Supplementary material for SyntheticDoc: A Large Synthetic Dataset for Document Unwarping and Illumination Correction

Daniel Woortmann*[1], Tanguy Magne*[1], and Olga Sorkine-Hornung[1]

ETH Zurich, Rämistrasse 101, 8092 Zurich, Switzerland
d.woortmann@gmail.com
{tanguy.magne,olga.sorkine}@inf.ethz.ch

1 Related works

Developability. Developability is a fundamental characteristic of paper, reflecting the physical constraint that it cannot stretch and can always be flattened to a plane without any stretching or shearing. A defining property of developable surfaces is their zero Gaussian curvature. In the discrete case, where surface 3D geometry is represented by a mesh, Gaussian curvature can be approximated by the angle defect at each vertex, computed as $2\pi - \sum_i \alpha_i$, where the α_i denote the angles formed by consecutive edges of the mesh faces incident on the vertex. We measure this quantity across all samples of several datasets.

Doc3D [2] was created by 3D scanning deformed physical documents and postprocessing the raw data. While the meshes are not explicitly provided, they can be extracted from the available 3D world coordinates. We utilize the 3D grids extracted by the authors of UVDoc [10] to compute the Gaussian curvature. UVDoc [10] itself is annotated with 3D point grids representing the samples that can be connected into meshes. LA-Doc [6] is another synthetic dataset generated similarly to SyntheticDoc. It is rendered using Blender [1] with simulated meshes. However, the dataset is limited to 90,000 samples and relies on a simplistic mass-spring system to generate the meshes, rather than an advanced physics simulation engine. Furthermore, only the generated meshes are publicly available, rather than the fully rendered images.

Fig. 1 presents the distribution of the angle defect for all interior vertices across these three datasets and our SyntheticDoc dataset. As shown, the simple simulation used for LA-Doc [6] produces meshes with large angle defects and consequently high Gaussian curvature. This indicates that the meshes are non-developable and fail to accurately represent physical paper. Doc3D [2] and UVDoc [10] exhibit similar Gaussian curvature distributions, which are overall slightly higher than that of our SyntheticDoc. This disparity stems from their data capture process, which slightly over-smooths the geometries while preserving the global shape. SyntheticDoc achieves the lowest overall angle defect, demonstrating the high physical fidelity of our mesh generation procedure.

* Equal contribution.

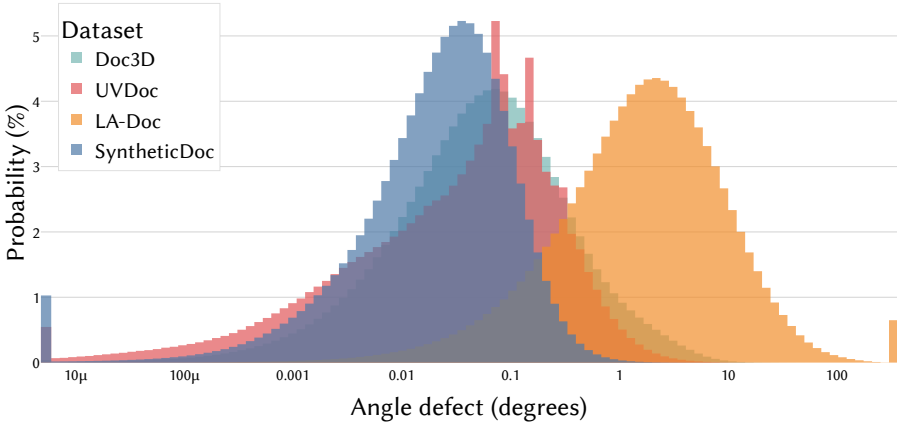


Fig. 1: Distribution of the angle defect across all interior vertices in the meshes from 3 existing datasets and our SyntheticDoc. The angle defect is a discrete approximation of Gaussian curvature, and paper, which is a developable surface, should have zero Gaussian curvature. The x-axis is on a logarithmic scale. Our dataset has the smallest mean and overall smaller angle defects, indicating superior developability and making our representation of paper more physically accurate.

