Perceptual Deep Depth Super-Resolution
Supplementary Material

Oleg Voynov[1], Alexey Artemov[1], Vage Egiazarian[1], Alexander Notchenko[1],
Gleb Bobrovskikh[1,2], Evgeny Burnaev[1], Denis Zorin[3,1]
[1]Skolkovo Institute of Science and Technology, [2]Higher School of Economics,
[3]New York University
{oleg.voinov, a.artemov, vage.egiazarian, alexandr.notchenko}@skoltech.ru,
bobrovskihg@gmail.com, e.burnaev@skoltech.ru, dzorin@cs.nyu.edu

adase.group/3ddl/projects/perceptual-depth-sr

1. Additional evaluation details

In the literature on range image processing, the term *depth* is used to denote three different types of range data:

- *disparity*, presented in, *e.g.*, Middlebury dataset, *i.e.*, the difference in image location of a feature within two stereo images;
- *orthogonal depth*, presented in, *e.g.*, SUN-RGBD dataset, *i.e.*, the distance from a point in the 3D-space to the image plane;
- *perspective depth*, presented in, *e.g.*, low-resolution scans in ToFMark dataset, *i.e.*, the distance from a point in the 3D-space to the camera.

We use the term *depth map* to denote any data of this kind, however, in our experiments we evaluated each super-resolution method on the range type that it was designed for. For evaluation of the disparity processing methods on the datasets that do not provide disparity maps, we calculated virtual disparity images with the baseline of 20 cm.

Here we describe the quality measures that we considered in addition to the ones discussed in the main text. We recall that we label the metrics that compare the depth values directly with subscript "d", and the visually-based metrics with subscript "v".

BadPix is the fraction of measurements with absolute deviation larger than a certain threshold τ

$$\mathrm{BadPix}_{\mathrm{d}}(\tau|d_1, d_2) = \frac{1}{N}\big|\{ij : \big|d_{1,ij} - d_{2,ij}\big| > \tau\}\big|,$$

or the fraction of measurements with relative deviation larger than a threshold

$$\mathrm{BadPix}_{\mathrm{d}}(\tau\%|d_1, d_2) = \frac{1}{N}\big|\{ij : \big|\frac{d_{1,ij} - d_{2,ij}}{d_{2,ij}}\big| > \frac{\tau}{100}\}\big|,$$

where d_1 and d_2 are the compared depth maps, ij represents individual pixels, and N is the number of pixels. We considered BadPix for depth map comparison with absolute thresholds of 1, 5, and 10 cm and relative thresholds of 1, 5, and 10%. We also considered this metric for comparison of depth map renderings with the absolute thresholds of 1, 5, and 10 each divided by 255 (which correspond to deviations by the respective numbers of shades of gray in 8-bit grayscale images).

Bumpiness, introduced in [4] for piece-wise planar regions and generalized in [3] for arbitrary smooth surfaces, is a measure of surface smoothness with respect to a reference. It is calculated as

$$\mathrm{Bumpiness}_{\mathrm{d}}(d_1, d_2) = \frac{1}{N}\sum_{ij} \min(0.05, \|\mathrm{H}_{d_1-d_2}(i,j)\|_{\mathrm{F}}) \cdot 100,$$

where $\|\cdot\|_{\mathrm{F}}$ is Frobenius norm and $\mathrm{H}_f(i,j)$ is the Hessian matrix of the function f, calculated at point (i,j). We used the original implementation of this metric. Since this metric includes some parameter values, presumably, specific for the original evaluation dataset, we converted the depth maps to disparity using the camera intrinsics of this dataset.

We used the implementation of SSIM from *scikit-image* [8] and the original implementation of LPIPS from [10].

In addition to our $\mathrm{RMSE}_{\mathrm{v}}$ we considered RMS difference of two rendered images without averaging over the basis renderings, *i.e.*, calculated for a single lighting condition. We denote this metric as $\mathrm{RMSE}^1_{\mathrm{v}}$: for a light direction $\boldsymbol{e}$ and a pair of normal maps $\boldsymbol{n}_1, \boldsymbol{n}_2$ it is calculated as

$$\mathrm{RMSE}^1_{\mathrm{v}}(d_1, d_2) = \sqrt{\frac{1}{N}\sum_{i,j} \|\boldsymbol{e}\cdot\boldsymbol{n}_{1,ij} - \boldsymbol{e}\cdot\boldsymbol{n}_{2,ij}\|_2^2}.$$

2. Comparison of quality measures

In Figures 5-10 we compare the relations between different subsets of quality measures. We present pair-wise correlations of the metrics in the form of scatter plots in the lower half of the figure and Pearson and Spearman correlation coefficients in the upper half of the figure. For reference, on the diagonal of the figure we also include kernel density estimates of metric value distributions for each super-resolution method. The distributions for the modified methods DIP-v and MSG-v are represented with the dashed black and solid black curves respectively.

On the depth maps with missing measurements, the methods that do not inpaint the regions with the missing measurements (including MSG-v) sometimes produced severe outliers around these regions. To minimize the influence of such outliers on the results of the metric comparison, we limited the value of $\mathrm{RMSE_d}$ to a maximum of 0.5 meters. Among the collected super-resolved images, 8% exceeded this threshold.

For each metric, applied to rendered images, we gathered the values of this metric for four different light directions, as described in Section 5.2 of the main text. We then calculated two additional values, the worst and the average of these four. We label the respective versions of the metric with suffixes $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$, $\boldsymbol{e}_4$, max and avg. For each metric, we found that these six versions are strongly correlated, as illustrated in Figures 5-7, so we further focused on the worst value of each metric.

We also found that different versions of $\mathrm{RMSE_v^1}$ produce very similar results to our $\mathrm{RMSE_v}$, as illustrated in Figure 5. It is consistent with the observation that if $\mathrm{RMSE_v}$ is bounded by a constant C, then for *any* choice of the light direction $\boldsymbol{e}$, $\mathrm{RMSE_v^1}$ is bounded by C, which can be easily seen from the fact that $\mathrm{RMSE_v}$ does not depend on the choice of the basis, so we can choose one of the basis light directions to be equal to $\boldsymbol{e}$.

In Figure 8 we compare the metrics of different types: pixel-wise $\mathrm{RMSE_d}$, $\mathrm{BadPix_d(5cm)}$ and $\mathrm{BadPix_d(5\%)}$ applied to depth directly; "worst" versions of pixel-wise $\mathrm{BadPix_v(5)}$ and perceptual $\mathrm{DSSIM_v}$ and $\mathrm{LPIPS_v}$, applied to rendered images; geometrical $\mathrm{Bumpiness_d}$ and our $\mathrm{RMSE_v}$. We found that all three pixel-wise metrics applied to depth directly demonstrate weak correlation with visual and geometrical metrics. Pixel-wise $\mathrm{BadPix_v(5)}$ applied to rendered images, although strongly correlated with perceptual metrics, is inappropriate for gradient-based optimization. Additional comparison of pixel-wise $\mathrm{BadPix_d}$ and $\mathrm{BadPix_v}$ with different thresholds to perceptual $\mathrm{DSSIM_v}$ and $\mathrm{LPIPS_v}$ (Figures 9 and 10) leads to the same conclusions. $\mathrm{Bumpiness_d}$ is also strongly correlated with perceptual metrics but only measures local curvature deviation, while the visual appearance of 3D surface is determined by its local orientation.

3. Comparison of super-resolution methods

In Tables 2-9 we present the results of quantitative evaluation of super-resolution methods on the datasets SimGeo, ICL-NUIM and Middlebury for Box downsampling model and scaling factors of 4 and 8. In Table 1 we present the average values. $\mathrm{RMSE_d}$ is in millimeters, BadPix is in percents, $\mathrm{DSSIM_v}$, $\mathrm{LPIPS_v}$ and $\mathrm{RMSE_v}$ are in thousandths. For all visual metrics except $\mathrm{RMSE_v}$ the presented value is of the "worst" version. For all metrics the lower value corresponds to the better result. The best results are highlighted in bold and the second best results are underlined.

In addition to metric values, the last three columns of the tables contain the results of the informal perceptual study collected over approximately 250 subjects. In this study, for each scene from SimGeo, ICL-NUIM and Middlebury datasets subjects were shown the renderings of super-resolved depth maps, shuffled randomly, and were asked to choose the renderings, the most and second most similar to ground truth. The renderings calculated with the fourth light direction were used. The values in the columns "User, 1st", "User, 2nd", and "Top 2" represent the percentages of the subjects who chose the rendering of the super-resolved depth map, produced by the method in the corresponding method, as the most similar, second most similar, or one of the two most similar to the ground truth respectively. We found that our $\mathrm{RMSE_v}$ is mostly consistent with human judgements.

4. Training with $\mathrm{MSE_v}$

Since optimization of $\mathrm{MSE_v}$ alone is an ill-posed problem, we used a regularization term that penalizes absolute depth deviation. We found that among different regularizers, including $\mathrm{MSE_d}$, $\mathrm{Lap_1}$ produces the best results. In general, we found that optimization leads to the best results if the terms are weighted in such way that geometrically corresponding depth error and angular normal error result in the same magnitudes of terms. The corresponding value of the weighing parameter w in Equation 4 of the main text is determined by the properties of the training data, such as depth map scaling or field of view of the camera.

