SGM-Nets: Semi-global matching with neural networks

We will show a brief summary of training algorithm and some results which are difficult to put in main paper due to space limitation. Sec. B and C deal with experiments on synthetic and real images, respectively.

A. Training process

Algorithm1 shows summary of training process. Here, we mention the standard parameterization.

Algorithm 1 Training of standard SGM-Net

Initialize SGM-Net $\mathbf{N}_0$.
while $i <$ maximum iterations **do**
  $\mathsf{E}[x]$ stores derivatives of the loss function at pixel x.
  for images for a backpropargation **do**
    Compute disparity map with SGM which penalties are given by SGM-Net $\mathbf{N}_{i-1}$
    for All directions $\mathbf{r} \in R$ **do**
      Deploy pixels $\mathbf{x}'$ for the neighbor cost.
      for All pixels $x'_k \in \mathbf{x}'$ **do**
        if x'_k is on border **then**
          Compute $\mathbf{e} = \frac{\partial E_{nb}}{\partial P}$ in Eq. (9, 10, and 11).
        else if x'_k is on slant **then**
          Compute $\mathbf{e} = \frac{\partial E_{ns}}{\partial P}$ in Eq. (9, 10, and 12).
        else if x'_k is on flat **then**
          Compute $\mathbf{e} = \frac{\partial E_{nf}}{\partial P}$ in Eq. (9, 10, and 13).
        end if
        if $\mathbf{e} \neq \mathbf{0}$ **then**
          Store $\mathbf{e}$ in $\mathsf{E}[x'_k]$ *.
        end if
      end for
      Deploy pixels $\mathbf{x}''$ for the path cost.
      for All pixels $x''_k \in \mathbf{x}''$ **do**
        Compute $\mathbf{e} = \xi \frac{\partial E_g}{\partial P}$ in Eq. (4) at the pixels over a path of the direction $\mathbf{r}$, and extract pixels x'''_k and store $\mathbf{e}$ in $\mathsf{E}[x'''_k]$ * when $\mathbf{e} \neq \mathbf{0}$.
      end for
    end for
  end for
  Update SGM-Net through backpropagation (x of $\mathsf{E}[x]$ for collection of the patches and their positions as input, $\mathsf{E}[x]$ for derivatives of the loss function) and get SGM-Net $\mathbf{N}_i$.
  $i \leftarrow i + 1$
end while

$\star$Search identical pixels over E on the same image, and accumulate them if found.

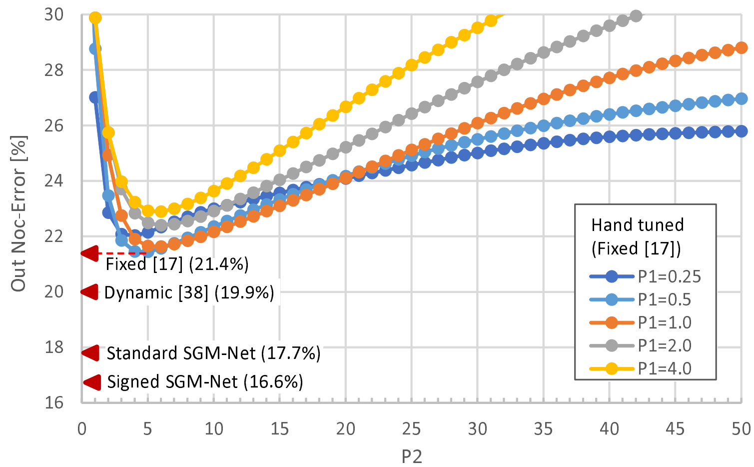


Figure 12. Overall Out-Noc error on synthetic train dataset with the parameter-sweep experiment. A matcher is ZNCC..

