

# On the uncertainty of self-supervised monocular depth estimation – Supplementary material

Matteo Poggi      Filippo Aleotti      Fabio Tosi      Stefano Mattoccia  
 Department of Computer Science and Engineering (DISI)  
 University of Bologna, Italy  
 {m.poggi, filippo.aleotti2, fabio.tosi5, stefano.mattoccia}@unibo.it

In this document, we provide more detailed results concerning the experiments reported in the paper “On the uncertainty of self-supervised monocular depth estimation”. As in the submitted paper, we often simplify the notation by referring to *self-supervision* as *supervision*.

## 1. Insights about sparsification over different metrics

In the paper, due to the lack of space, we choose to show sparsification performance over three metrics, respectively Abs Rel, RMSE and  $\delta \geq 1.25$ . The first two metrics concern with the sparsification of an average error over a single depth map. In other words, this explains how good is our uncertainty modelling at finding pixels with the highest errors in magnitude and thus how much we can reduce the overall average error by removing them accordingly. The difference between the two consists in the fact that Abs Rel is normalized over the ground truth depth value, *i.e.* the magnitude of the error decreases for points farther from the camera, while the RMSE is independent of the depth in the scene.

Differently,  $\delta \geq 1.25$  metric selects a set of pixels (*i.e.* for which estimated depth is greater/smaller than  $1.25 \times$  the ground-truth value) as outliers. By sparsification, according to this metric, we aim at reducing the percentage of outliers in the depth map.

## 2. Detailed depth evaluation

In this document, we report the complete evaluation of each Monodepth2 variant on the seven metrics traditionally adopted in this field [2], obtained as follows:

$$\text{Abs Rel} = \frac{1}{|\mathcal{I}|} \sum_{p \in \mathcal{I}} \frac{|d(p) - d^*(p)|}{d^*(p)} \quad (1)$$

$$\text{Sq Rel} = \frac{1}{|\mathcal{I}|} \sum_{p \in \mathcal{I}} \frac{(d(p) - d^*(p))^2}{d^*} \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{1}{|\mathcal{I}|} \sum_{p \in \mathcal{I}} (d(p) - d^*(p))^2} \quad (3)$$

$$\text{RMSE log} = \sqrt{\frac{1}{|\mathcal{I}|} \sum_{p \in \mathcal{I}} (\log d(p) - \log d^*(p))^2} \quad (4)$$

$$\delta < 1.25^k = \frac{1}{|\mathcal{I}|} \sum_{p \in \mathcal{I}} \max\left(\frac{d}{d^*}, \frac{d^*}{d}\right) < 1.25^k \quad (5)$$

with  $d, d^*$  respectively estimated and ground truth depth maps,  $p$  a single pixel from input image  $\mathcal{I}$  and  $|\mathcal{I}|$  the total amount of pixels in  $\mathcal{I}$ . Tables 1, 2 and 3 exhaustively collect results on the Eigen test split [2] using the improved ground truth made available in [6], respectively when using monocular (M), stereo (S) or both (MS) (self-)supervisions. Since the ground truth is not provided for all 697 images, we reduce this split to 652 according to previous works [1, 5, 7].

Method	Sup	#Trn	#Par	#Fwd	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Monodepth2 [5]	M	1×	1×	1×	0.090	0.545	3.942	0.137	0.914	<b>0.983</b>	<b>0.995</b>
Monodepth2-Post [5]	M	1×	1×	2×	0.088	0.508	3.843	0.134	0.917	<b>0.983</b>	<b>0.995</b>
Monodepth2-Drop	M	1×	1×	N×	0.101	0.596	4.148	0.150	0.892	0.976	0.994
Monodepth2-Boot	M	N×	N×	1×	0.092	<b>0.505</b>	3.823	0.136	0.911	0.982	<b>0.995</b>
Monodepth2-Snap	M	1×	N×	1×	0.091	0.532	3.923	0.137	0.912	<b>0.983</b>	<b>0.995</b>
Monodepth2-Repr	M	1×	1×	1×	0.092	0.543	3.936	0.138	0.912	0.981	<b>0.995</b>
Monodepth2-Log	M	1×	1×	1×	0.091	0.588	4.053	0.139	0.911	0.980	<b>0.995</b>
Monodepth2-Self	M	(1+1)×	1×	1×	<b>0.087</b>	0.514	3.827	<b>0.133</b>	<b>0.920</b>	<b>0.983</b>	<b>0.995</b>
Monodepth2-Boot+Log	M	N×	N×	1×	0.092	0.509	3.852	0.137	0.910	0.982	<b>0.995</b>
Monodepth2-Boot+Self	M	(1+N)×	N×	1×	0.088	0.507	<b>3.800</b>	<b>0.133</b>	0.918	<b>0.983</b>	<b>0.995</b>
Monodepth2-Snap+Log	M	1×	1×	1×	0.092	0.564	3.961	0.139	0.911	0.981	0.994
Monodepth2-Snap+Self	M	(1+1)×	1×	1×	0.088	0.518	3.833	<b>0.133</b>	0.919	<b>0.983</b>	<b>0.995</b>

Table 1. **Depth evaluation for monocular (M) supervision.** Evaluation on the Eigen test split [2] with improved ground truth [6].

Method	Sup	#Trn	#Par	#Fwd	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Monodepth2 [5]	S	1×	1×	1×	0.085	0.537	3.942	0.139	0.912	0.979	0.993
Monodepth2-Post [5]	S	1×	1×	2×	<b>0.084</b>	<b>0.504</b>	3.777	0.137	<b>0.915</b>	0.980	<b>0.994</b>
Monodepth2-Drop	S	1×	1×	N×	0.129	0.791	4.908	0.187	0.819	0.959	0.990
Monodepth2-Boot	S	N×	N×	1×	0.085	0.511	<b>3.772</b>	0.137	0.914	0.980	<b>0.994</b>
Monodepth2-Snap	S	1×	N×	1×	0.085	0.535	3.849	0.139	0.912	0.980	0.993
Monodepth2-Repr	S	1×	1×	1×	0.085	0.532	3.873	0.140	0.913	0.979	0.993
Monodepth2-Log	S	1×	1×	1×	0.085	0.535	3.860	0.140	<b>0.915</b>	0.979	0.993
Monodepth2-Self	S	(1+1)×	1×	1×	<b>0.084</b>	0.524	3.835	0.137	<b>0.915</b>	0.980	0.993
Monodepth2-Boot+Log	S	N×	N×	1×	0.085	0.511	3.777	0.137	0.913	0.980	<b>0.994</b>
Monodepth2-Boot+Self	S	(1+N)×	N×	1×	0.085	0.510	3.792	<b>0.135</b>	0.914	<b>0.981</b>	<b>0.994</b>
Monodepth2-Snap+Log	S	1×	1×	1×	<b>0.084</b>	0.529	3.833	0.138	0.914	0.980	<b>0.994</b>
Monodepth2-Snap+Self	S	(1+1)×	1×	1×	0.086	0.532	3.858	0.138	0.912	0.980	<b>0.994</b>

Table 2. **Depth evaluation for stereo (S) supervision.** Evaluation on the Eigen test split [2] with improved ground truth [6].

Method	Sup	#Trn	#Par	#Fwd	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Monodepth2 [5]	MS	1×	1×	1×	0.084	0.494	3.739	0.132	0.918	0.983	<b>0.995</b>
Monodepth2-Post [5]	MS	1×	1×	2×	<b>0.082</b>	<b>0.470</b>	<b>3.666</b>	<b>0.129</b>	<b>0.919</b>	<b>0.984</b>	<b>0.995</b>
Monodepth2-Drop	MS	1×	1×	N×	0.172	1.074	5.886	0.237	0.679	0.933	0.982
Monodepth2-Boot	MS	N×	N×	1×	0.086	0.497	3.787	0.136	0.910	0.981	<b>0.995</b>
Monodepth2-Snap	MS	1×	N×	1×	0.085	0.504	3.803	0.134	0.914	0.983	<b>0.995</b>
Monodepth2-Repr	MS	1×	1×	1×	0.084	0.500	3.829	0.134	0.913	0.982	<b>0.995</b>
Monodepth2-Log	MS	1×	1×	1×	0.083	0.518	3.789	0.132	0.916	<b>0.984</b>	<b>0.995</b>
Monodepth2-Self	MS	(1+1)×	1×	1×	0.083	0.485	3.682	0.130	<b>0.919</b>	<b>0.984</b>	<b>0.995</b>
Monodepth2-Boot+Log	MS	N×	N×	1×	0.086	0.497	3.771	0.135	0.911	0.981	<b>0.995</b>
Monodepth2-Boot+Self	MS	(1+N)×	N×	1×	0.085	0.486	3.704	0.131	0.915	0.983	<b>0.995</b>
Monodepth2-Snap+Log	MS	1×	1×	1×	0.084	0.512	3.828	0.134	0.914	0.982	<b>0.995</b>
Monodepth2-Snap+Self	MS	(1+1)×	1×	1×	0.085	0.497	3.714	0.131	0.916	0.983	<b>0.995</b>

Table 3. **Depth evaluation for monocular+stereo (MS) supervision.** Evaluation on the Eigen test split [2] with improved ground truth [6].

### 3. Sparsification curves

Figures 1, 2 and 3 show Sparsification Error curves computed for all the three metrics evaluated in the submitted paper, *i.e.* Abs Rel, RMSE and  $\delta \geq 1.25$ , respectively for M, S and MS supervisions. The curves highlight consistent behaviour on each metric, confirming that Self-Teaching strategies (blue) outperform traditional log-likelihood maximization (green) on M and MS, with these latter ones yielding better results with S supervision.