5. Noisy depth measurements in the input

SimGeo, ICL-NUIM and Middlebury datasets were our primary evaluation sets, yielding the most pronounced outcomes, however, these datasets contain only noise-free scenes. As we were interested in evaluation of our approach on a diverse set of RGBD images, we included twelve scenes from SUN RGBD dataset and three scenes from ToFMark dataset that feature real-world noise patterns in our evaluation data. We observed that increased levels of noise are extremely harmful to all non over-smoothing methods, including those modified with our loss, as they fail

	$\mathrm{RMSE_d}$		$\mathrm{BadPix_d}$(5cm)		$\mathrm{BadPix_v}$(5)		$\mathrm{DSSIM_v}$		$\mathrm{LPIPS_v}$		$\mathrm{Bumpiness_d}$		$\mathrm{RMSE_v}$		User, 1st	User, 2nd	Top 2
	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x4	x4
Average performance on SimGeo dataset																	
Bicubic	55	79	4.1	7.9	<u>23.2</u>	<u>38.3</u>	197	301	320	427	0.70	0.98	193	234	0.5	7.8	8.3
SRfS [2]	61	88	7.5	14.3	74.2	77.1	711	729	869	865	1.48	1.69	311	328	0.0	0.0	0.0
EG [9]	53		2.2		33.1		<u>168</u>		306		<u>0.54</u>		<u>136</u>		0.2	3.9	4.2
PDN [6]	162	211	99.4	99.1	39.2	45.1	224	<u>264</u>	<u>278</u>	<u>407</u>	0.63	<u>0.79</u>	165	201	1.6	<u>12.3</u>	13.8
DG [1]	54	84	3.0	6.4	35.2	39.1	293	316	420	437	0.69	0.82	171	190	0.2	2.9	3.2
DIP [7]	52	59	8.5	12.5	90.5	92.0	887	880	893	915	2.21	2.77	395	475	0.6	0.9	1.5
MSG [5]	<u>39</u>	<u>39</u>	<u>1.5</u>	3.3	51.9	69.3	374	544	569	713	0.79	0.97	194	242	0.4	3.7	4.0
DIP-v	**33**	41	1.7	<u>2.3</u>	49.7	67.1	313	491	524	598	0.60	0.88	147	<u>174</u>	<u>8.3</u>	**59.4**	<u>67.8</u>
MSG-v	96	**29**	**0.7**	**1.5**	**14.2**	**34.6**	**95**	**206**	**194**	**367**	**0.34**	**0.46**	**99**	**129**	**88.1**	9.1	**97.2**
Average performance on ICL-NUIM dataset																	
Bicubic	34	54	2.8	5.5	<u>59.3</u>	<u>64.2</u>	<u>431</u>	<u>490</u>	558	668	1.15	1.32	210	252	5.0	<u>28.3</u>	33.3
SRfS [2]	42	62	5.5	11.0	73.5	76.1	641	664	636	660	1.72	1.83	287	314	0.0	0.0	0.0
PDN [6]	135	165	93.8	82.9	66.2	70.2	480	509	623	650	<u>1.14</u>	<u>1.24</u>	237	264	2.6	10.5	13.1
DG [1]	36	58	4.3	6.4	64.4	65.5	497	505	663	689	1.28	1.32	234	259	0.6	5.5	6.1
DIP [7]	43	56	10.6	14.2	83.6	83.4	812	806	690	690	2.73	2.58	394	389	1.1	0.9	2.0
MSG [5]	<u>25</u>	**36**	<u>1.6</u>	<u>3.5</u>	64.1	69.0	489	557	<u>510</u>	<u>534</u>	1.27	1.46	210	255	1.1	7.2	8.3
DIP-v	28	<u>40</u>	2.6	3.9	67.8	69.6	516	548	**407**	**503**	1.45	1.56	<u>209</u>	<u>236</u>	<u>9.6</u>	**31.9**	<u>41.4</u>
MSG-v	**24**	41	**1.3**	**3.1**	**56.3**	**61.1**	**387**	**437**	527	602	**0.94**	**1.06**	**157**	**192**	**79.9**	11.8	**91.7**
Average performance on Middlebury dataset																	
Bicubic	843	1139	10.8	13.9	**71.5**	**76.7**	**648**	**748**	<u>575</u>	720	**0.87**	**0.76**	**344**	**386**	4.1	<u>25.3</u>	29.4
SRfS [2]	100	145	21.4	33.6	86.4	89.5	780	810	669	704	1.32	<u>1.28</u>	428	461	0.0	0.0	0.0
PDN [6]	173	225	85.3	76.5	83.4	86.4	744	790	653	711	1.38	1.67	405	467	9.9	**28.1**	<u>37.9</u>
DG [1]	266	330	15.0	24.5	81.8	84.1	765	784	728	740	1.54	1.73	421	442	0.7	10.6	11.3
DIP [7]	<u>72</u>	<u>104</u>	19.6	24.4	92.4	93.4	927	947	737	717	2.82	2.90	565	592	1.2	5.6	6.8
MSG [5]	228	426	10.8	13.1	81.8	87.2	774	858	649	696	1.96	2.19	477	525	0.2	1.6	1.8
DIP-v	**56**	**87**	**6.4**	<u>10.6</u>	83.1	87.4	728	821	**506**	**568**	1.34	1.56	<u>353</u>	<u>409</u>	**72.3**	18.2	**90.5**
MSG-v	96	133	<u>7.3</u>	**9.2**	<u>73.3</u>	<u>79.0</u>	<u>667</u>	<u>757</u>	639	<u>690</u>	<u>1.20</u>	1.35	376	431	<u>10.8</u>	9.9	20.7
Average performance on the scenes without missing measurements (SimGeo, ICL-NUIM, Vintage)																	
	$\mathrm{RMSE_d}$		$\mathrm{BadPix_d}$(5cm)		$\mathrm{BadPix_v}$(5)		$\mathrm{DSSIM_v}$		$\mathrm{LPIPS_v}$		$\mathrm{Bumpiness_d}$		$\mathrm{RMSE_v}$		User, 1st	User, 2nd	Top 2
	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x4	x4
Bicubic	46	69	3.5	6.9	<u>43.7</u>	<u>53.3</u>	<u>333</u>	415	<u>452</u>	561	0.97	1.19	206	248	3.0	<u>18.9</u>	21.9
SRfS [2]	55	81	7.3	14.1	74.6	77.4	680	701	743	753	1.60	1.75	303	326	0.0	0.0	0.0
PDN [6]	148	187	94.4	90.1	55.0	59.8	378	<u>412</u>	482	<u>542</u>	<u>0.93</u>	<u>1.06</u>	210	241	1.9	10.5	12.4
DG [1]	47	73	3.9	6.7	52.1	54.4	416	430	561	584	1.02	1.10	209	230	0.4	4.0	4.4
DIP [7]	50	62	10.7	15.9	87.6	88.2	857	853	801	808	2.59	2.79	414	452	0.8	0.8	1.7
MSG [5]	<u>33</u>	<u>39</u>	<u>1.7</u>	3.6	59.8	70.3	454	569	547	622	1.08	1.26	210	258	0.7	5.8	6.4
DIP-v	**31**	43	2.2	<u>3.3</u>	60.8	69.9	444	548	474	560	1.09	1.32	<u>191</u>	<u>223</u>	<u>10.2</u>	**45.5**	<u>55.8</u>
MSG-v	38	**38**	**1.1**	**2.6**	**38.0**	**50.1**	**264**	**346**	**385**	**501**	**0.69**	**0.81**	**135**	**169**	**82.7**	10.9	**93.6**
Average performance on the scenes with missing measurements (Middlebury excluding Vintage)																	
Bicubic	972	1313	11.8	14.7	**71.3**	**76.6**	**663**	**765**	<u>570</u>	718	**0.77**	**0.61**	<u>358</u>	**400**	3.8	<u>24.7</u>	28.5
SRfS [2]	100	145	22.2	33.8	86.9	89.8	790	820	676	716	1.26	<u>1.21</u>	441	474	0.0	0.0	0.0
PDN [6]	178	234	88.3	76.1	83.5	86.6	757	803	644	713	1.36	1.69	419	487	<u>11.5</u>	**32.8**	<u>44.3</u>
DG [1]	298	367	16.3	26.9	82.2	84.7	781	803	716	724	1.55	1.76	442	465	0.9	12.2	13.1
DIP [7]	<u>72</u>	<u>102</u>	18.8	20.7	92.2	93.3	923	943	708	691	2.62	2.68	549	577	1.2	6.4	7.7
MSG [5]	259	488	12.1	14.1	82.0	87.7	785	870	673	711	2.02	2.24	507	552	0.2	0.2	0.5
DIP-v	**58**	**91**	**7.1**	<u>11.4</u>	82.7	87.2	716	811	**494**	**550**	1.23	1.41	**354**	<u>405</u>	**80.1**	13.8	**93.9**
MSG-v	107	145	<u>8.1</u>	**9.8**	<u>73.6</u>	<u>79.2</u>	<u>688</u>	<u>776</u>	634	<u>688</u>	<u>1.19</u>	1.34	404	458	1.3	8.8	10.1
Average performance on SimGeo, ICL-NUIM, Middlebury																	
Bicubic	339	462	6.1	9.3	<u>52.4</u>	<u>60.6</u>	<u>437</u>	<u>525</u>	489	611	<u>0.91</u>	<u>1.00</u>	254	296	3.3	<u>20.7</u>	24.0
SRfS [2]	69	101	12.0	20.3	78.5	81.3	715	738	722	741	1.50	1.58	347	372	0.0	0.0	0.0
PDN [6]	157	202	92.4	85.7	64.0	68.2	498	536	533	596	1.07	1.26	276	319	5.0	17.5	22.5
DG [1]	126	166	7.8	13.1	61.6	64.0	531	548	610	628	1.19	1.31	283	305	0.5	6.6	7.1
DIP [7]	<u>57</u>	75	13.3	17.4	89.1	89.8	878	881	771	771	2.60	2.76	457	491	1.0	2.6	3.6
MSG [5]	104	181	5.0	6.9	66.8	75.8	559	664	587	650	1.37	1.57	304	350	0.5	4.0	4.6
DIP-v	**40**	**58**	<u>3.8</u>	<u>5.9</u>	67.7	75.4	530	631	<u>481</u>	**557**	1.14	1.34	<u>242</u>	<u>280</u>	<u>32.3</u>	**35.5**	**67.8**
MSG-v	60	<u>72</u>	**3.3**	**4.8**	**49.2**	**59.3**	**398**	**482**	**464**	<u>560</u>	**0.85**	**0.98**	**220**	**260**	**57.0**	10.2	<u>67.3</u>

Table 1: Quantitative evaluation summary. The best result is in bold, the second best in underlined.

