Hand-written Text Removal from Papers

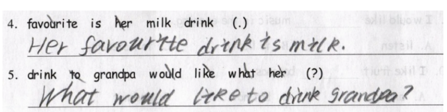
Jiangyi Liu
jiangyi.liu@wisc.edu

Venkata Abhijeeth Balabhadruni
balabhadruni@wisc.edu

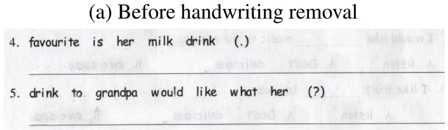
Sujay Chandra Shekara Sharma
schandrashe5@wisc.edu

1. Introduction

The current maturity of computer vision algorithms enables its real-world application. Therefore, in this project, we also want to work on a problem that originated in our daily lives. In short, we focus on the problem of *removing hand-written text from papers*: Consider a sheet of paper (Figure 1a) with both printed text and hand-written text (e.g., exam papers or application forms). Given a photo of the paper, how can we remove hand-written text while still keeping the printed text? It is expected that ink from handwriting be replaced in a context-aware way (Figure 1b), so that the paper can be recovered to its state before being written by hand.



(a) Before handwriting removal



(b) After handwriting removal

Figure 1. Example of hand-written text removal.

This problem has potential applications in privacy protection, education, and office work, e.g., redacting personal information from a scanned file.

2. Related Works

In the literature of Computer Vision, the *text removal* problem is closely related to the problem of removing hand-written text. Before the emergence of deep neural networks, manual marking of hand-written text is required. The algorithm then uses interpolation to overwrite hand-written areas with blanks. Such a method requires significant human intervention and effort and performs poorly when the background is complex or handwriting appears on images[2].

However, deep learning brings new possibilities to the text removal problem. Deep neural networks can understand the semantic information of images and can be used for locating texts from an image [6]. The classical approach for this task [5] uses a two-step structure: (1) apply techniques in object detection to identify areas with texts, and (2) use image inpainting algorithms to erase these areas. In some works, Generative Adversarial Networks[4] or sliding window [3] is used, but the main idea of a two-step structure remains unchanged. On the other hand, [7] is the first work that introduces an end-to-end network for this task, directly generating results after text removal.

The task of handwriting removal is similar to the text removal problem but slightly more difficult. While the objective of text removal is removing all text, handwriting removal requires first classifying handwritten and printed text and then only removing the former. There is only limited research in this area. [1] is the first work (to our knowledge) that attempts to solve the handwriting removal problem, introducing a "stroke mask" based on characteristics of hand-written text. However, its robustness is not good and it requires high-quality photos of paper.

Another issue is that only one dataset [1] is openly available for the handwriting removal problem, hindering related research.

3. Tentative Approach and Evaluation

We will use deep learning to solve the task of hand-written text removal because the task inherently requires splitting printed and hand-written text. The bounding box level precision provided by traditional approaches is not enough.

We plan to start by first implementing state-of-the-art methods like [7] and learn about their network structure (encoders, GANs, etc.). If possible, based on these works, we plan to implement stroke-level detection [1] – for each pixel, use a network to directly predict its possibility of being in a stroke in handwriting. This is similar to the semantic segmentation of images. Due to the fact that font size varies in handwritten text, we will also develop a multi-scale network that adapts all sorts of handwriting. Data augmentation will also be applied for better robustness.

Evaluation will loosely follow [1], using (1) Peak signal-to-noise ratio, (2) Multi-scale Structural Similarity (MSSIM), (3) MSE, (4) AGE, (5) pEPs, (6) pCEPs, and (7) precision/recall/F-score.

References

[1] Liufeng Huang, Bangdong Chen, Chongyu Liu, Dezhi Peng, Weiying Zhou, Yaqiang Wu, Hui Li, Hao Ni, and Lianwen Jin. Ensexam: A dataset for handwritten text erasure on examination papers. In *International Conference on Document Analysis and Recognition*, pages 470–485. Springer, 2023. 1

[2] Max Jaderberg, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Reading text in the wild with convolutional neural networks. *International journal of computer vision*, 116:1–20, 2016. 1

[3] Toshiki Nakamura, Anna Zhu, Keiji Yanai, and Seiichi Uchida. Scene text eraser. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, pages 832–837. IEEE, 2017. 1

[4] Siyang Qin, Jiahui Wei, and Roberto Manduchi. Automatic semantic content removal by learning to neglect. *arXiv preprint arXiv:1807.07696*, 2018. 1

[5] Osman Tursun, Simon Denman, Sabesan Sivapalan, Sridha Sridharan, Clinton Fookes, and Sandra Mau. Component-based attention for large-scale trademark retrieval. *IEEE Transactions on Information Forensics and Security*, 17: 2350–2363, 2019. 1

[6] Qixiang Ye and David Doermann. Text detection and recognition in imagery: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 37(7):1480–1500, 2014. 1

[7] Shuaitao Zhang, Yuliang Liu, Lianwen Jin, Yaoxiong Huang, and Songxuan Lai. Ensnet: Ensconce text in the wild. In *Proceedings of the AAAI conference on artificial intelligence*, pages 801–808, 2019. 1