M-FUSE: Multi-frame Fusion for Scene Flow Estimation
Supplementary Material

In this supplementary material, we provide a visualization of our fusion U-Net, additional ablations and additional qualitative results.

1. Architecture of the U-Net

Figure 1 shows the architecture of our 3-level U-Net with residual connections.

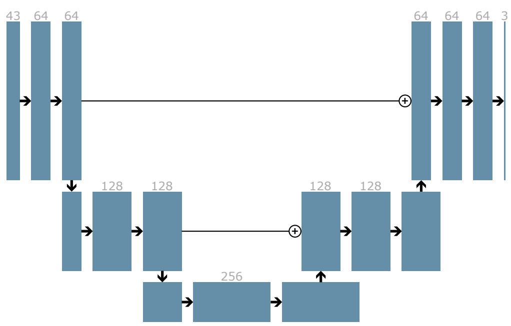

Figure 1. Architecture of our fusion U-Net.

2. Additional ablations

In addition to the ablations in the main paper, we conducted two more experiments as shown in Table 1.

In-between convolutions. As can be seen in Figure 1, in every depth level for the contracting as well as the expanding part one additional in-between convolutional layer is used to process information. Thus, we performed an ablation over several options: completely omitting this layer (none), having one (1 conv) or two (2 convs) convolutions, or using a residual block [1] (resblock). The results for none, one or two convolutional layers are inconclusive, with no significant best option. As a compromise, we chose one convolutional layer for our method since it is most similar to other U-Nets in the literature. Finally, despite being most closely related to the two convolutions, the residual block slightly decreases the quality compared to all other cases.

Image features. Finally, we compare two options to encode image-related features guiding our fusion module. The first option is to utilize the learned correlation cost from our baseline, which is upsampled from 1/8th of the resolution.

Table 1. Additional ablations. We show 4-fold cross validation results on KITTI *train* in terms of the D2, Fl and SF errors as well as the number of parameters in millions.

	D2	Fl	SF	#param
two-frame	1.81	3.67	4.07	
In-between convs				
none	**1.99**	3.33	3.89	1.42
1 conv (ours)	**1.99**	3.21	**3.82**	2.38
2 convs	2.06	**3.19**	3.84	3.34
resblock	2.08	3.47	3.99	3.34
Image features				
corrCost (ours)	**1.99**	**3.21**	**3.82**	2.38
BCE	2.00	3.33	3.96	2.38

The second option is a full-resolution brightness constancy error map [2] as the $L2$ distance between the warped and original image. As one can see, the learned correlation features outperform the brightness constancy maps slightly – although the former are upsampled from lower resolution.

3. Additional qualitative results

We show additional visual results from the KITTI benchmark in Figures 2–5.

References

[1] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

[2] Z. Ren, O. Gallo, D. Sun, M. Yand, E. B. Sudderth, and J. Kautz. A fusion approach for multi-frame optical flow estimation. In *WACV*, 2019.