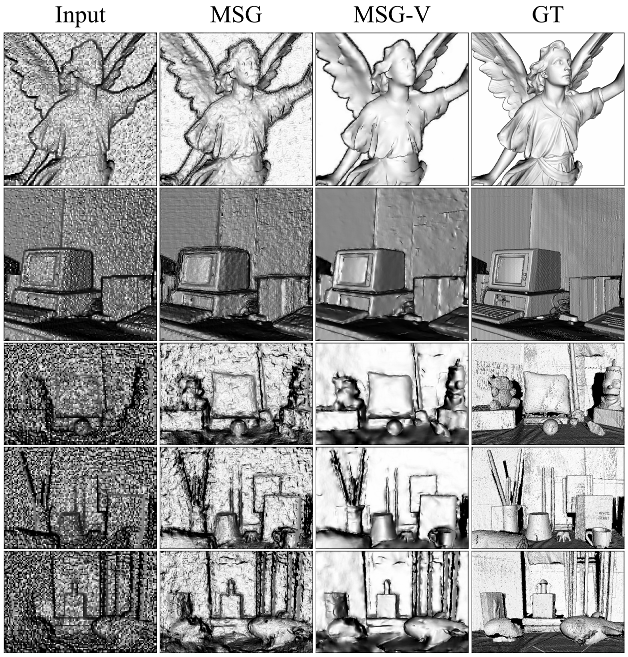

Figure 1: $\times 4$ super-resolution results produced by the original MSG and MSG-V with our loss, both trained on noisy data. The upper two samples contain synthetic noise, while the lower three from ToFMark dataset represent real noisy ToF measurements. Best viewed in large scale.

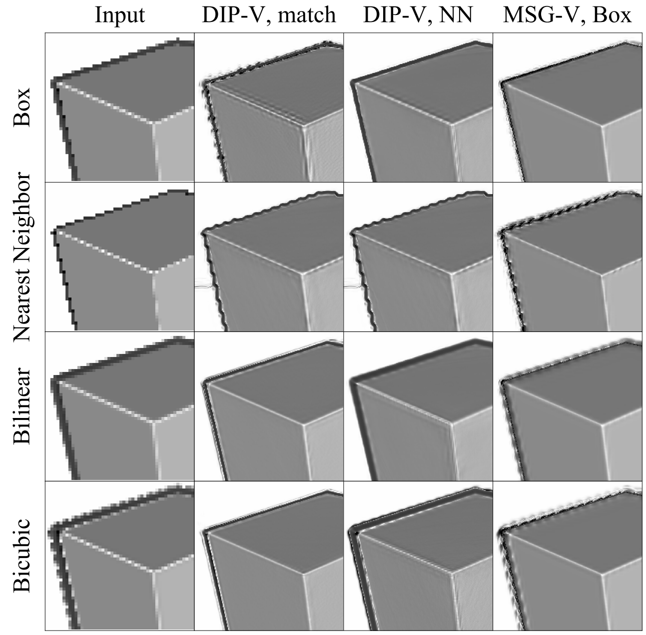

Figure 2: $\times 4$ super-resolution results for different input downsampling models produced by DIP-V with a matching downsampling model, DIP-V with Nearest Neighbor downsampling model and MSG-V with Box downsampling model. Best viewed in large scale.

to produce reasonable super-resolution results, as illustrated in Figures 3-4. To demonstrate that this is not a limitation of our approach, in Figure 1 we present the super-resolution results produced by modified and unmodified versions of MSG, trained on the data with synthetic multiplicative gaussian noise.

6. Different downsampling models

In Figure 2 we present the results for different downsampling models, used for calculation of low-resolution input. We found that the visual quality remains high when the downsampling model used during training and that of the input match; if this is not the case, the quality deteriorates, as expected.

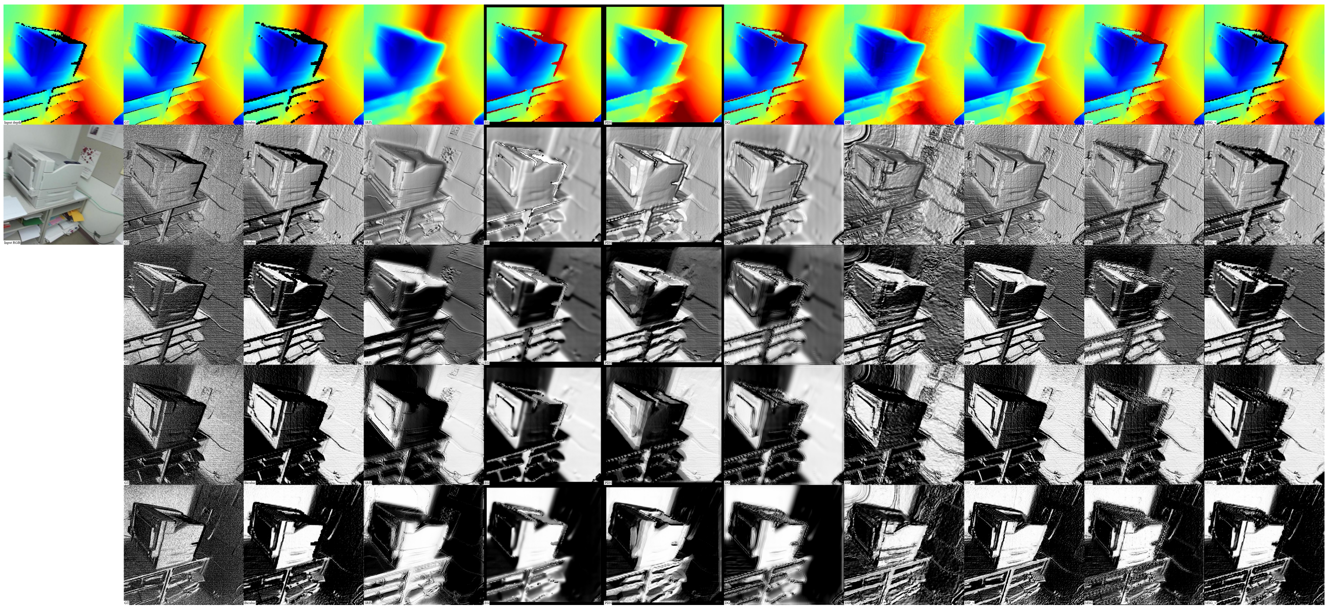

Figure 3: x4 super-resolution results on a Kinect v2 RGBD scan from SUN RGBD dataset. Each visualization is labeled in the bottom left corner. Ground truth is in the 2nd column, DIP-v is in the third from the right, MSG-v in the last one.

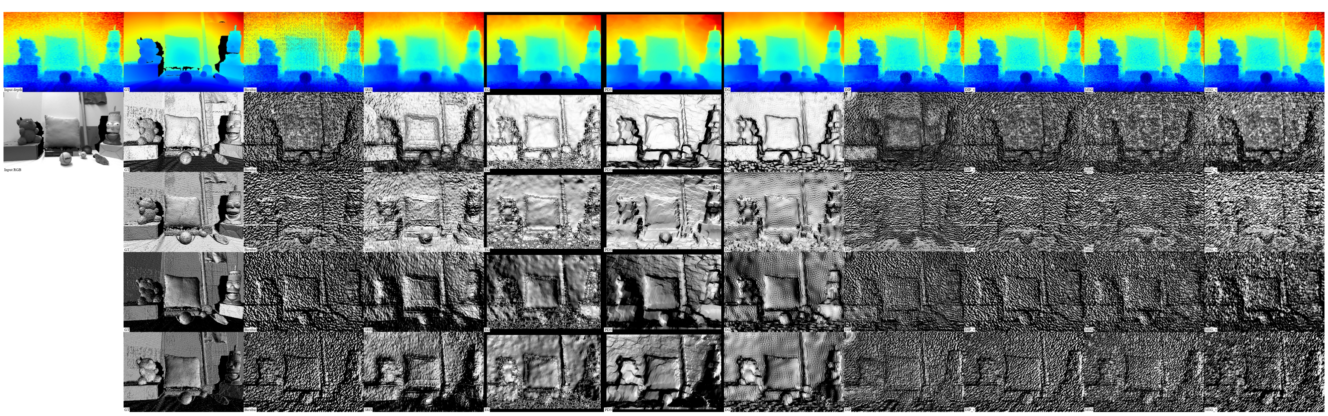

Figure 4: x4 super-resolution results on “Devil” from ToFMark dataset. Each visualization is labeled in the bottom left corner. Ground truth is in the 2nd column, DIP-v is in the third from the right, MSG-v in the last one.