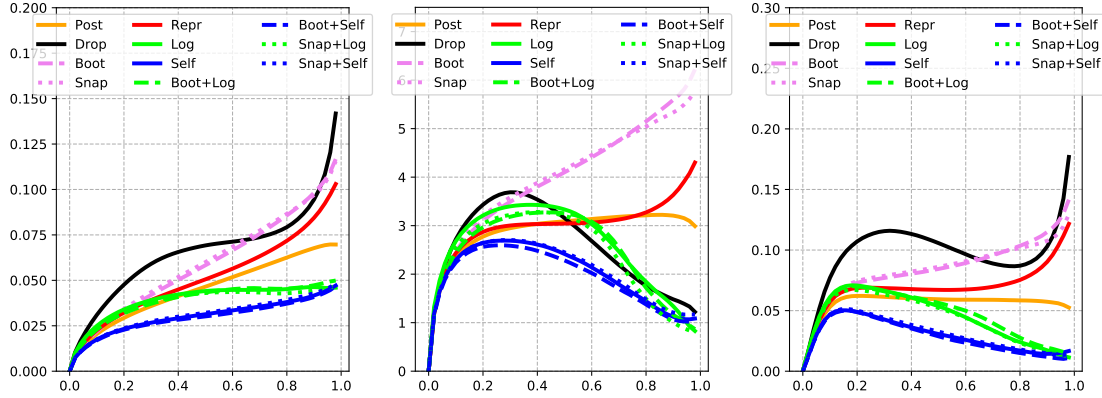


Figure 1. Sparsification Error curves for monocular (M) supervision. From left to right, Abs Rel, RMSE and  $\delta \geq 1.25$ .

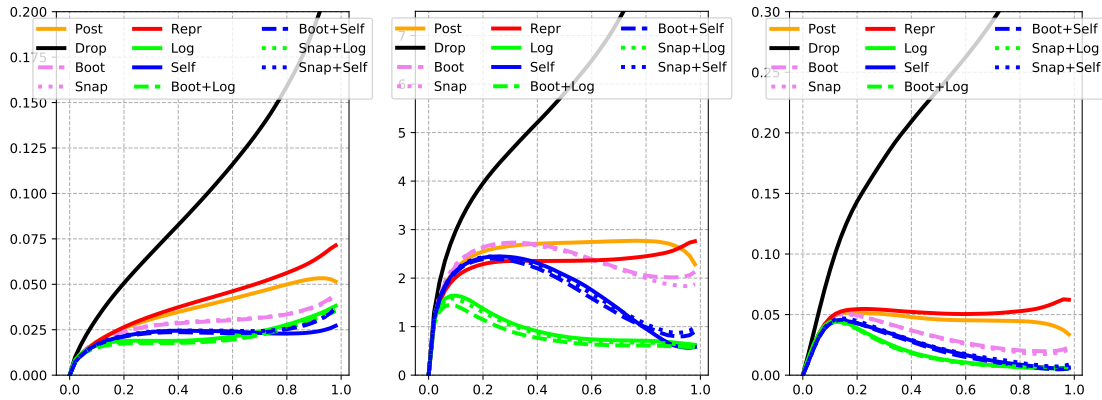


Figure 2. Sparsification Error curves for stereo (S) supervision. From left to right, Abs Rel, RMSE and  $\delta \geq 1.25$ .

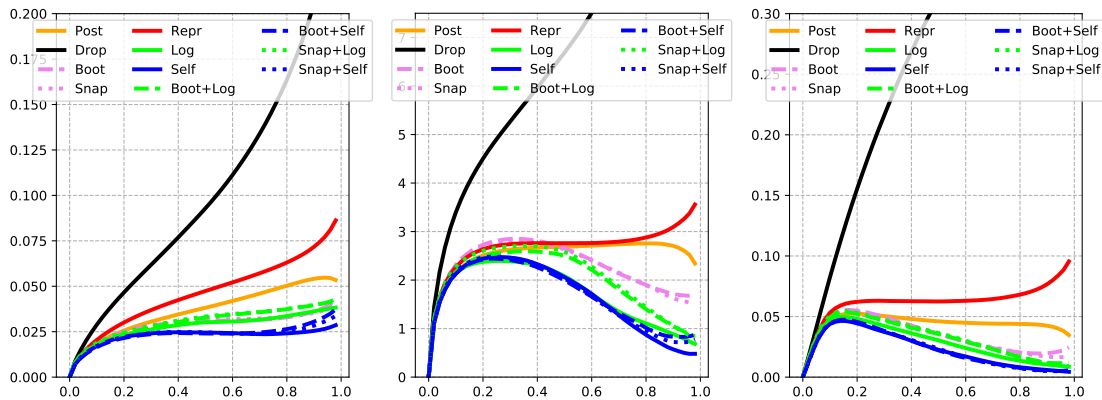


Figure 3. Sparsification Error curves for monocular+stereo (MS) supervision. From left to right, Abs Rel, RMSE and  $\delta \geq 1.25$ .

## 4. Depth evaluation – 50 meters cap

Consistently with previous works [3, 4], we also report results obtained by capping depth range to 50 meters. We can notice how, conversely to the evaluation carried out on raw LiDAR traditionally performed by existing works [3, 4], the margin between evaluating at 80 or 50 meters is much lower.

Method	Sup	#Trn	#Par	#Fwd	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Monodepth2 [5]	M	1×	1×	1×	0.090	0.502	3.979	0.137	0.911	0.983	0.995
Monodepth2-Post [5]	M	1×	1×	2×	0.088	<b>0.476</b>	3.914	0.134	0.914	<b>0.984</b>	<b>0.996</b>
Monodepth2-Drop	M	1×	1×	N×	0.101	0.580	4.239	0.151	0.889	0.976	0.994
Monodepth2-Boot	M	N×	N×	1×	0.092	0.494	3.958	0.138	0.907	0.982	<b>0.996</b>
Monodepth2-Snap	M	1×	N×	1×	0.091	0.502	3.997	0.137	0.909	0.983	<b>0.996</b>
Monodepth2-Repr	M	1×	1×	1×	0.091	0.504	3.991	0.138	0.909	0.982	0.995
Monodepth2-Log	M	1×	1×	1×	0.091	0.588	4.053	0.139	0.911	0.980	0.995
Monodepth2-Self	M	(1+1)×	1×	1×	<b>0.086</b>	0.477	3.898	<b>0.133</b>	<b>0.916</b>	0.983	0.995
Monodepth2-Boot+Log	M	N×	N×	1×	0.092	0.495	3.970	0.138	0.907	0.982	0.995
Monodepth2-Boot+Self	M	(1+N)×	N×	1×	0.088	0.477	<b>3.885</b>	<b>0.133</b>	0.914	0.983	<b>0.996</b>
Monodepth2-Snap+Log	M	1×	1×	1×	0.091	0.526	4.011	0.139	0.908	0.981	0.994
Monodepth2-Snap+Self	M	(1+1)×	1×	1×	0.088	0.478	3.890	<b>0.133</b>	0.915	0.983	<b>0.996</b>

Table 4. **Depth evaluation for monocular (M) supervision.** Evaluation on the Eigen test split [2] with improved ground truth [6]. Maximum depth reduced to 50 meters.

Method	Sup	#Trn	#Par	#Fwd	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Monodepth2 [5]	S	1×	1×	1×	0.084	0.491	3.907	0.139	0.908	0.980	0.994
Monodepth2-Post [5]	S	1×	1×	2×	<b>0.083</b>	<b>0.470</b>	<b>3.848</b>	0.137	<b>0.911</b>	<b>0.981</b>	0.994
Monodepth2-Drop	S	1×	1×	N×	0.129	0.777	4.960	0.187	0.817	0.960	0.990
Monodepth2-Boot	S	N×	N×	1×	0.084	0.481	3.869	0.137	0.910	<b>0.981</b>	<b>0.995</b>
Monodepth2-Snap	S	1×	N×	1×	0.085	0.491	3.901	0.139	0.908	0.980	0.994
Monodepth2-Repr	S	1×	1×	1×	0.084	0.491	3.926	0.140	0.909	0.980	0.994
Monodepth2-Log	S	1×	1×	1×	0.084	0.494	3.906	0.140	<b>0.911</b>	0.980	0.994
Monodepth2-Self	S	(1+1)×	1×	1×	<b>0.083</b>	0.475	3.854	0.137	<b>0.911</b>	0.980	0.994
Monodepth2-Boot+Log	S	N×	N×	1×	0.084	0.483	3.873	0.137	0.909	<b>0.981</b>	<b>0.995</b>
Monodepth2-Boot+Self	S	(1+N)×	N×	1×	0.084	0.472	3.852	<b>0.136</b>	0.910	<b>0.981</b>	<b>0.995</b>
Monodepth2-Snap+Log	S	1×	1×	1×	0.084	0.488	3.894	0.138	<b>0.911</b>	<b>0.981</b>	0.994
Monodepth2-Snap+Self	S	(1+1)×	1×	1×	0.085	0.490	3.899	0.138	0.908	<b>0.981</b>	0.994

Table 5. **Depth evaluation for stereo (S) supervision.** Evaluation on the Eigen test split [2] with improved ground truth [6]. Maximum depth reduced to 50 meters.

