Supplementary Material: Learning the Distribution of Errors in Stereo Matching for Joint Disparity and Uncertainty Estimation

Liyan Chen Weihan Wang Philippos Mordohai
Stevens Institute of Technology

In this document, we present more quantitative and qualitative results extending Section 4 of the main paper. Ablation studies on the configuration of the histograms and on different settings of inlier filters are provided in Section 1. A detailed version of the synthetic-to-real transfer from Virtual KITTI 2 (VK2) to the DrivingStereo weather subsets (DS-Weather) is presented in Section 2. Section 3 includes additional qualitative results, such as disparity, error and uncertainty maps.

1. Ablation Studies

We conduct ablation studies to explore the impact of the number and spacing between the histogram bins, as well as the inlier threshold. Table S.1 summarizes the quantitative results obtained from different baselines and our proposed method trained on the three datasets. (We selected the best variant of each method and presented a brief version of this table as Table 3 in the main paper.)

The first ablation study is to explore the configuration of the histograms, i.e. the α and m in Section 3.2 in the main paper, with respect to the *Bins* and *Scale* under *Loss* category in all tables. We performed this experiment on the Scene Flow dataset. When training SEDNet, we varied the number of bins and toggled between linear and logarithmic scaling. The results reveal that using bins defined in log space is better than linear space. This is due to the fact that the distributions of errors and uncertainty are approximately Laplace distributions, and as a result, most samples are concentrated around the means. See the two rows corresponding to SEDNet with 11 bins in the *Scene Flow* section of Table S.1 for a comparison of linearly and logarithmically spaced bins. The distributions corresponding to these two rows are also plotted in Figure S.1. Moreover, we find that increasing the number of bins for soft-histogramming does not increase the accuracy but only the computational cost. See the last four rows in the *Scene Flow* section in Table S.1 that shows SEDNet results with 11, 20 or 50 bins.

The second ablation study revolves around the selection of the inlier threshold, i.e. the *Inliers* category in all tables. We first tried GwcNet, $\mathcal{L}_{log}$ and SEDNet with a fixed threshold (EPE$<$5) and then an adaptive threshold ($<\mu_\epsilon$+1b_ϵ) when training on Scene Flow and VK2. We find that GwcNet and $\mathcal{L}_{log}$ are more compatible with a fixed threshold, while SEDNet works better when using an adaptive threshold. We also changed the threshold of SEDNet to be μ_ϵ+1b_ϵ and μ_ϵ+5b_ϵ when training models on VK2.

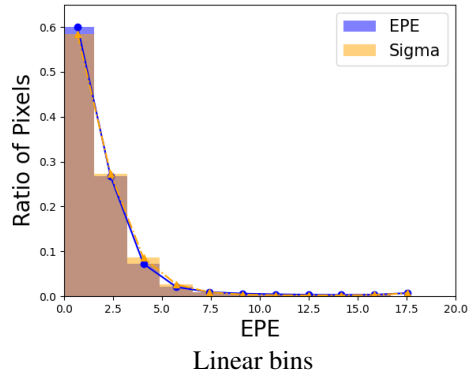

Figure S.1. Distributions of error and predicted uncertainty using different scaling of bins. The distributions correspond to the two rows of SEDNet with 11 bins in the *Scene Flow* section of Table S.1.

In most cases, models with the threshold at $3b_\epsilon$ are better than others. This does not hold for the evaluation on DS-Weather, where the network lacks knowledge of the unseen domain. Information from the error maps available when the threshold is larger may support a better prediction.

We also computed the precise percentage of the inliers in each experiment, see *Pct* in Table S.1. The results indicate that the threshold, fixed or adaptive, should not be too restrictive because the network does not improve if back-propagation only occurs on pixels with small errors. At the same time, outliers contaminate the solution and sometimes hinder convergence.

2. Generalization from Synthetic to Real Data

Here, we present more quantitative results on the synthetic to real data experiments supplementing Table 3 in the main paper. Table S.2 presents results from multiple variants of each method, including with and without inlier filtering, and different fixed or adaptive thresholds. The results are consistent with the findings in the ablation studies.

It is worth noting that on *DS-Rainy*, SEDNet with an inlier threshold of $\mu_\epsilon + 1b_\epsilon$ exhibits very poor performance due to the fact that only 85.82% of the pixels are considered inliers. An overly restrictive inlier threshold, fixed or adaptive, is harmful to the performance of the network, since the back-propagation only occurs on pixels with small errors and the network does not benefit from hard examples.

3. Additional Qualitative Results

In the following pages, we provide more qualitative results to demonstrate the effectiveness of the proposed method. We select examples from the three datasets mentioned in Section 4.1. The main differences are highlighted in red boxes.

Scene Flow. Figure S.2 and Figure S.3 show two examples from the Flying3D dataset. Unlike VK2 and DrivingStereo that include images of street views, this dataset provides image pairs of indoor objects. We find that SEDNet is good at capturing the uncertainty of the boundaries of overlapping objects, and at predicting more accurate disparity for textureless objects and objects with complicated structure such as holes.

VK2. Figure S.4 presents a comparison of disparity estimation under different weather conditions on the synthetic datasets. To further illustrate the strength of SEDNet in predicting disparity as well as accurate uncertainty, we pick two hard examples, in Figures S.5 and S.6, from the foggy subset. SEDNet still performs very well considering the surrounding environment is hazy, while $\mathcal{L}_{log}$ fails to figure out the background.

DS-Weather. Real data acquired under adverse weather conditions exhibit more challenges than synthetic data under simulated similar weather conditions. In addition to poor illumination and opacity, real data also suffer from reflections and the Tyndall effect. Another large challenge for this dataset is that the LIDAR ground truth is sparse. Figure S.7 presents uncertainty estimates by the different methods under diverse illumination conditions. Figures S.8, S.9 and S.10 further illustrate how SEDNet outperforms the baselines under adverse weather.