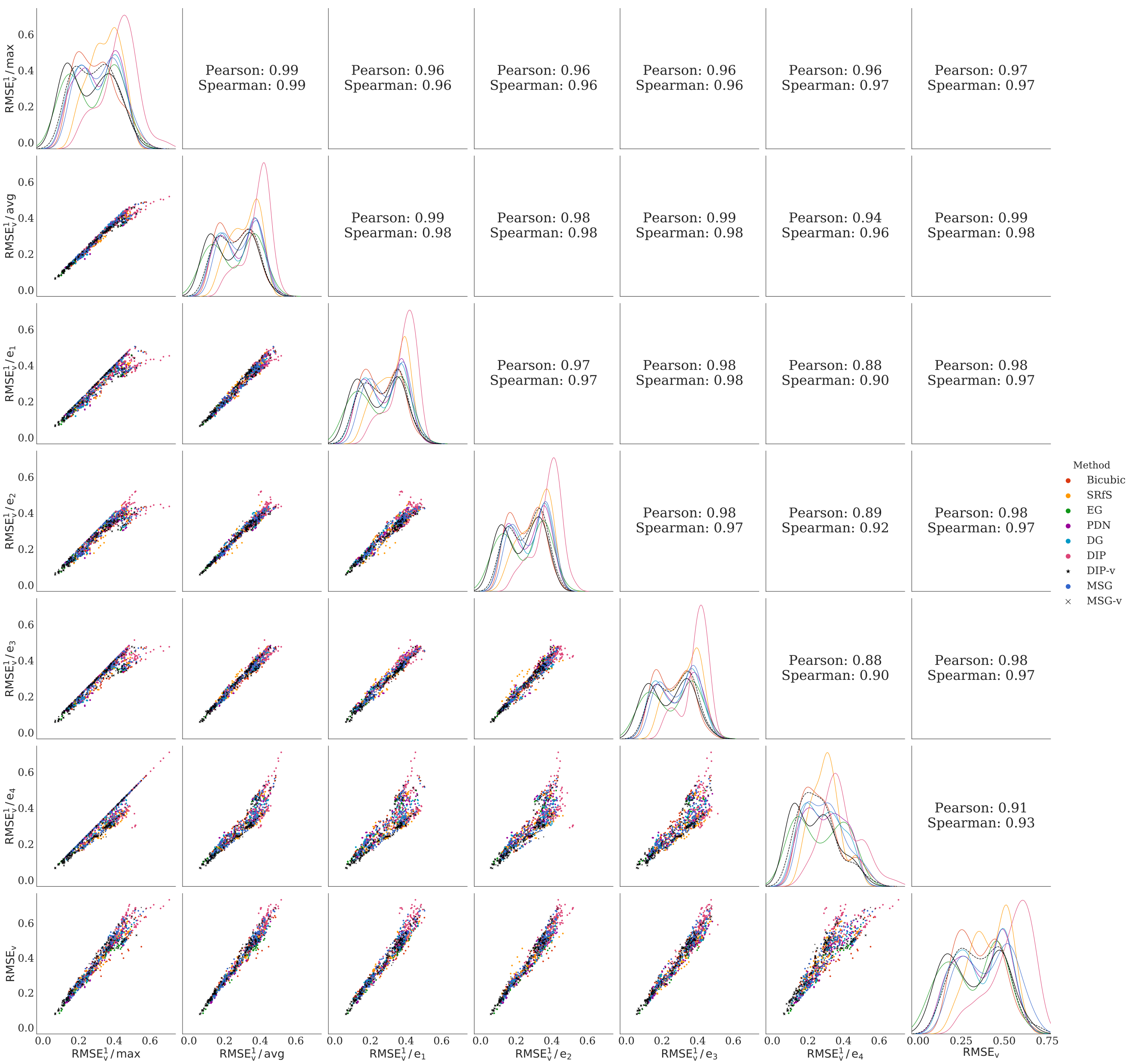

Figure 5: Comparison of different versions of $\mathrm{RMSE}_\mathrm{v}^1$ metric and RMSE_v metric. Best viewed in large scale and in color.

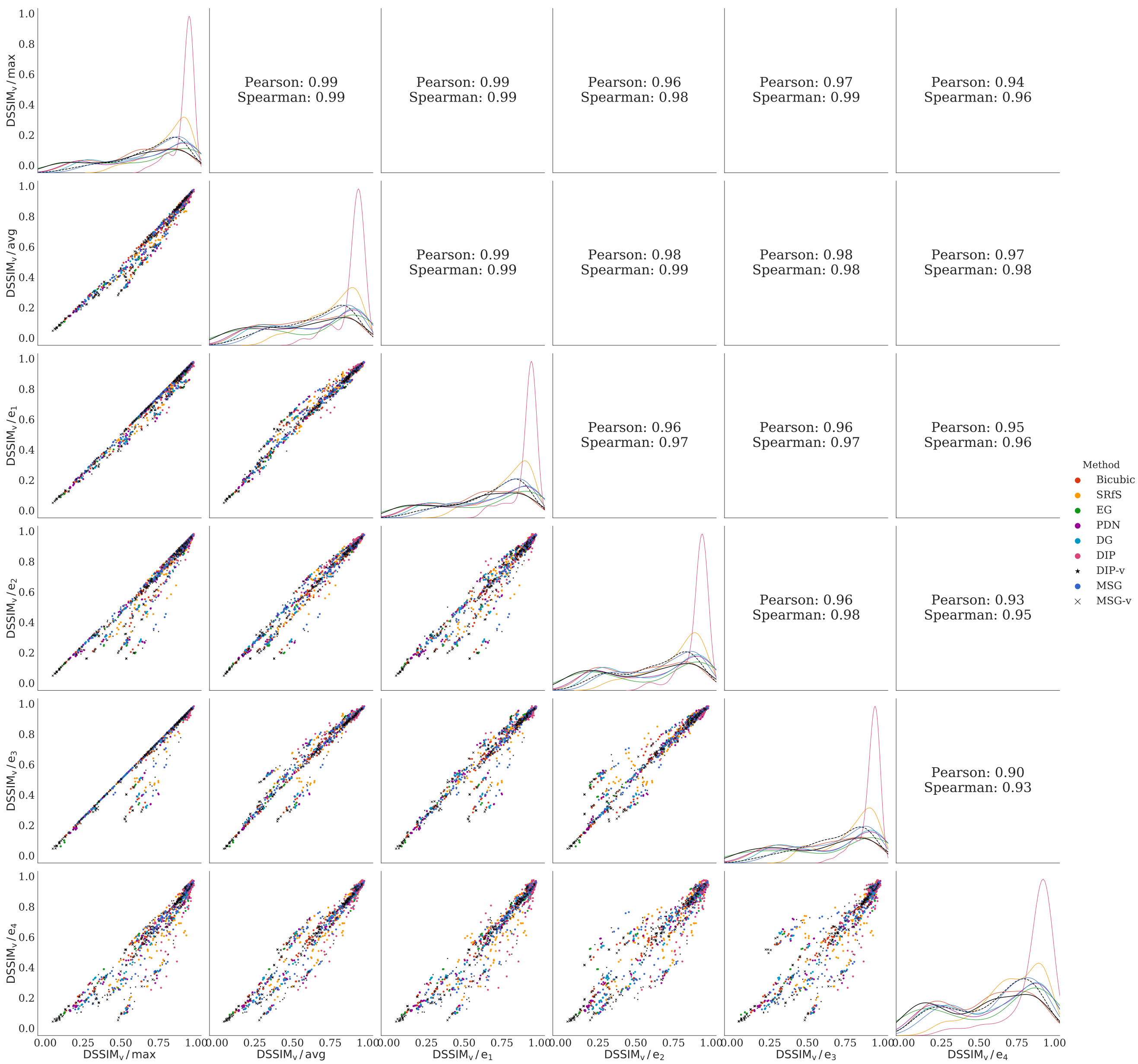

Figure 6: Comparison of different versions of $\mathrm{DSSIM_v}$ metric. Best viewed in large scale and in color.

