Supplementary Material - Rethinking Planar Homography Estimation Using Perspective Fields

Rui Zeng[0000−0003−0155−1288], Simon Denman[0000−0002−0983−5480], Sridha Sridharan[0000−0003−4316−9001], and Clinton Fookes[0000−0002−8515−6324]

Queensland University of Technology, Brisbane, Australia
{r5, s.denman, s.sridharan, c.fookes}@qut.edu.au

In this document, we provide additional materials to supplement our main submission. In Section 1, we compare the performance of PFNet when optimised with the smooth-l_1 and l_2 loss to further demonstrate the effect of the loss choice. Section 2 further compares the performance of FRCN, FRCN-IB, and PFNet in terms of l_2 loss to supplement the evalution using the smooth-l_1 loss in the main submission. In Section 3, additional qualitative experiments on real world data are presented, which have been conducted to demonstrate the applicability of PFNet for "in the wild" scenes.

1 l_2 Loss V.S. Smooth l_1 Loss

In main submission, we claim that the l_2 loss is not suitable for PFNet because the l_2 loss places emphasis on outliers. This property is contrary to the goal of PFNet, which aims to regress dominant inliers. A small quantity of outliers does not adversely affect the accuracy of the PFNet as we use RANSAC to further process predictions generated by PFNet. To verify this hypothesis, we train PFNet using smooth-l_1 and l_2 loss respectively. Both of the training phases contains 300 epochs and the learning rate is divided by 10 after each 100 epochs. A sufficiently large number of epochs can help us examine loss trend during training.

Figure 1 shows the MACE trend training and validation regarding the two losses. We find that the smooth-l_1 consistently converges faster than the l_2 in both training and validation phases. The smooth-l_1 trained PFNet needs far fewer epochs to reach the same accuracy as the l_2 trained PFNet.

2 Comparison between FRCN, FRCN-IB, and PFNet

To thoroughly verify our hypothesis on the difference between these networks, we provide an extra set of experiments to compare the performance of FRCN, FRCN-IB, and PFNet. Regarding experimental setting, we use 300 epochs training and divide the learning rate by 10 after each 100 epochs. The loss function in this case is set to the l_2 loss to supplement the smooth-l_1 loss evaluation presented in main submission. All other parameters remain the same. The 300 epoch

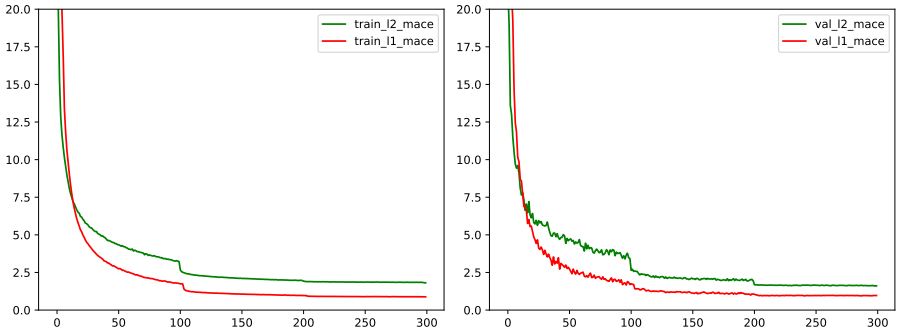

Fig. 1. The training and validation MACE of PFNet regarding smooth-l_1 and l_2 loss.

Method	MACE
FRCN	3.21
FRCN-IB	2.39
PFNet	**1.46**

Table 1. Mean average corner error (MACE) comparison between FRCN, FRCN-IB, and PFNet regarding l_2 loss training.

training ensures that training for all networks has converged to best identify the most effectively network for homograph estimation.

Table 1 reports the results of these three networks. We find that PFNet greatly outperforms the other two networks. The greater the number of training epochs, the more noticeable performance gap between these three networks.

Observing Figure 2, we can see that the FCRN validation loss oscillates significantly during the first 100 epochs, suggesting that the network does not generalise as well to unseen data as the proposed approach. With respect to training and validation metric, PFNet consistently outperforms the other two networks.

