

Supplementary Material: Inv3D: A High-Resolution 3D Invoice Dataset for Template-Guided Single-Image Document Unwarping

Felix Hertlein^{1,2*}, Alexander Naumann^{1,2} and Patrick Philipp¹

¹FZI Research Center for Information Technology, Haid-und-Neu-Str. 10–14, Karlsruhe, 76131, Baden-Württemberg, Germany.

²KIT Karlsruhe Institute of Technology, Kaiserstraße 12, Karlsruhe, 76131, Baden-Württemberg, Germany.

*Corresponding author(s). E-mail(s): hertlein@fzi.de;
Contributing authors: anaumann@fzi.de; philipp@fzi.de;

Evaluation of state of the art with respect to modifications

Tables 1 and 2 show an in depth inspection of DewarpNet (w/o refinement network) [Das et al (2019)] and GeoTr [Feng et al (2021)] on the Inv3DReal benchmark split by the modification type. Both networks were trained on our proposed Inv3D dataset.

Both approaches agree upon *perspective* and *curled* deformation, as well as the *bright* setting as the easiest modifications. Similar to our model GeoTrTemplateLarge, the metric local distortion (LD) yields best values with crumples and *shadow*.

References

- Das S, Ma K, Shu Z, et al (2019) Dewarpnet: Single-image document unwarping with stacked 3d and 2d regression networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 131–140
- Feng H, Wang Y, Zhou W, et al (2021) Doctr: Document image transformer for geometric unwarping and illumination correction. In: Proceedings of the 29th ACM International Conference on Multimedia, pp 273–281

Modification	↑MS-SSIM	↓LD	↓LPIPS	↓ED	↓CER
perspective	0.61 (0.10)	29.23 (11.52)	0.37 (0.12)	357 (185)	0.56 (0.27)
curled	0.60 (0.09)	29.18 (10.77)	0.38 (0.10)	344 (154)	0.54 (0.22)
fewfold	0.58 (0.10)	26.46 (9.51)	0.40 (0.11)	377 (191)	0.58 (0.26)
multifold	0.53 (0.09)	25.87 (9.33)	0.44 (0.11)	404 (167)	0.62 (0.21)
crumples easy	0.55 (0.11)	21.51 (7.75)	0.41 (0.11)	377 (177)	0.59 (0.27)
crumples hard	0.45 (0.09)	19.76 (10.68)	0.55 (0.10)	462 (163)	0.71 (0.17)
bright	0.60 (0.11)	24.70 (9.80)	0.35 (0.10)	324 (170)	0.50 (0.23)
color	0.57 (0.09)	29.15 (11.04)	0.48 (0.13)	369 (184)	0.57 (0.25)
shadow	0.48 (0.09)	22.16 (9.67)	0.44 (0.09)	468 (140)	0.73 (0.18)

Table 1: Detailed evaluation of DewarpNet (w/o refinement network) [Das et al (2019)] on our new benchmark Inv3DReal split by modification category. Values in brackets denote standard deviations.

Modification	↑MS-SSIM	↓LD	↓LPIPS	↓ED	↓CER
perspective	0.61 (0.10)	26.64 (10.57)	0.36 (0.12)	342 (192)	0.53 (0.27)
curled	0.60 (0.09)	23.69 (9.67)	0.36 (0.10)	318 (156)	0.50 (0.23)
fewfold	0.58 (0.10)	24.32 (9.72)	0.38 (0.11)	372 (203)	0.57 (0.26)
multifold	0.53 (0.10)	24.32 (11.13)	0.43 (0.11)	363 (161)	0.56 (0.21)
crumples easy	0.56 (0.10)	18.88 (7.54)	0.39 (0.10)	360 (196)	0.57 (0.30)
crumples hard	0.46 (0.09)	19.03 (8.63)	0.53 (0.10)	434 (158)	0.67 (0.18)
bright	0.60 (0.11)	21.81 (10.27)	0.34 (0.10)	297 (180)	0.46 (0.24)
color	0.58 (0.09)	26.18 (9.81)	0.46 (0.13)	345 (180)	0.53 (0.25)
shadow	0.49 (0.09)	20.45 (8.98)	0.42 (0.10)	453 (145)	0.70 (0.19)

Table 2: Detailed evaluation of GeoTr [Feng et al (2021)] on our new benchmark Inv3DReal split by modification category. Values in brackets denote standard deviations.