Dataset	Method	Loss						Inliers		Disparity↓		APE↓		AUC↓	
		BCE	L1	Log	KL	Bins	Scale	Def.	Pct(%)	EPE	D1(%)	Avg.	Median	Opt.	Est.
Scene Flow	GwcNet	-	√	-	-	-	-	-	-	0.7758	4.127	-	-	10.9291	-
	GwcNet	-	√	-	-	-	-	EPE<5	97.09	0.7799	3.940	-	-	8.3413	-
	GwcNet	-	√	-	-	-	log	EPE<μ_ϵ+3b_ϵ	98.41	0.7981	4.072	-	-	9.6451	-
	+LAF	√	-	-	-	-	-	-	-	0.7758	4.127	-	-	10.9291	20.0813
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	-	-	0.7445	4.522	0.7133	0.0795	6.2567	10.9635
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	EPE<5	96.96	0.7611	4.131	0.6999	0.0728	5.7449	12.1121
	+$\mathcal{L}_{log}$	-	√	√	-	-	log	EPE<μ_ϵ+3b_ϵ	98.33	0.7890	4.428	0.7047	0.0869	6.6069	12.4265
	+SEDNet	-	√	√	√	-	-	EPE<5	96.57	1.0046	4.455	0.9092	0.1327	7.1444	16.0036
	+SEDNet	-	√	√	√	11	lin	EPE<μ_ϵ+3b_ϵ	98.42	0.6827	4.022	0.5877	0.0450	5.1258	9.0113
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+3b_ϵ	98.42	0.6754	3.963	0.5797	0.0432	4.9134	8.7195
	+SEDNet	-	√	√	√	20	log	EPE<μ_ϵ+3b_ϵ	98.42	0.6762	3.966	0.5821	0.0433	4.9845	8.8103
	+SEDNet	-	√	√	√	50	log	EPE<μ_ϵ+3b_ϵ	98.42	0.6894	4.016	0.5931	0.0426	5.0216	8.9412
VK2-S6	GwcNet	-	√	-	-	-	-	-	-	0.4125	1.763	-	-	6.0962	-
	GwcNet	-	√	-	-	-	-	EPE<5	98.82	0.4339	1.634	-	-	6.0598	-
	GwcNet	-	√	-	-	-	log	EPE<μ_ϵ+3b_ϵ	99.50	0.4597	1.847	-	-	6.5506	-
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	-	-	0.4360	1.957	0.5876	0.1123	4.9389	10.2407
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	EPE<5	98.86	0.3899	1.584	0.4136	0.1753	4.6872	12.5320
	+$\mathcal{L}_{log}$	-	√	√	-	-	log	EPE<μ_ϵ+3b_ϵ	99.52	0.4079	1.673	0.4549	0.2261	5.1675	13.3036
	+SEDNet	-	√	√	√	-	-	EPE<5	98.59	0.5197	1.905	0.5382	0.1779	5.2803	13.1777
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+1b_ϵ	98.30	0.3896	1.582	0.3876	0.1330	4.4569	11.8449
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+3b_ϵ	99.24	0.3109	1.392	0.5234	0.1454	4.1726	9.7637
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+5b_ϵ	99.68	0.3236	1.427	0.3561	0.1096	4.2767	9.9843
VK2-S6-Moving	GwcNet	-	√	-	-	-	-	-	-	0.4253	1.689	-	-	5.9184	-
	GwcNet	-	√	-	-	-	-	EPE<5	98.85	0.4543	1.593	-	-	5.7884	-
	GwcNet	-	√	-	-	-	log	EPE<μ_ϵ+3b_ϵ	99.60	0.4708	1.707	-	-	6.2409	-
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	-	-	0.4618	1.930	0.5592	0.1176	4.6928	9.3604
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	EPE<5	98.91	0.4231	1.537	0.4575	0.1890	4.3663	11.3532
	+$\mathcal{L}_{log}$	-	√	√	-	-	log	EPE<μ_ϵ+3b_ϵ	99.63	0.4280	1.632	0.4885	0.2406	4.7709	12.5484
	+SEDNet	-	√	√	√	-	-	EPE<5	98.67	0.5581	1.823	0.5806	0.1910	4.9404	12.1068
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+1b_ϵ	98.77	0.4220	1.567	0.4277	0.1442	4.1986	10.8037
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+3b_ϵ	99.62	0.3577	1.389	0.5958	0.1573	3.9012	8.8339
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+5b_ϵ	99.76	0.3862	1.420	0.4002	0.1164	4.0423	9.0631
DrivingStereo	+$\mathcal{L}_{log}$(FT)	-	√	√	-	-	-	-	-	0.5332	0.2641	0.3449	0.2297	21.7002	45.7096
	+SEDNet(FT)	-	√	√	√	11	log	EPE<μ_ϵ+5b_ϵ	99.86	0.5264	0.2439	0.3324	0.2267	21.2856	44.3297
DS-Weather	GwcNet	-	√	-	-	-	-	-	-	1.6962	8.313	-	-	44.4896	-
	GwcNet	-	√	-	-	-	-	EPE<5	73.05	18.2891	30.856	-	-	358.1623	-
	GwcNet	-	√	-	-	-	log	EPE<μ_ϵ+3b_ϵ	97.87	14.3547	31.835	-	-	315.8265	-
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	-	-	1.9458	8.700	6.2375	0.8295	43.3146	127.4829
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	EPE<5	95.78	2.3944	6.666	2.1443	0.4383	41.1909	95.4264
	+$\mathcal{L}_{log}$	-	√	√	-	-	log	EPE<μ_ϵ+3b_ϵ	98.32	5.4850	15.136	5.1234	0.7290	90.5541	175.9993
	+SEDNet	-	√	√	√	-	-	EPE<5	94.33	3.7375	9.297	3.3794	0.5249	49.9049	124.0092
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+1b_ϵ	95.54	6.9335	9.346	6.1682	0.4946	53.5638	99.2891
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+3b_ϵ	98.95	1.5637	6.508	2.3406	0.5309	38.4871	86.1118
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+5b_ϵ	99.41	1.7051	6.057	1.5842	0.6104	39.8057	87.1882

Table S.1. Quantitative results: (1) *within-domain* on SceneFlow, VK2-S6 and VK2-S6-Moving; (2) after finetuning (FT) on DrivingStereo; (3) *cross-domain* on DS-Weather. The best results in each category in each experiment are highlighted in dark blue. The top-performing variant of SEDNet, namely SEDNet with EPE<μ_ϵ+3b_ϵ, outperforms the baselines with respect to disparity and uncertainty metrics in the majority of experiments.

