Revisiting Document Image Dewarping by Grid Regularization
Supplementary Material

1. Qualitative Results for CER Comparison

Fig. 1 qualitatively compared our proposed method to DewarpNet. Although DewarpNet could obtain horizontally-aligned rectification results, the local distortions in the text regions (marked in red rectangles) limited the performance of readability. For our method, benefiting from the structural regularization of the text lines and the document boundaries, we handles the local distortion of paper-sheets very well. As a result, our method reduced the CER score by large margins.

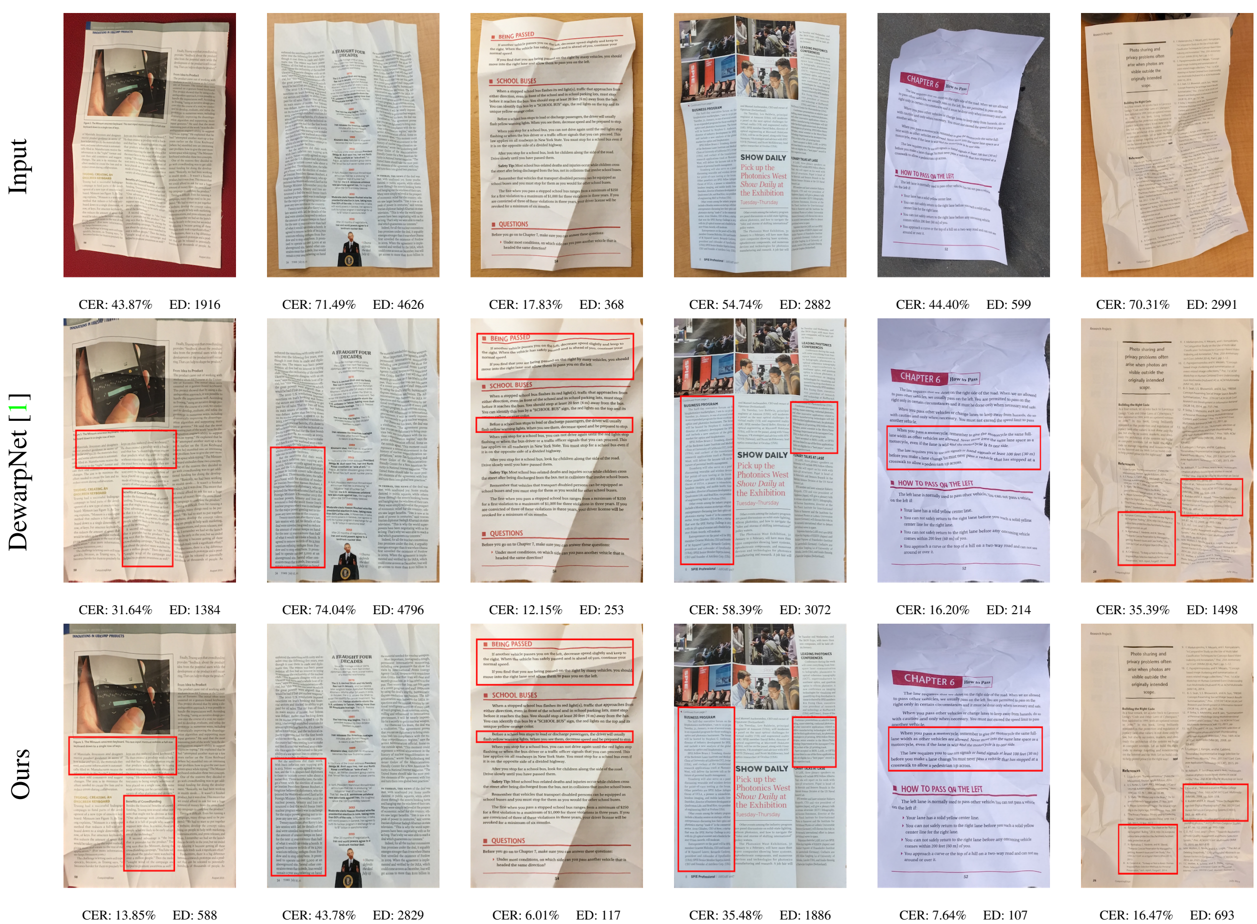


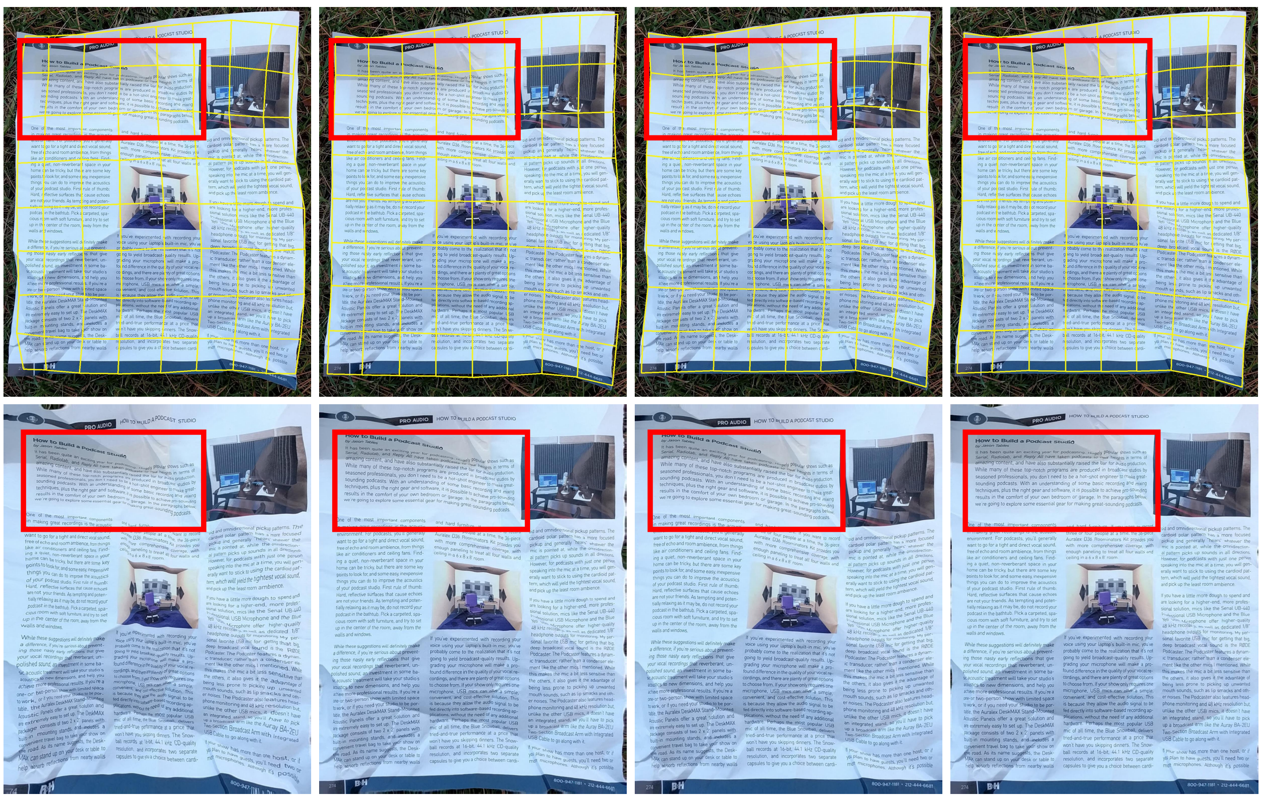
Figure 1. Qualitative comparison with the metric of CER. From top to bottom, we show the input images, the results of DewarpNet and our results with the CER and ED scores.

2. Comparison for the Different Deformation Grid Generation

As discussed in our submission, the deformation grids played an important role for the final dewarping results. Therefore, we visualize the deformation grids generated by the different approaches including (DocUNet [3], DewarpNet [1], TransFinite Interpolation (TFI) [2] and our method) and the corresponding dewarping results in Fig. 2.

Specifically, DocUNet [3] pays too much attention to the value of each point in the deformation grid while ignoring the property of the grid, thereafter resulting in the distortion of text lines in the rectified images. For the methods of DewarpNet [1] and TFI [2] that focus on the uniformly-distributed grids, they ignore the deformation in the interior of the document image. Compared to TFI [2], the DewarpNet has better rectification results in the internal regions.

Different from the prior art that learns deformation fields [1, 3], our method exploited the geometric regularization for the document boundaries and the text lines. As a result, the grid generated by our method is tightly attached to the text lines for better dewarping results (Fig. **??**).