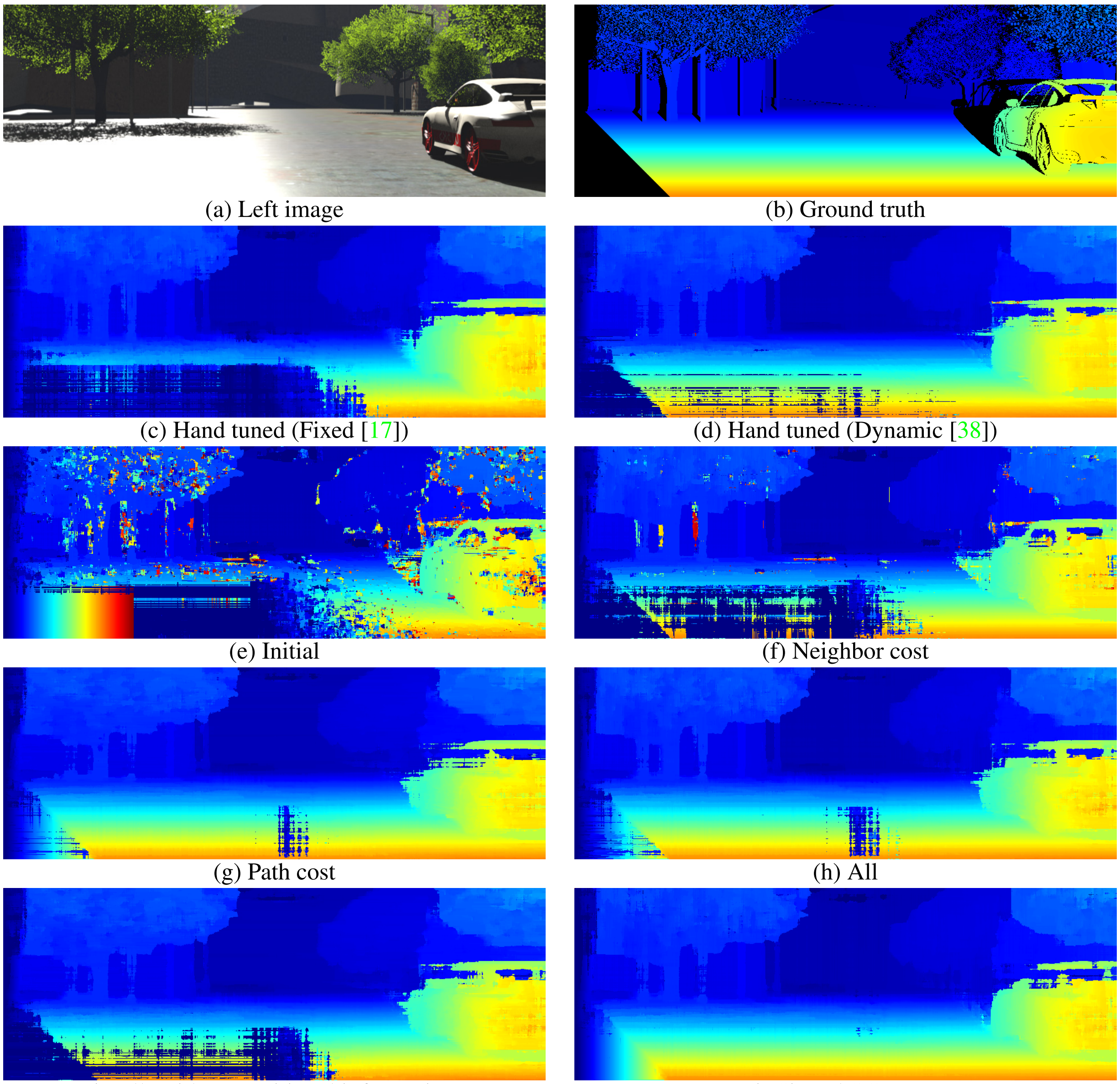
B. Synthetic images

B.1. Penalty choise

Sensitivity to the penalty choice is depending on the quality of stereo correspondence. If all pixels are correctly estimated (e.g. well textured, noiseless images), the winner-takes-all strategy at each pixel may be sufficient and SGM penalty is less sensitive. Figure 12 shows Out-Noc error with a parameter-sweep for P_1 and P_2 of hand tuned method [17]. The setting is the same as ZNCC in Table 1. The graph shows the penalty choice is important to achieve lower error. Whilst parameters of SGM-Nets are randomly initialized and they output around 1.0 for P_1 and P_2 at initialization time. The error of disparity map (around 30 % at the initialization) becomes 17.7 % (Standard SGM-Net) or 16.6 % (Signed SGM-Net) by training SGM-Nets with the proposed loss function. Important thing is that even though the penalty is tuned (Fixed or Dynamic), they are unable to achieve the same accuracy as SGM-Nets.

