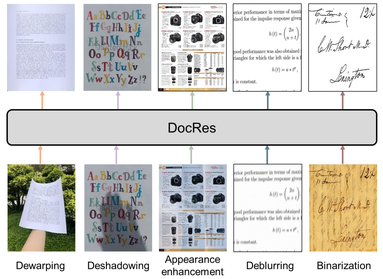
DocRes: A Generalist Model Toward Unifying Document Image Restoration Tasks

Jiaxin Zhang[1,2], Dezhi Peng[1], Chongyu Liu[1], Peirong Zhang[1], Lianwen Jin[1,2,*]
[1] South China University of Technology
[2] INTSIG-SCUT Joint Lab on Document Analysis and Recognition

CVPR SEATTLE, WA JUNE 17-21, 2024

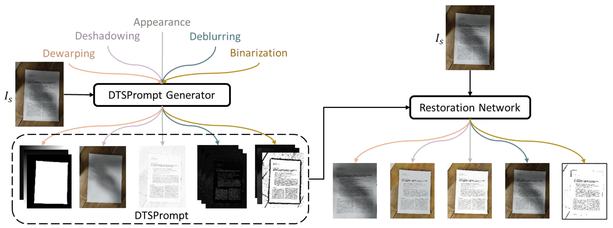
A generalist model for document image restoration



- Prevailing methods address distinct document restorations independently
- Task-specific models result in intricate systems and the incapability to harness the potential synergies of multi-task learning.
- To overcome this challenge, we propose DocRes, a generalist model that unifies five document image restoration tasks.

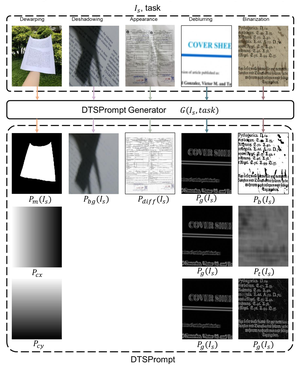
Methodology

1. Overall



- Obtain DTSPrompt based on the distorted input image I_S and given task: $DTSPrompt = G(I_s, \text{task})$, $G()$: DTSPrompt Generator
- Concatenate the prompt and distorted images together and input them into a pre-existing restoration network.

2. Dynamic task specific prompt (DTSPrompt)



$P_m(\ )$: *document region mask*
$P_{cx}(i,j) = i$: *x − coordinate values of the pixel at* (i,j)
$P_{cy}(i,j) = j$: *y − coordinate values of the pixel at* (i,j)
$P_{bg}(\ )$: *document background extracted by dilation and median filtering*
$P_g(\ )$: *gradient map extracted by Sobel filtering*
$P_b(\ )$: *initial binarization result from Sauvola algorithm*
$P_b(\ )$: *threshold map from Sauvola algorithm*

$G(I_s, \text{“dewarp”}) = [P_m(I_s), P_{cx}, P_{cy}]$
$G(I_s, \text{“deshadow”}) = P_{bg}(I_s)$
$G(I_s, \text{“appearance”}) = 255 - abs(I_s - P_{bg}(I_s))$
$G(I_s, \text{“deblur”}) = [P_g(I_s), P_g(I_s), P_g(I_s)]$
$G(I_s, \text{“binarize”}) = [P_b(I_s), P_t(I_s), P_g(I_s)]$

- Comprises distinct prior features based on the characteristics of each task.
- Serve as (1) an cue to determine which task to perform and (2) supplementary information to enhance restoration performance for the corresponding task.
- A vision-centric prompt that has the same shape as distorted image. It can effectively adapt to variable and high resolutions, as well as patch training and whole-image testing in low-level tasks.

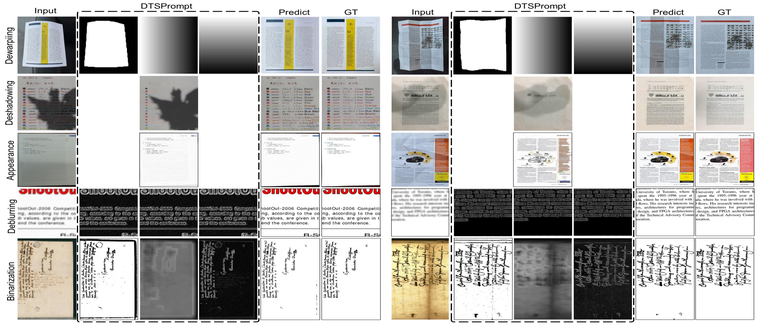
Comparisons with SOTA

Metrics	Datasets	DocGeo [11] *ECCV'22*	Li et al. [23] *ICCV'23*	BGSNet [71] *CVPR'23*	DocShadow [25] *ICCV'23*	UDoc-GAN [59] *MM'22*	GCDRNet [70] *TAI'23*	GDB [62] *PR'24*	DE-GAN [50] *TPAMI'20*	DocDiff [61] *MM'23*	DocRes (ours)
MSSSIM↑	DIR300 [11]	**0.6380**	0.6070								0.6264
AD↓		0.242	0.244	-	-	-	-	-	-	-	**0.241**
LD↓		**6.40**	7.68								6.83
SSIM↑ PSNR↑	Kligler et al. [19]	-	-	**0.9480** **29.17**	0.9088 25.12	-	-	-	-	-	0.9005 27.14
	Jung et al. [18]	-	-	0.9040 17.34	0.9005 21.05	-	-	-	-	-	**0.9089** **23.02**
	OSR [55]	-	-	**0.9388** **22.64**	0.9023 18.25	-	-	-	-	-	0.9370 21.64
SSIM↑ PSNR↑	DocUNet* [33]	-	-	-	-	0.6833 14.29	**0.7658** 17.09	-	-	-	0.7592 **17.55**
	RealDAE [70]	-	-	-	-	0.7558 16.43	**0.9423** 24.42	-	-	-	0.9168 **24.62**
SSIM↑ PSNR↑	TDD [14]	-	-	-	-	-	-	-	0.9226 22.24	0.9559 24.00	**0.9723** **27.35**
FM↑ pFM↑ PSNR↑	DIBCO'18 [46]	-	-	-	-	-	-	**91.09** **94.57** **19.92**	77.59 85.74 16.16	88.11 90.43 17.92	89.82 94.33 19.35

- Competes with existing task-specific SOTA and surpasses them in several instances.
- Achieves new records in certain metrics for benchmark related to dewarping, deshadowing, deblurring, and appearance enhancement.
- While dedicated SOTA model GDB still holds the top position for binarization, DocRes exhibits performance closely trailing behind it.

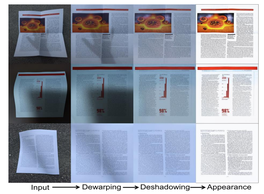
Visualization

1. On benchmark dataset



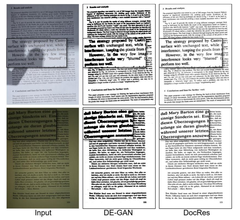
2. Control ability and potentials for end-to-end enhancement

Visualization results for DTSPrompt's control ability



Visualization results for DocRes's application on end-to-end enhancement

3. Generization



Applying DocRes to perform binarization on photographed documents, which were merely trained on scanned ancient document binarization data.

Visualization results when applying models to perform the deblurring task on photographed documents, which were merely trained on clean document blurring data.

Contact & Codes

Jiaxin Zhang
Email: eejxzhang@gmail.com