

Bridging Stereo Matching and Optical Flow via Spatiotemporal Correspondence

Supplementary Materials

Hsueh-Ying Lai¹ Yi-Hsuan Tsai² Wei-Chen Chiu¹
¹National Chiao Tung University, Taiwan ²NEC Laboratories America

1. Network Architecture

Figure 1 shows the architecture of the proposed single model for both stereo matching and flow estimation. We concatenate an image pair as the input no matter they are stereo pairs or flow pairs on the color channel. The output is the correspondence map at 4 different scales and each is a 2-channel map indicating vertical and horizontal correspondence.

2. More Visual Results

Figure 2 shows the depth maps of different variants of our models trained on KITTI stereo pairs, KITTI stereo and flow pairs, and our full model with the 2-warp consistency, respectively. Our full model can deal with challenging cases such as the boundary and thin/small objects. Figure 3 shows our optical flow results on the KITTI 2012 set which consists of rigid scenes, in which the sparse ground truths are provided as the reference.

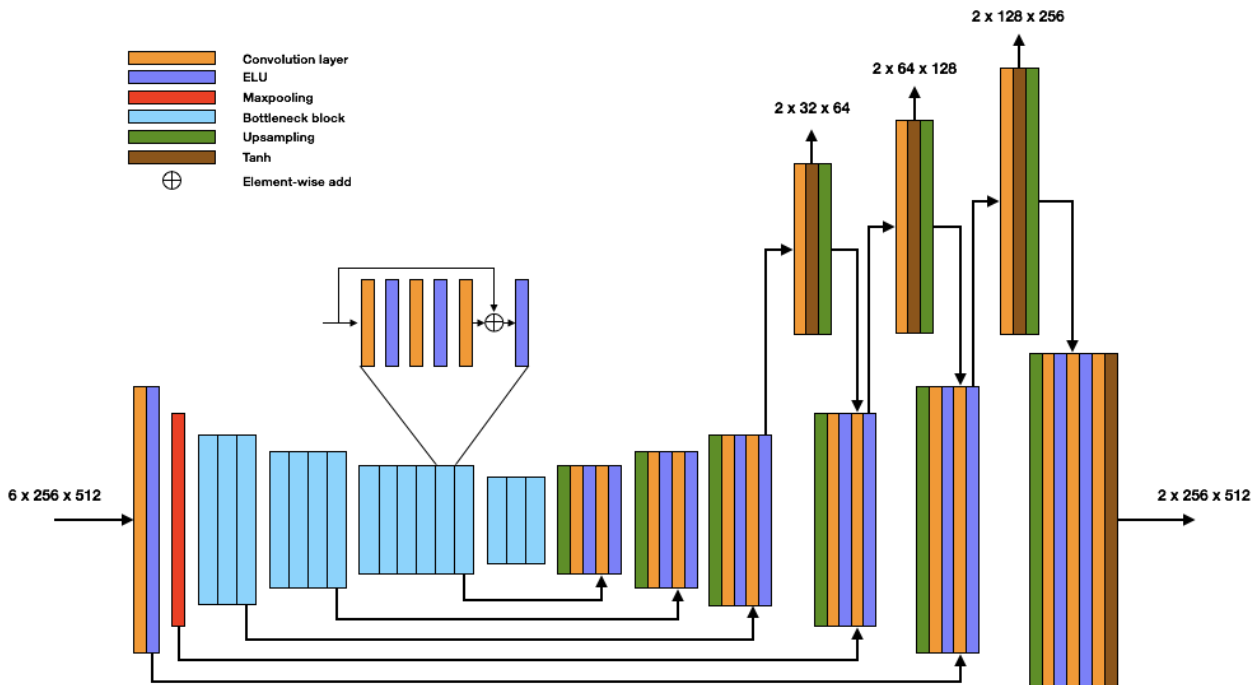


Figure 1: The detailed structure of our model. We use ResNet50 as the encoder, where there are four bottleneck blocks with different spatial size (denoted as blue blocks). The last convolution layer of each block would halve the spatial size by stride = 2. In addition, each bottleneck block has its own skip connection with the element-wise addition before the last activation function. The decoder outputs four estimated maps at different scales and up-samples them as the input toward the next decoder block.