B.2. Comparison of loss functions

Figure 13 shows estimated disparity maps. Only SGM penalties are different for estimating the disparity maps. The penalties of hand tuned methods are tuned so as to mark the minimum error over the training images. The hand tuned method[17] fails disparities in saturated pixels on the road (Fig. 13(c)). The hand tuned method[38] improves the disparities on the pixels, however erroneous pixels are remaining (Fig. 13(d)). SGM-Net trained with neighbor loss (Fig. 13(f)) gives better disparity map than initial parameters of SGM-Net (Fig. 13(e)). SGM-Net trained with the path cost (Fig. 13(g)) is likely to lose details such as the trunks of trees on left side, but disparities at saturated pixels are well estimated. By using all loss functions, the details tend to be preserved as shown in Fig. 13(h). All loss functions without positional information loses accuracy (Fig. 13(i)). Signed SGM-Net looks the best of the other disparity maps (Fig. 13(j)).



(a) Left image
(b) Ground truth
(c) Hand tuned (Fixed [17])
(d) Hand tuned (Dynamic [38])
(e) Initial
(f) Neighbor cost
(g) Path cost
(h) All
(i) All (without positional information)
(j) Signed SGM-Net

Figure 13. Disparity maps on the synthetic image. (e)-(i) are estimated with standard SGM-Net.

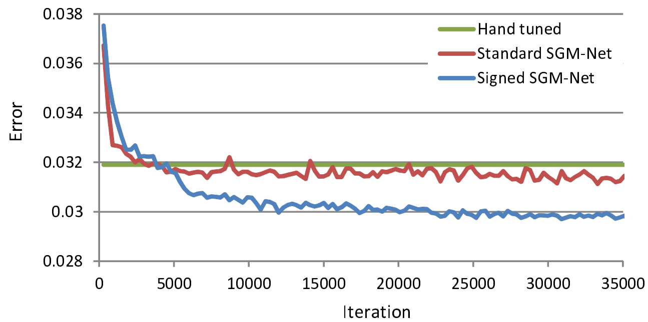


Figure 14. Errors during training.

C. Real images

C.1. Training of SGM-Nets

Figure 14 shows training errors on KITTI. Green, red, and blue lines mean errors on training images by hand tuned, standard, and signed SGM-Net respectively. In this dataset, standard SGM-Net converged faster than signed SGM-Net. SGM-Nets outperform hand tuned method over 5,000 iterations.

C.2. SGM penalties by SGM-Net

Figure 15 and 16 show the penalties by standard and signed SGM-Nets. The color of P_1 and P_2 encodes strength of respective values, blue and red mean small and large, respectively. As we explained in main paper, P_1 on road region is larger in a horizontal direction than that in a vertical direction due to its shape, i.e. flat in a horizontal direction . And P_2 on the edges orthogonal to an aggregation direction has smaller value.

C.3. Benchmark on KITTI datasets

C.3.1 Training images

Figure 17 shows Out-Noc error improvement compared with hand tuned method[38] on every images. The images are sorted in descending order with respect to the error. As can see, in the most of the frames, SGM-Nets estimate more accurate disparity map than the hand tuned method.

C.3.2 Testing images

Figure 18 shows estimated disparity map and their errors on KITTI 2012. Here, we cited results of MC-CNN because the following rank method, PBCP[28], uses additional information “stereo confidence”. Errors on the side walls of vehicles and buildings are removed properly.

Figure 19 shows results on KITTI 2015. As you see, some erroneous pixels appear at upper part of the disparity map, where is out of evaluation area. Ground truth on the part isn’t provided and the penalties on the part are more likely to be influenced by synthetic images.