RealDAE dataset. As mentioned in the main text, RealDAE [16] is the only dataset of real-world documents dedicated to illumination correction. Because its ground-truth images were obtained through manual corrections in Adobe Photoshop, a laborious annotation process, it is restricted to only 450 training samples. We highlight a critical issue regarding the training split of this dataset: 12 of its training samples are directly taken from the DocUNet benchmark [9]. Since DocUNet serves as a standard evaluation benchmark for both document unwarping and illumination correction, the inclusion of these samples in a training set introduces severe data leakage. Illumination correction methods are commonly evaluated on the DocUNet benchmark after unwarping via DocAligner [15], while the samples included in the RealDAE training set are still warped documents. Nevertheless, this remains an issue because the underlying shadow patterns are identical. More importantly, recent multi-task models [17, 19] simultaneously address several document enhancement tasks and are trained on diverse datasets and evaluated on multiple benchmarks. Therefore, including RealDAE in the training mixture, as done by the DocRes [17] and Uni-DocDiff [19] methods, inadvertently compromises the integrity of the benchmark, potentially inflating evaluation scores. We thus argue that these overlapping samples should be removed from the RealDAE dataset prior to training any illumination correction model, if the DocUNet benchmark is to be used for evaluation.

2 Experiments

Metrics. As discussed in the main text, we assess the performance of our model using a combination of image similarity and OCR metrics.

For image similarity, we employ various metrics: MS-SSIM, PSNR, LD and AD. The structural similarity index measure SSIM [13] evaluates visual resemblance between two images by comparing the mean and variance of pixel values across image patches. MS-SSIM [12] is a multi-scale extension of this approach, where the metric is computed across various scales of a Gaussian pyramid and averaged across these scales. We adopt the original weighting scheme [12] for this cross-scale averaging. Peak signal-to-noise ratio (PSNR) provides a logarithmic measure of reconstruction fidelity by calculating the ratio between the maximum possible signal power and the mean squared error (MSE) of the approximation. Local distortion (LD) [14] relies on a dense SIFT flow mapping between the ground-truth and unwarped images [7]. It measures the mean L_2 distance between corresponding pixels to quantify the local distortion remaining in the unwarped image. Aligned distortion (AD), introduced by Ma et al. [8], improves upon LD by being more robust to global translation and scaling, focusing primarily on text and visual content rather than the background.

To evaluate optical character recognition (OCR) performance, we first extract the reference text from the flat ground-truth scans and compute the ED and CER metrics. The edit distance (ED) is a standard Levenshtein distance [5], measuring the minimum number of character substitutions, insertions and deletions required to transform the OCR prediction of the unwarped document into the reference text. The character error rate (CER) is the ratio between the ED and the total number of characters in the reference text.

As previously noted by Verhoeven et al. [10], OCR metrics are not well suited for unwarped documents that retain their original shadows. We want to highlight this limitation, as we observed these metrics to be extremely volatile throughout a single training run. Applying an OCR engine to shadowed images can result in drastically different predictions even when the inputs appear visually identical, as demonstrated in Fig. 2. In addition, because the OCR metrics for the DocUNet benchmark [9] are computed on a restricted subset of only 50 samples, the resulting scores are highly sensitive to extreme values. This volatility is compounded by the fact that the edit distance (ED) is an unnormalized metric, meaning its average is highly influenced by text-heavy samples. In such documents, the effective text resolution is significantly lower, causing the OCR engine to produce unstable predictions. In contrast, while the character error rate (CER) remains somewhat sensitive to these text-heavy images, it is normalized by the number of characters, and its average across all samples provides a more reliable representation of the performance.

We would therefore like to emphasize that results on these OCR-related metrics should be interpreted with caution.

