U-Net	261.2
ST-CGAN	294.8
Mask-ShadowNet	262.9
AEFNet	292.4
BEDSR-Net	189.0
Ours	**178.6**

Table 2: OCR result

Conclusion

In conclusion, the ShadocNet model proposed in this paper presents a competitive Transformer-based approach for document shadow removal. By leveraging feature extraction and color rendering modules, as well as shadow context encoding and decoding in both shadow and shadow-free regions, ShadocNet is able to improve the visual quality and legibility of digital copies of documents. The model's ability to extract shadow area tokens and shadow-free region tokens with the assistance of a detection module and Transformer encoder, as well as its two-stage coarse-to-fine approach, make it more resilient than previous methods for document shadow removal.

Acknowledgment

This work was supported in part by the University of Macau under Grant MYRG2022-00190-FST, in part by the Science and Technology Development Fund, Macau SAR, under Grant 0034/2019/AMJ, Grant 0087/2020/A2 and Grant 0049/2021/A, in part by the National Natural Science Foundations of China under Grants 62172403 and in part by the Distinguished Young Scholars Fund of Guangdong under Grant 2021B1515020019.

References

[1] Xiaodong Cun, Chi-Man Pun, and Cheng Shi. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan. In *AAAI*.

Future Research

ShadocNet will continue to increase in size, quality, and variety as a result of our future study.

In addition, given the size of the present dataset is inadequate, we will attempt to create a new massive dataset on document shadow removal. Similar to the natural shadow dataset, we will supply image triplets of shadow, shadow-free, and shadow mask.

Contact Information

Web https://www.cis.um.edu.mo/ cmpun/
Email cmpun@umac.mo
Phone +853 8822 4369