DeSTNet: Densely Fused Spatial Transformer Networks – Supplementary Material

Roberto Annunziata
roberto.annunziata@onfido.com
Christos Sagonas
christos.sagonas@onfido.com
Jacques Calì
jacques.cali@onfido.com

Onfido Research
3 Finsbury Avenue
London, UK

1 Additional Results for Section 4.1

Figures 1 and 2 show additional alignment results obtained by the proposed DeSTNet model on GTSRB [3] and MNIST [1] datasets, respectively.

Figure 1: Sample alignment results produced by the DeSTNet-4 model on the GTSRB dataset. Row 1: input image. Rows 2-4: results produced after each one of the four levels.

2 Architectures and Additional Results for Section 4.2

Table 1 reports the architectures of the compared CSTN-5 [2] and DeSTNet-5 models for the task of planar image alignment.

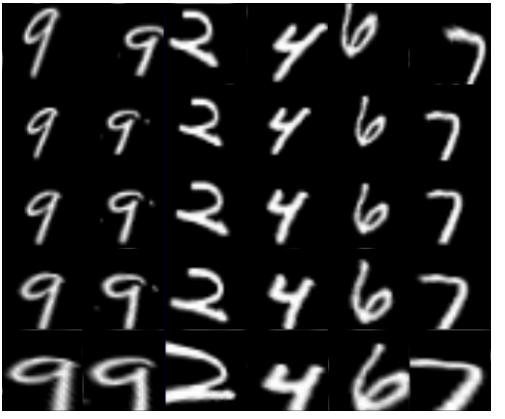

Figure 2: Sample alignment results produced by the DeSTNet-4 model on the MNIST dataset. Row 1: input image. Rows 2-4: results produced after each one of the four levels.

Model	Architecture
CSTN-5	[conv3-64(2) ǀ conv3-128(2) ǀ conv3-256(2) ǀ FC8]×5
DeSTNet-5	$\mathcal{F}$\{[conv3-64(2) ǀ conv3-128(2) ǀ conv3-256(2) ǀ FC8]×5\}

Table 1: Architectures utilized by CSTN-5 and DeSTNet-5. convD_1-D_2(D_3): convolution layer with $D_1 \times D_1$ receptive field, D_2 channels and D_3 stride, FC: fully connected layer, $\mathcal{F}$: fusion operation used in DeSTNet for fusing the parameters updates.

Additional qualitative results obtained by the CSTN-5 and DeSTNet-5 on the IDocDB database are provided in Figs. 3, 4. These results confirm that the proposed DeSTNet is more accurate than the CSTN and show better robustness against partial-occlusions, clutter and low-light conditions.

References

[1] Yann LeCun. The mnist database of handwritten digits. *http://yann. lecun. com/exdb/mnist/*, 1998.

[2] Chen-Hsuan Lin and Simon Lucey. Inverse compositional spatial transformer networks. In *Proceedings of IEEE International Conference on Computer Vision & Pattern Recognition (CVPR)*, pages 2568–2576, 2017.

[3] Johannes Stallkamp, Marc Schlipsing, Jan Salmen, and Christian Igel. The german traffic sign recognition benchmark: a multi-class classification competition. In *Proceedings of International Joint Conference on Neural Networks (IJCNN)*, pages 1453–1460, 2011.

Figure 3: Qualitative results obtained with **CSTN-5** and **DeSTNet-5** on IDocDB. (Results are best viewed on a digital screen)

Figure 4: Qualitative results obtained with **CSTN-5** and **DeSTNet-5** on IDocDB. (Results are best viewed on a digital screen)