Method	Sup	#Trn	#Par	#Fwd	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Monodepth2 [5]	MS	1×	1×	1×	0.083	0.461	3.830	0.132	0.914	<b>0.984</b>	<b>0.996</b>
Monodepth2-Post [5]	MS	1×	1×	2×	<b>0.082</b>	<b>0.445</b>	3.790	<b>0.130</b>	<b>0.915</b>	<b>0.984</b>	<b>0.996</b>
Monodepth2-Drop	MS	1×	1×	N×	0.172	1.074	5.921	0.237	0.678	0.933	0.982
Monodepth2-Boot	MS	N×	N×	1×	0.086	0.485	3.925	0.137	0.906	0.981	0.995
Monodepth2-Snap	MS	1×	N×	1×	0.085	0.476	3.899	0.135	0.910	0.983	<b>0.996</b>
Monodepth2-Repr	MS	1×	1×	1×	0.084	0.470	3.905	0.134	0.909	0.983	0.995
Monodepth2-Log	MS	1×	1×	1×	0.083	0.471	3.832	0.133	0.912	<b>0.984</b>	<b>0.996</b>
Monodepth2-Self	MS	(1+1)×	1×	1×	0.083	0.455	<b>3.781</b>	<b>0.130</b>	<b>0.915</b>	<b>0.984</b>	<b>0.996</b>
Monodepth2-Boot+Log	MS	N×	N×	1×	0.086	0.481	3.903	0.136	0.907	0.981	0.995
Monodepth2-Boot+Self	MS	(1+N)×	N×	1×	0.085	0.462	3.815	0.132	0.911	0.983	<b>0.996</b>
Monodepth2-Snap+Log	MS	1×	1×	1×	0.084	0.481	3.900	0.134	0.911	0.982	<b>0.996</b>
Monodepth2-Snap+Self	MS	(1+1)×	1×	1×	0.084	0.467	3.810	0.132	0.912	0.983	<b>0.996</b>

Table 6. **Depth evaluation for monocular+stereo (MS) supervision.** Evaluation on the Eigen test split [2] with improved ground truth [6]. Maximum depth reduced to 50 meters.

## 5. Uncertainty evaluation – 50 meters cap

To complete the experiments from the previous section, we also evaluate uncertainty modelling by assuming a maximum depth of 50 meters. Tables 7, 8 and 9 collect the results of this evaluation for M, S and MS supervisions. Compared with Tables 1, 2 and 3 from the main paper, we highlight that the same behaviour occurs regardless of the maximum depth set to 80 or 50 meters.

Method	Abs Rel		RMSE		$\delta \geq 1.25$	
	AUSE	AURG	AUSE	AURG	AUSE	AURG
Monodepth2-Post	0.044	0.012	2.967	0.378	0.059	0.021
Monodepth2-Drop	0.065	0.000	2.621	0.980	0.098	0.003
Monodepth2-Boot	0.059	-0.000	4.259	-0.890	0.094	-0.008
Monodepth2-Snap	0.059	-0.001	4.159	-0.748	0.091	-0.007
Monodepth2-Repr	0.051	0.007	3.085	0.321	0.072	0.012
Monodepth2-Log	0.038	0.020	2.547	0.908	0.047	0.039
Monodepth2-Self	0.030	0.026	2.136	1.207	0.033	0.045
Monodepth2-Boot+Log	0.038	0.020	2.605	0.778	0.050	0.036
Monodepth2-Boot+Self	<b>0.029</b>	<b>0.028</b>	<b>2.053</b>	<b>1.267</b>	<b>0.031</b>	<b>0.049</b>
Monodepth2-Snap+Log	0.037	0.022	2.482	0.949	0.046	0.039
Monodepth2-Snap+Self	0.030	0.026	2.160	1.166	0.034	0.045

Table 7. **Quantitative results for monocular (M) supervision.** Evaluation on the Eigen test split [2] with improved ground truth [6]. Maximum depth cap to 50 meters.

Method	Abs Rel		RMSE		$\delta \geq 1.25$	
	AUSE	AURG	AUSE	AURG	AUSE	AURG
Monodepth2-Post	0.036	0.019	2.662	0.663	0.048	0.033
Monodepth2-Drop	0.102	-0.029	6.276	-2.234	0.240	-0.086
Monodepth2-Boot	0.028	0.029	2.555	0.790	0.037	0.046
Monodepth2-Snap	0.028	0.029	2.450	0.925	0.035	0.049
Monodepth2-Repr	0.040	0.016	2.386	1.011	0.052	0.031
Monodepth2-Log	0.022	0.035	0.980	2.401	0.019	0.063
Monodepth2-Self	0.022	0.034	1.858	1.479	0.026	0.055
Monodepth2-Boot+Log	<b>0.020</b>	<b>0.037</b>	<b>0.847</b>	<b>2.504</b>	<b>0.018</b>	<b>0.066</b>
Monodepth2-Boot+Self	0.023	0.034	1.795	1.533	0.025	0.058
Monodepth2-Snap+Log	0.021	0.036	0.929	2.443	0.019	0.064
Monodepth2-Snap+Self	0.023	0.034	1.863	1.506	0.027	0.057

Table 8. **Quantitative results for stereo (S) supervision.** Evaluation on the Eigen test split [2] with improved ground truth [6]. Maximum depth cap to 50 meters.

Method	Abs Rel		RMSE		$\delta \geq 1.25$	
	AUSE	AURG	AUSE	AURG	AUSE	AURG
Monodepth2-Post	0.036	0.018	2.650	0.621	0.048	0.031
Monodepth2-Drop	0.103	-0.027	7.188	-2.621	0.307	-0.085
Monodepth2-Boot	0.029	0.029	2.470	0.914	0.038	0.049
Monodepth2-Snap	0.028	0.028	2.413	0.947	0.037	0.047
Monodepth2-Repr	0.046	0.010	2.781	0.587	0.065	0.018
Monodepth2-Log	0.028	0.028	<b>1.818</b>	<b>1.495</b>	0.030	0.051
Monodepth2-Self	<b>0.023</b>	<b>0.033</b>	1.870	1.391	<b>0.027</b>	0.052
Monodepth2-Boot+Log	0.030	0.028	2.112	1.256	0.035	0.052
Monodepth2-Boot+Self	<b>0.023</b>	<b>0.033</b>	1.880	1.407	<b>0.027</b>	<b>0.056</b>
Monodepth2-Snap+Log	0.030	0.026	2.152	1.218	0.035	0.048
Monodepth2-Snap+Self	<b>0.023</b>	<b>0.033</b>	1.880	1.403	<b>0.027</b>	0.054

Table 9. **Quantitative results for monocular+stereo (MS) supervision.** Evaluation on the Eigen test split [2] with improved ground truth [6]. Maximum depth cap to 50 meters.

## 6. Sparsification curves – 50 meters cap

To conclude the evaluation at 50 meters, we report sparsification curves. Figures 4, 5 and 6 confirm that for M, S and MS the same behavior observed for the 80 meters evaluation is kept.

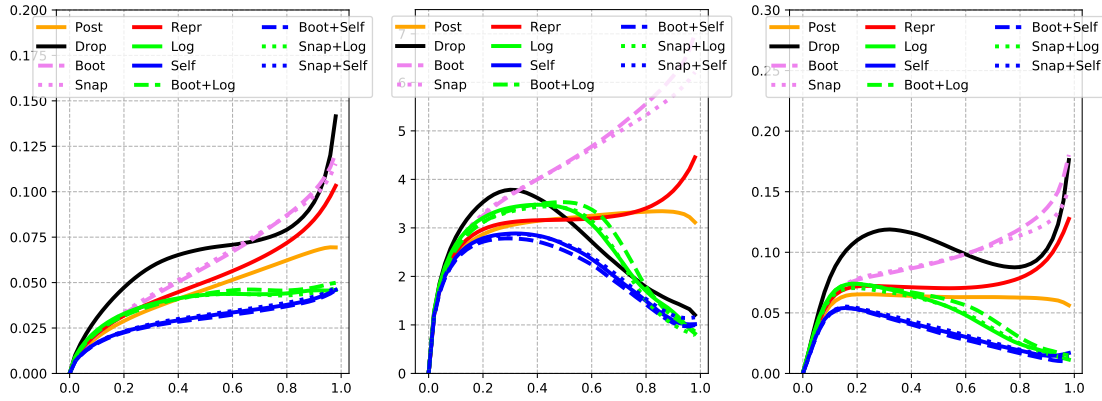


Figure 4. Sparsification Error curves for monocular (M) supervision. From left to right, Abs Rel, RMSE and  $\delta \geq 1.25$ .

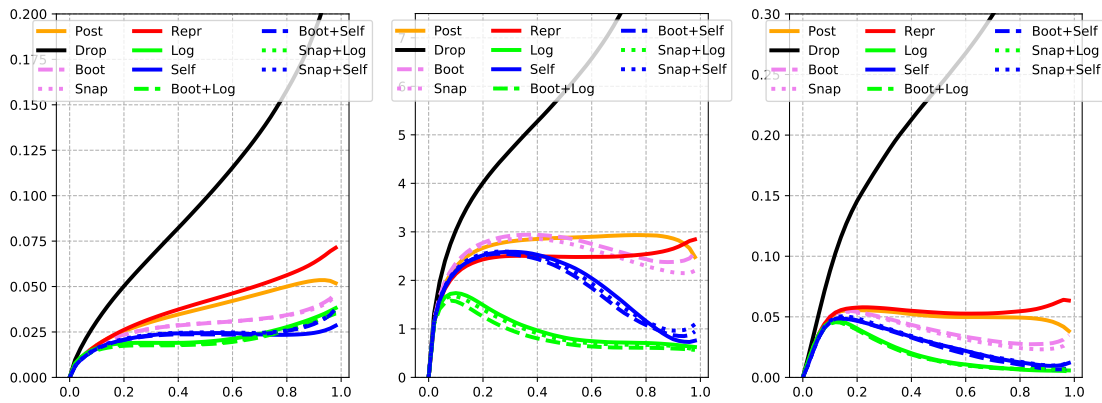


Figure 5. Sparsification Error curves for stereo (S) supervision. From left to right, Abs Rel, RMSE and  $\delta \geq 1.25$ .

