# DocRes: A Generalist Model Toward Unifying Document Image Restoration Tasks

## A. Efficiency

As shown in Table 1, we compared the number of parameters and computational complexities of our method with those of other methods. It can be observed that even when compared to certain task-specific models, our method maintains an advantage in both the number of parameters and computational complexity. This advantage is particularly significant when considering that the number of parameters of these task-specific models would multiply several times over if they were to support multitasks. In contrast, our method does not require additional parameter increments.

Table 1. Model size and computational complexities for each method.

| Methods | Li et al. [3] | DocShadow [5] | UDoc-GAN [8] | DE-GAN [7] | DocRes (ours) |
|---|---|---|---|---|---|
| Params (M) ↓ | 63.5 | 29.3 | 19.6 | 30 | **15.2** |
| GFLOPs ↓ | 262.4 | 9.7 | **2.0** | 109.0 | 183.0 |

## B. Additional ablation study

To quantify the individual contributions of task synergies and DTSPrompt, we conducted an additional ablation experiment and presented the results in Table 2. From the table, it can be observed that both DTSPrompt and multitask synergy lead to improvements. The improvement brought by multitask synergy may be attributed to the relatively small amount of binarization training data. The incorporation of other task data aids the model in avoiding overfitting to the limited binarization training data.

Table 2. For binarization, we gradually add DTSPrompt and multi-task to provide further insights of these factors.

| DTSPrompt | Multi-task | DIBCO'18 (FM↑/pFM↑/PSNR↑) |
|---|---|---|
| ✗ | ✗ | 76.5 / 79.5 / 14.4 |
| ✓ | ✗ | 87.1 / 90.6 / 18.6 |
| ✓ | ✓ | **89.8 / 94.3 / 19.3** |

## C. Comparison with more SOTA

We provide quantitative comparisons with more state-of-the-art (SOTA) methods for dewarping task in Table 3, which demonstrates the superior performance of DocRes when compared with some of these task-specific methods.

Table 3. Quantitative comparison with existing SOTA task-specific document dewarping methods on the DIR300 [2] benchmark.

| Methods | Venue | MS-SSIM↑ | LD↓ | AD↓ |
|---|---|---|---|---|
| DewarpNet [1] | ICCV'19 | 0.492 | 13.94 | 0.254 |
| DDCP [9] | ICDAR'21 | 0.552 | 10.95 | 0.331 |
| DocTr [1] | MM'21 | 0.616 | 7.21 | 0.254 |
| DocGeoNet [2] | ECCV'22 | <u>0.638</u> | <u>6.40</u> | 0.242 |
| PaperEdge [6] | SIGGRAPH'22 | 0.583 | 8.00 | 0.255 |
| Li et al. [3] | ICCV'23 | 0.607 | 7.68 | 0.244 |
| LA-DocFlatten [4] | ACM TOG'23 | **0.651** | **5.70** | **0.195** |
| DocRes (ours) | - | 0.626 | 6.83 | <u>0.241</u> |

## D. More visualized results

In Fig. 1, we present additional visualization results of DocRes across multiple tasks. Additionally, in Fig. 2, we illustrate the potential issue of error accumulation when applying DocRes for end-to-end camera-captured document image enhancement. Due to the iterative nature of DocRes for end-to-end tasks, where each forward pass utilizes the output of the previous pass as input, errors in one task can accumulate and affect the final result. Exploring methods that can accomplish multiple document restoration tasks in a single forward pass would be a meaningful avenue for future research.

## E. Further discussions about DTSPrompt

In fact, from the ablation experiments in Table 2 of the main text, it is evident that compared to fixed prompts or task-specific models, DTPrompt does not demonstrate significant advantages and may even perform worse in certain tasks, such as shadow removal and deblurring tasks. This is mainly because document image restoration tasks like deblurring lack recognized prior features that definitively enhance performance. In such tasks, DTPrompt primarily serves as a discriminative cue rather than significantly boosting model performance. As for shadow removal tasks, while shadow maps have been widely proven to be useful, our extraction operations are very simple, involving basic image processing techniques rather than the conventional