3 Additional qualitative results

We provide additional qualitative results with a focus on examining the applicability of the PFNet to real world scenes. Considering there are no publicly available real world homography datasets, we use photos captured from an IPhone 6S as the experimental materials. To generate I_{A}, we capture a random scene on a university campus. Regarding the generation of I_{B}, we capture the same scene but randomly perturb the pose of the camera to artificially generate a perspective transform. All images are then resized to 320×240

Figure 3 shows the qualitative results of the in-the-wild homography estimation. One can clearly see that the warped image generated by the predicted

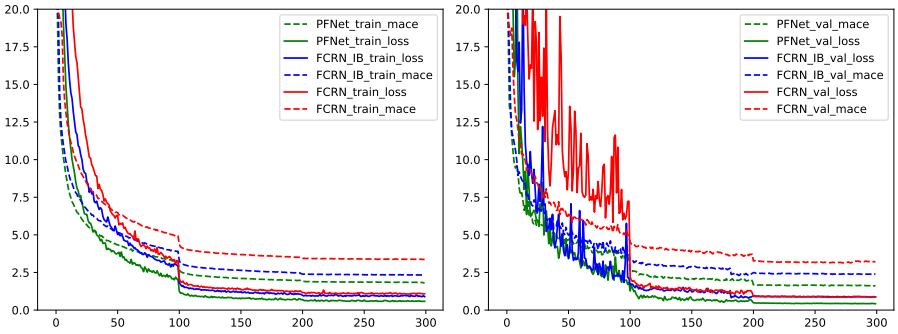

Fig. 2. The MACE and the loss value in the training and validation phases for FCRN, FCRN-IB, and PFNet regarding l_2 loss.

homography (see the last column of Figure 3) is almost the same as the I_{B} (the second column of Figure 3). The high level of performance provides further evidence that the PFNet, which is trained on a large-scale image dataset, has good applicability to real world scenes. More examples can be found in Figures 4 and 5.

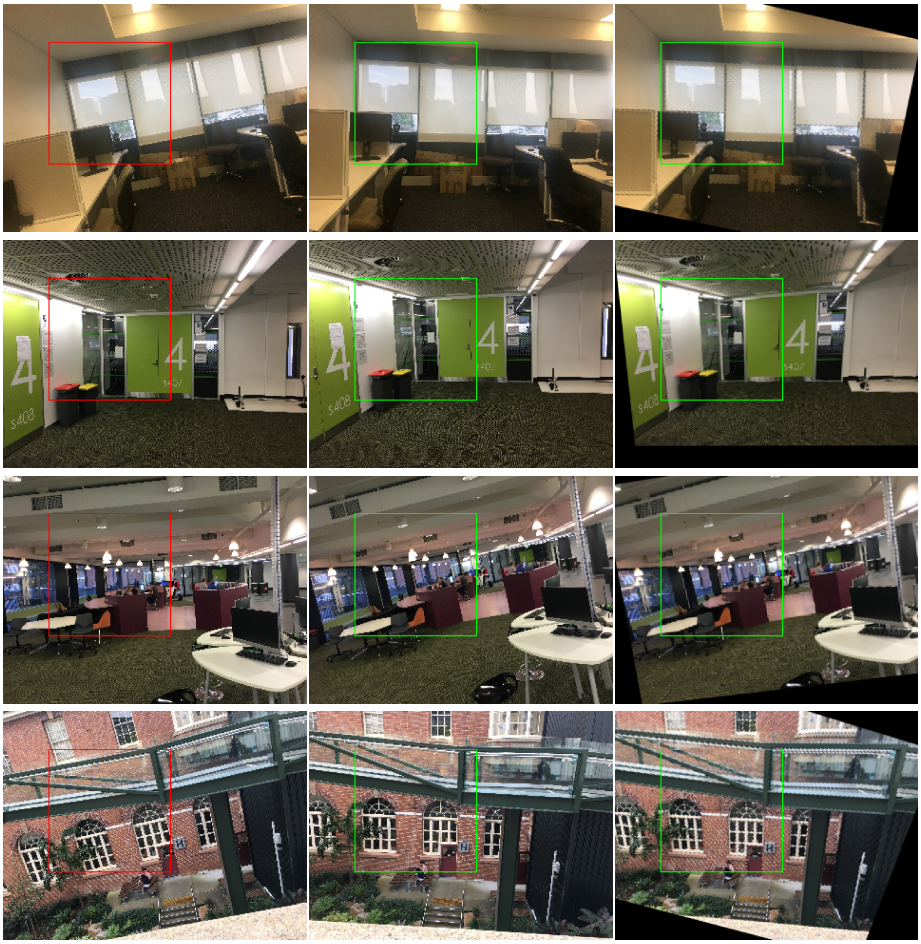

Fig. 3. Demonstration of the applicability of the PFNet on real-world scenes. From left to right: I_{A}, I_{B} (captured from the same scene using a randomly perturbed camera pose), the image generated by warping I_{A} using the predicted homography. Every square has the same place in the image. The squares in the first two columns of the figure are used to generated the input tensor for PFNet. The image content in the green square in the last column is used to qualitatively compare with that shown in the second column. We clearly see that the image content contained in the squares of the second and third column are almost the same. These results demonstrate that our method has a good applicability to real-world scenes.