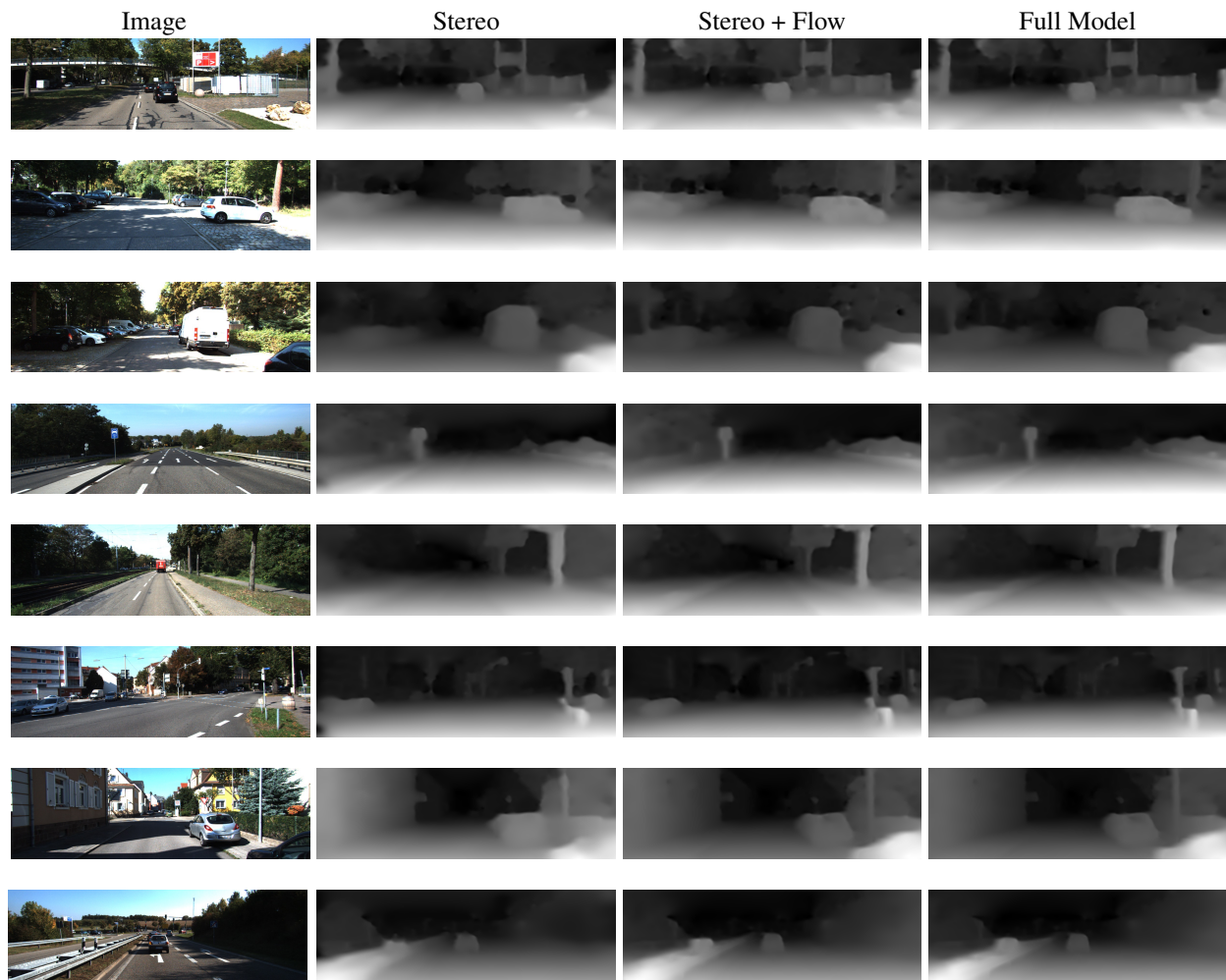


Figure 2: Qualitative results on the Eigen test split. The boundary is more clear and accurate as we add flow pairs and the proposed 2-warp consistency during training.



Figure 3: Estimated optical flow results on KITTI 2015. There are 6 examples and from top to bottom, they are real gray-scale images, ground truths and our estimated maps.