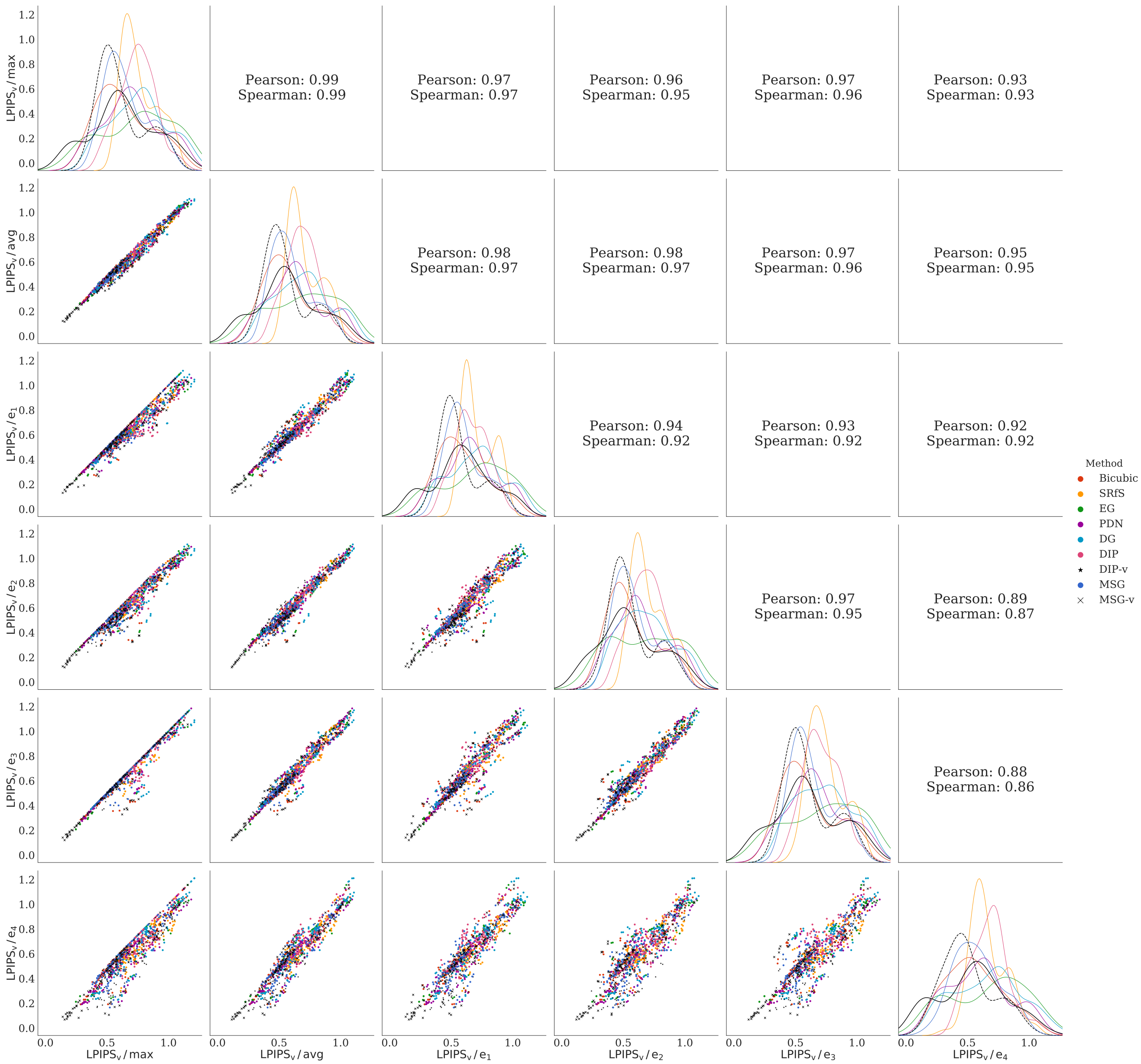

Figure 7: Comparison of different versions of $\mathrm{LPIPS_v}$ metric. Best viewed in large scale and in color.

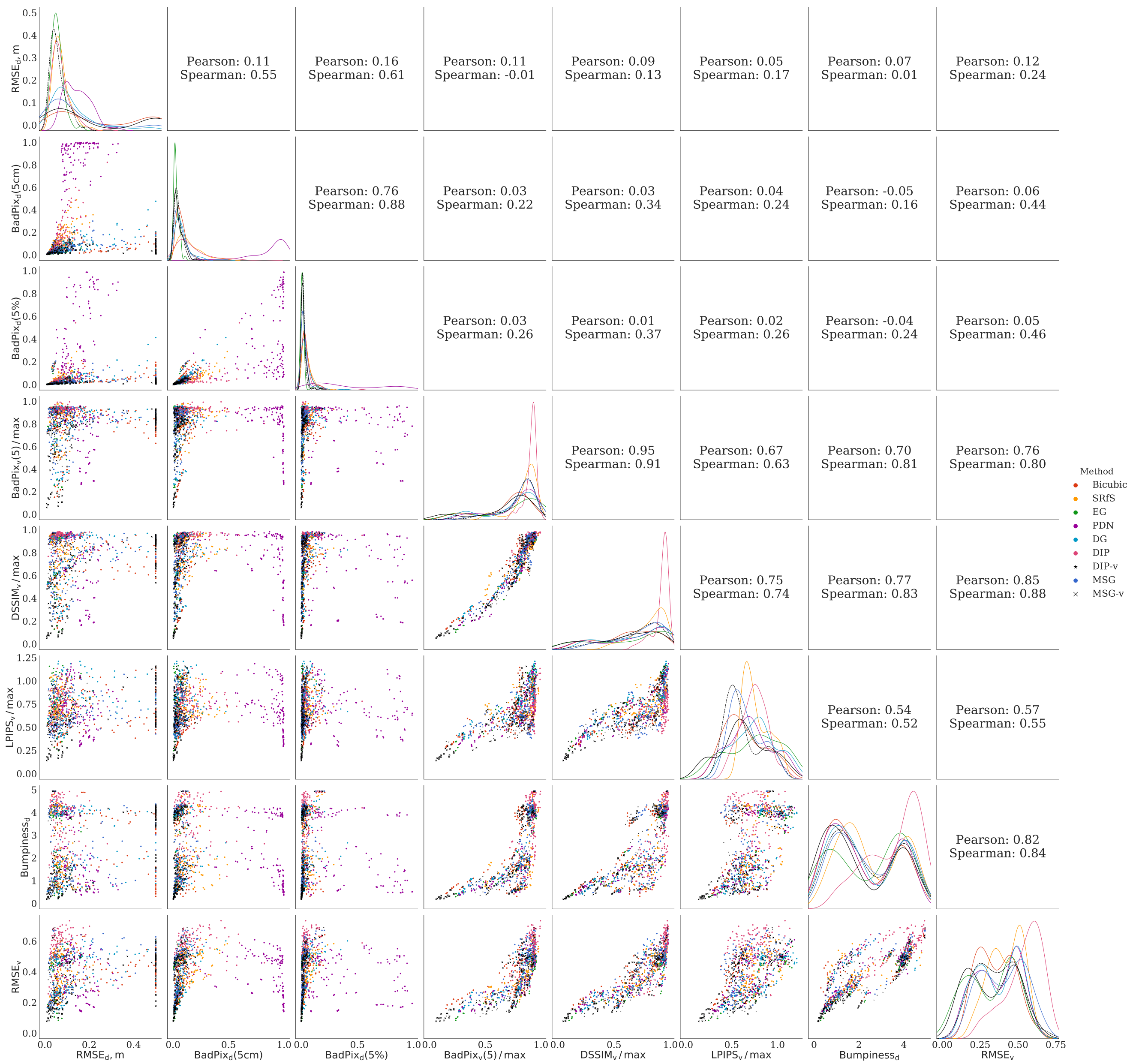

Figure 8: Comparison of metrics of different types. Best viewed in large scale and in color.

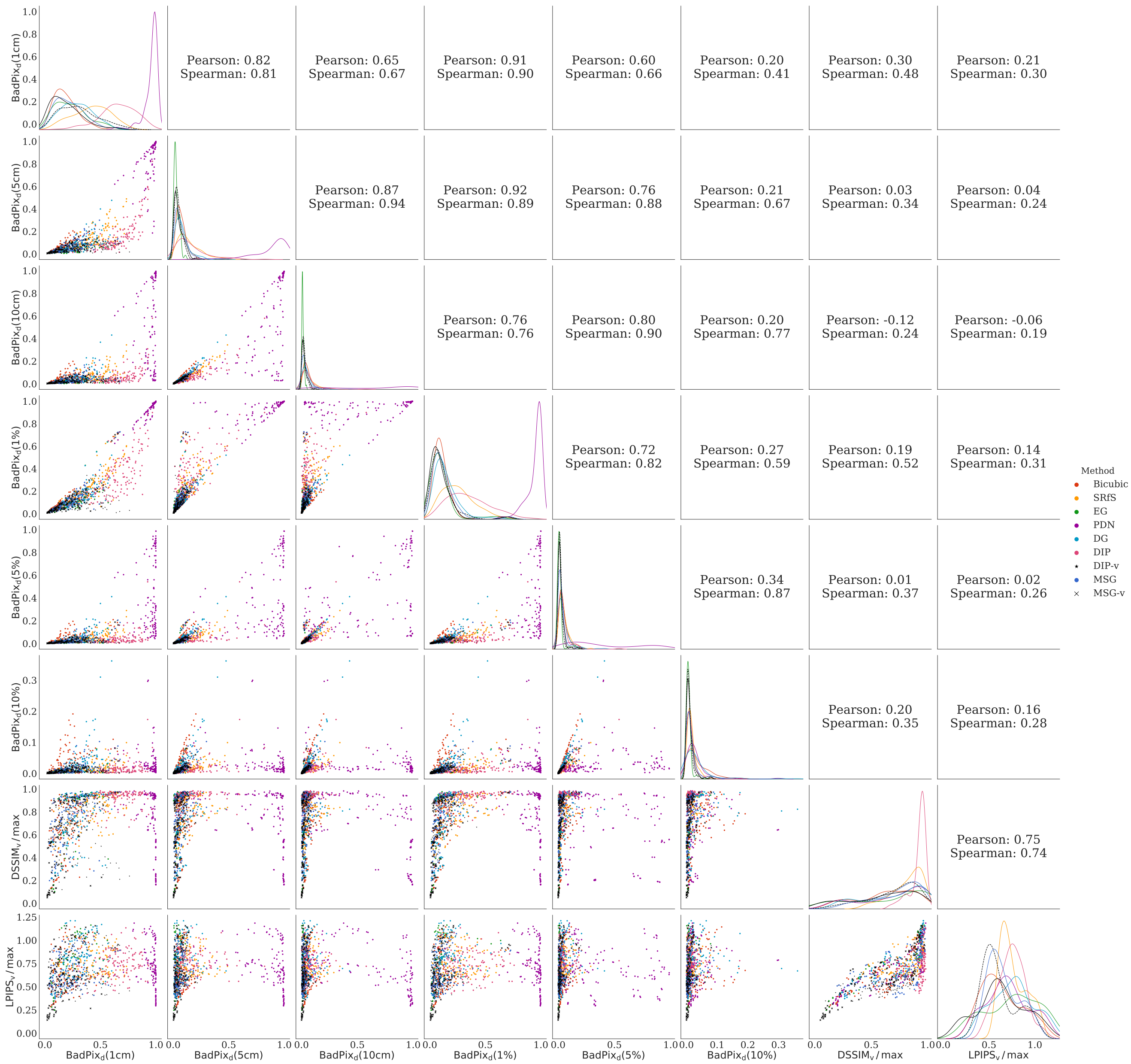

Figure 9: Comparison of different pixel-wise metrics applied to depth directly and perceptual metrics. Best viewed in large scale and in color.