Dataset	Method	Loss						Inliers		Disparity↓		APE↓		AUC↓	
		BCE	L1	Log	KL	Bins	Scale	Def.	Pct(%)	EPE	D1(%)	Avg.	Median	Opt.	Est.
VK2-S6-Morning	GwcNet	-	√	-	-	-	-	-	-	0.4642	1.740	-	-	6.1845	-
	GwcNet	-	√	-	-	-	-	EPE<5	98.79	0.5107	1.649	-	-	6.0792	-
	GwcNet	-	√	-	-	-	log	EPE<μ_ϵ+3b_ϵ	99.57	0.5065	1.742	-	-	6.4706	-
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	-	-	0.5117	1.998	0.6616	0.1162	5.0563	10.0704
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	EPE<5	98.82	0.4774	1.624	0.5067	0.1872	4.6698	12.5192
	+$\mathcal{L}_{log}$	-	√	√	-	11	log	EPE<μ_ϵ+3b_ϵ	99.59	0.4571	1.614	0.5063	0.2231	4.8135	12.2921
	+SEDNet	-	√	√	√	-	-	EPE<5	98.58	0.6136	1.928	0.6300	0.1897	5.2304	13.1250
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+1b_ϵ	98.83	0.4774	1.626	0.4768	0.1431	4.4936	11.8164
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+3b_ϵ	99.62	0.4003	1.442	0.6183	0.1553	4.1847	9.4063
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+5b_ϵ	99.71	0.4265	1.481	0.4356	0.1150	4.3216	9.6694
VK2-S6-Sunset	GwcNet	-	√	-	-	-	-	-	-	0.4810	1.825	-	-	6.6907	-
	GwcNet	-	√	-	-	-	-	EPE<5	98.76	0.5222	1.701	-	-	6.5110	-
	GwcNet	-	√	-	-	-	log	EPE<μ_ϵ+3b_ϵ	99.57	0.5345	1.795	-	-	6.9508	-
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	-	-	0.5112	2.040	0.6134	0.1137	5.4426	11.1299
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	EPE<5	98.84	0.4863	1.627	0.5060	0.1827	5.0075	13.7848
	+$\mathcal{L}_{log}$	-	√	√	-	11	log	EPE<μ_ϵ+3b_ϵ	99.62	0.4678	1.646	0.5046	0.2151	5.2551	13.8256
	+SEDNet	-	√	√	√	-	-	EPE<5	98.54	0.6506	1.981	0.6558	0.1827	5.7348	14.6926
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+1b_ϵ	98.81	0.4871	1.664	0.4764	0.1356	4.9970	13.4379
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+3b_ϵ	99.61	0.4108	1.475	0.6189	0.1509	4.5840	10.7946
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+5b_ϵ	99.72	0.4422	1.505	0.4399	0.1103	4.8745	11.0680
VK2-S6-Fog	GwcNet	-	√	-	-	-	-	-	-	0.4660	1.812	-	-	6.8355	-
	GwcNet	-	√	-	-	-	-	EPE<5	98.91	0.4817	1.556	-	-	6.4733	-
	GwcNet	-	√	-	-	-	log	EPE<μ_ϵ+3b_ϵ	99.66	0.5073	1.810	-	-	7.1501	-
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	-	-	0.4919	1.859	0.5574	0.1174	5.3870	11.1750
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	EPE<5	98.98	0.4425	1.448	0.4609	0.1865	4.8983	12.1305
	+$\mathcal{L}_{log}$	-	√	√	-	11	log	EPE<μ_ϵ+3b_ϵ	99.69	0.4330	1.490	0.4671	0.2211	5.2211	12.5494
	+SEDNet	-	√	√	√	-	-	EPE<5	98.88	0.5410	1.579	0.5398	0.1880	5.4246	12.2987
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+1b_ϵ	98.90	0.4657	1.533	0.4459	0.1415	4.9125	12.1241
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+3b_ϵ	99.71	0.3731	1.288	0.5517	0.1547	4.4200	9.9380
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+5b_ϵ	99.77	0.4108	1.339	0.4162	0.1156	4.5310	10.0341
VK2-S6-Rain	GwcNet	-	√	-	-	-	-	-	-	0.4618	1.700	-	-	6.6774	-
	GwcNet	-	√	-	-	-	-	EPE<5	98.84	0.5030	1.633	-	-	6.5983	-
	GwcNet	-	√	-	-	-	log	EPE<μ_ϵ+3b_ϵ	99.61	0.5197	1.835	-	-	7.1739	-
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	-	-	0.5160	1.989	0.6064	0.1199	5.3429	10.8433
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	EPE<5	98.88	0.4707	1.571	0.4899	0.1861	4.9351	13.3214
	+$\mathcal{L}_{log}$	-	√	√	-	11	log	EPE<μ_ϵ+3b_ϵ	99.66	0.4543	1.565	0.4914	0.2198	5.1625	13.7641
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+1b_ϵ	98.82	0.4701	1.611	0.4710	0.1413	4.7792	12.8168
	+SEDNet	-	√	√	√	-	-	EPE<5	98.75	0.5868	1.751	0.5833	0.1865	5.5200	14.5175
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+3b_ϵ	99.69	0.3873	1.356	0.6685	0.1537	4.4013	10.3362
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+5b_ϵ	99.74	0.4383	1.461	0.4557	0.1171	4.5840	10.6757
DS-Cloudy	GwcNet	-	√	-	-	-	-	-	-	1.3413	5.229	-	-	37.4263	-
	GwcNet	-	√	-	-	-	-	EPE<5	84.75	7.3789	18.559	-	-	71.6779	-
	GwcNet	-	√	-	-	-	log	EPE<μ_ϵ+3b_ϵ	97.19	7.8976	21.581	-	-	88.6479	-
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	-	-	1.8379	6.731	3.3032	0.6810	37.8238	159.3730
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	EPE<5	97.48	1.4780	3.948	1.2617	0.3513	34.4488	82.5380
	+$\mathcal{L}_{log}$	-	√	√	-	11	log	EPE<μ_ϵ+3b_ϵ	98.78	1.9784	5.719	1.7444	0.3785	39.2511	91.5997
	+SEDNet	-	√	√	√	-	-	EPE<5	96.96	2.0096	4.844	1.7492	0.3930	37.8752	95.2040
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+1b_ϵ	97.11	4.0438	6.101	3.5618	0.3936	40.3969	82.5675
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+3b_ϵ	98.83	1.3183	4.414	1.5260	0.4021	33.9037	73.6330
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+5b_ϵ	99.41	1.1901	3.772	1.0795	0.4736	32.5237	69.6368
DS-Sunny	GwcNet	-	√	-	-	-	-	-	-	1.5448	6.991	-	-	38.7386	-
	GwcNet	-	√	-	-	-	-	EPE<5	88.95	5.2974	14.292	-	-	56.5606	-
	GwcNet	-	√	-	-	-	log	EPE<μ_ϵ+3b_ϵ	97.61	4.8873	17.035	-	-	62.4926	-
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	-	-	1.5645	6.039	3.1431	0.8429	36.3650	76.7900
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	EPE<5	97.08	1.4837	4.631	1.2806	0.3835	35.5226	85.8715
	+$\mathcal{L}_{log}$	-	√	√	-	11	log	EPE<μ_ϵ+3b_ϵ	98.57	1.7809	6.131	1.5580	0.3828	38.5942	97.1704
	+SEDNet	-	√	√	√	-	-	EPE<5	95.43	2.6221	7.001	2.3380	0.4468	41.8861	108.8739
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+1b_ϵ	97.24	3.5945	6.274	3.6359	0.3894	38.2571	85.1056
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+3b_ϵ	98.64	1.5548	5.878	3.0025	0.4808	35.6523	83.2573
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+5b_ϵ	99.27	1.4164	5.219	1.3549	0.6110	34.0465	83.2316
DS-Foggy	GwcNet	-	√	-	-	-	-	-	-	1.5476	8.859	-	-	51.4640	-
	GwcNet	-	√	-	-	-	-	EPE<5	83.04	10.1839	22.130	-	-	105.1603	-
	GwcNet	-	√	-	-	-	log	EPE<μ_ϵ+3b_ϵ	97.34	5.0526	20.534	-	-	80.6149	-
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	-	-	1.6435	9.694	4.0449	0.8879	49.2533	95.5706
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	EPE<5	94.89	2.9553	9.015	2.6923	0.5556	48.7136	101.7025
	+$\mathcal{L}_{log}$	-	√	√	-	11	log	EPE<μ_ϵ+3b_ϵ	98.93	1.6931	8.979	1.4534	0.6143	50.7000	106.2925
	+SEDNet	-	√	√	√	-	-	EPE<5	94.64	3.3875	11.676	3.0463	0.6894	57.0915	137.2929
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+1b_ϵ	96.15	6.1046	9.996	5.1343	0.6046	56.8943	104.6067
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+3b_ϵ	99.27	1.5398	7.357	2.4109	0.7023	47.7932	97.8627
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+5b_ϵ	99.50	1.6536	7.145	1.5196	0.7310	44.5539	99.5714
DS-Rainy	GwcNet	-	√	-	-	-	-	-	-	3.1918	17.356	-	-	68.0346	-
	GwcNet	-	√	-	-	-	-	EPE<5	35.47	49.6661	68.441	-	-	1199.2505	-
	GwcNet	-	√	-	-	-	log	EPE<μ_ϵ+3b_ϵ	99.32	39.5814	68.191	-	-	1031.5507	-
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	-	-	3.6950	17.079	24.7441	1.0236	67.2717	253.0992
	+$\mathcal{L}_{log}$	-	√	√	-	-	-	EPE<5	98.79	5.3539	12.501	4.9480	0.5759	59.3952	146.8906
	+$\mathcal{L}_{log}$	-	√	√	-	11	log	EPE<μ_ϵ+3b_ϵ	96.98	16.4877	39.713	15.7376	1.5403	233.6712	408.9347
	+SEDNet	-	√	√	√	-	-	EPE<5	90.28	6.9306	13.668	6.3839	0.5705	62.7668	154.6661
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+1b_ϵ	85.82	22.9318	22.408	20.1601	0.8772	219.9731	278.7730
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+3b_ϵ	99.10	2.2165	11.020	2.6599	0.6722	50.8103	110.8360
	+SEDNet	-	√	√	√	11	log	EPE<μ_ϵ+5b_ϵ	98.95	3.6734	10.975	3.4255	0.7346	54.1441	129.2394

Table S.2. Quantitative results of synthetic to real evaluation. This is the extended version of Table 3 in the main paper. The best results in each category in each experiment are in dark blue. The top-performing Variant of SEDNet, namely SEDNet with EPE<μ_ϵ+3b_ϵ, outperforms the baselines in the majority of experiments, especially in uncertainty estimation on real data and under adverse weather (i.e., foggy and rainy).

