High-Fidelity Document Stain Removal via A Large-Scale Real-World Dataset and A Memory-Augmented Transformer

Mingxian Li[1] Hao Sun[1] Yingtie Lei[2] Xiaofeng Zhang[3] Yihang Dong[4] Yilin Zhou[1] Zimeng Li[5] Xuhang Chen[1]

[1]Huizhou University [2]The Ohio State University [3]Shanghai Jiao Tong University [4]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences [5]Shenzhen Polytechnic University

Abstract

Document images are often marred by stains, reducing readability and complicating digitization and analysis. The lack of a comprehensive dataset has hindered effective stain removal while preserving details. To overcome this, we introduce StainDoc, a high-resolution dataset for stain removal, featuring over 5,000 pairs of stained and clean document images with diverse stains. This dataset aids in training robust stain removal algorithms. We also propose StainRestorer, a Transformer-based approach using a memory-augmented Transformer to capture hierarchical stain representations via the DocMemory module. The Stain Removal Transformer (SRTransformer) employs enhanced spatial and channel attention mechanisms for precise stain removal while maintaining content integrity. StainRestorer outperforms existing methods on the StainDoc dataset, setting a new benchmark. Our work underscores memory-augmented Transformers' potential in this field and contributes a valuable dataset for future research.

Main Contributions

- **StainDoc** We present StainDoc, the first large-scale, high-resolution (2145×2245) dataset for document stain removal. It includes over 5,000 pairs of stained and clean documents from various scenes, stains, severities, and backgrounds, addressing a critical need in document enhancement.
- **DocMemory** Our DocMemory module captures hierarchical stain representations through Memory Units, extracting and analyzing deep features at different granularities.
- **SRTransformer** The Stain Removal Transformer (SRTransformer) improves spatial mapping for precise stain removal, distinguishing stains from document details, ensuring high-quality results while preserving content integrity.

StainDoc Dataset

(a) (b) (c) (d) (e)

Figure 1. Example from StainDoc dataset: (a) stained document image, (b) clean original image. Dataset creation involves: (a) applying stains and preparing documents, (b) controlled-condition photography, (c) post-processing and standardization.

Methodology

StainRestorer, as shown in Fig. 2, uses a hierarchical method for stain representation and removal. Initially, the DocMemory module captures diverse stain representations at multiple levels: part, instance, and semantic. This hierarchy enhances understanding of stain characteristics. Then, the SRTransformer module, influenced by X-Restormer [1], uses these representations for accurate removal. Combining DocMemory's hierarchical learning with SRTransformer's spatial mapping ensures high-quality stain removal while preserving document content.