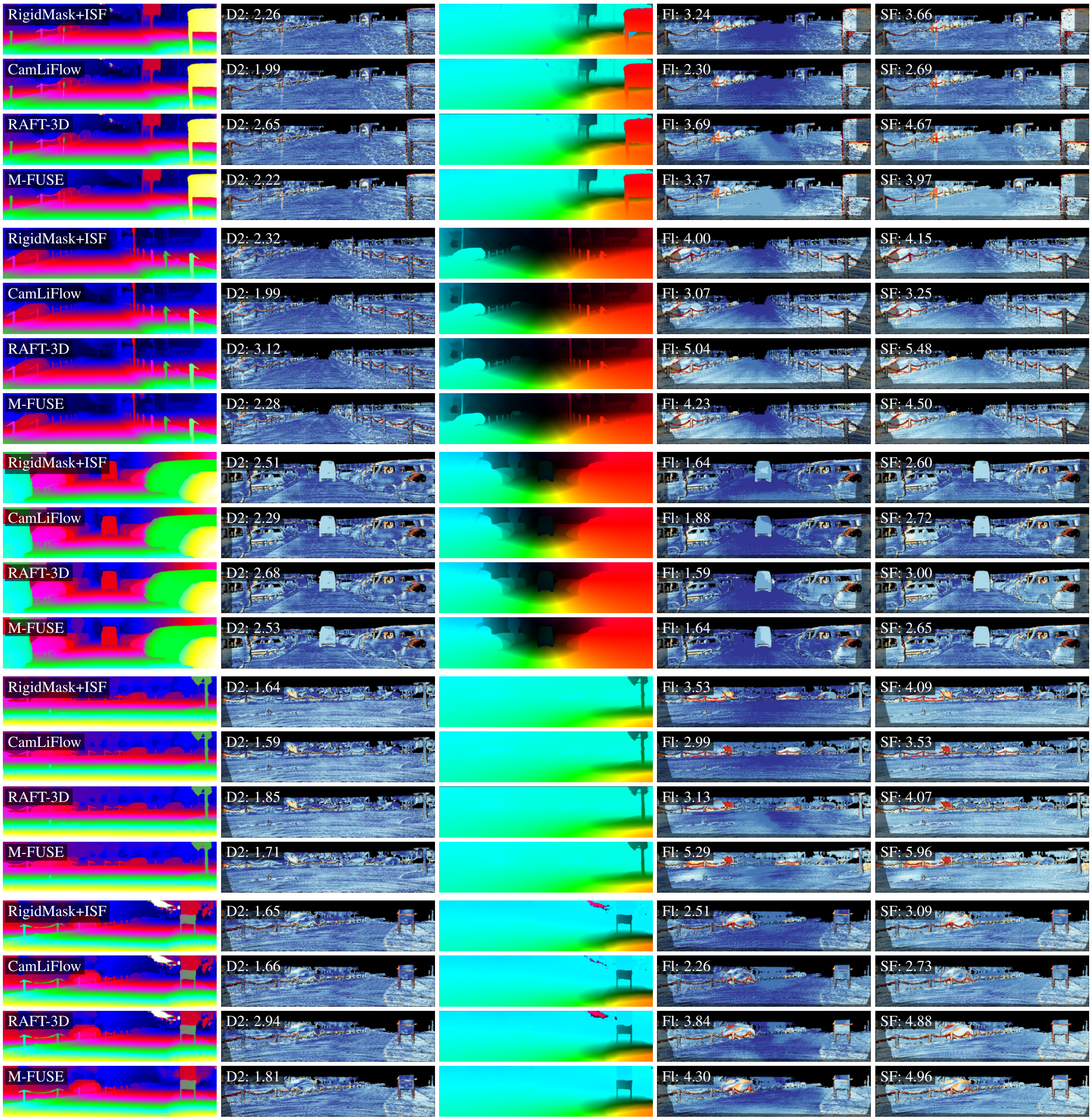

Figure 2. Qualitative comparison of our method, the original RAFT-3D, as well as the two top-performing approaches from the literature using the visualizations provided by the KITTI benchmark. *From left to right:* Target disparity visualization, corresponding *D2* error plot, optical flow visualization, corresponding *Fl* error plot, combined *SF* error plot.