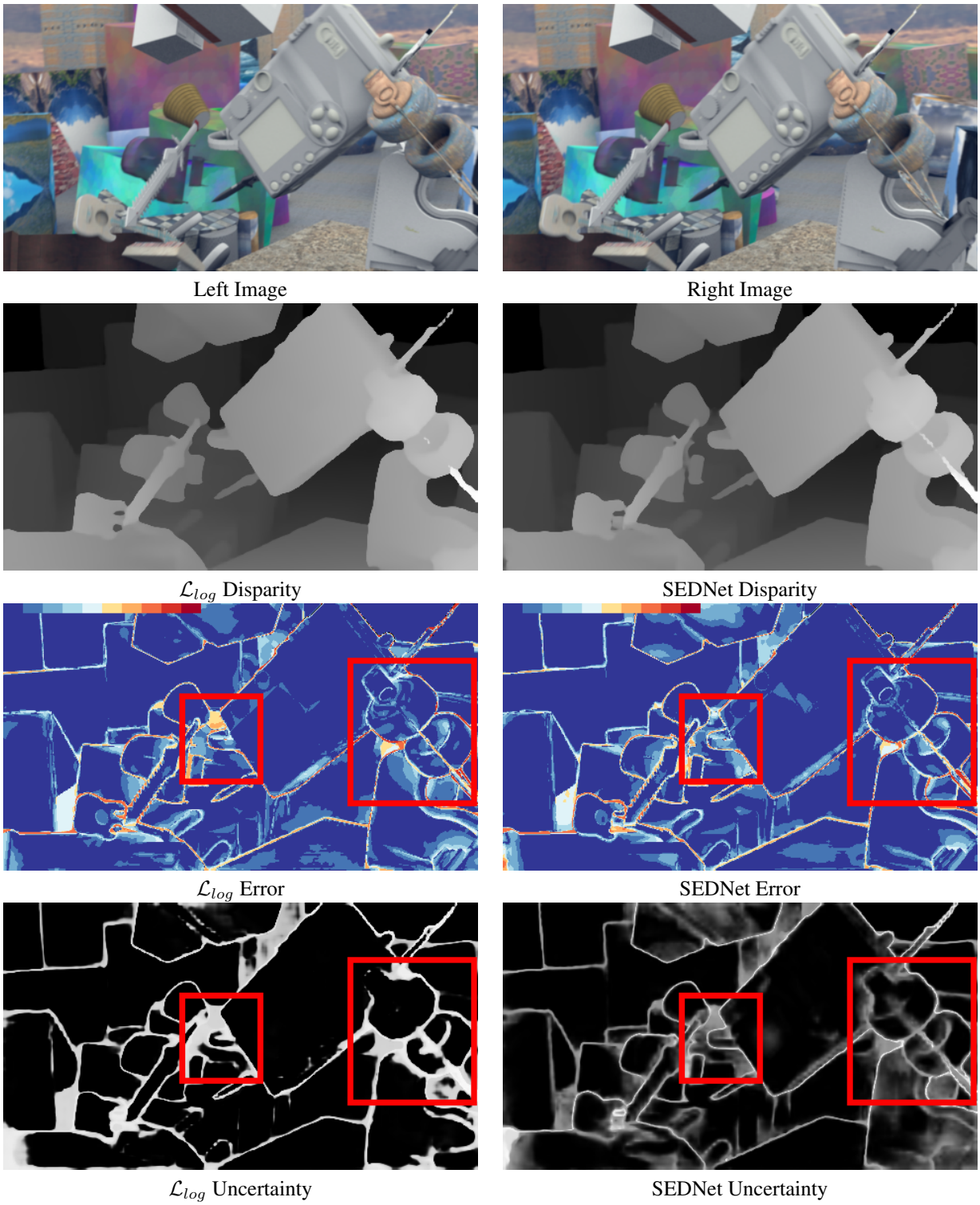

Figure S.2. Example from Scene Flow. When objects overlap with each other and depth ordering is unclear, SEDNet captures the uncertainty more precisely according to the error map. In the regions outlined in red, SEDNet successfully detects the pull ring of the camera and the lid of the wheel, while the $\mathcal{L}_{log}$ model fails to estimate their disparity.