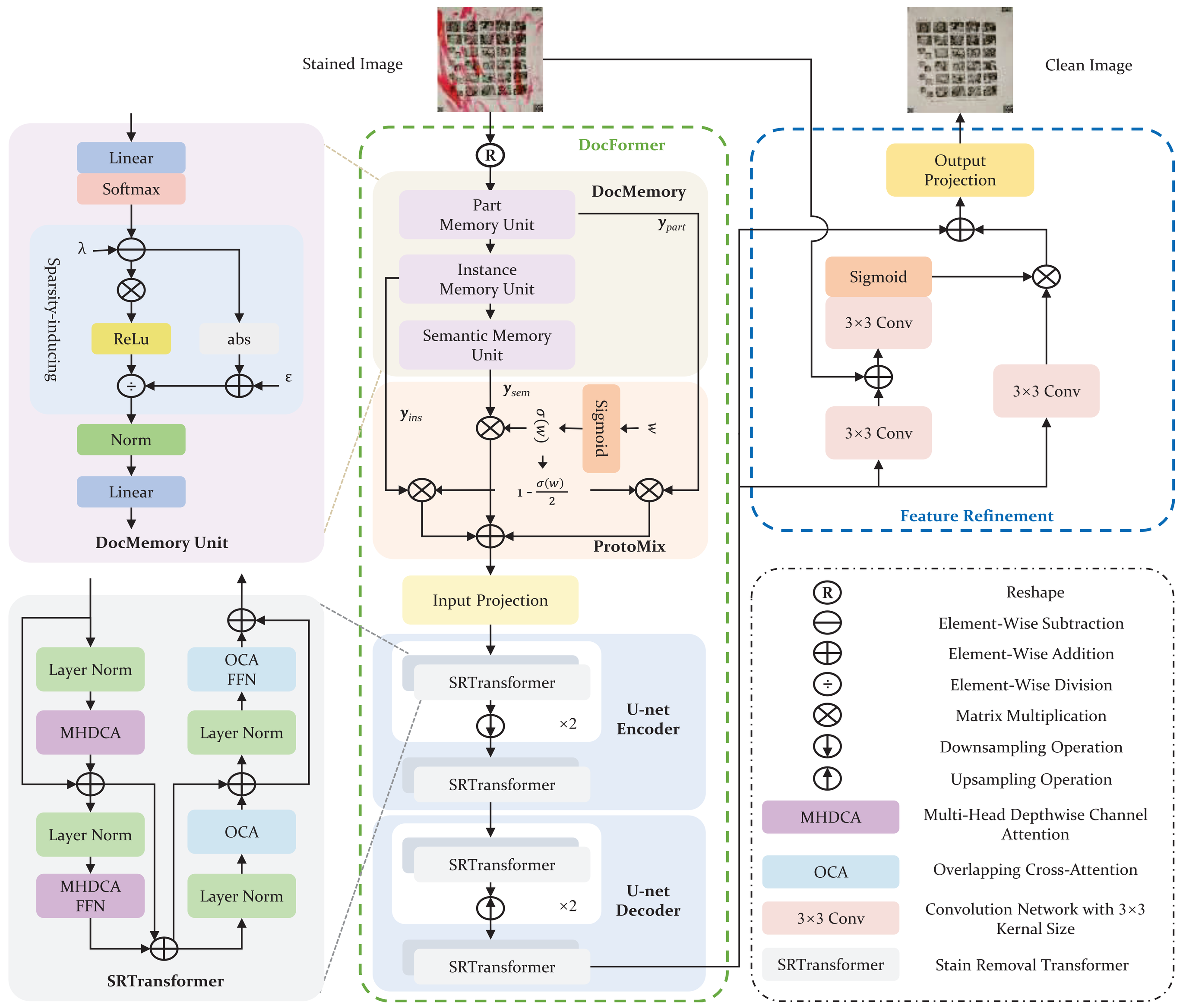


Figure 2. StainRestorer's architecture features a DocMemory module for hierarchical stain representation and an SRTransformer for accurate removal. DocMemory captures part, instance, and semantic-level stain features, which are fused using ProtoMix. SRTransformer uses these to precisely remove stains while preserving content.

StainRestorer is trained using a composite loss function that balances pixel-wise accuracy and the preservation of structural information. This is achieved by combining the Mean Squared Error (MSE) loss $\mathcal{L}_{MSE}$ and the Structural Similarity Index Measure (SSIM) loss $\mathcal{L}_{SSIM}$.

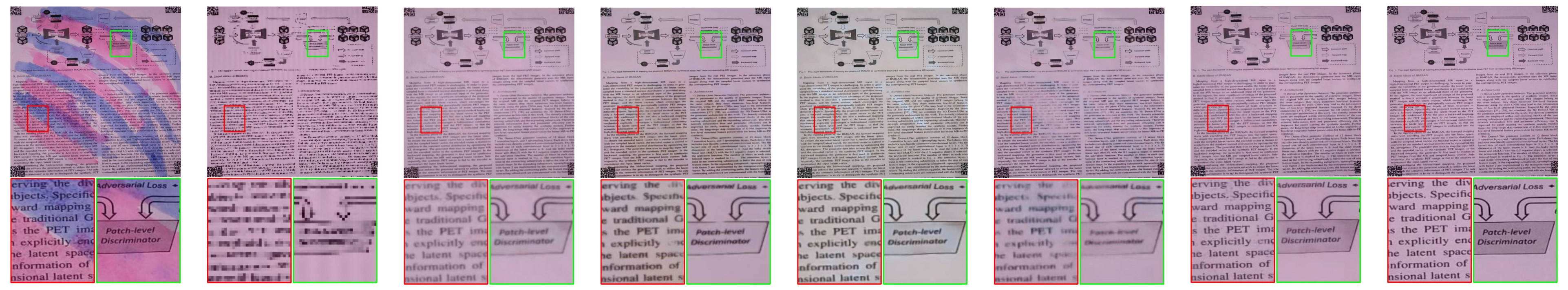
$$\mathcal{L}_{total} = \mathcal{L}_{MSE}(\hat{I}, I) + \alpha \cdot \mathcal{L}_{SSIM}(\hat{I}, I), \quad (1)$$