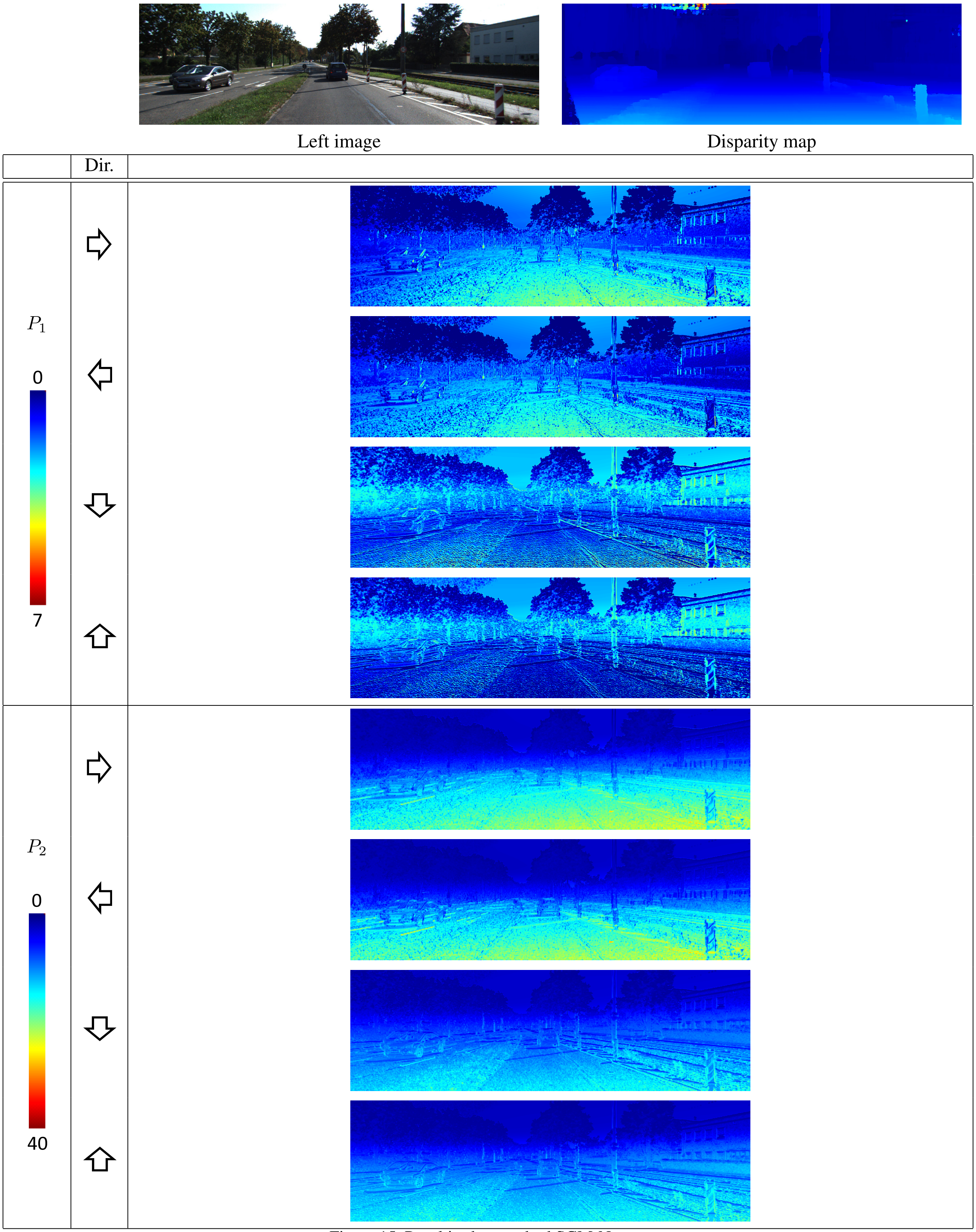


Figure 15. Penalties by standard SGM-Net.