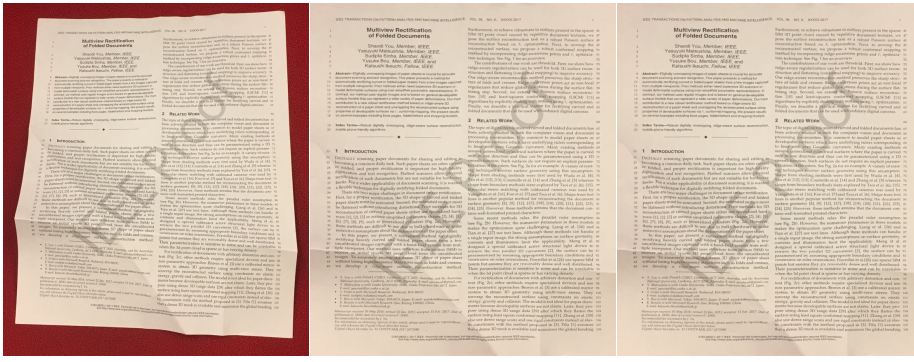
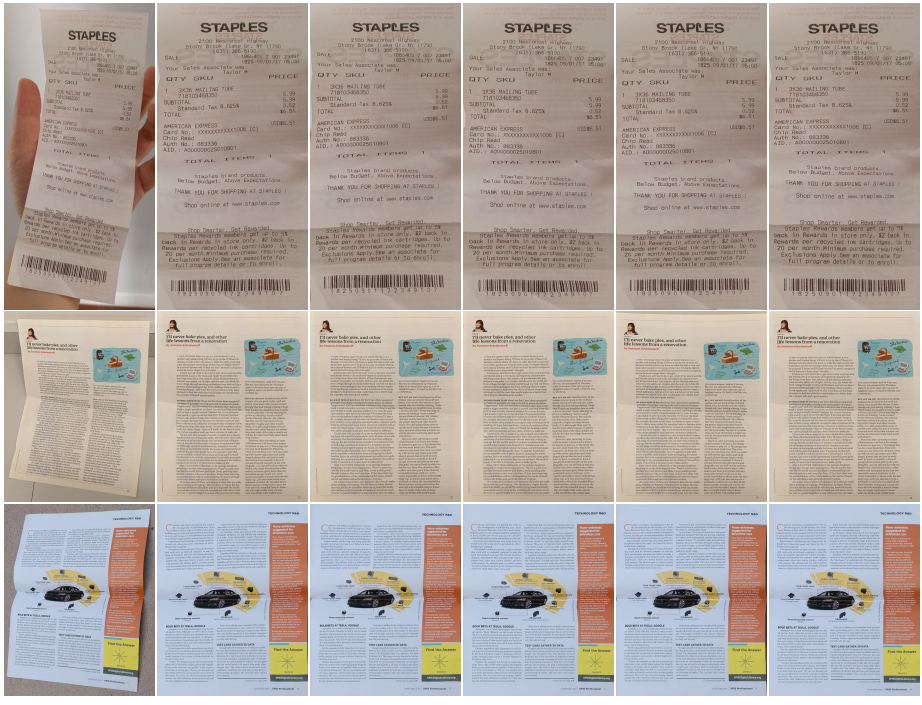


Fig. 2: Sample 49_2 from the DocUNet benchmark [9], unwarped using two versions of our model separated by a single training epoch. While the resulting images are visually nearly identical (achieving an MS-SSIM of 0.969 and an LD of 0.215 between them), their OCR metrics vary drastically. The prediction in the middle yields an ED of 1363 and a CER of 0.163, whereas the prediction on the right results in an ED of 6030 and a CER of 0.720 when compared to the ground-truth scan.



Input UVDoc [10] DocScanner [4] DvD [18] AADD [11] Ours

Fig. 3: Qualitative comparisons of unwarping-only results produced by various document unwarping methods on the DocUNet benchmark.

Input Ours (shading) DocTr [3] DocRes [17] Ours Ground-Truth

Fig. 4: Qualitative comparisons of results produced by various document unwarping plus illumination correction methods on the DocUNet benchmark. The second column presents the shading predicted by our network.

