A NORMAL ESTIMATION SUB-NET

Our Normal Estimation Sub-Net is illustrated in Figure 11. We directly use the shared encoder part to provide the feature and adopt the network structure of DispNetC[1] to estimate the surface normal in a coarse-to-fine manner. For residual learning(Residual Blocks in blue), we use a shallow U-Net architecture to take features at the current scale together with upsampled predicted surface normal at the previous scale and left-right image pairs as input to predict the surface normal residual at the current scale. Besides, we normalized the output of each scale directly on the feature channels(3), making it more in line with the definition of the surface normal. Figure 9 illustrate visualization results on SceneFlow and KITTI 2015 and MiddleBurry 2014 datasets.

B NON-LOCAL DISPARITY PROPAGATION AT DIFFERENT SCALES.

In this section, by visualizing the disparities estimation results at different resolutions, we show how our proposed NDP module improves the quality of the predicted disparity through spatial propagation, especially in thin structures, edges, and occluded regions. As is shown in Figure 12, the non-local disparity propagation based on surface normal and other semi-context information performs well at thin structures and edges especially at low scales to alleviate the disparity discontinuity and blurring issues. After spatial propagation, our disparity preserves more structural details, which facilitates more accurate depth results for downstream tasks and thus constructs more accurate AR/VR applications.

C SPATIAL ATTENTION AND AFFINITIES DISTRIBUTION IN ARL MODULE

As mentioned in the main paper, we have proposed an ARL module that manages to optimize disparity estimation results at the feature level. There are two main tricks in the ARL module for efficient residual learning. One is to apply the dynamic spatial attention mechanism to generate attention maps at different refinement scales. The attention maps imply where should be highlighted as well as where should be ignored. It promotes the network to efficiently optimize the disparity estimation results for different locations at different scales.

As shown in Figure 10, the dynamic spatial attention mechanism tends to assist residual learning at different levels and focus on regions with different attributes. The attention distribution at the 1/4 scale mainly lies in the background objects, while the distribution of attention at 1/2 resolution is highly consistent with that of the occlusion mask. This enables the ARL module to be occlusion-aware when refining the disparity. Besides, the attention map at full scale shows bigger attention to foreground object and planes that is texture-less, which benefit the refinement in such regions.

The other trick of the ARL module is the local affinity filtering module whose intermediate affinity visualization result of the local 8 neighbors can be seen in Figure 13. As the figure illustrated, the red arrow in reference images shows the plane direction, as we can see that the affinity value at the left-top value is bigger than the other direction, which indicates the direction of the feature aggregation should be. By applying the local affinity filtering module, our ARL module can make full use of surface normal information for better feature aggregation.

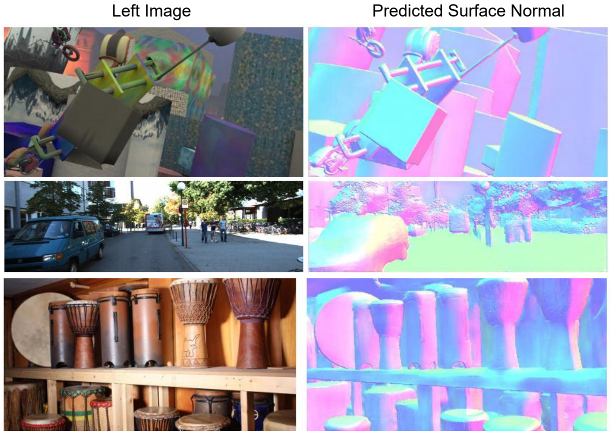

Figure 9: Normal Estimation Result on SceneFlow,KITTI2015 and MiddleBurry testing set.

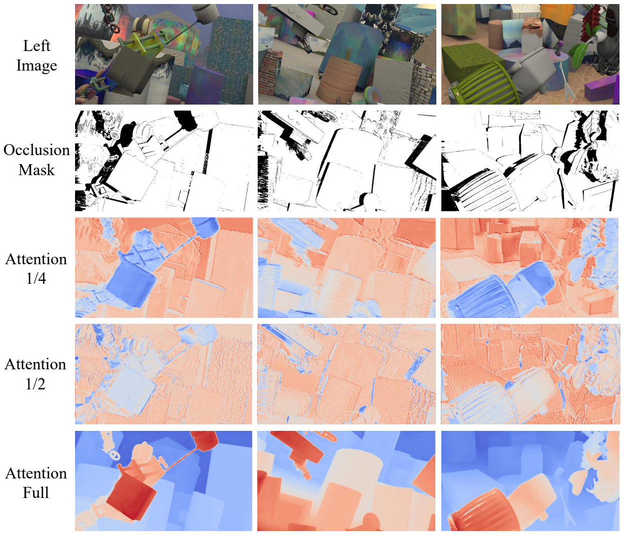

Figure 10: Attention Maps at different scales(Range from 1/4 full scale). Different scales pay different attention to different regions, such as foreground, texture-less planes, occluded areas and background.