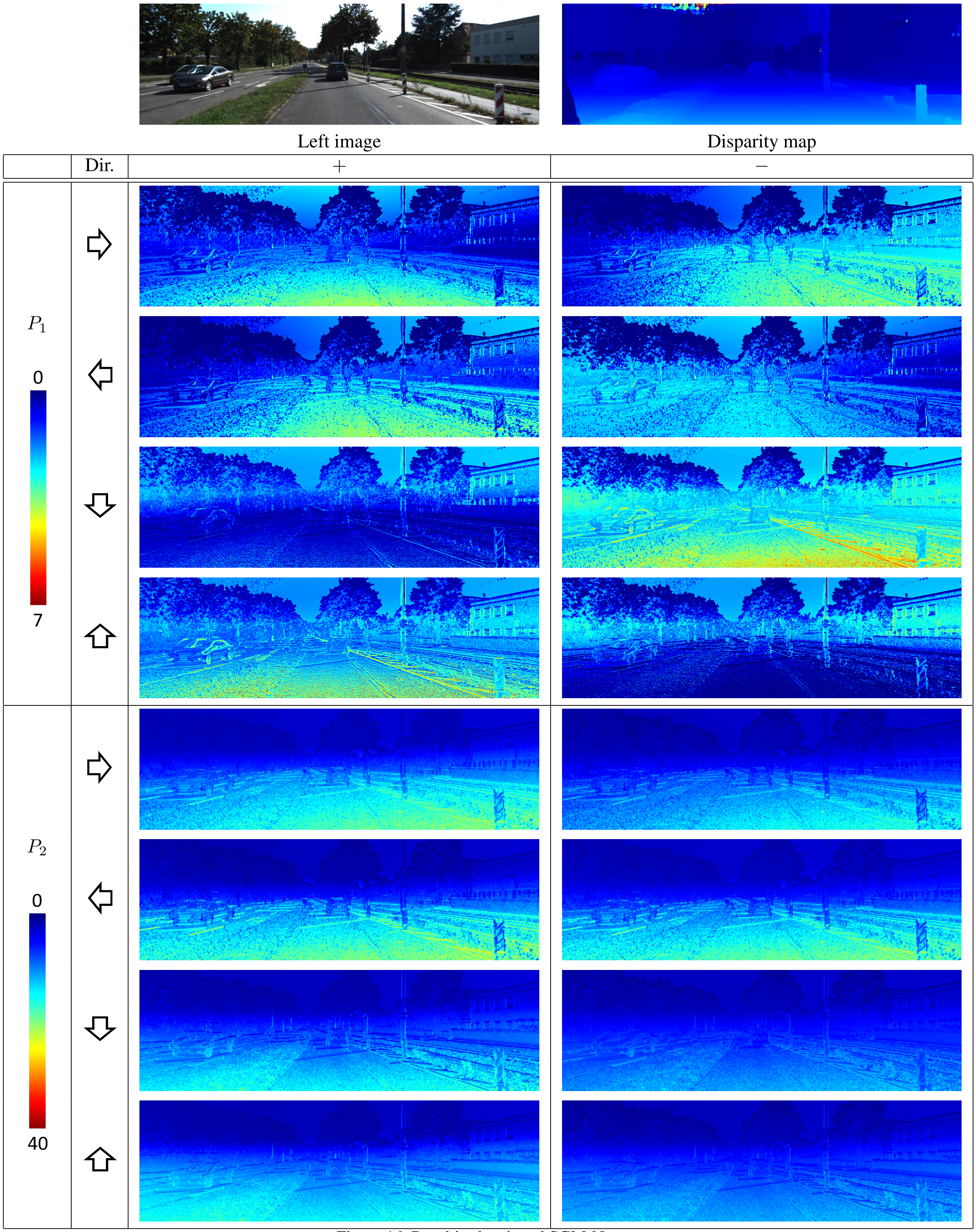


Figure 16. Penalties by signed SGM-Net.

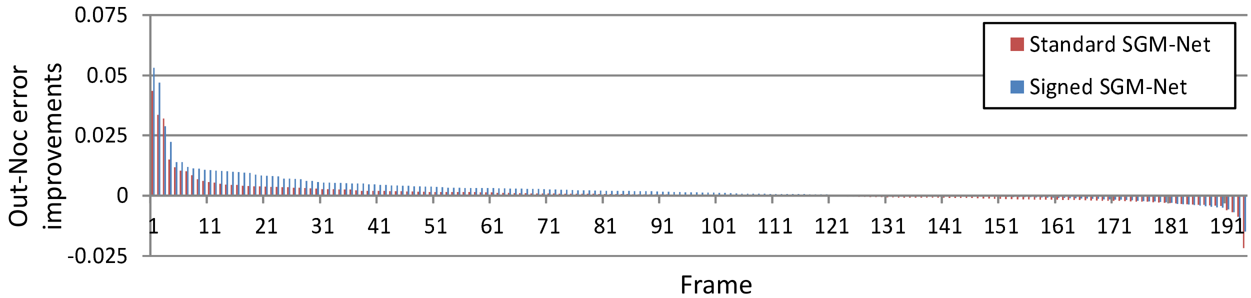


Figure 17. Out-Noc error improvement compared with hand tuned method[38] on KITTI 2012.

