# Supplemental Material for Exploring Stereovision-Based 3-D Scene Reconstruction for Augmented Reality

Anonymous

## 1 Architecture Details of SLED

Tab. 1 shows the layer-wise definition of single long encoder-decoder which is used to replace the stacked hourglass networks in PSM-Net [1].

Table 1: Details of Cost Volume Regularization and Refinement

| Type | K | S | P | D | Output | Dim. | I/O | Input |
|---|---|---|---|---|---|---|---|---|
| Cost Volume | | | | | | | | |
| Concat | – | – | – | – | CVol. | 32/64 | 4/4 | **LVol.,RVol.** |
| Res. | 3 | 1 | 1 | 1 | Cls1_1 | 64/32 | 4/4 | CVol. |
| Res. | 3 | 1 | 1 | 1 | Cls1_2 | 32/1 | 4/4 | Cls1_1 |
| UpS. | – | – | – | – | Cls1_3 | 1/1 | 4/1 | Cls1_2 |
| S.Arg. | – | – | – | – | Output1 | 1/1 | 1/1 | Cls1_3 |
| Cost Volume Regularization | | | | | | | | |
| Res. | 3 | 1 | 1 | 1 | 3Dconv0_1 | 64/64 | 4/4 | CVol. |
| Res. | 3 | 1 | 1 | 1 | 3Dconv0_2 | 64/64 | 4/4 | 3Dconv0_1 |
| Res. | 3 | 1 | 1 | 1 | Econv1_1 | 64/64 | 4/4 | 3Dconv0_2 |
| Res. | 3 | 1 | 1 | 1 | Econv1_2 | 64/64 | 4/4 | Econv1_1 |
| Res. | 2 | 2 | 0 | – | Epool1 | 64/64 | 4/8 | Econv1_2 |
| Res. | 3 | 1 | 1 | 1 | Econv2_1 | 64/64 | 8/8 | Epool1 |
| Res. | 3 | 1 | 1 | 1 | Econv2_2 | 64/64 | 8/8 | Econv2_1 |
| AvgP. | 2 | 2 | 0 | – | Epool2 | 64/64 | 8/16 | Econv2_2 |
| Res. | 3 | 1 | 1 | 1 | Econv3_1 | 64/64 | 16/16 | Epool2 |
| Res. | 3 | 1 | 1 | 1 | Econv3_2 | 64/64 | 16/16 | Econv3_1 |
| AvgP. | 2 | 2 | 0 | – | Epool3 | 64/64 | 16/32 | Econv3_2 |
| Res. | 3 | 1 | 1 | 1 | Econv4_1 | 64/64 | 32/32 | Epool3 |
| Res. | 1 | 1 | 0 | 1 | Econv4_2 | 64/64 | 32/32 | Econv4_1 |
| UpS. | – | – | – | – | up_4t3 | 64/64 | 32/16 | Econv4_2 |
| Res. | 3 | 1 | 2 | 2 | Dconv3 | 64/64 | 16/16 | up_4t3+Econv3_2 |
| UpS. | – | – | – | – | up_3t2 | 64/64 | 16/8 | Dconv3 |
| Res. | 3 | 1 | 2 | 2 | Dconv2 | 64/64 | 8/8 | up_3t2+Econv2_2 |
| UpS. | – | – | – | – | up_2t1 | 64/64 | 8/4 | Dconv2 |
| Res. | 3 | 1 | 2 | 2 | Dconv1 | 64/64 | 4/4 | up_2t1+Econv1_2 |
| Res. | 3 | 1 | 1 | 1 | Cls2_1 | 64/32 | 4/4 | Dconv1 |
| Res. | 3 | 1 | 1 | 1 | Cls2_2 | 32/1 | 4/4 | Cls2_1 |
| UpS. | – | – | – | – | Cls2_3 | 1/1 | 4/1 | Cls2_2+Cls1_2 |
| S.Arg. | – | – | – | – | Output | 1/1 | 1/1 | Cls1_3 |

**K, S, P, D**: kernel size, stride, padding, and dilation of convolutional layer; **Dim.**: dimension of input/output feature maps; **I/O**: scale of input/output feature maps; Symbol "**+/-**": element-wise summation/subtraction operation; **Res.**: the basic residual module; **S.Arg.**: Softmax Argmin function for regress the disparity map from the volume; **AvgP.**: average pooling operation. **UpS.**: upsample operation.

## References

[1] J.-R. Chang and Y.-S. Chen. Pyramid stereo matching network. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 5410–5418, 2018. 1