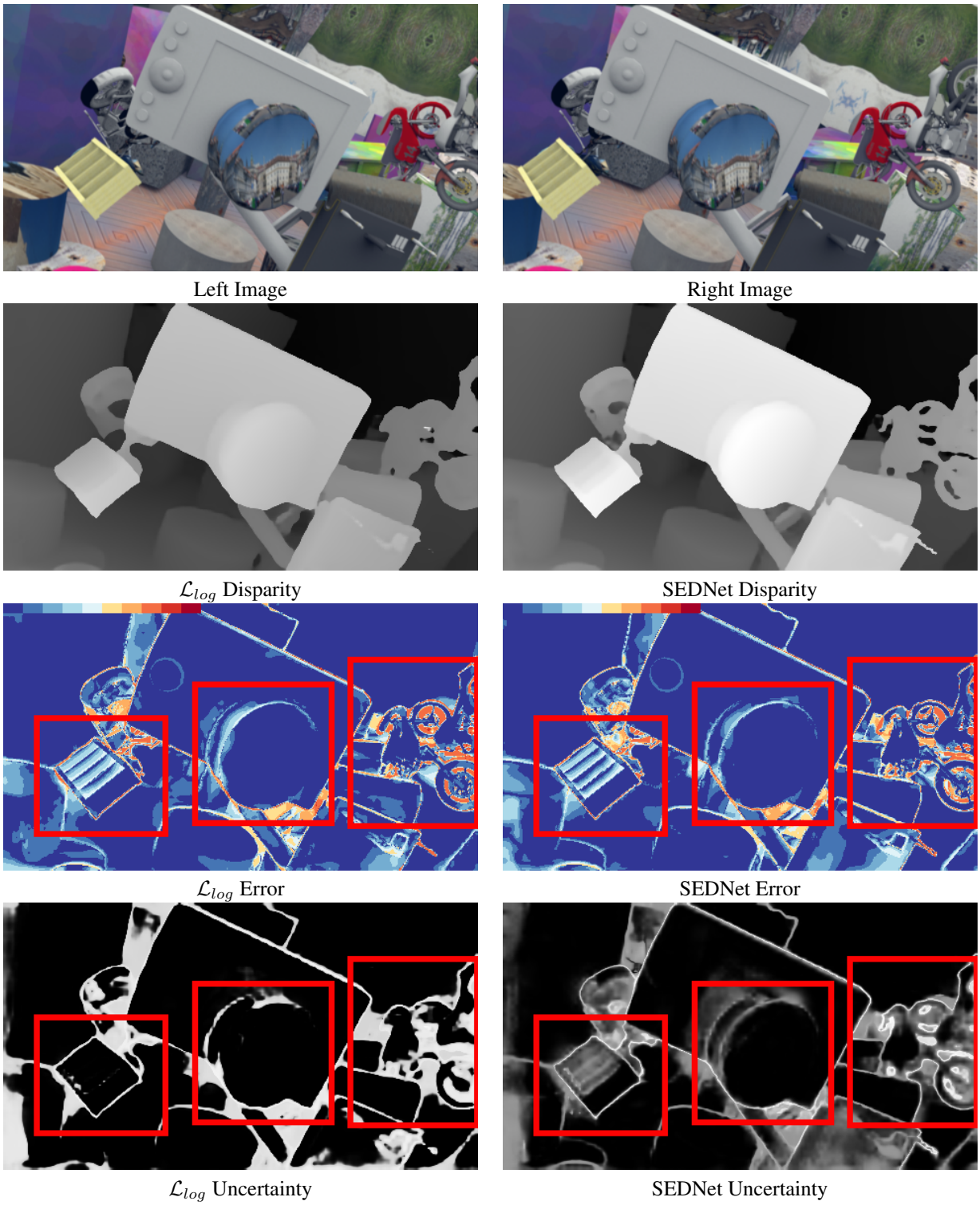

Figure S.3. Example from Scene Flow. The disparity and uncertainty maps of SEDNet include some structure information of the objects, such as the bookcase on the left side which has several openings, the cylinder which has many intersecting surfaces, and the wheels of the motorcycle. The prediction of $\mathcal{L}_{log}$ lacks these details.

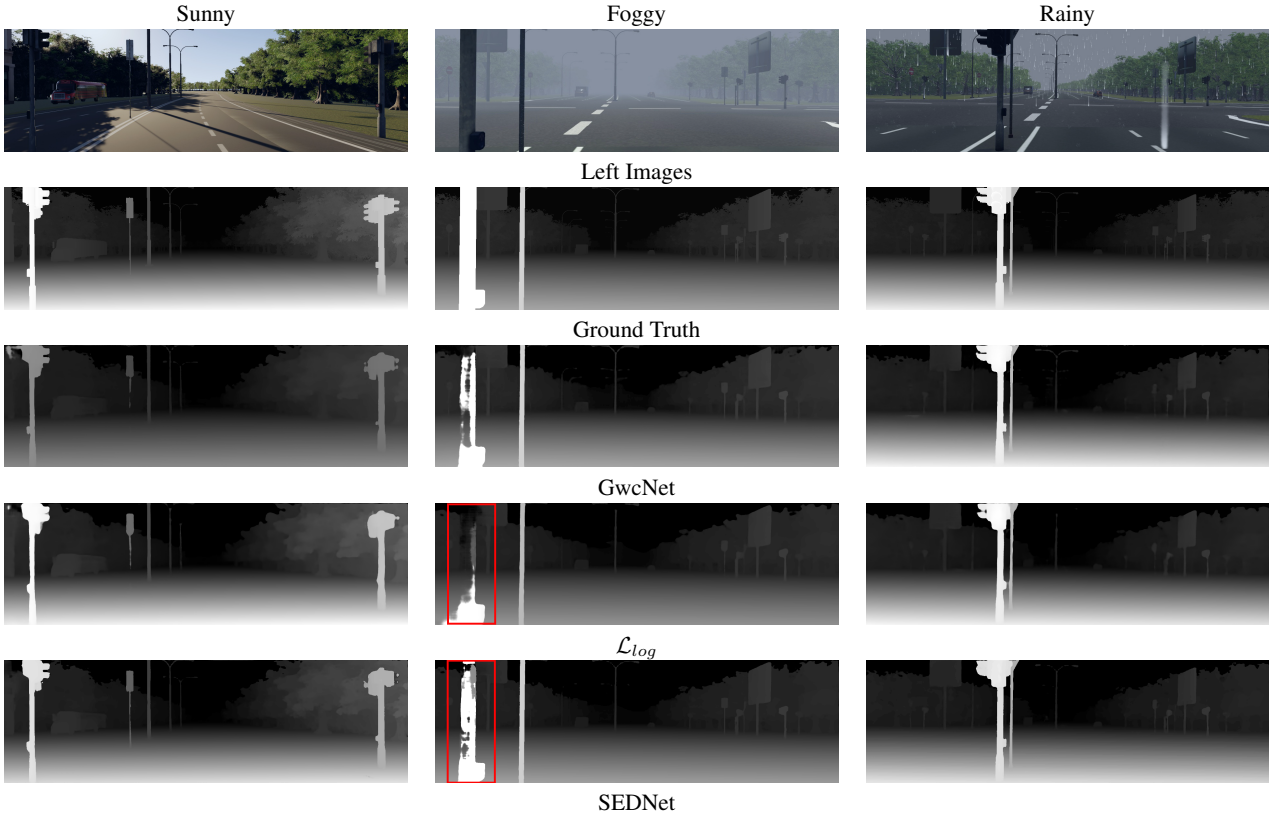

Figure S.4. Examples of driving in different weather conditions from VK2-S6. We pick the best $\mathcal{L}_{log}$ and SEDNet based on EPE in Table S.1. (Please zoom in to see details.) SEDNet is better at predicting fine, challenging details in the disparity maps, such as the precise shape of the traffic light and the car under the left shadow in the sunny picture, the traffic sign behind the light post and the street light in the middle of the image in the fog.