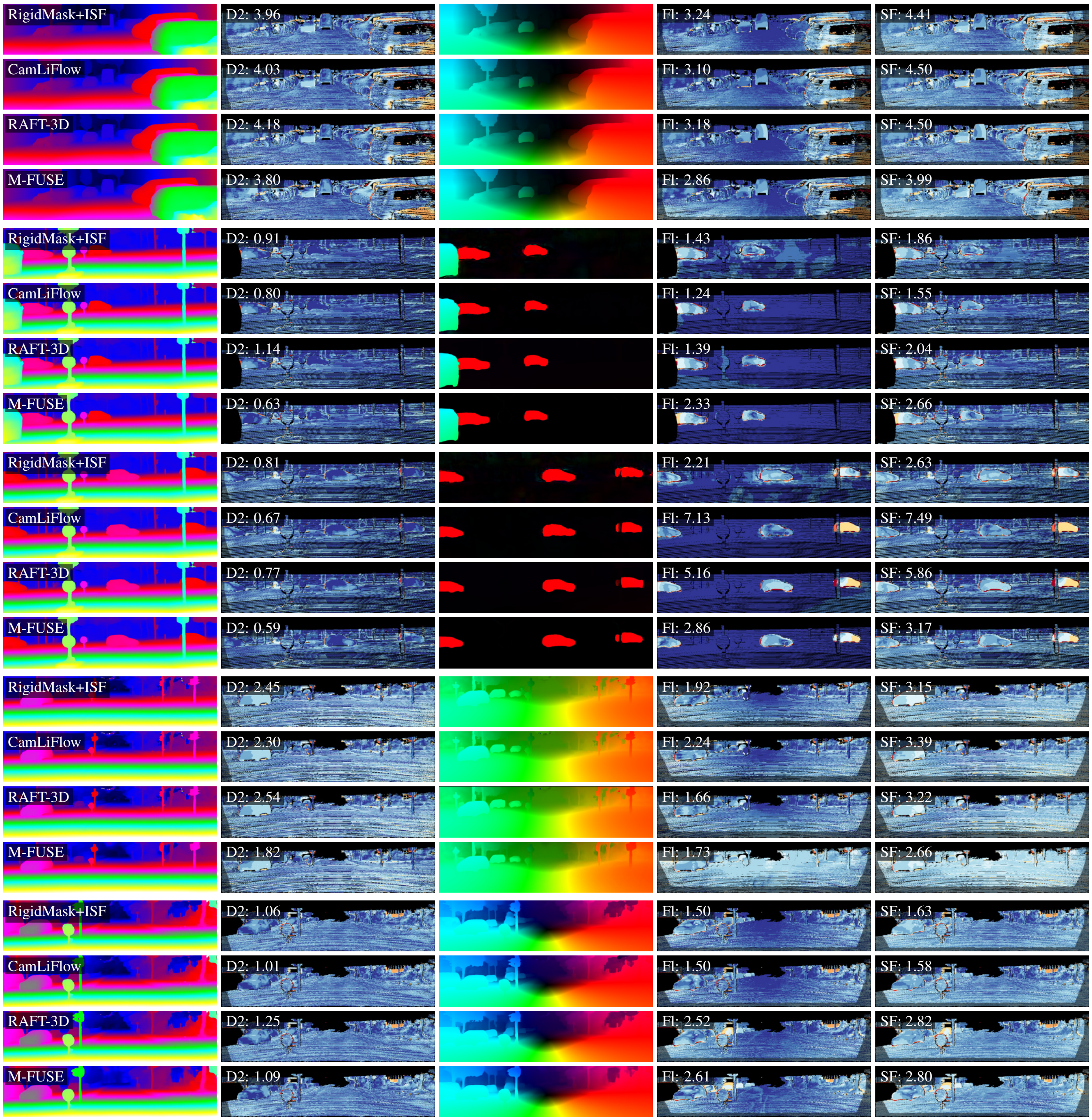

Figure 3. Qualitative comparison of our method, the original RAFT-3D, as well as the two top-performing approaches from the literature using the visualizations provided by the KITTI benchmark. *From left to right:* Target disparity visualization, corresponding *D2* error plot, optical flow visualization, corresponding *Fl* error plot, combined *SF* error plot.

