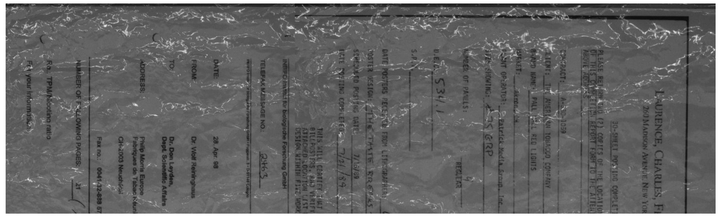
SHDocs: A dataset, benchmark, and method to efficiently generate high-quality, real-world specular highlight data with near-perfect alignment

Jovin Leong, Ming Di Koa, Benjamin Cham, Shaun Heng

Background

Specular highlights are bright reflections of light that appear as bright spots of glare which occlude computer vision models in tasks such as OCR and robotics where specularity is persistent yet challenging to reconstruct.



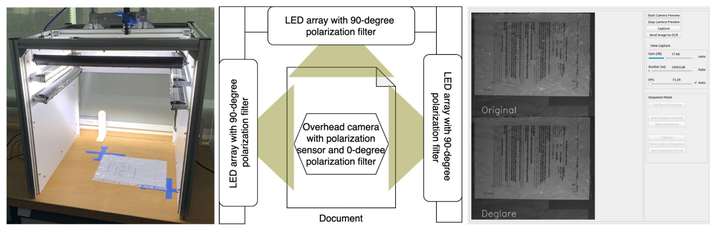
Specular highlights on documents

The lack of quality datasets in this space motivated us to develop:

1. A process to generate quality, aligned specular highlight data cheaply
2. A benchmark for specular highlight removal based on impact to OCR
3. The SHDocs dataset: A dataset of real-world specular highlights on document images based with aligned ground truth annotations

Building the SHDocs dataset

We built an enclosure illuminated by LED arrays attached to polarizing filters at fixed polarization angles orthogonal to a polarization sensor. The polarization sensor captures and filters out specularity from the polarized lights to generate the aligned, deglared counterfactuals. The enclosed setup helps to reduce specularity from unpolarized light.



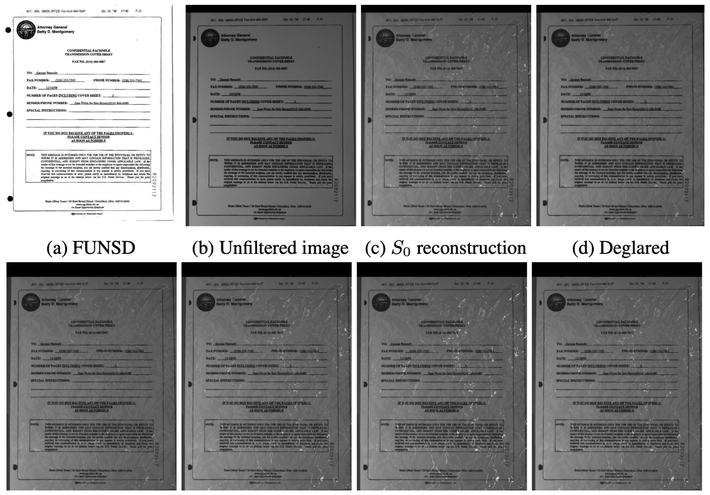
(a) Physical setup (b) Setup diagram (c) GUI

SHDocs data collection setup and data collection application interface

We used the FUNSD dataset of 199 annotated documents as our base set of document data which we printed and inserted into transparent filing pockets to generate specular highlights through their reflective surfaces. We developed a custom application to facilitate image capture; for each image captured, our polarization sensor obtains 4 images corresponding to 4 different polarization angles.

The SHDocs dataset

The SHDocs dataset contains 3000+ document scenes with 19000+ images comprising of the raw image captures from the polarization sensor and the reconstructions of the S_0 normalized image with the deglared counterfactual.



(a) FUNSD (b) Unfiltered image (c) S_0 reconstruction (d) Deglared

(e) i_0 (f) i_{45} (g) i_{90} (h) i_{135}

The SHDocs benchmark

We found that conventional quantitative image quality assessments with Pixel-Signal-to-Noise ratio (PSNR), Structural Similarity Index Measure (SSIM), and Universal Image Quality Index (UIQI) had limited evaluative utility when mapping outcomes back to human observation.

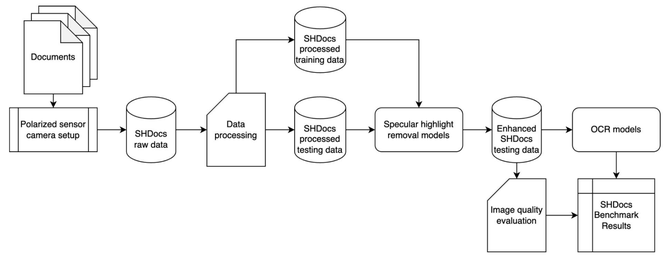


Figure 1: Process pipeline for data collection, model development, and benchmarking

As such, we devised a benchmark that evaluates specular highlight removal models by passing their enhanced images through OCR engines to assess image enhancement outcomes through downstream text recovery performance.

We found that our benchmark constitutes a more reflective discriminator of specular highlight removal models for text recovery as it tangibly relates enhancement outcomes to the downstream use of the images in OCR.

Table 3: OCR performance on SHDocs evaluation dataset

	Textract			EasyOCR			Tesseract		
Baseline	WER	CER	LED	WER	CER	LED	WER	CER	LED
Original FUNSD	0.343	0.096	2.12	0.656	0.260	5.65	0.470	0.270	5.99
No filter	0.415	0.134	3.06	0.907	0.583	12.7	0.7314	0.586	12.9
No enhancement	0.593	0.358	7.89	0.950	0.723	15.8	0.846	0.686	15.1
Model									
Fu et al. [11]	0.960	0.865	18.8	1.00	0.998	21.6	1.00	0.960	20.8
M2-Net [18]	0.698	0.468	10.3	0.995	0.906	19.7	0.933	0.794	17.3
TSHRNet [9]	0.591	0.353	7.77	0.958	0.744	16.2	0.847	0.683	15.0
Hu et al. [17]	0.708	0.484	10.6	0.996	0.914	19.8	0.941	0.817	17.8
MIMO-UNetPlus [5]	**0.582**	**0.342**	**7.53**	**0.942**	**0.686**	**15.0**	**0.824**	**0.653**	**14.3**

The SHDocs dataset and our benchmark enabled us to effectively discern between specular highlight removal models in terms of their ability to enhance images that more accurately maps to their real-world utility. We found that even a generic U-Net model trained our dataset is able to impactfully improve image enhancement performance over leading models in the space. Further generalization study of SHDocs gave evidence for general utility across hardware and specular highlight removal tasks.

Conclusions and future work

Our work demonstrates limitations in existing methods of specular highlight removal for textual data and we provide the means to advance research in the image enhancement space through our dataset, benchmark, and method of specular highlight data collection.

Future work can leverage our dataset and benchmark to develop impactful and practical image enhancement models that are better suited to textual images. Additionally, our work supports the development of text-aware image enhancement models and vision-text foundation models through our aligned and annotated dataset. Finally, by including all unprocessed frames from 4 polarization angles, adjacent domains such as such as light modeling and simulation can similarly leverage our work.

https://go.gov.sg/shdocs