1. Introduction

In this supplementary material, we have included the specification of DispFulNet (our first-stage network) and some miscellaneous training details of our network. We have also showcased more visual results of the proposed cascade residual learning (CRL) scheme.

2. Specification of DispFulNet

The detailed specification of the first stage, DispFulNet, is provided in Table 1. The disparity images are produced by the convolution layers with prefix pr at multiple scales, they are supervised by the ground-truth by computing the $\ell_1$ loss. The final disparity prediction of DispFulNet, $d_1$, is given by the layer pr$_1$; this layer is named as pr$_{s1}$ in our paper (s1 means stage 1). Note that different from [4], the proposed DispFulNet has an output resolution equals to the input resolution, gives rise to disparity images with extra details.

Having obtained the first-stage output $d_1$, it is then concatenated with the left image $I_L$, the right image $I_R$, the synthesized left image (using the warping layer) $\tilde{I}_L$, and the error image $e_L = |I_L - \tilde{I}_L|$. The obtained 3D array is then fed to the second-stage network (DispResNet) to further refine the disparity.

3. Miscellaneous Training Details

In general, we train our network in a way similar to that of [1, 4]. Nevertheless, for simplicity, we have assigned different (fixed) weights to different $\ell_1$ loss layers, in contrast to the loss weight schedule of [4]. Specifically, when training either the first stage (DispFulNet) or the second stage (DispResNet), we assign the highest resolution loss, e.g. pr$_1$ in Table 1, with fixed weight 1. For other losses we let their weights be 0.2. When finetuning the overall network, we assign the highest resolution loss at the second stage with a weight 1; for other losses of the second stage and the loss of pr$_1$ (the highest resolution loss of Disp-
FulNet), their losses have weights 0.2. We let the weights of the rest of the loss layers in DispFulNet be 0.1.

4. More Visual Results

Our paper has shown many disparity segments produced by the proposed method, so as to demonstrate its superior characteristics. As supplement, we hereby showcase a few intact disparity images. Fig. 1 and Fig. 2 show three groups of visual results of the FlyingThings3D dataset [4] and the Middlebury 2014 dataset [6], respectively. DispNetC [4] is adopted as a baseline method for visual comparisons in these two figures. We remind that the Middlebury dataset has not been applied for training in this work. From the disparity images and the error images, one can see that our CRL scheme not only produces sharp edges but also provides more accurate disparity values at the inherently ill-posed regions.

Fig. 3 presents four groups of visual results on the test set of KITTI stereo 2015 dataset [5]. Our approach, CRL, ranks first (with a D1-all of 2.67%) in the KITTI 2015 stereo benchmark; while another approach, GC-NET [3], ranks second (with a D1-all of 2.87%). As a result, we compare our results with those of GC-NET here. In Fig. 3, the disparity images and the error images are obtained from the KITTI evaluation website, where the disparity images are shown using the color coding scheme of [2]. For the error images, warmer color indicate larger error. Again, our proposed CRL scheme produces sharp disparity images with high accuracy.

References


Figure 1. Three groups of results from the FlyingThings3D dataset are shown. The six images in each group, from left to right and from top to bottom, are the left image, the disparity images given by DispNetC and CRL (ours), the ground-truth disparity, the error images of DispNetC and CRL, respectively.
Figure 2. Three groups of results from the Middlebury 2104 dataset are shown. The six images in each group, from left to right and from top to bottom, are the left image, the disparity images given by DispNetC and CRL (ours), the ground-truth disparity, the error images of DispNetC and CRL, respectively.
Figure 3. Four groups of results from the KITTI stereo 2015 dataset are shown. The five images in each group, from top to bottom, are the left image, the disparity image given by GC-NET and its error image, the disparity image given by the proposed CRL and its error image, respectively. For the error images, warmer colors indicate larger error.