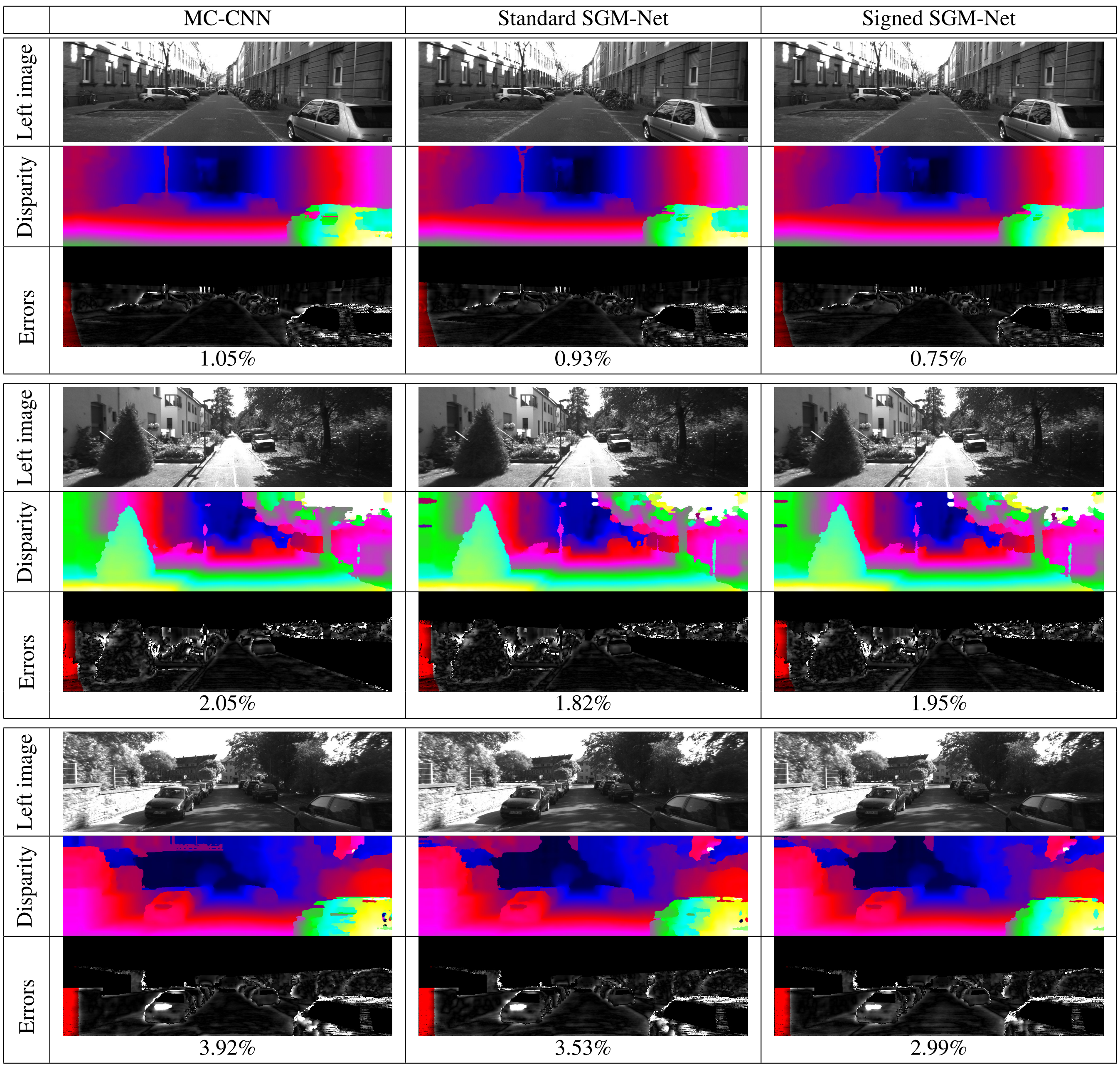


Figure 18. Estimated disparity maps and their errors on KITTI 2012. Number under error image indicates Noc-All error.