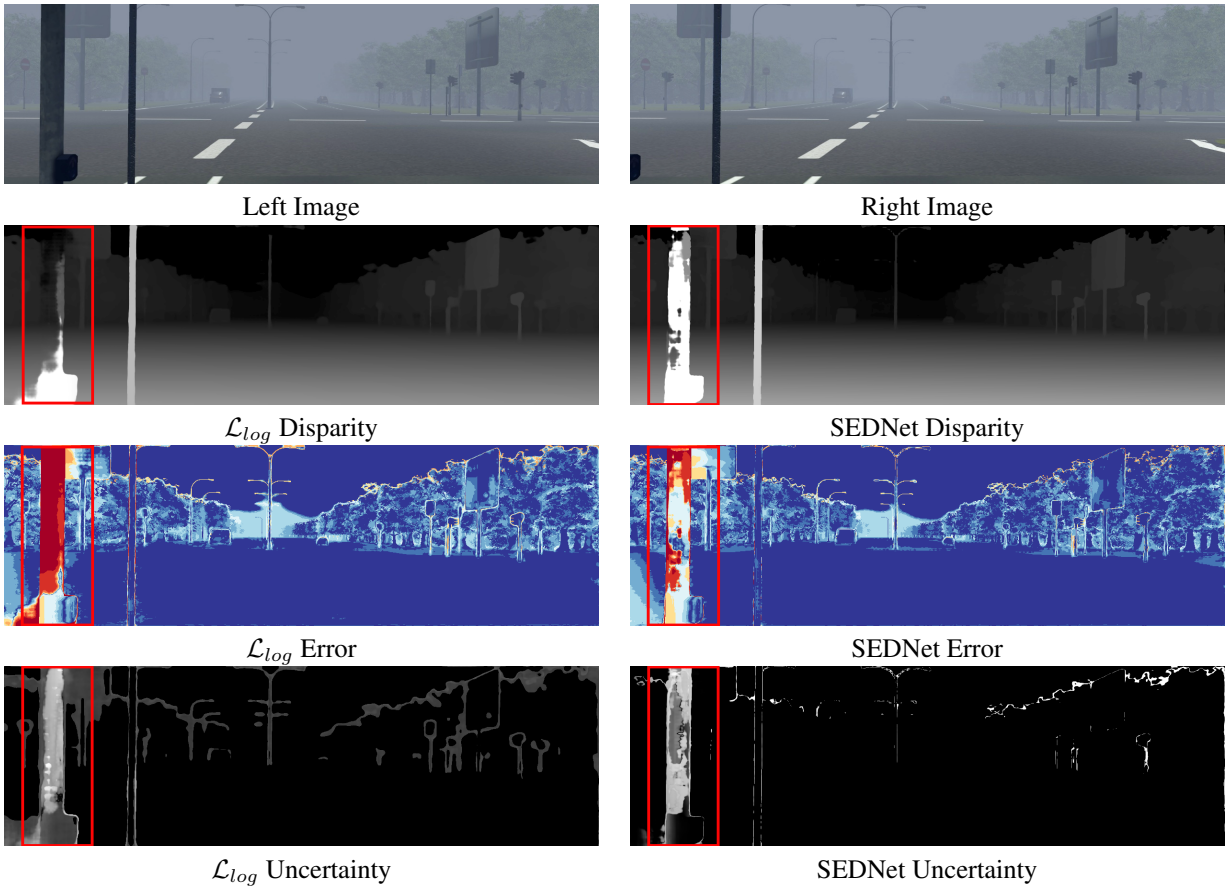

Figure S.5. Example from VK2-S6-Fog. The post only appears in the left image, $\mathcal{L}_{log}$ fails to predict its disparity, but SEDNet does. The uncertainty map of SEDNet matches the error map better than that of $\mathcal{L}_{log}$.

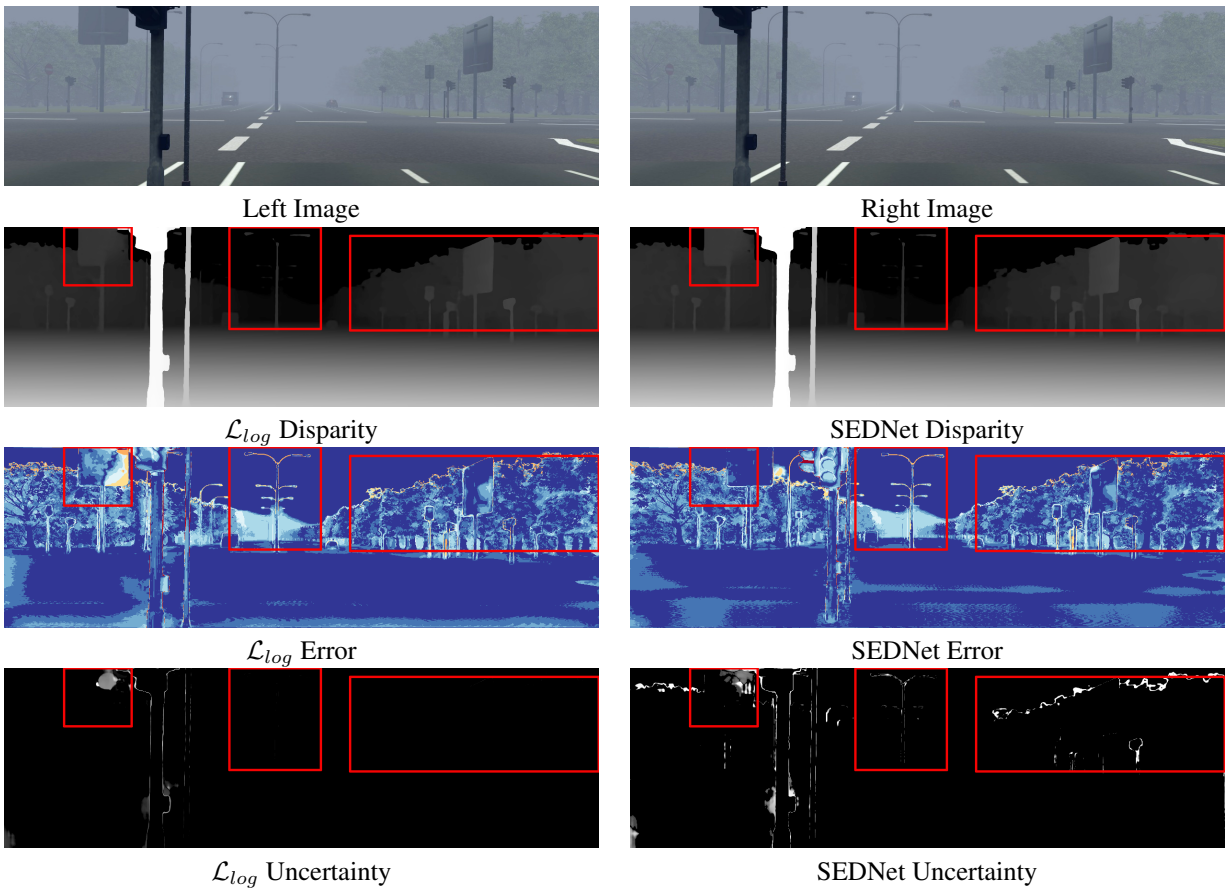

Figure S.6. Example from VK2-S6-Fog. Contrasting error and uncertainty maps, we observe that SEDNet predicts more accurate uncertainty in the regions covered by fog, such as the edges of the traffic signs and the trees in the background.