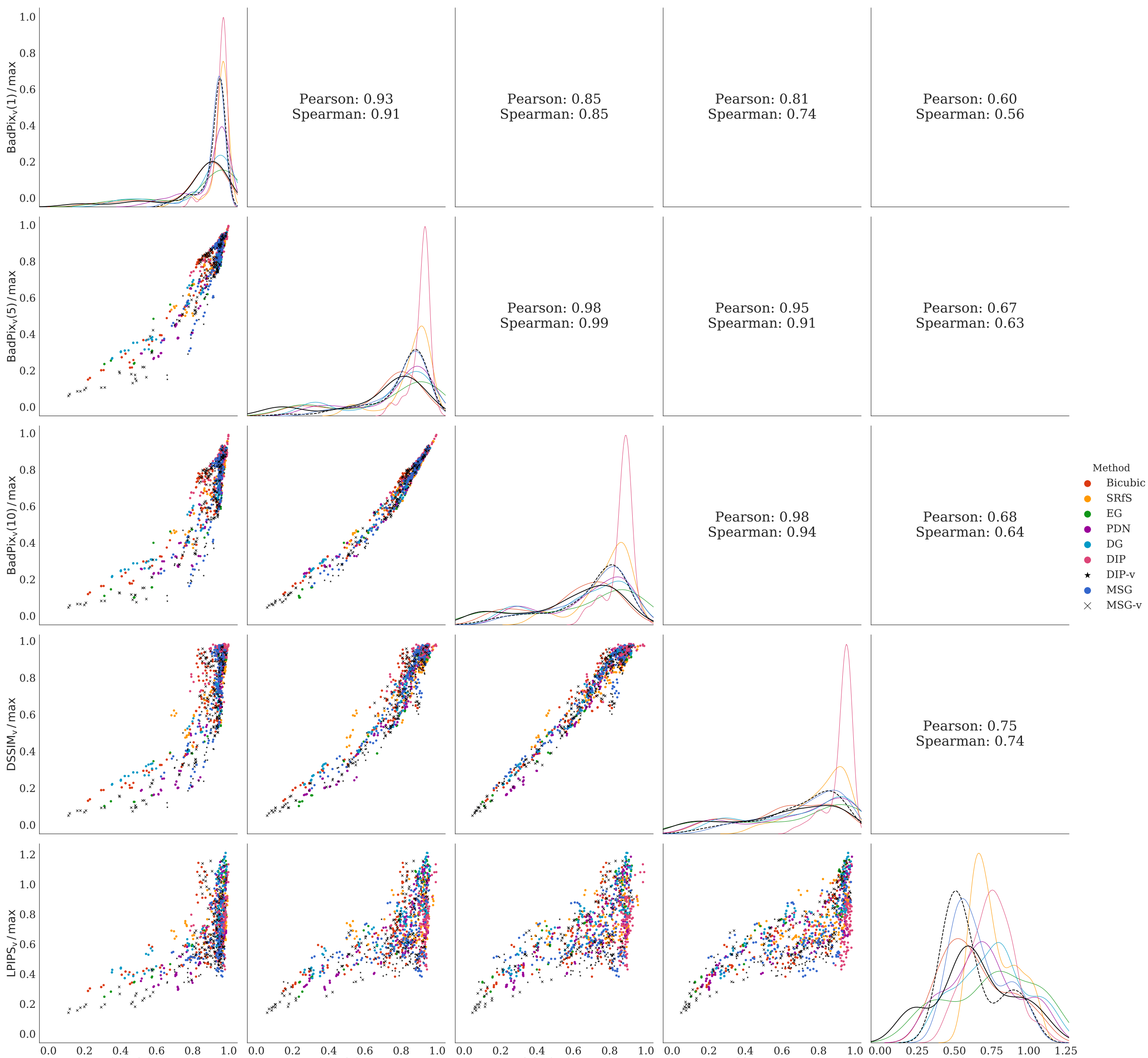

Figure 10: Comparison of different pixel-wise metrics applied to rendered images and perceptual metrics. Best viewed in large scale and in color.

	$\mathrm{RMSE_d}$		$\mathrm{BadPix_d(5cm)}$		$\mathrm{BadPix_v(5)}$		$\mathrm{DSSIM_v}$		$\mathrm{LPIPS_v}$		$\mathrm{Bumpiness_d}$		$\mathrm{RMSE_v}$		User, 1st	User, 2nd	Top 2
	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x4	x4
Cube, high-frequency texture																	
Bicubic	44	63	2.7	5.2	15.0	**27.3**	131	204	287	395	0.43	0.61	160	188	0.7	13.2	14.0
SRfS [2]	52	75	6.2	12.1	89.2	80.3	934	818	1036	938	1.73	1.67	361	339	0.0	0.0	0.0
EG [9]	43		1.2		25.4		113		214		0.35		105		0.7	9.6	10.3
PDN [6]	164	219	99.6	99.4	27.5	29.5	156	**186**	250	**368**	0.39	0.49	145	171	0.7	4.4	5.1
DG [1]	44	67	1.9	4.2	26.4	30.1	218	240	411	437	0.44	0.55	139	159	0.7	7.4	8.1
DIP [7]	45	48	6.4	8.5	93.5	92.5	963	947	906	918	2.98	2.50	530	494	0.0	0.7	0.7
MSG [5]	29	38	1.0	2.6	60.1	77.9	445	653	687	877	0.79	0.98	176	233	0.0	0.0	0.0
DIP-v	**26**	36	0.8	1.6	56.2	60.8	352	413	613	653	0.64	0.89	146	162	5.9	**58.1**	64.0
MSG-v	102	**20**	**0.3**	**0.7**	**9.3**	51.0	**70**	316	**179**	676	**0.20**	**0.39**	**77**	**125**	**91.2**	6.6	**97.8**
Cube, no texture																	
Bicubic	44	63	2.7	5.2	15.0	27.3	131	204	287	395	0.43	0.61	160	188	0.0	5.9	5.9
SRfS [2]	43	63	2.1	4.5	53.4	51.7	516	476	754	728	0.67	0.89	219	228	0.0	0.0	0.0
EG [9]	43		1.2		25.4		128		282		0.35		105		0.0	2.9	2.9
PDN [6]	164	219	99.6	99.4	26.3	29.3	162	185	314	353	0.38	0.49	145	171	0.7	4.4	5.1
DG [1]	44	67	1.9	4.2	26.4	30.1	218	240	411	437	0.44	0.55	139	159	0.0	2.2	2.2
DIP [7]	72	56	23.2	17.3	94.3	99.1	912	980	1026	1133	2.05	4.22	434	683	0.0	0.7	0.7
MSG [5]	29	26	1.0	1.7	30.7	49.8	199	314	509	642	0.42	0.47	157	171	0.0	6.6	6.6
DIP-v	26	35	0.8	1.4	15.1	45.4	95	237	347	478	0.28	0.35	111	107	1.5	**76.5**	77.9
MSG-v	**9**	**19**	**0.3**	**0.4**	**6.0**	**13.0**	**50**	**73**	**141**	**213**	**0.17**	**0.21**	**77**	**82**	**97.8**	0.7	**98.5**

Table 2: Quantitative evaluation on “Cube” with different RGBs from SimGeo dataset. The best result is in bold, the second best is underlined.