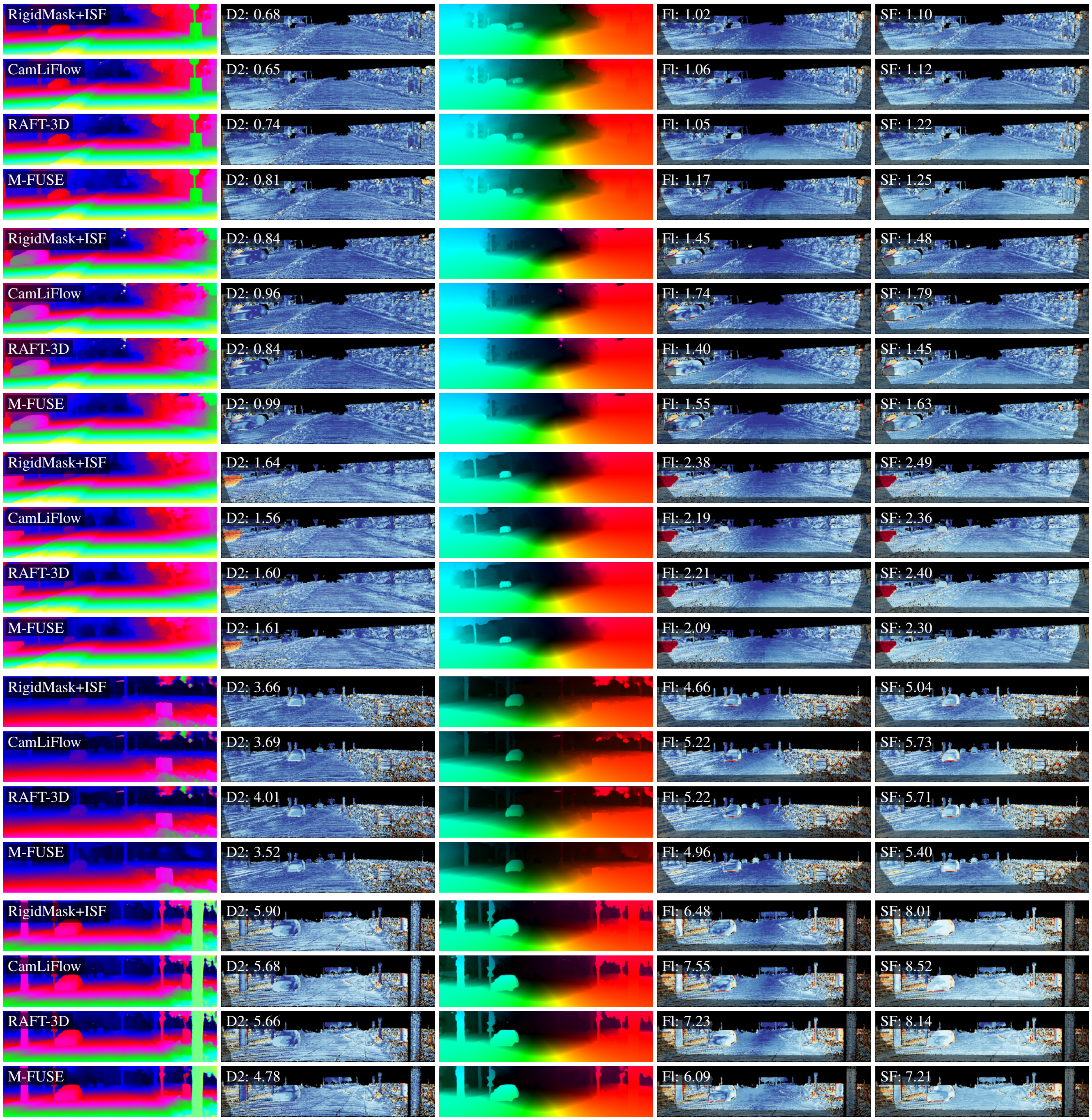

Figure 4. Qualitative comparison of our method, the original RAFT-3D, as well as the two top-performing approaches from the literature using the visualizations provided by the KITTI benchmark. *From left to right:* Target disparity visualization, corresponding *D2* error plot, optical flow visualization, corresponding *Fl* error plot, combined *SF* error plot.