References

1. Blender Foundation: Blender 4.5 (2026), https://www.blender.org/ 1
2. Das, S., Ma, K., Shu, Z., Samaras, D., Shilkrot, R.: Dewarpnet: Single-image document unwarping with stacked 3d and 2d regression networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (October 2019) 1
3. Feng, H., Wang, Y., Zhou, W., Deng, J., Li, H.: Doctr: Document image transformer for geometric unwarping and illumination correction. In: Proceedings of the 29th ACM International Conference on Multimedia. p. 273–281. MM '21, Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3474085.3475388 5
4. Feng, H., Zhou, W., Deng, J., Tian, Q., Li, H.: Docscanner: Robust document image rectification with progressive learning. Int. J. Comput. Vision **133**(8), 5343–5362 (May 2025). https://doi.org/10.1007/s11263-025-02431-5, https://doi.org/10.1007/s11263-025-02431-5 4
5. Levenshtein, V.I.: Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady **10**(8), 707–710 (1966) 3
6. Li, P., Quan, W., Guo, J., Yan, D.M.: Layout-aware single-image document flattening. ACM Trans. Graph. **43**(1) (Nov 2023). https://doi.org/10.1145/3627818 1
7. Liu, C., Yuen, J., Torralba, A., Sivic, J., Freeman, W.T.: Sift flow: Dense correspondence across different scenes. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) Computer Vision – ECCV 2008. pp. 28–42. Springer Berlin Heidelberg, Berlin, Heidelberg (2008) 3
8. Ma, K., Das, S., Shu, Z., Samaras, D.: Learning from documents in the wild to improve document unwarping. In: ACM SIGGRAPH 2022 Conference Proceedings. SIGGRAPH '22, Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3528233.3530756 3
9. Ma, K., Shu, Z., Bai, X., Wang, J., Samaras, D.: Docunet: Document image unwarping via a stacked u-net. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2018) 2, 3, 4
10. Verhoeven, F., Magne, T., Sorkine-Hornung, O.: Uvdoc: Neural grid-based document unwarping. In: SIGGRAPH Asia 2023 Conference Papers. SA '23, Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3610548.3618174 1, 3, 4
11. Wang, C., Shen, I.C., Igarashi, T., Jiang, C.: Axis-aligned document dewarping (2025), https://arxiv.org/abs/2507.15000 4
12. Wang, Z., Simoncelli, E., Bovik, A.: Multiscale structural similarity for image quality assessment. In: The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. vol. 2, pp. 1398–1402 Vol.2 (2003). https://doi.org/10.1109/ACSSC.2003.1292216 3
13. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004). https://doi.org/10.1109/TIP.2003.819861 3
14. You, S., Matsushita, Y., Sinha, S., Bou, Y., Ikeuchi, K.: Multiview rectification of folded documents. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(2), 505–511 (2018). https://doi.org/10.1109/TPAMI.2017.2675980 3
15. Zhang, J., Chen, B., Cheng, H., Guo, F., Ding, K., Jin, L.: Docaligner: Annotating real-world photographic document images by simply taking pictures (2023), https://arxiv.org/abs/2306.05749 2

16. Zhang, J., Liang, L., Ding, K., Guo, F., Jin, L.: Appearance enhancement for camera-captured document images in the wild. IEEE Transactions on Artificial Intelligence **5**(5), 2319–2330 (2024). https://doi.org/10.1109/TAI.2023.3321257 2
17. Zhang, J., Peng, D., Liu, C., Zhang, P., Jin, L.: Docres: A generalist model toward unifying document image restoration tasks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 15654–15664 (June 2024) 2, 5
18. Zhang, W., Lu, H., Ning, M., Huang, X., Wang, W., Huang, K., Wang, Q.: Dvd: Unleashing a generative paradigm for document dewarping via coordinates-based diffusion model. In: Proceedings of the SIGGRAPH Asia 2025 Conference Papers. SA Conference Papers '25, Association for Computing Machinery, New York, NY, USA (2025). https://doi.org/10.1145/3757377.3763913 4
19. Zhao, F., Zeng, W., Li, Z., Yang, D., Li, B., Bi, X., Zhou, Y.: Uni-docdiff: A unified document restoration model based on diffusion. In: Proceedings of the 33rd ACM International Conference on Multimedia. p. 8204–8213. MM '25, Association for Computing Machinery, New York, NY, USA (2025). https://doi.org/10.1145/3746027.3755362 2