	$\mathrm{RMSE_d}$		$\mathrm{BadPix_d(5cm)}$		$\mathrm{BadPix_v(5)}$		$\mathrm{DSSIM_v}$		$\mathrm{LPIPS_v}$		$\mathrm{Bumpiness_d}$		$\mathrm{RMSE_v}$		User, 1st	User, 2nd	Top 2
	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x4	x4
Sphere and cylinder, high-frequency texture																	
Bicubic	57	82	4.1	8.1	20.1	36.7	189	294	313	420	0.67	0.98	189	234	0.0	0.7	0.7
SRfS [2]	70	102	12.1	24.6	91.9	91.8	887	865	1025	1008	2.43	2.41	417	403	0.0	0.0	0.0
EG [9]	55		2.4		30.4		143		326		0.50		130		0.0	1.5	1.5
PDN [6]	157	197	99.3	98.9	40.7	54.1	198	**242**	295	461	0.60	0.77	150	187	1.5	9.6	11.0
DG [1]	56	87	3.2	6.3	30.9	**35.2**	265	285	372	**386**	0.66	0.77	166	180	0.0	1.5	1.5
DIP [7]	46	69	3.9	27.2	97.0	99.2	965	975	1062	1014	4.01	4.80	548	696	1.5	2.9	4.4
MSG [5]	41	41	1.4	3.6	72.6	85.6	626	820	859	960	0.98	1.43	229	314	0.0	0.0	0.0
DIP-v	**28**	43	1.2	2.4	69.2	86.0	560	850	766	832	0.56	1.45	142	242	32.4	**52.9**	85.3
MSG-v	99	**37**	**0.6**	**2.0**	**14.3**	53.0	**94**	334	**267**	583	**0.29**	**0.55**	**96**	**164**	**64.7**	30.9	**95.6**
Sphere and cylinder, no texture																	
Bicubic	57	82	4.1	8.1	20.2	36.8	189	294	325	437	0.67	0.98	190	233	0.0	0.7	0.7
SRfS [2]	59	85	4.6	8.6	51.4	70.8	430	619	657	766	0.77	1.25	193	256	0.0	0.0	0.0
EG [9]	56		2.4		30.9		160		383		0.50		128		0.0	1.5	1.5
PDN [6]	157	197	99.3	98.9	38.0	44.1	202	218	294	386	0.58	0.76	150	186	5.9	17.6	23.5
DG [1]	57	87	3.2	6.4	31.0	35.3	265	284	396	409	0.66	0.78	165	180	0.7	2.2	2.9
DIP [7]	49	56	5.0	5.5	85.6	81.6	856	662	927	723	1.01	0.96	244	249	1.5	0.0	1.5
MSG [5]	40	37	1.4	3.1	45.6	64.5	288	444	509	610	0.65	0.76	183	218	0.7	0.7	1.5
DIP-v	35	39	1.4	1.8	41.0	72.6	210	523	517	643	0.47	0.70	130	141	9.6	**64.7**	74.3
MSG-v	**14**	**27**	**0.7**	**1.3**	**8.5**	**18.0**	**77**	**93**	**174**	**200**	**0.27**	**0.32**	**96**	**110**	**81.6**	12.5	**94.1**
Sphere and cylinder, low-frequency texture																	
Bicubic	57	82	4.1	8.1	20.1	36.7	189	294	313	420	0.67	0.98	189	234	0.0	2.2	2.2
SRfS [2]	62	91	6.8	14.9	74.9	81.0	691	738	961	956	1.38	1.65	311	335	0.0	0.0	0.0
EG [9]	54		2.4		30.4		160		377		0.50		129		0.7	7.4	8.1
PDN [6]	157	197	99.3	98.9	37.9	44.5	202	219	299	397	0.58	0.76	150	186	0.7	36.0	36.8
DG [1]	56	87	3.2	6.3	30.9	35.2	265	285	372	386	0.66	0.77	166	180	0.0	3.7	3.7
DIP [7]	49	52	8.0	4.9	85.5	84.7	796	812	821	924	1.19	1.18	267	250	0.0	0.0	0.0
MSG [5]	41	41	1.3	3.0	39.6	66.2	264	458	493	612	0.64	0.74	181	213	0.0	1.5	1.5
DIP-v	38	42	1.7	2.2	48.0	60.4	238	351	456	516	0.50	0.61	128	152	0.7	**47.8**	48.5
MSG-v	**16**	**26**	**0.7**	**1.2**	**8.5**	**17.5**	**76**	**92**	**156**	**181**	**0.27**	**0.31**	**97**	**100**	**97.8**	1.5	**99.3**

Table 3: Quantitative evaluation on “Sphere and cylinder” with different RGBs from SimGeo dataset. The best result is in bold, the second best is underlined.

	Lucy																
	$\mathrm{RMSE_d}$		$\mathrm{BadPix_d(5cm)}$		$\mathrm{BadPix_v(5)}$		$\mathrm{DSSIM_v}$		$\mathrm{LPIPS_v}$		$\mathrm{Bumpiness_d}$		$\mathrm{RMSE_v}$		User, 1st	User, 2nd	Top 2
	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x4	x4
Bicubic	72	103	6.8	13.0	48.8	65.0	355	519	398	497	1.37	1.74	267	328	2.2	24.3	26.5
SRfS [2]	82	113	13.2	20.8	84.6	87.1	811	857	781	792	1.90	2.28	367	407	0.0	0.0	0.0
EG [9]	69		3.5		56.2		357		426		1.05		220		0.0	0.7	0.7
PDN [6]	173	234	99.0	98.8	64.9	68.9	456	535	368	480	1.24	1.47	251	303	0.0	1.5	1.5
DG [1]	69	108	4.9	11.0	65.5	68.6	523	562	558	565	1.28	1.50	249	281	0.0	0.7	0.7
DIP [7]	53	75	4.7	11.4	87.4	95.2	827	908	615	778	2.02	2.93	344	478	0.7	0.7	1.5
MSG [5]	54	53	2.7	5.4	62.9	71.7	444	577	480	578	1.30	1.42	259	306	1.5	13.2	14.7
DIP-v	**44**	55	4.6	4.4	69.0	77.5	421	574	446	468	1.15	1.27	223	239	0.0	**56.6**	56.6
MSG-v	74	**47**	**1.6**	**3.7**	**38.8**	**55.0**	**205**	**325**	**251**	**348**	**0.82**	**0.96**	**156**	**195**	**95.6**	2.2	**97.8**

Table 4: Quantitative evaluation on "Lucy" from SimGeo dataset. The best result is in bold, the second best is underlined.

	Painting																
	$\mathrm{RMSE_d}$		$\mathrm{BadPix_d(5cm)}$		$\mathrm{BadPix_v(5)}$		$\mathrm{DSSIM_v}$		$\mathrm{LPIPS_v}$		$\mathrm{Bumpiness_d}$		$\mathrm{RMSE_v}$		User, 1st	User, 2nd	Top 2
	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x4	x4
Bicubic	28	47	2.5	5.6	57.1	64.1	423	514	544	649	0.95	1.15	213	265	4.4	**47.8**	52.2
SRfS [2]	39	60	6.5	15.9	78.4	81.2	707	722	612	661	1.47	1.55	308	337	0.0	0.0	0.0
EG [9]	36		3.1		61.9		481		720		0.94		231		0.0	3.7	3.7
PDN [6]	151	215	99.3	99.2	65.2	70.2	488	532	669	709	0.89	1.01	237	275	4.4	10.3	14.7
DG [1]	31	49	2.4	5.5	61.9	63.9	503	506	678	700	1.08	1.13	232	272	0.7	3.7	4.4
DIP [7]	30	37	4.0	4.7	80.4	79.5	802	766	630	612	2.18	1.82	362	341	0.0	0.0	0.0
MSG [5]	21	**29**	1.2	2.2	63.7	67.9	495	570	475	507	0.97	1.12	203	243	2.2	5.1	7.4
DIP-v	22	32	2.3	3.2	70.1	70.3	567	564	**386**	**501**	1.07	1.12	210	239	2.9	21.3	24.3
MSG-v	**17**	34	**0.9**	**1.8**	**51.4**	**58.0**	**354**	**410**	532	607	**0.67**	**0.77**	**142**	**170**	**85.3**	8.1	**93.4**
	Sofa																
Bicubic	38	58	1.8	3.6	75.4	77.0	566	616	704	764	2.12	2.33	212	250	3.7	15.4	19.1
SRfS [2]	39	58	2.0	3.5	82.3	88.1	715	832	631	743	2.97	3.45	310	405	0.0	0.0	0.0
EG [9]	42		2.5		79.0		598		767		2.28		213		0.0	8.8	8.8
PDN [6]	86	91	71.0	70.8	83.3	83.0	641	658	784	763	2.40	2.50	260	264	0.7	3.7	4.4
DG [1]	41	63	3.2	4.4	77.7	77.9	624	632	823	855	2.30	2.33	255	263	0.0	5.1	5.1
DIP [7]	45	57	7.1	12.7	93.1	94.0	928	946	758	738	3.91	3.99	518	560	0.0	0.0	0.0
MSG [5]	**27**	**36**	1.2	2.3	80.6	85.7	718	791	606	610	2.71	3.22	254	316	0.0	0.7	0.7
DIP-v	**27**	43	0.9	2.0	79.1	82.5	645	718	**414**	**585**	2.67	3.07	215	266	19.1	**47.8**	66.9
MSG-v	35	44	**0.7**	**1.6**	**74.0**	**75.7**	**537**	**585**	710	759	**1.96**	**2.10**	**165**	**196**	**76.5**	18.4	**94.9**
	Plant																
Bicubic	38	58	3.7	6.4	75.9	79.9	562	610	688	763	1.58	1.79	249	290	1.5	22.1	23.5
SRfS [2]	46	65	5.8	9.5	82.9	85.0	658	692	632	649	1.96	2.13	280	309	0.0	0.0	0.0
EG [9]	43		4.5		82.2		568		677		1.64		255		0.0	0.7	0.7
PDN [6]	88	89	94.5	37.8	79.5	82.5	574	612	659	699	1.46	1.60	269	305	4.4	7.4	11.8
DG [1]	40	63	3.9	6.7	79.5	81.1	611	622	745	785	1.67	1.70	268	291	2.2	11.0	13.2
DIP [7]	38	47	6.9	6.1	93.9	92.8	919	880	764	723	4.33	3.95	490	437	0.0	0.7	0.7
MSG [5]	31	44	2.3	**3.7**	78.0	81.8	571	645	582	**495**	1.62	1.84	234	285	0.0	11.8	11.8
DIP-v	31	**40**	4.7	4.8	83.5	84.1	694	707	**463**	555	2.25	2.21	262	276	11.0	**33.1**	44.1
MSG-v	**27**	44	**1.8**	3.9	**74.3**	**77.8**	**524**	**575**	639	720	**1.31**	**1.47**	**194**	**236**	**80.9**	13.2	**94.1**

Table 5: Quantitative evaluation on RGBD frames from ICL-NUIM "Living Room" sequence. The best result is in bold, the second best is underlined.