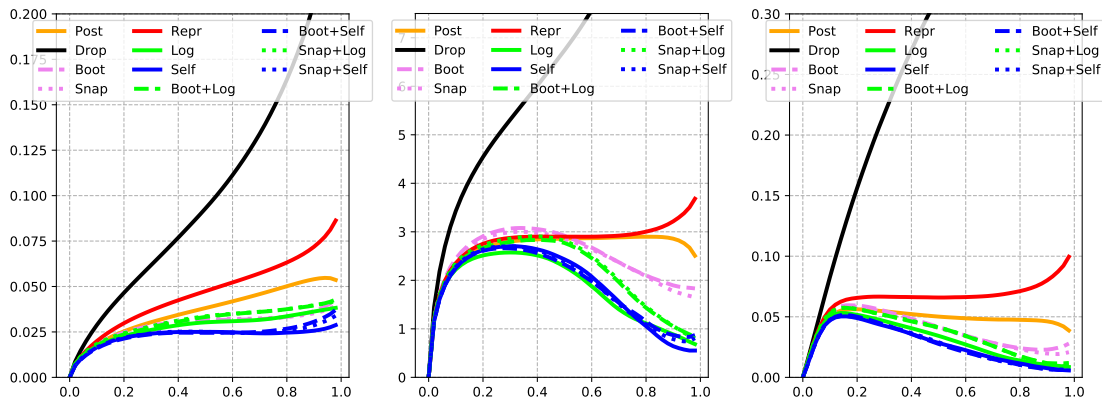


Figure 6. Sparsification Error curves for monocular+stereo (MS) supervision. From left to right, Abs Rel, RMSE and  $\delta \geq 1.25$ .

## 7. Depth evaluation – raw LiDAR (80 meters)

To ease comparison with previous works in literature [5], we also report the same evaluation carried out in the main paper by assuming the raw LiDAR depth measurements as ground truth. Tables 10, 11 and 12 collect the outcome of this evaluation.

Method	Sup	#Trn	#Par	#Fwd	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Monodepth2 [5]	M	1×	1×	1×	0.115	0.903	4.863	0.193	0.877	0.959	0.981
Monodepth2-Post [5]	M	1×	1×	2×	0.112	0.852	4.755	0.190	<b>0.881</b>	0.960	0.981
Monodepth2-Drop	M	1×	1×	N×	0.126	0.895	4.911	0.198	0.850	0.952	0.982
Monodepth2-Boot	M	N×	N×	1×	0.114	0.803	<b>4.639</b>	0.187	0.875	<b>0.961</b>	<b>0.983</b>
Monodepth2-Snap	M	1×	N×	1×	0.114	0.865	4.787	0.190	0.877	0.960	0.982
Monodepth2-Repr	M	1×	1×	1×	0.116	0.914	4.853	0.193	0.875	0.958	0.981
Monodepth2-Log	M	1×	1×	1×	0.113	0.928	4.919	0.192	0.876	0.958	0.981
Monodepth2-Self	M	(1+1)×	1×	1×	<b>0.111</b>	0.863	4.756	0.188	<b>0.881</b>	<b>0.961</b>	0.982
Monodepth2-Boot+Log	M	N×	N×	1×	0.114	<b>0.797</b>	4.661	0.186	0.872	0.960	<b>0.983</b>
Monodepth2-Boot+Self	M	(1+N)×	N×	1×	<b>0.111</b>	0.826	4.667	<b>0.184</b>	0.880	<b>0.961</b>	<b>0.983</b>
Monodepth2-Snap+Log	M	1×	1×	1×	0.117	0.900	4.838	0.192	0.873	0.958	0.981
Monodepth2-Snap+Self	M	(1+1)×	1×	1×	0.112	0.871	4.747	0.187	0.880	<b>0.961</b>	0.982

Table 10. **Depth evaluation for monocular (M) supervision.** Evaluation on the Eigen test split [2] with raw LiDAR.

Method	Sup	#Trn	#Par	#Fwd	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Monodepth2 [5]	S	1×	1×	1×	0.109	0.873	4.960	0.209	0.864	0.948	0.975
Monodepth2-Post [5]	S	1×	1×	2×	0.108	0.842	4.892	0.207	0.866	0.949	0.976
Monodepth2-Drop	S	1×	1×	N×	0.151	1.110	5.780	0.244	0.764	0.926	0.970
Monodepth2-Boot	S	N×	N×	1×	0.108	0.822	4.809	0.201	0.866	0.951	0.977
Monodepth2-Snap	S	1×	N×	1×	0.109	0.868	4.918	0.206	0.864	0.949	0.976
Monodepth2-Repr	S	1×	1×	1×	0.109	0.876	4.975	0.210	0.862	0.948	0.975
Monodepth2-Log	S	1×	1×	1×	0.110	0.876	4.952	0.209	0.865	0.948	0.975
Monodepth2-Self	S	(1+1)×	1×	1×	0.109	0.858	4.920	0.206	<b>0.867</b>	0.949	0.976
Monodepth2-Boot+Log	S	N×	N×	1×	<b>0.107</b>	0.811	<b>4.796</b>	0.200	0.866	<b>0.952</b>	<b>0.978</b>
Monodepth2-Boot+Self	S	(1+N)×	N×	1×	<b>0.107</b>	<b>0.806</b>	4.798	<b>0.199</b>	0.866	<b>0.952</b>	<b>0.978</b>
Monodepth2-Snap+Log	S	1×	1×	1×	0.108	0.851	4.894	0.204	<b>0.867</b>	0.951	0.976
Monodepth2-Snap+Self	S	(1+1)×	1×	1×	0.109	0.848	4.895	0.204	0.864	0.950	0.977

Table 11. **Depth evaluation for stereo (S) supervision.** Evaluation on the Eigen test split [2] with raw LiDAR.

Method	Sup	#Trn	#Par	#Fwd	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Monodepth2 [5]	MS	1×	1×	1×	0.106	0.818	4.750	0.196	0.874	0.957	0.979
Monodepth2-Post [5]	MS	1×	1×	2×	<b>0.104</b>	0.787	4.687	0.194	<b>0.876</b>	<b>0.958</b>	0.980
Monodepth2-Drop	MS	1×	1×	N×	0.201	1.421	6.704	0.295	0.593	0.896	0.962
Monodepth2-Boot	MS	N×	N×	1×	0.109	0.787	4.747	0.195	0.866	0.956	0.980
Monodepth2-Snap	MS	1×	N×	1×	0.109	0.828	4.815	0.198	0.869	0.956	0.979
Monodepth2-Repr	MS	1×	1×	1×	0.109	0.820	4.830	0.199	0.869	0.955	0.979
Monodepth2-Log	MS	1×	1×	1×	0.107	0.839	4.792	0.197	0.873	0.956	0.979
Monodepth2-Self	MS	(1+1)×	1×	1×	<b>0.104</b>	0.797	4.686	0.192	<b>0.876</b>	0.957	0.980
Monodepth2-Boot+Log	MS	N×	N×	1×	0.108	0.784	4.735	0.194	0.866	0.955	0.980
Monodepth2-Boot+Self	MS	(1+N)×	N×	1×	0.105	<b>0.766</b>	<b>4.638</b>	<b>0.189</b>	0.873	<b>0.958</b>	<b>0.982</b>
Monodepth2-Snap+Log	MS	1×	1×	1×	0.108	0.824	4.821	0.196	0.870	0.956	0.980
Monodepth2-Snap+Self	MS	(1+1)×	1×	1×	0.105	0.795	4.682	0.191	0.875	0.957	0.981

Table 12. **Depth evaluation for monocular+stereo (MS) supervision.** Evaluation on the Eigen test split [2] with raw LiDAR.

## 8. Uncertainty evaluation – raw LiDAR (80 meters)

We report the evaluation of uncertainty modelling adopting the raw LiDAR as well. Tables 13, 14 and 15 resume the outcome showing how the same behaviour occurs, *i.e.* *Self* solutions are much better when dealing with M and MS supervisions, while *Log* outperforms *Self* approach training with S.

Method	Abs Rel		RMSE		$\delta \geq 1.25$	
	AUSE	AURG	AUSE	AURG	AUSE	AURG
Monodepth2-Post	0.053	0.020	3.322	0.734	0.069	0.040
Monodepth2-Drop	0.083	0.001	3.044	1.142	0.133	0.001
Monodepth2-Boot	0.064	0.010	4.266	-0.334	0.091	0.023
Monodepth2-Snap	0.068	0.007	4.391	-0.315	0.095	0.018
Monodepth2-Repr	0.058	0.019	3.282	0.855	0.080	0.034
Monodepth2-Log	0.051	0.027	3.097	1.188	0.060	0.056
Monodepth2-Self	0.036	0.038	2.292	1.779	0.037	0.072
Monodepth2-Boot+Log	0.046	0.028	2.830	1.119	0.060	0.057
Monodepth2-Boot+Self	<b>0.033</b>	<b>0.040</b>	<b>2.124</b>	<b>1.857</b>	<b>0.033</b>	<b>0.077</b>
Monodepth2-Snap+Log	0.047	0.030	2.837	1.281	0.058	0.059
Monodepth2-Snap+Self	0.036	0.038	2.331	1.725	0.036	0.073

Table 13. **Quantitative results for monocular (M) supervision.** Evaluation on the Eigen test split [2] with raw LiDAR.

Method	Abs Rel		RMSE		$\delta \geq 1.25$	
	AUSE	AURG	AUSE	AURG	AUSE	AURG
Monodepth2-Post	0.044	0.029	3.104	1.119	0.063	0.057
Monodepth2-Drop	0.114	-0.027	6.821	-2.088	0.266	-0.071
Monodepth2-Boot	0.032	0.041	2.640	1.504	0.041	0.080
Monodepth2-Snap	0.034	0.041	2.654	1.589	0.042	0.080
Monodepth2-Repr	0.045	0.029	2.647	1.650	0.069	0.055
Monodepth2-Log	0.026	0.048	1.144	3.126	0.031	0.091
Monodepth2-Self	0.026	0.048	1.931	2.321	0.029	0.091
Monodepth2-Boot+Log	<b>0.024</b>	<b>0.049</b>	<b>0.988</b>	<b>3.151</b>	0.028	<b>0.093</b>
Monodepth2-Boot+Self	0.025	0.048	1.813	2.326	<b>0.027</b>	<b>0.093</b>
Monodepth2-Snap+Log	0.025	<b>0.049</b>	1.082	3.148	0.029	0.092
Monodepth2-Snap+Self	0.026	0.048	1.937	2.286	0.030	0.092

Table 14. **Quantitative results for stereo (S) supervision.** Evaluation on the Eigen test split [2] with raw LiDAR.