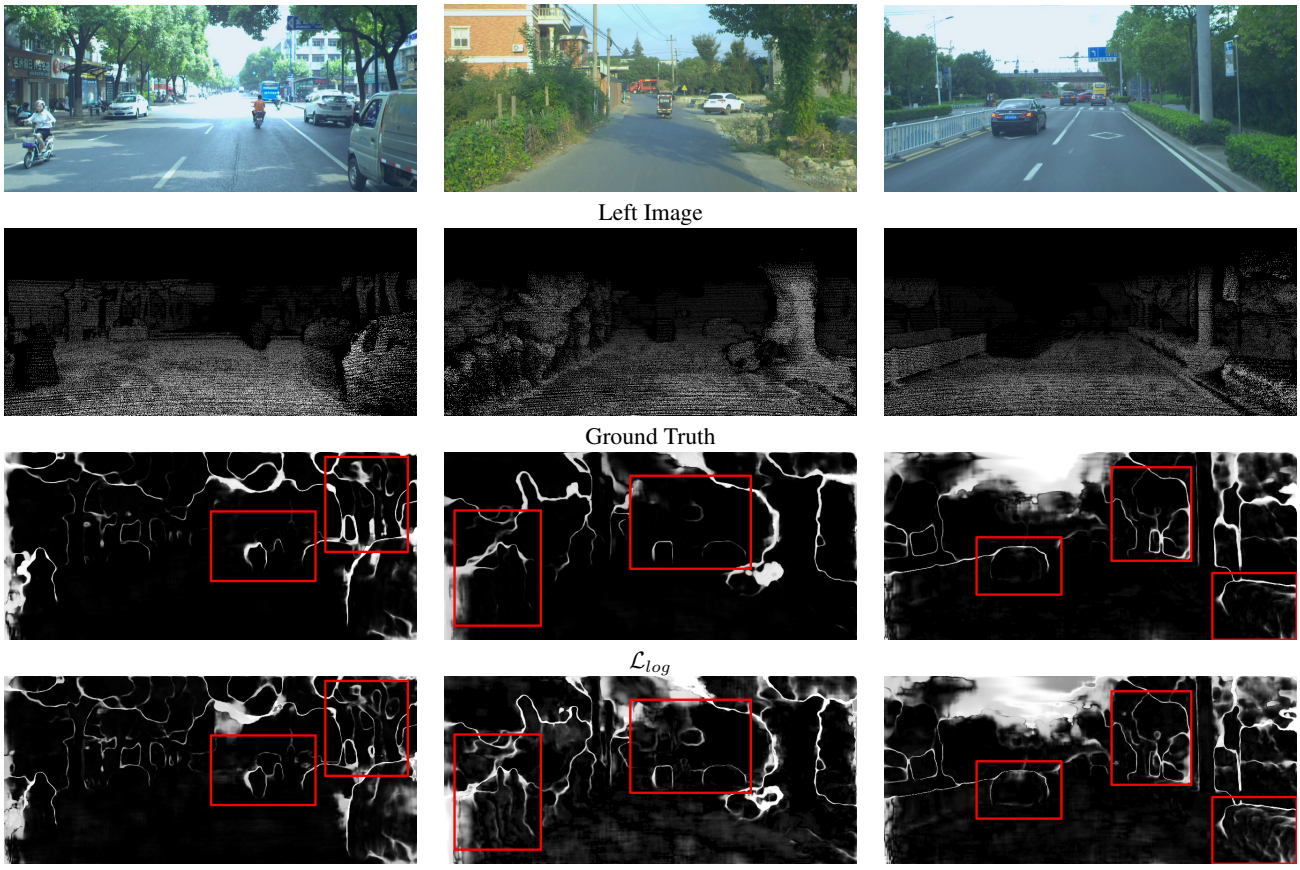

Figure S.7. Examples of uncertainty estimation on *DrivingStereo* also shown in Table S.1. SEDNet captures details more faithfully. For example, the cars, pedestrians and trees at different depths, even in overexposed parts of the images.

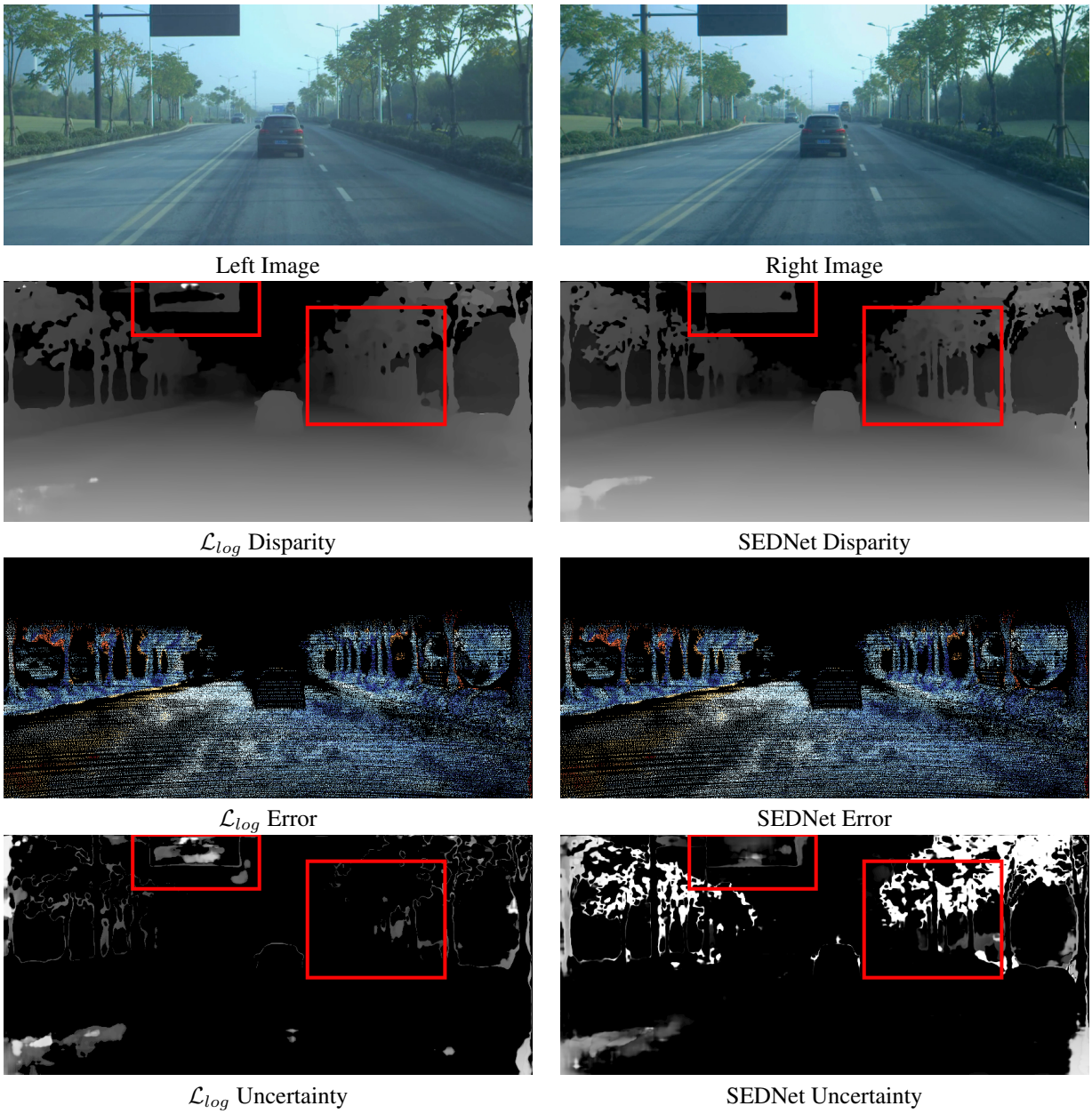

Figure S.8. Example from DS-Foggy. This subset is more challenging than the corresponding synthetic data due to unmodeled sources of noise. In this example, the thin tree trunk almost blends in with the background. Also, the billboard is very dark. $\mathcal{L}_{log}$ fails to predict the disparity of the textureless part of the billboard, as well as the space between the tree trunks on the right-hand side. The prediction of SEDNet is more accurate on these challenging parts.