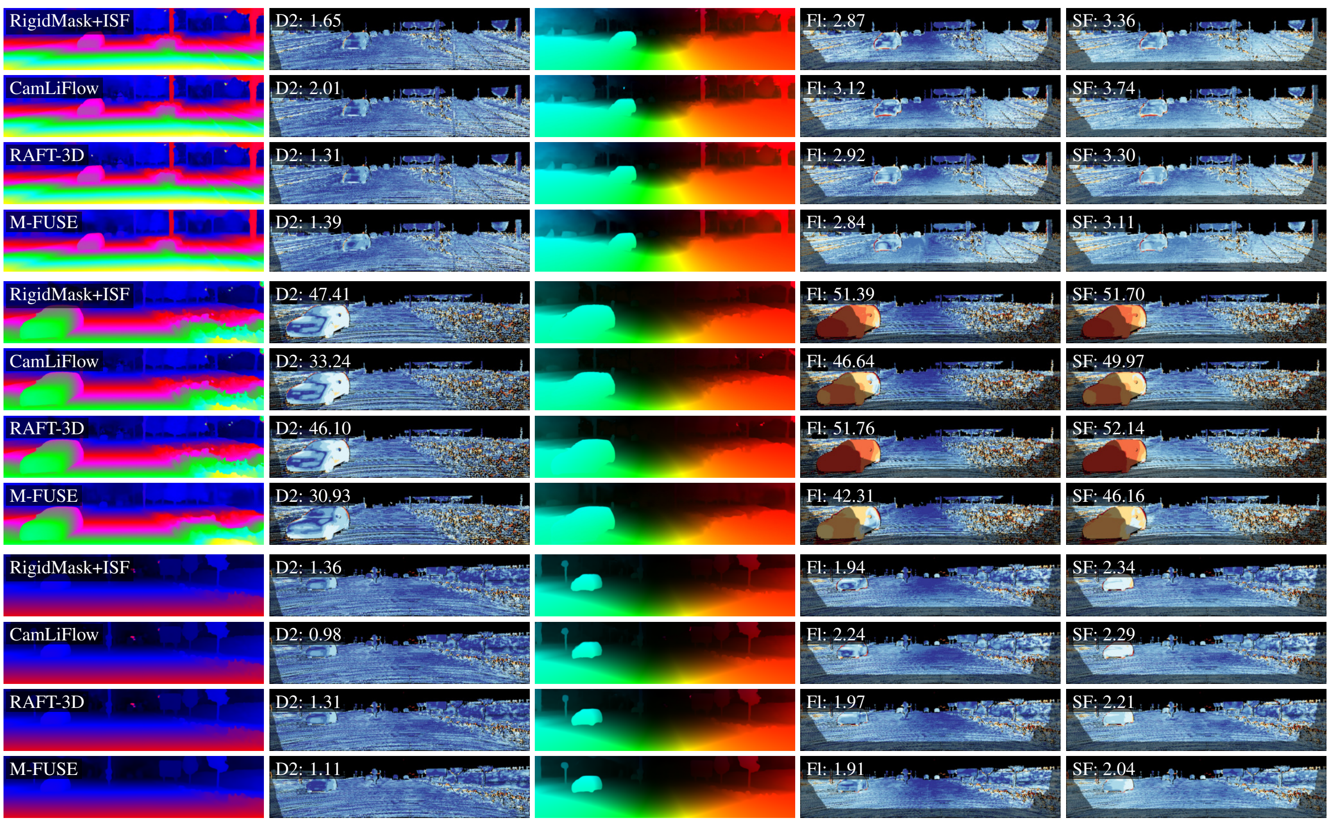

Figure 5. Qualitative comparison of our method, the original RAFT-3D, as well as the two top-performing approaches from the literature using the visualizations provided by the KITTI benchmark. *From left to right:* Target disparity visualization, corresponding *D2* error plot, optical flow visualization, corresponding *Fl* error plot, combined *SF* error plot.