Method	Abs Rel		RMSE		$\delta \geq 1.25$	
	AUSE	AURG	AUSE	AURG	AUSE	AURG
Monodepth2-Post	0.043	0.028	3.021	1.024	0.060	0.053
Monodepth2-Drop	0.110	-0.022	7.450	-2.370	0.321	-0.065
Monodepth2-Boot	0.032	0.040	2.590	1.477	0.043	0.078
Monodepth2-Snap	0.034	0.039	2.654	1.482	0.044	0.075
Monodepth2-Repr	0.051	0.021	3.026	1.120	0.078	0.039
Monodepth2-Log	0.032	0.040	2.093	2.039	0.040	0.075
Monodepth2-Self	<b>0.025</b>	0.046	1.896	<b>2.153</b>	0.029	0.083
Monodepth2-Boot+Log	0.034	0.039	2.331	1.735	0.045	0.076
Monodepth2-Boot+Self	<b>0.025</b>	<b>0.047</b>	<b>1.859</b>	2.131	<b>0.028</b>	<b>0.087</b>
Monodepth2-Snap+Log	0.035	0.037	2.431	1.719	0.046	0.072
Monodepth2-Snap+Self	<b>0.025</b>	<b>0.047</b>	1.903	2.137	0.029	0.085

Table 15. **Quantitative results for monocular+stereo (MS) supervision.** Evaluation on the Eigen test split [2] with raw LiDAR.



## 9. Sparsification curves – raw LiDAR (80 meters)

We also report Sparsification Error curves to perceive the behaviour of the modelled uncertainties better. Figures 7, 8 and 9 highlight once more how the variants based on *Self* outperforms *Log* ones on M and MS, with these latter yielding better results on S in particular when considering RMSE sparsification.

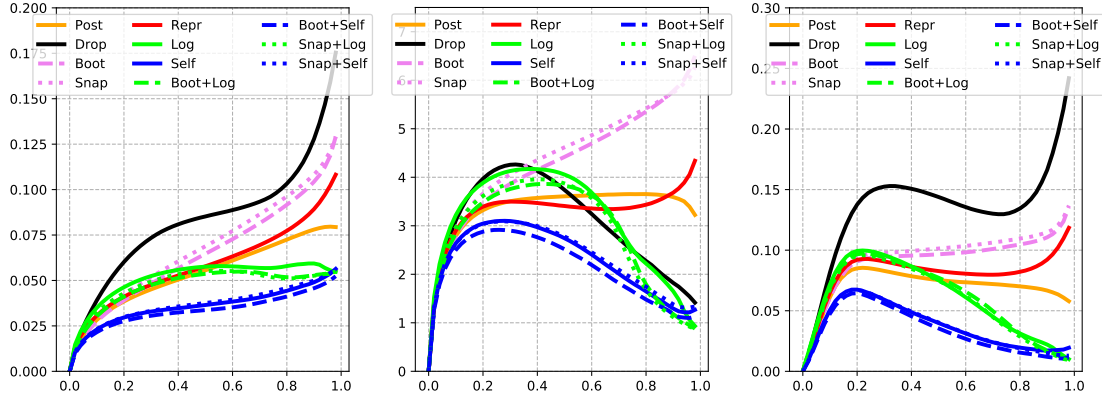


Figure 7. Sparsification Error curves for monocular (M) supervision. From left to right, Abs Rel, RMSE and  $\delta \geq 1.25$ .

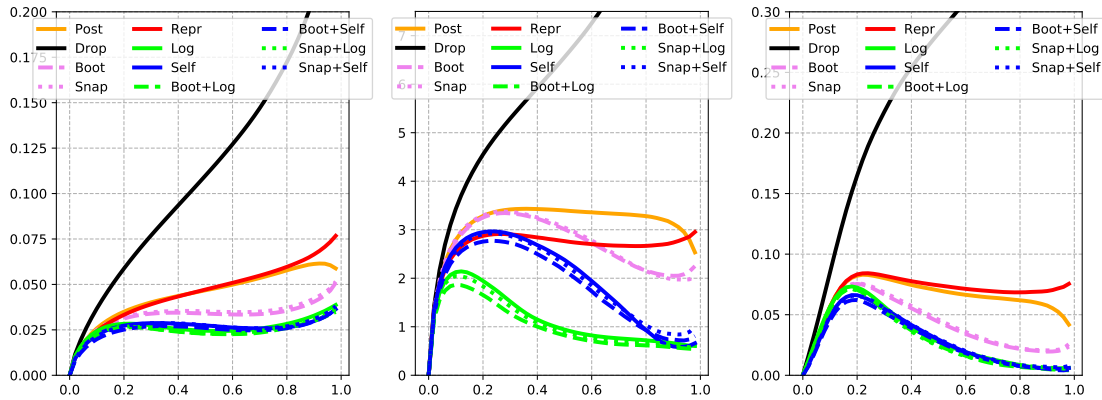


Figure 8. Sparsification Error curves for stereo (S) supervision. From left to right, Abs Rel, RMSE and  $\delta \geq 1.25$ .

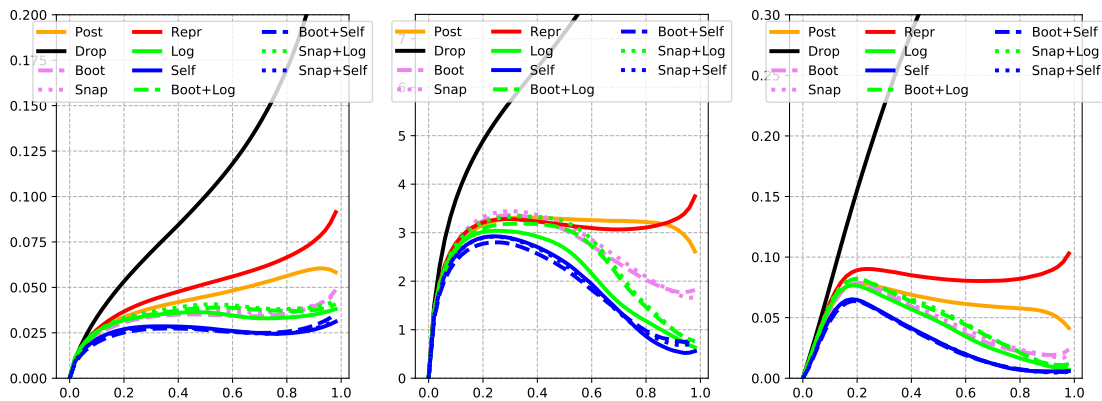


Figure 9. Sparsification Error curves for monocular+stereo (MS) supervision. From left to right, Abs Rel, RMSE and  $\delta \geq 1.25$ .

## 10. Qualitative results

Finally, we report some qualitative examples of both depth and uncertainty maps obtained by the different methods evaluated in the paper. Given the high amount of images produced by all the considered variants, at first, we introduce the notation to ease readability.

### 10.1. Colormap encodings

To show qualitative examples obtained by our framework, we adopt colormap **magma** for depth maps and colormap **hot** for uncertainty. Figure 10 shows the adopted colormaps and how they range from far to close depth and from low to high uncertainty.



Figure 10. **Colormap encodings for depth and uncertainty.** We choose colormap **magma** (on left) to encode depth maps and colormap **hot** (on right) for uncertainty. Best viewed with colors.

### 10.2. Results topology

In our paper, we evaluated eleven different strategies to obtain depth and corresponding uncertainty maps. Thus, we report both outcomes for each of the considered variants, organized as shown in Figure 11.

Monodepth2	Reference Image	Post	Post
Drop	Drop	Boot	Boot
Snap	Snap	Repr	Repr
Log	Log	Self	Self
Boot+Log	Boot+Log	Boot+Self	Boot+Self
Snap+Log	Snap+Log	Snap+Self	Snap+Self

Figure 11. **Legend for qualitative results.** Each cell in the table shows what each of the qualitative figure reported in the reminder represents.

We will show results on three images taken from the Eigen test split [2], respectively **2011\_09\_26\_drive\_0002\_sync/0000000021**, **2011\_09\_26\_drive\_0013\_sync/0000000045** and **2011\_09\_26\_drive\_0101\_sync/0000000114**. For each one, we report results for the network trained with monocular (M), stereo (S) or both (MS) supervision strategies.