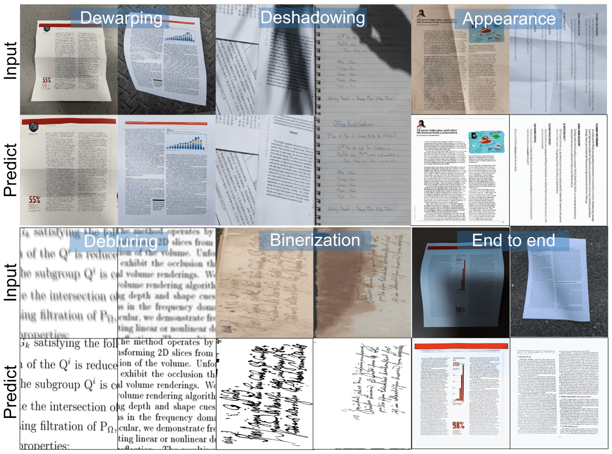


Figure 1. More visualized results from DocRes. Zoom in for best view.

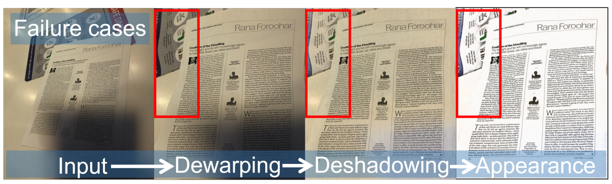


Figure 2. Failure case from DocRes when applying it for end-to-end camera-captured document image enhancement. Zoom in for best view.

approach [10, 11] of using deep models for prediction.

Based on these observations and analyses, future improvements for DTSPrompt could focus on two directions: **1.** Exploring more effective prior features for specific tasks: This entails delving deeper into identifying prior features that are more conducive to enhancing performance for particular tasks, such as deblurring. **2.** Employing a trainable prediction module for the DTSPrompt generator: This would enhance the prior feature extraction capabilities. Importantly, there's still no necessity to designate a separate DTSPrompt generator for each task. Instead, a single DTSPrompt generator with shared parameters could simultaneously output multiple prior features. During input to the restoration network, different prior features can be chosen for task guidance based on the specific task at hand.

## References

[1] Hao Feng, Yuechen Wang, Wengang Zhou, Jiajun Deng, and Houqiang Li. DocTr: Document image transformer for geometric unwarping and illumination correction. In *ACM MM*, pages 273–281, 2021. 1

[2] Hao Feng, Wengang Zhou, Jiajun Deng, Yuechen Wang, and Houqiang Li. Geometric representation learning for document image rectification. In *ECCV*, pages 475–492, 2022. 1

[3] Heng Li, Xiangping Wu, Qingcai Chen, and Qianjin Xiang. Foreground and text-lines aware document image rectification. In *CVPR*, pages 19574–19583, 2023. 1

[4] Pu Li, Weize Quan, Jianwei Guo, and Dong-Ming Yan. Layout-aware single-image document flattening. *ACM Transactions on Graphics*, 2023. 1

[5] Zinuo Li, Xuhang Chen, Chi-Man Pun, and Xiaodong Cun. High-resolution document shadow removal via a large-scale real-world dataset and a frequency-aware shadow erasing net. In *CVPR*, pages 12449–12458, 2023. 1

[6] Ke Ma, Sagnik Das, Zhixin Shu, and Dimitris Samaras. Learning from documents in the wild to improve document unwarping. In *ACM SIGGRAPH*, pages 1–9, 2022. 1

[7] Mohamed Ali Souibgui and Yousri Kessentini. DE-GAN: A conditional generative adversarial network for document enhancement. *IEEE TPAMI*, 44(3):1180–1191, 2020. 1

[8] Yonghui Wang, Wengang Zhou, Zhenbo Lu, and Houqiang Li. UDoc-GAN: Unpaired document illumination correction with background light prior. In *ACM MM*, pages 5074–5082, 2022. 1

[9] Guo-Wang Xie, Fei Yin, Xu-Yao Zhang, and Cheng-Lin Liu. Document dewarping with control points. In *ICDAR*, pages 466–480, 2021. 1

[10] Jiaxin Zhang, Lingyu Liang, Kai Ding, Fengjun Guo, and Lianwen Jin. Appearance enhancement for camera-captured document images in the wild. *IEEE Transactions on Artificial Intelligence*, 2023. 2

[11] Ling Zhang, Yinghao He, Qing Zhang, Zheng Liu, Xiaolong Zhang, and Chunxia Xiao. Document image shadow removal guided by color-aware background. In *CVPR*, pages 1818–1827, 2023. 2