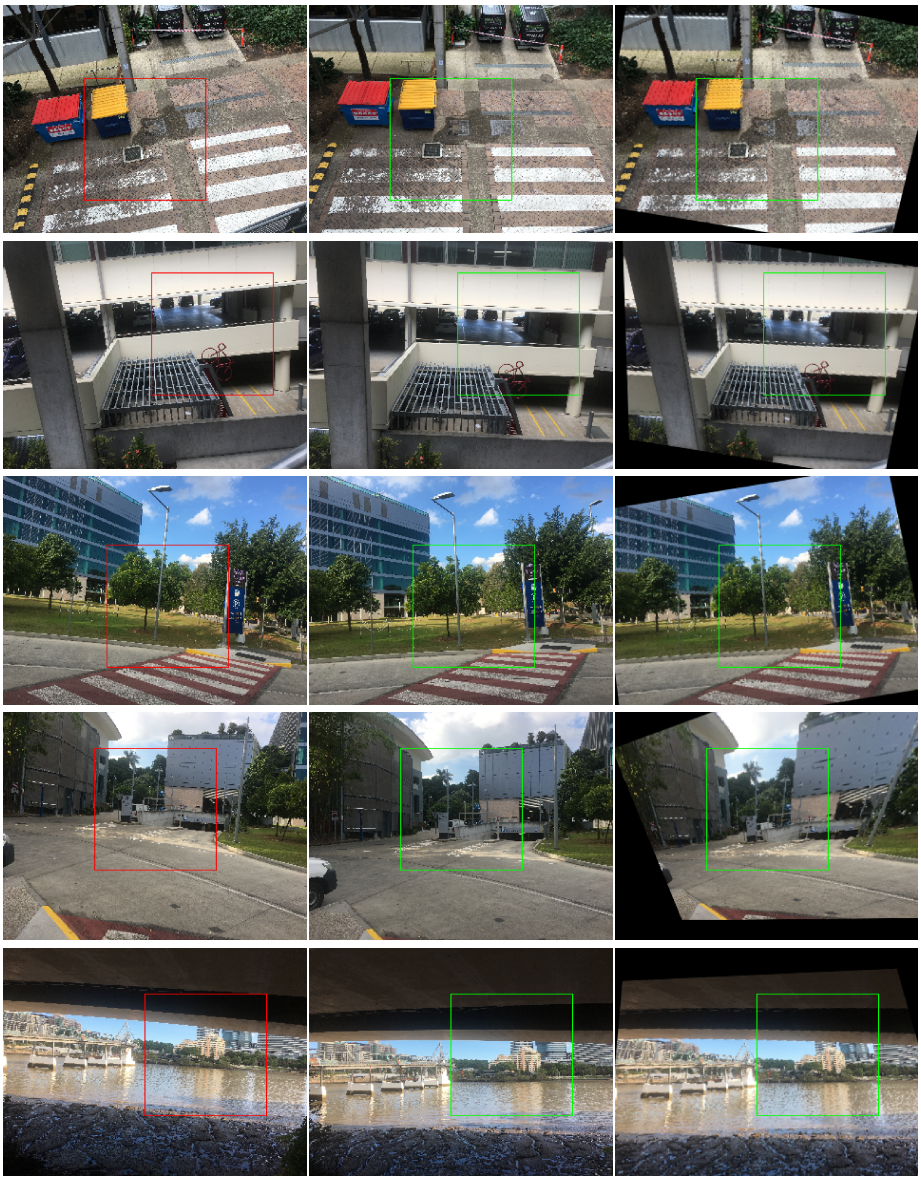

Fig. 4. Extra examples of predictions by PFNet for real-world scenes. Column and red/green bounding box descriptions are as per Figure 3.

Fig. 5. Extra examples of predictions by PFNet for real-world scenes. Column and red/green bounding box descriptions are as per Figure 3.