### 10.3. Image 2011\_09\_26\_drive\_0002\_sync/0000000021

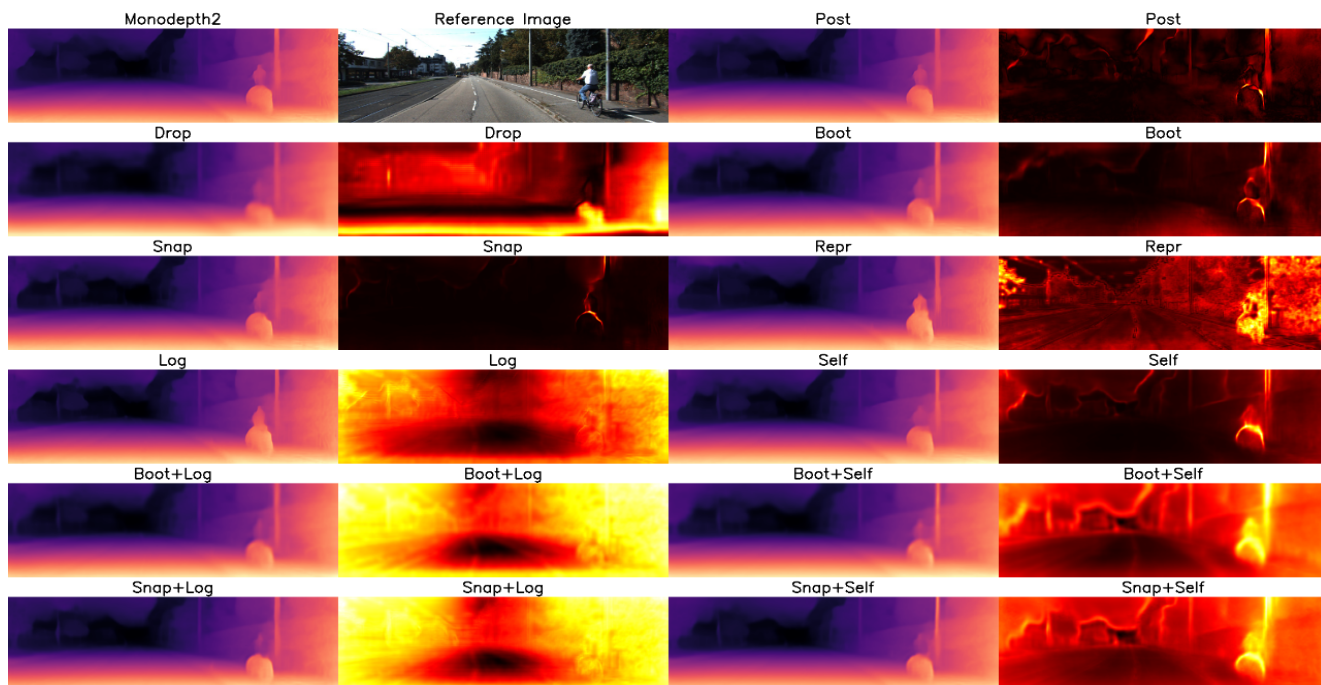


Figure 12. **Qualitative results on image 2011\_09\_26\_drive\_0002\_sync/0000000021 from the Eigen test split [2].** Each row shows depth and uncertainty maps from one of the considered variants trained with monocular (M) supervision.

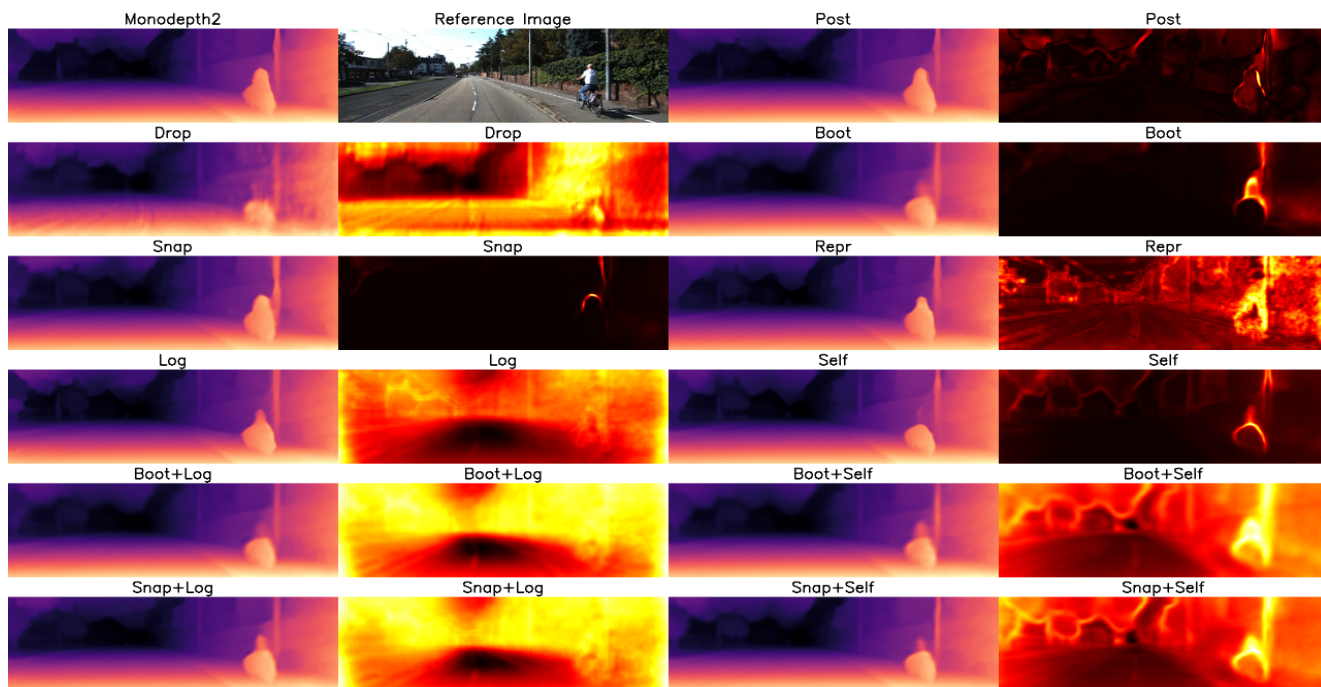


Figure 13. **Qualitative results on image 2011\_09\_26\_drive\_0002\_sync/0000000021 from the Eigen test split [2].** Each row shows depth and uncertainty maps from one of the considered variants trained with stereo (S) supervision.

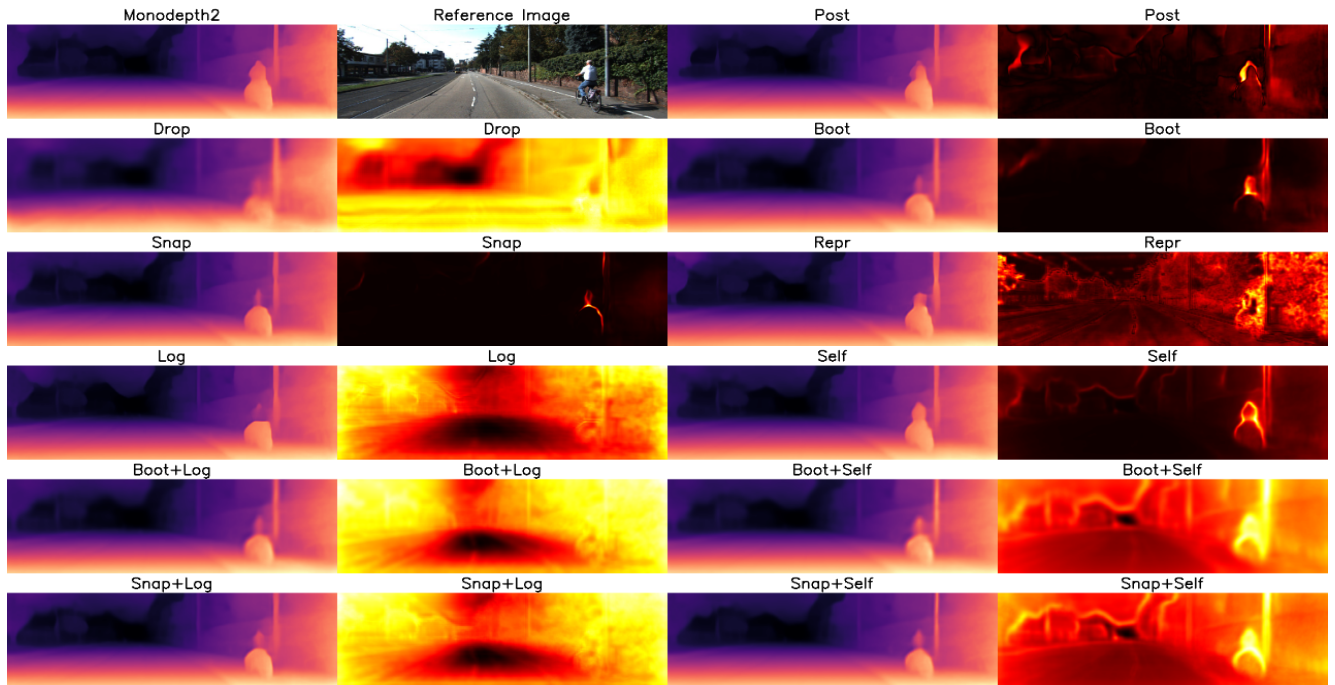


Figure 14. Qualitative results on image 2011\_09\_26\_drive\_0002\_sync/000000021 from the Eigen test split [2]. Each row shows depth and uncertainty maps from one of the considered variants trained with mono+stereo (MS) supervision.