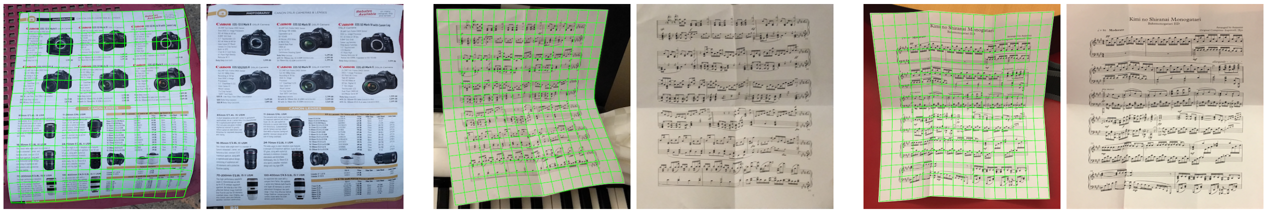


DocUNet DewarpNet Interpolation Proposed

Figure 2. Deformation grid on the original image for different approaches. Compared to the prior art, the deformation grid by our approach is generated by the guidance of boundaries and text lines of the document image.

3. Failure cases

We show the results of our method in Fig. 3 for your mentioned cases. Although some required inputs of our method would be inaccurate or missing, our method is able to obtain decent dewarping results benefiting from our design that couples the learned cues and the robust optimization method in a reasonable way. Since we use DocUNet to directly regress the boundary bm(backward mapping) instead of segmentation, the boundary information will certainly be output. When the learned boundaries are not accurate, the internal regions still provide strong constraints based on the text information with only a marginal degeneration obtained. When an image has no text lines, our method degenerates into a boundary + TFI (TPS) mode and still obtains pretty good results (As reported in Table 4). If the text line is disconnected, the disconnected text lines are regarded as separate text lines and constrained separately, which basically has no effect on the result. This can be seen in the fourth example in Fig. 6 of our main paper.



(a) Boundary failure (b) No text (c) Sparse text

Figure 3. Illustration of the failure cases. [Best view in zoom-in]

References

[1] Sagnik Das, Ke Ma, Zhixin Shu, Dimitris Samaras, and Roy Shilkrot. Dewarpnet: Single-image document unwarping with stacked 3d and 2d regression networks. October 2019.

[2] William J Gordon and Charles A Hall. Transfinite element methods: blending-function interpolation over arbitrary curved element domains. *Numerische Mathematik*, 21(2):109–129, 1973.

[3] Ke Ma, Zhixin Shu, Xue Bai, Jue Wang, and Dimitris Samaras. Docunet: Document image unwarping via a stacked u-net. In *CVPR*, June 2018.