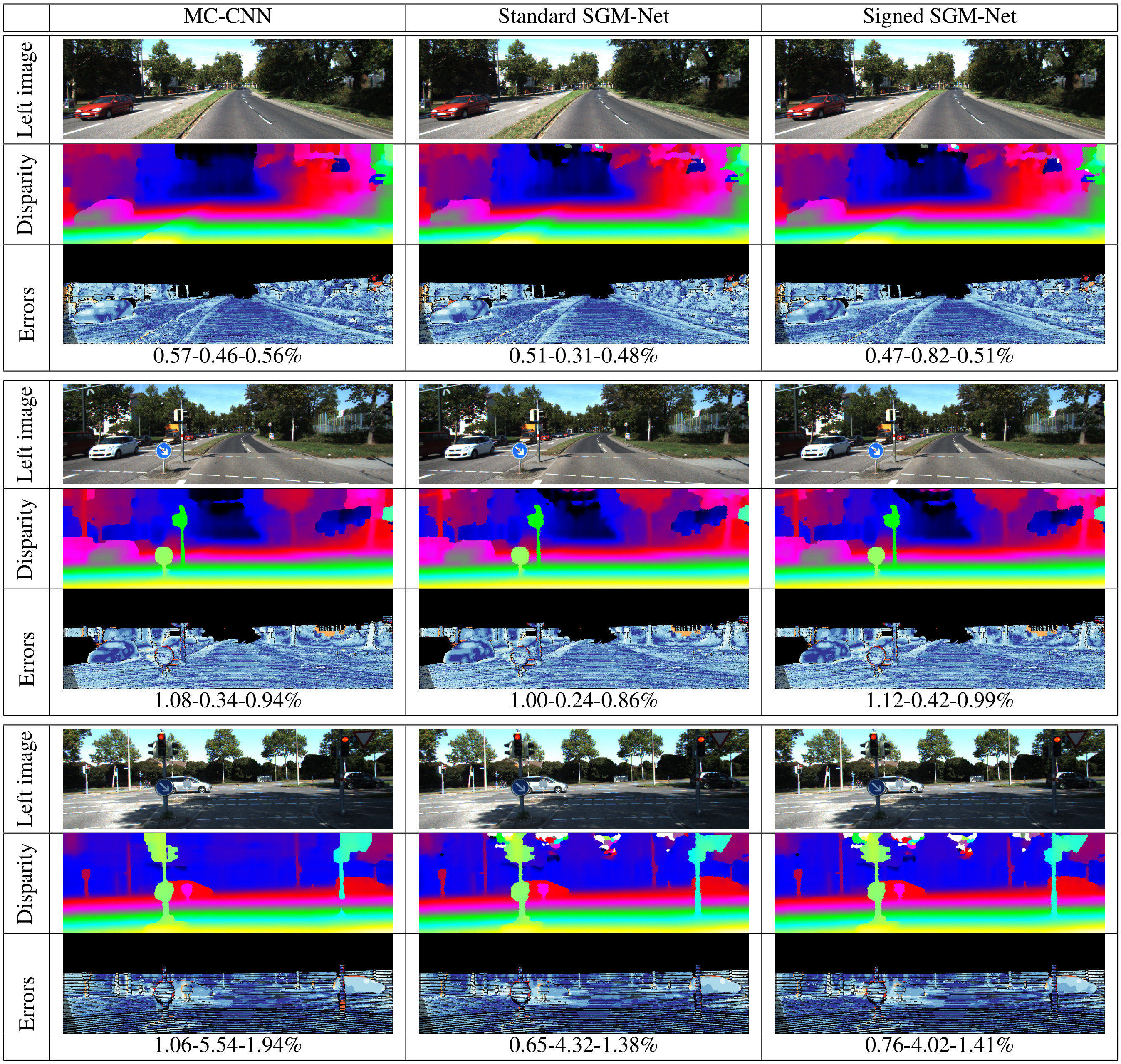


Figure 19. Estimated disparity maps and their errors on KITTI 2015. Numbers under error image indicate Noc-All error of background, foreground, and all, respectively.