#### 10.4. Image 2011\_09\_26\_drive\_0013\_sync/0000000045

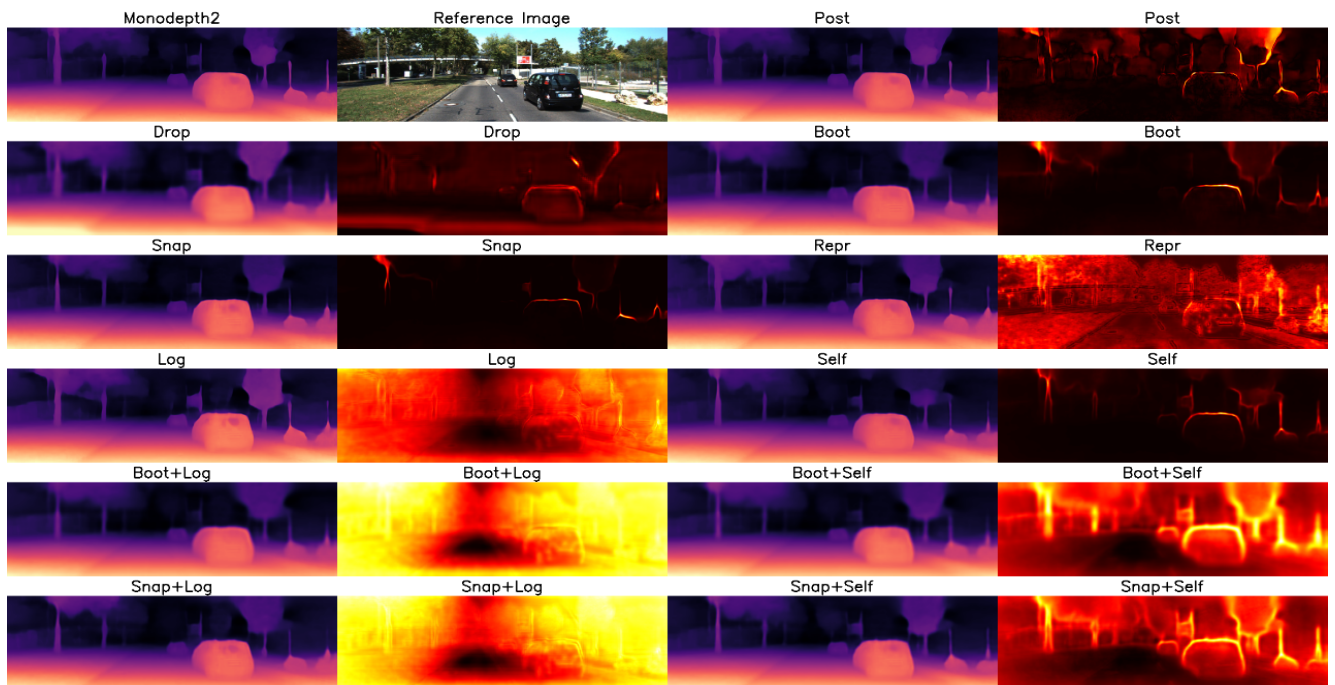


Figure 15. Qualitative results on image 2011\_09\_26\_drive\_0013\_sync/0000000045 from the Eigen test split [2]. Each row shows depth and uncertainty maps from one of the considered variants trained with monocular (M) supervision.

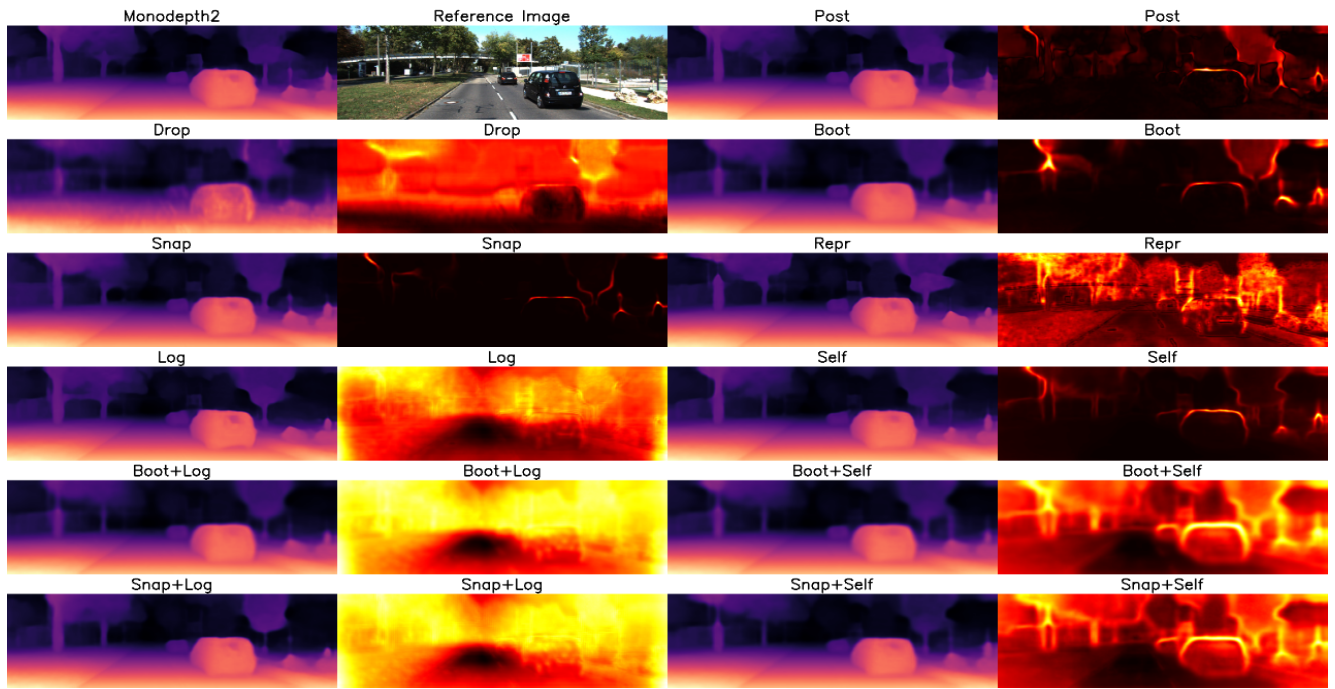


Figure 16. **Qualitative results on image 2011\_09\_26\_drive\_0013\_sync/0000000045 from the Eigen test split [2].** Each row shows depth and uncertainty maps from one of the considered variants trained with stereo (S) supervision.

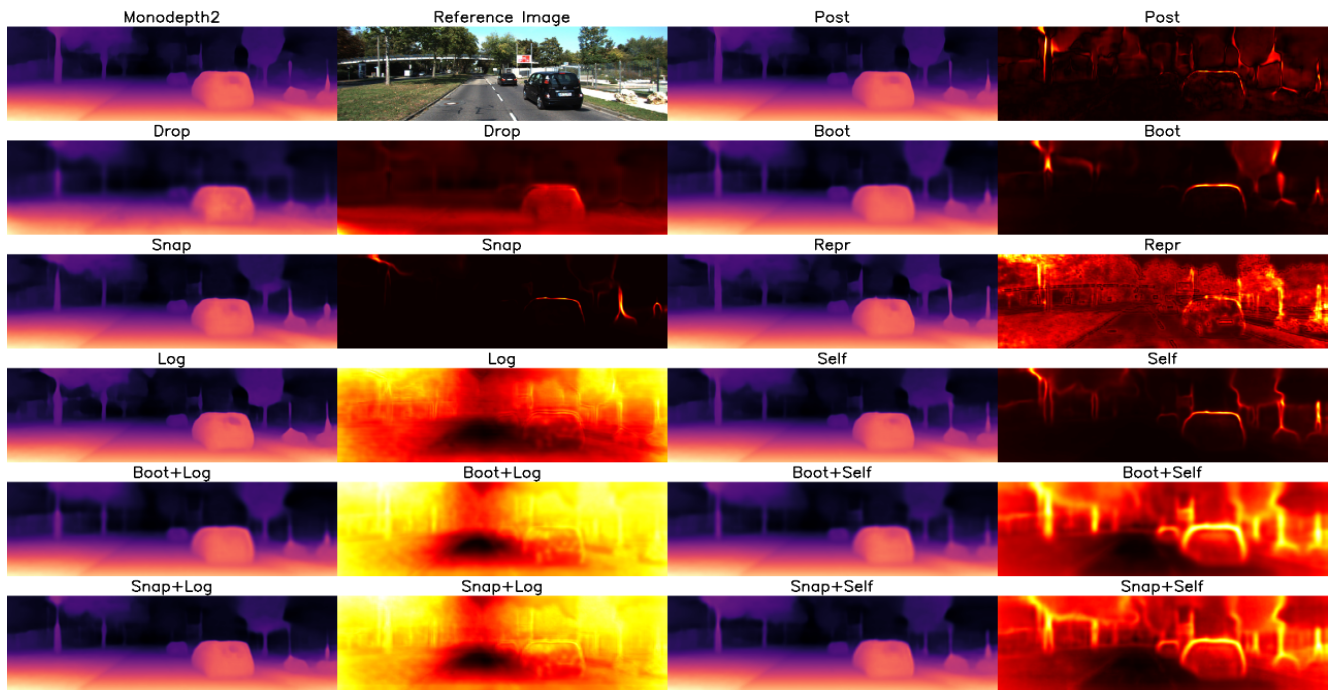


Figure 17. **Qualitative results on image 2011\_09\_26\_drive\_0013\_sync/0000000045 from the Eigen test split [2].** Each row shows depth and uncertainty maps from one of the considered variants trained with mono+stereo (MS) supervision.

### 10.5. Image 2011\_09\_26\_drive\_0101\_sync/0000000114

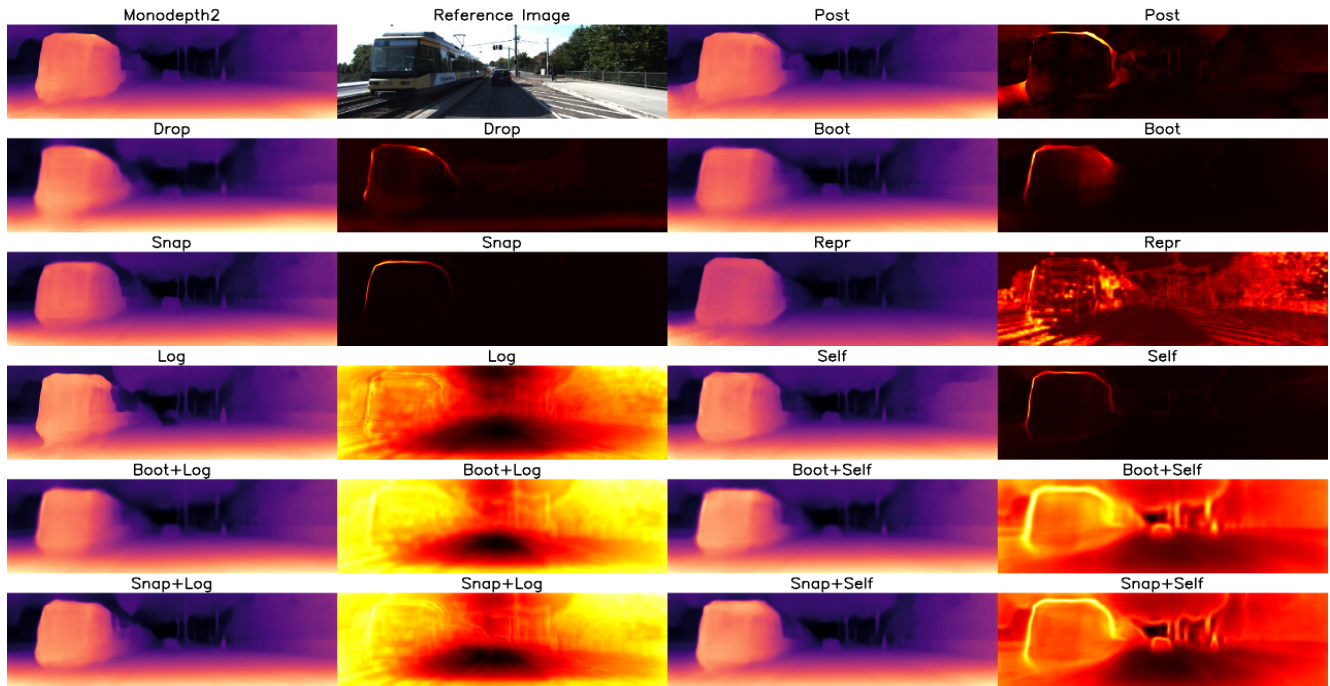


Figure 18. Qualitative results on image 2011\_09\_26\_drive\_0101\_sync/0000000114 from the Eigen test split [2]. Each row shows depth and uncertainty maps from one of the considered variants trained with monocular (M) supervision.