where α is a weighting factor controlling the relative importance of the SSIM loss. Empirically, α is set to 0.2.

Results

Method	StainDoc				StainDoc_Mark				StainDoc_Seal			
	PSNR ↑	SSIM ↑	MAE ↓	LPIPS ↓	PSNR ↑	SSIM ↑	MAE ↓	LPIPS ↓	PSNR ↑	SSIM ↑	MAE ↓	LPIPS ↓
Input	18.372	0.757	17.59	0.194	22.198	0.703	11.277	0.389	24.875	0.904	4.852	0.143
DocTr++	14.012	0.43	32.057	0.692	14.927	0.348	29.097	0.767	16.861	0.518	18.634	0.534
DocEnTR	15.173	0.514	34.955	0.84	16.095	0.545	29.643	0.775	17.424	0.558	20.126	0.686
TextDIAE	15.953	0.541	24.836	0.656	16.253	0.497	24.397	0.703	16.863	0.558	20.918	0.622
Kligler	16.285	0.713	26.313	0.234	18.255	0.631	22.611	0.429	20.219	0.852	16.323	0.194
DocDiff	16.662	0.607	25.609	0.359	17.646	0.658	23.842	0.435	12.891	0.458	44.894	0.673
illtrtemplate	17.374	0.616	16.919	0.391	17.832	0.593	20.844	0.601	17.919	0.61	17.246	0.521
GAN_HTR	17.377	0.724	23.261	0.374	24.809	0.848	12.321	0.191	10.583	0.672	60.194	0.514
DocTr	17.87	0.597	20.812	0.601	16.866	0.613	16.736	0.38	16.879	0.613	16.734	0.383
DeepOtsu	20.747	0.767	16.278	0.25	22.803	0.829	9.468	0.238	22.187	0.863	14.296	0.214
DocProj	22.127	0.781	12.996	0.227	21.048	0.702	16.435	0.432	27.052	0.901	6.604	0.113
DocNLC	22.45	0.781	11.378	0.232	22.031	0.72	13.866	0.407	27.854	0.912	5.569	0.097
DocRes	22.612	0.792	11.378	0.198	23.159	0.751	11.667	0.343	27.696	0.908	5.646	0.098
GCDRNet	22.619	0.802	11.077	0.187	26.194	0.893	6.629	0.121	23.675	0.86	12.141	0.184
DE-GAN	22.62	0.819	11.209	0.154	23.421	0.842	8.947	0.213	22.123	0.844	12.774	0.202
UDoc-GAN	22.834	0.803	10.608	0.189	25.94	0.88	6.888	0.142	26.391	0.881	8.399	0.142
Ours	**23.372**	**0.822**	**9.265**	**0.085**	**34.191**	**0.955**	**2.786**	**0.067**	**33.298**	**0.968**	**2.441**	**0.024**

Table 1. Quantitative comparison of different methods. The best results are highlighted in bold, and the second best results are underlined.



(a) Input (b) illtrtemplate (c) DocNLC (d) GCDRNet (e) DEGAN (f) UDoc-GAN (g) Ours (h) Target

Figure 3. Qualitative comparison of stain removal performance across different models.

References

[1] Xiangyu Chen, Zheyuan Li, Yuandong Pu, Yihao Liu, Jiantao Zhou, Yu Qiao, and Chao Dong. A comparative study of image restoration networks for general backbone network design. In *European Conference on Computer Vision*, pages 74–91. Springer, 2025.

[2] Xiao Feng Zhang, Chao Chen Gu, and Shan Ying Zhu. Memory augment is all you need for image restoration. *arXiv preprint arXiv:2309.01377*, 2023.