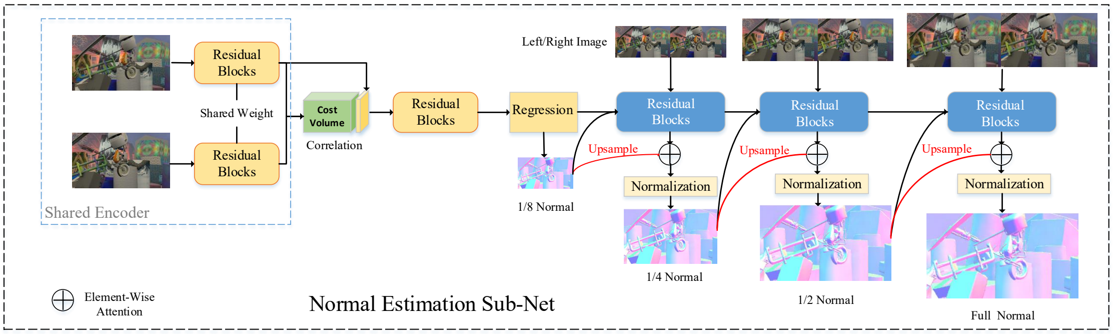

Figure 11: Normal Estimation Sub-Net Architecture. The feature encoder is shared with the disparity estimation branch. The whole Sub-Net adopts the DispNetC [12] architecture with a residual learning module. We normalized the predicted disparity to satisfy the geometry constrain.

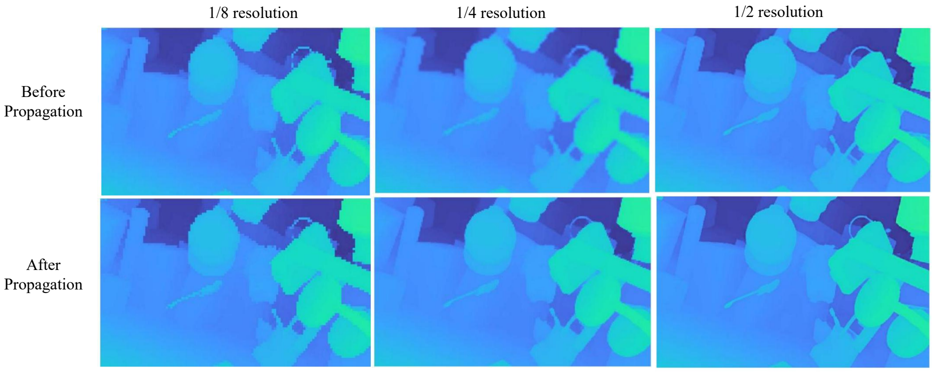

Figure 12: Disparity refinement at different scales. It clearly shows that our proposed non-local propagation witness a clear edges and structures at different disparity scales, which alleviates the blurring and breakage issues at the edges of the image.

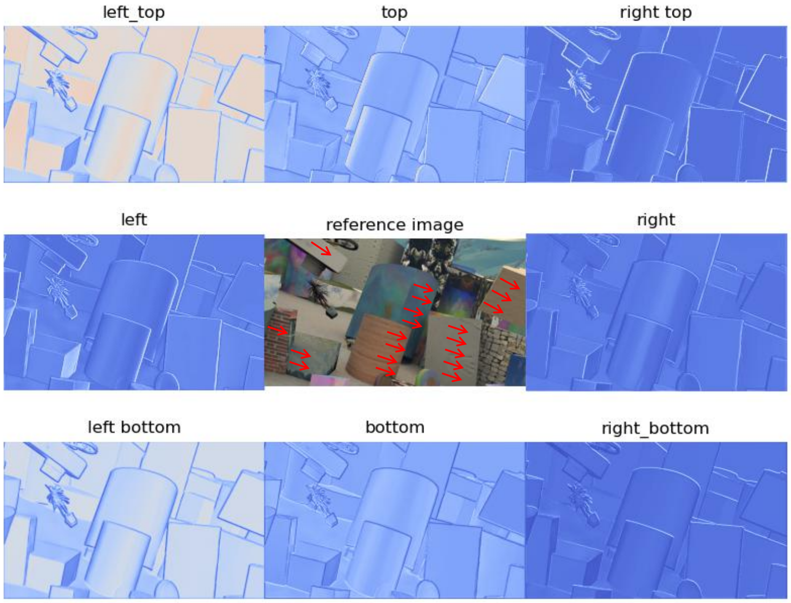

Figure 13: A sample of affinity distribution of local 8 neighbors in ARL module.