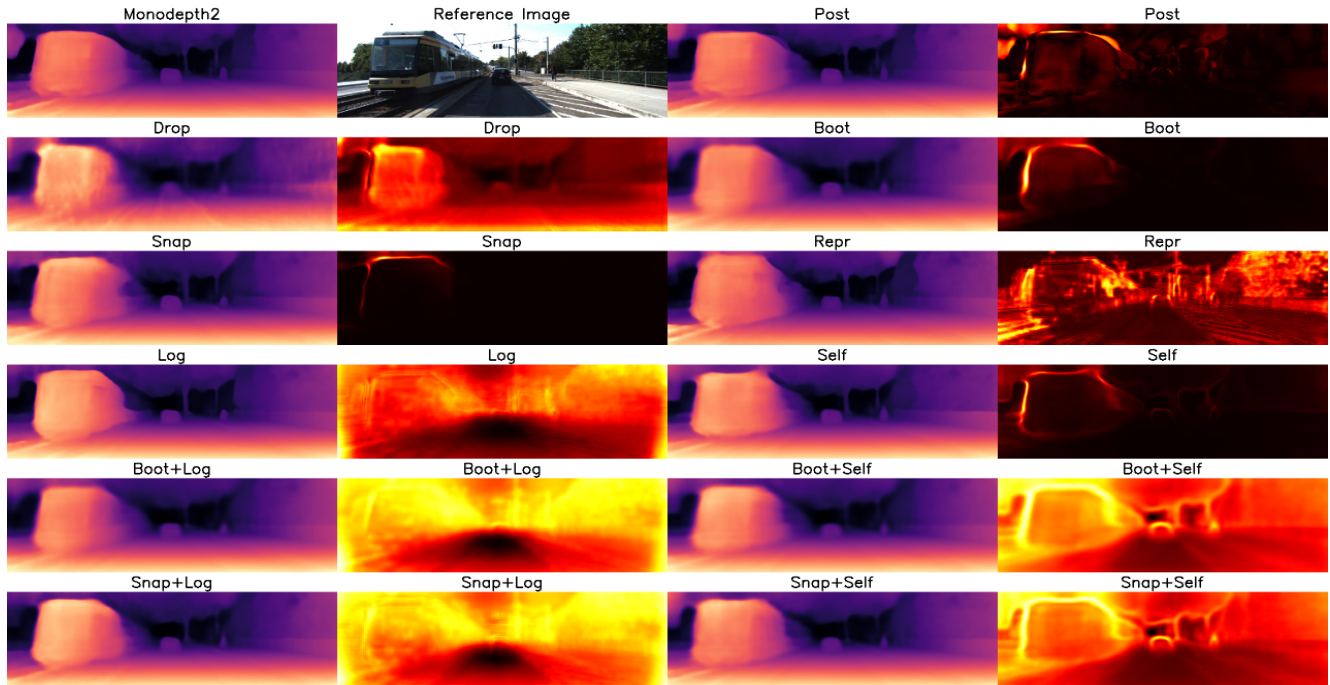


Figure 19. Qualitative results on image 2011\_09\_26\_drive\_0101\_sync/0000000114 from the Eigen test split [2]. Each row shows depth and uncertainty maps from one of the considered variants trained with stereo (S) supervision.

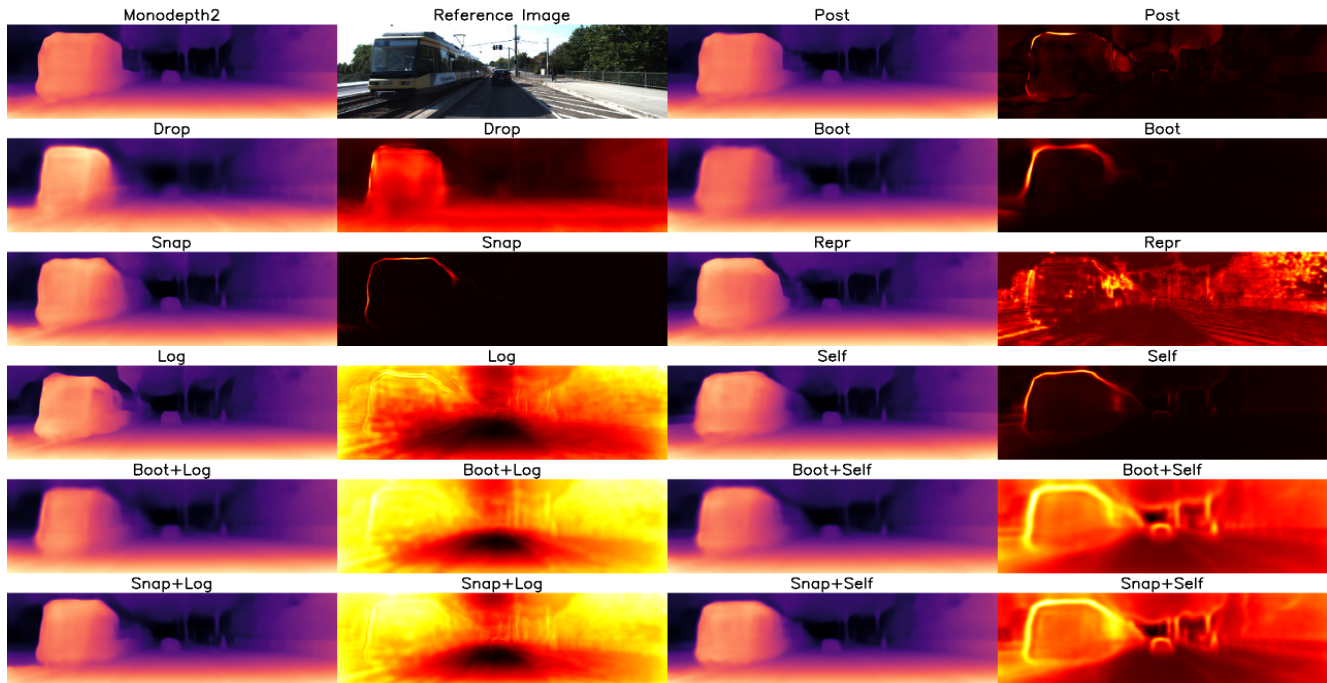


Figure 20. **Qualitative results on image 2011.09.26\_drive.0101\_sync/0000000114 from the Eigen test split [2].** Each row shows depth and uncertainty maps from one of the considered variants trained with mono+stereo (MS) supervision.

## 10.6. Qualitative video sequence

Finally, we refer the reader to the supplementary video available at [www.youtube.com/watch?v=bxVPXqf4zt4](http://www.youtube.com/watch?v=bxVPXqf4zt4), featuring the **2011.09.26\_drive.0101\_sync** sequence from the KITTI dataset and showing in order results for M, S and MS supervisions. From the video, we can perceive some of the behaviours highlighted in the submitted paper and this document. Specifically, we can observe how *Drop* provides reasonable uncertainty estimation when trained with M while it fails with S and MS. Moreover, we can notice how *Log* estimates are much more defined when dealing with S supervision compared to M and MS. Finally, the video also highlights how *Self* solutions are much more selective at providing high uncertainties compared to *Log* ones.

## References

- [1] Filippo Aleotti, Fabio Tosi, Matteo Poggi, and Stefano Mattoccia. Generative adversarial networks for unsupervised monocular depth prediction. In *15th European Conference on Computer Vision (ECCV) Workshops*, 2018. 1
- [2] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network. In *Advances in neural information processing systems*, pages 2366–2374, 2014. 1, 2, 4, 5, 7, 8, 10, 11, 12, 13, 14, 15
- [3] Ravi Garg, Vijay Kumar BG, Gustavo Carneiro, and Ian Reid. Unsupervised cnn for single view depth estimation: Geometry to the rescue. In *European Conference on Computer Vision*, pages 740–756. Springer, 2016. 4
- [4] Clément Godard, Oisín Mac Aodha, and Gabriel J Brostow. Unsupervised monocular depth estimation with left-right consistency. In *CVPR*, pages 270–279, 2017. 4
- [5] Clément Godard, Oisín Mac Aodha, Michael Firman, and Gabriel J. Brostow. Digging into self-supervised monocular depth estimation. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019. 1, 2, 4, 7
- [6] Jonas Uhrig, Nick Schneider, Lukas Schneider, Uwe Franke, Thomas Brox, and Andreas Geiger. Sparsity invariant cnns. In *International Conference on 3D Vision (3DV)*, pages 11–20. IEEE, 2017. 1, 2, 4, 5
- [7] Jamie Watson, Michael Firman, Gabriel J. Brostow, and Daniyar Turmukhambetov. Self-supervised monocular depth hints. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019. 1