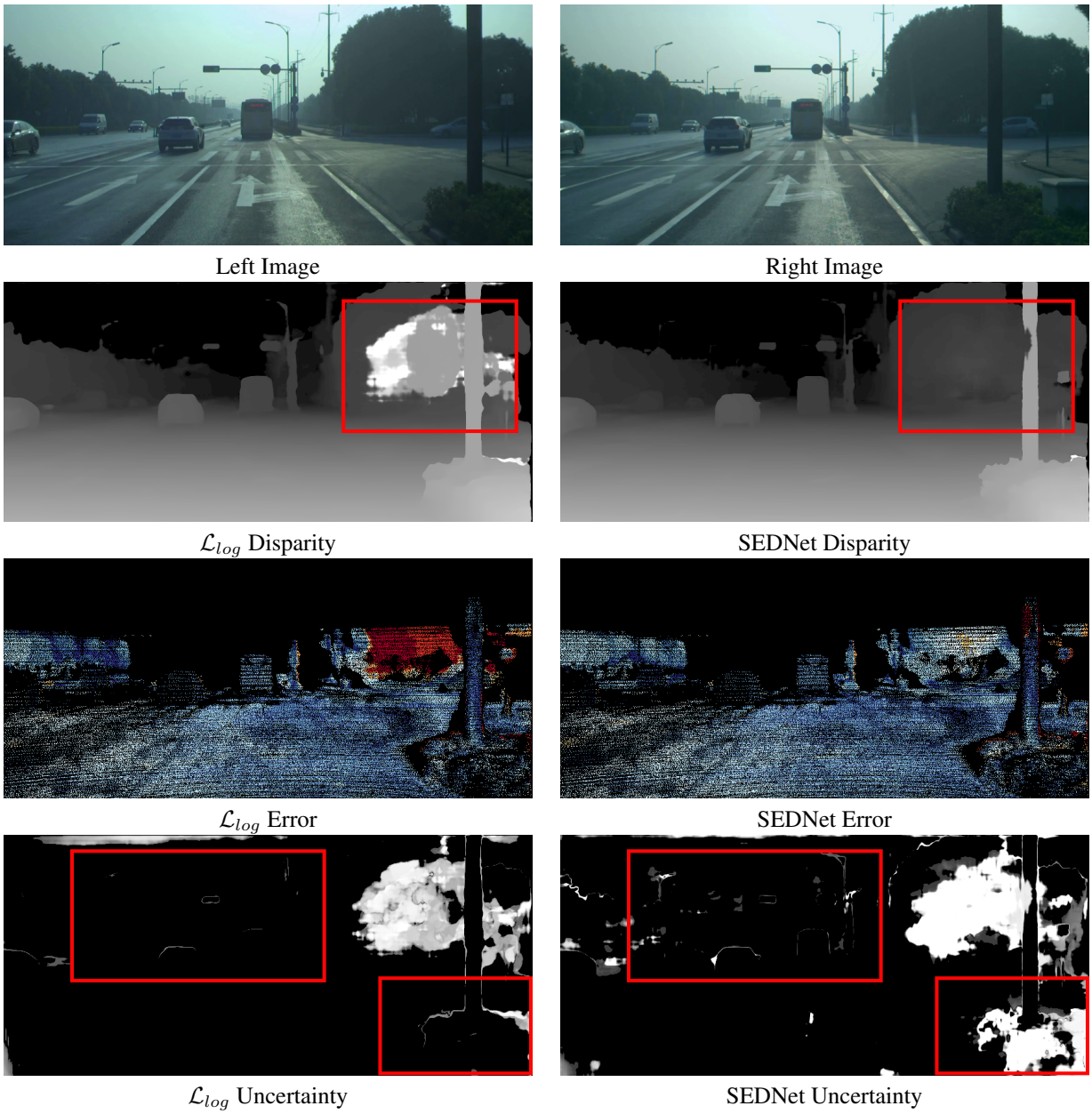

Figure S.9. Example from DS-Foggy. Similar to the synthetic data, images from the foggy day subset are usually very dark, which makes distinguishing objects in the shadow difficult. In this example, $\mathcal{L}_{log}$ makes a mistake in predicting the disparity of the trees on the right side, since they have similar color to the post. On the other hand, the prediction of SEDNet is more accurate. The uncertainty maps are dominated by the right backlit regions, making it hard to see the other parts. However, zooming in the figures reveals that SEDNet still performs better in predicting the uncertainty of the objects far from the camera and in the bottom right dark corner.

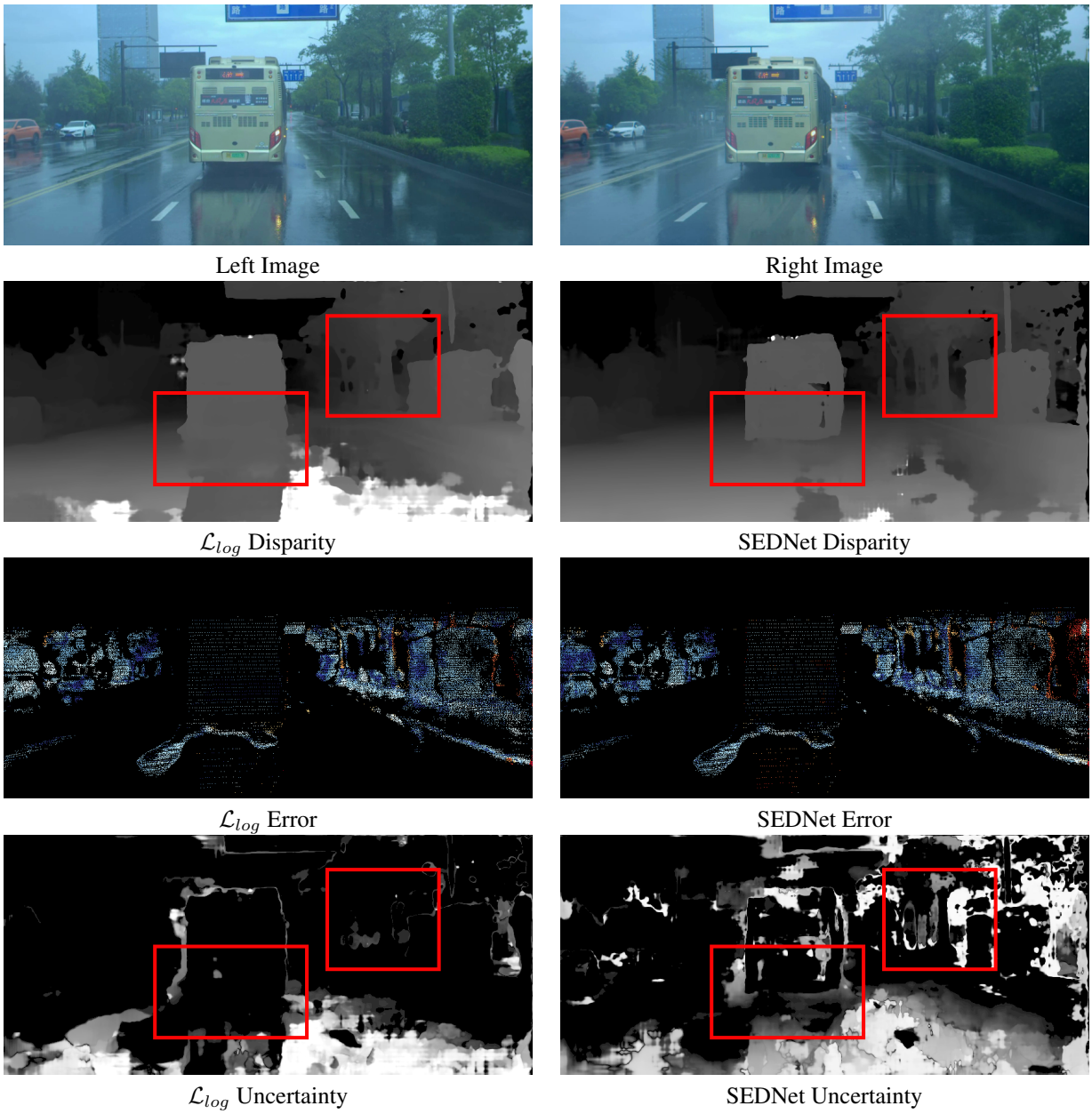

Figure S.10. Example from DS-Rainy. Unlike the synthetic data, the rainy-day real images do not only suffer from poor illumination, but also face challenges due to reflections in the water. In this example, the road is like a mirror, misleading the $\mathcal{L}_{log}$ model. Recall that the LIDAR ground truth disparity is very sparse, and is even sparser in reflective regions. Zooming in is required to see the recorded disparity errors on the road in the error maps. The disparity map of $\mathcal{L}_{log}$ fails to distinguish the car and the reflection, but SEDNet is able to estimate the correct disparity and the uncertainty of the car.