	RMSE$_d$		BadPix$_d$(5cm)		BadPix$_v$(5)		DSSIM$_v$		LPIPS$_v$		Bumpiness$_d$		RMSE$_v$		User, 1st	User, 2nd	Top 2
	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x4	x4
Office																	
Bicubic	47	80	4.0	7.8	24.4	34.1	216	285	412	594	0.81	0.95	208	254	19.9	**44.1**	64.0
SRfS [2]	49	89	5.8	14.4	53.4	54.4	595	593	690	636	1.71	1.66	298	302	0.0	0.0	0.0
PDN [6]	185	185	99.3	90.5	36.5	50.2	250	294	457	518	0.76	0.92	234	272	0.7	3.7	4.4
DG [1]	49	85	9.0	11.9	36.5	37.6	319	330	534	571	1.03	1.05	240	266	0.0	0.7	0.7
DIP [7]	76	109	30.3	48.2	72.1	73.9	726	819	690	797	2.45	2.70	372	408	1.5	1.5	2.9
MSG [5]	35	**48**	2.4	6.8	35.4	44.5	263	360	415	543	0.83	0.95	199	247	2.2	3.7	5.9
DIP-v	40	65	3.8	7.4	45.4	47.9	311	352	414	504	1.08	1.18	205	235	17.6	25.0	42.6
MSG-v	**32**	65	**1.9**	**5.3**	**19.3**	**29.6**	**157**	**224**	**313**	**432**	**0.59**	**0.72**	**151**	**198**	**58.1**	21.3	**79.4**
Coat rack																	
Bicubic	13	20	1.5	3.0	73.1	75.3	507	539	537	651	0.54	0.60	171	196	0.0	19.1	19.1
SRfS [2]	24	28	3.8	5.4	82.3	80.5	672	556	650	612	0.83	0.57	237	203	0.0	0.0	0.0
EG [9]	13		1.2		77.8		541		550		0.55		186		0.7	7.4	8.1
PDN [6]	140	191	99.6	99.9	77.1	78.0	544	557	621	631	0.48	0.50	178	193	5.1	**28.7**	33.8
DG [1]	13	20	1.4	3.2	74.4	75.6	530	532	593	621	0.54	0.58	166	201	0.0	9.6	9.6
DIP [7]	15	24	1.9	3.5	85.5	85.4	766	701	625	624	1.15	0.97	256	246	4.4	2.2	6.6
MSG [5]	**11**	**17**	0.9	**1.6**	73.3	76.1	522	546	523	554	0.51	0.55	165	189	2.2	16.2	18.4
DIP-v	13	**17**	1.8	2.0	75.2	75.5	543	542	**422**	**463**	0.62	0.60	171	181	0.7	12.5	13.2
MSG-v	**11**	18	**0.8**	2.2	**71.4**	**74.2**	**482**	**502**	516	563	**0.42**	**0.48**	**136**	**161**	**86.8**	4.4	**91.2**
Displays																	
Bicubic	41	63	3.2	6.4	49.9	54.9	315	374	460	585	0.92	1.08	208	256	0.7	21.3	22.1
SRfS [2]	53	75	9.0	17.3	61.9	67.3	500	591	599	659	1.35	1.60	288	328	0.0	0.0	0.0
EG [9]	46		5.9		66.7		388		587		0.94		216		0.0	2.9	2.9
PDN [6]	159	220	99.2	99.0	55.4	57.2	381	403	547	580	0.85	0.95	242	275	0.0	9.6	9.6
DG [1]	43	66	5.8	6.7	56.5	56.7	395	406	606	601	1.06	1.10	243	265	0.7	2.9	3.7
DIP [7]	52	60	13.4	9.7	76.9	74.6	732	724	672	645	2.36	2.06	365	344	0.7	0.7	1.5
MSG [5]	26	**42**	1.7	4.4	53.9	58.0	367	430	461	493	0.97	1.08	204	251	0.0	5.9	5.9
DIP-v	32	45	2.4	4.0	53.7	57.6	336	407	**344**	**409**	1.00	1.18	191	221	5.9	**51.5**	57.4
MSG-v	**23**	43	**1.4**	**3.5**	**47.2**	**51.0**	**271**	**324**	451	531	**0.69**	**0.80**	**152**	**190**	**91.9**	5.1	**97.1**

Table 6: Quantitative evaluation on RGBD frames from ICL-NUIM “Office Room” sequence. The best result is in bold, the second best is underlined.

	RMSE$_d$		BadPix$_d$(5cm)		BadPix$_v$(5)		DSSIM$_v$		LPIPS$_v$		Bumpiness$_d$		RMSE$_v$		User, 1st	User, 2nd	Top 2
	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x4	x4
Vintage																	
Bicubic	67	98	4.6	9.0	72.8	**77.3**	558	649	602	729	1.51	1.64	258	302	5.9	28.7	34.6
SRfS [2]	101	145	16.8	32.3	83.7	87.2	721	749	631	634	1.64	1.68	346	382	0.0	0.0	0.0
PDN [6]	140	174	67.6	79.0	82.3	85.7	663	714	706	700	1.51	1.57	319	350	0.0	0.0	0.0
DG [1]	72	107	7.1	10.4	79.4	80.1	666	669	796	840	1.50	1.52	290	300	0.0	0.7	0.7
DIP [7]	74	117	24.8	46.9	93.6	94.2	953	965	910	872	4.01	4.16	656	687	0.7	0.7	1.5
MSG [5]	41	**59**	3.2	6.8	80.6	84.6	708	785	**510**	**610**	1.62	1.85	292	364	0.0	9.6	9.6
DIP-v	42	67	2.7	**5.9**	85.2	88.8	804	884	579	674	1.94	2.48	343	435	25.7	**44.1**	69.9
MSG-v	**33**	65	**2.5**	**5.9**	**71.4**	77.6	**536**	**643**	670	702	**1.29**	**1.43**	**211**	**268**	**67.6**	16.2	**83.8**
Recycle																	
Bicubic	587	880	9.2	16.6	**70.6**	**78.6**	**575**	721	474	576	**1.23**	**1.17**	329	398	0.0	11.0	11.0
SRfS [2]	47	72	10.2	22.1	86.1	88.8	715	772	610	623	1.68	1.81	376	410	0.0	0.0	0.0
PDN [6]	95	128	90.5	79.8	84.0	85.7	635	**701**	523	589	1.66	2.18	364	457	0.0	6.6	6.6
DG [1]	39	82	3.4	11.7	81.6	83.6	696	719	602	617	1.75	1.99	328	383	2.9	**65.4**	68.4
DIP [7]	29	45	3.9	9.3	91.0	91.8	871	923	576	605	2.95	3.31	434	500	1.5	5.9	7.4
MSG [5]	106	1182	5.8	11.9	82.8	89.6	741	869	624	661	2.60	3.01	485	550	0.7	0.0	0.7
DIP-v	**20**	**34**	**1.5**	**4.2**	78.9	85.0	**575**	735	**388**	**485**	1.56	1.86	**273**	**332**	**94.9**	3.7	**98.5**
MSG-v	51	76	3.9	7.9	73.9	82.1	603	737	520	564	1.66	2.02	368	473	0.0	7.4	7.4

Table 7: Quantitative evaluation on samples with small number of missing measurements from Middlebury dataset. The best result is in bold, the second best is underlined.

	$\mathrm{RMSE_d}$		$\mathrm{BadPix_d(5cm)}$		$\mathrm{BadPix_v(5)}$		$\mathrm{DSSIM_v}$		$\mathrm{LPIPS_v}$		$\mathrm{Bumpiness_d}$		$\mathrm{RMSE_v}$		User, 1st	User, 2nd	Top 2
	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x4	x4
Umbrella																	
Bicubic	1013	1507	6.9	12.1	**77.7**	**80.9**	**749**	837	747	886	**0.60**	**0.60**	323	380	5.9	**35.3**	41.2
SRfS [2]	148	217	19.4	35.5	87.5	90.8	843	853	797	831	0.71	0.78	397	443	0.0	0.0	0.0
PDN [6]	220	287	94.9	89.1	86.6	88.1	799	**828**	847	882	0.79	1.13	367	452	3.7	22.8	26.5
DG [1]	365	507	9.1	20.3	84.6	87.3	846	878	781	856	0.92	1.36	399	457	0.0	0.7	0.7
DIP [7]	138	145	48.5	21.6	90.5	93.2	915	953	737	722	1.19	1.65	467	528	2.9	16.2	19.1
MSG [5]	292	555	7.4	12.4	84.3	88.1	834	896	678	787	1.27	1.47	442	496	0.0	0.7	0.7
DIP-v	**91**	**129**	**3.4**	**5.7**	83.4	85.3	796	854	**604**	**598**	0.67	0.79	**318**	**352**	**82.4**	8.1	**90.4**
MSG-v	129	218	5.2	9.7	79.1	82.3	778	842	800	890	0.72	0.89	348	427	5.1	16.2	21.3
Classroom1																	
Bicubic	966	1371	6.7	9.0	**75.8**	**78.3**	**636**	**728**	581	784	**0.41**	**0.30**	268	**295**	12.5	**37.5**	50.0
SRfS [2]	135	202	18.5	28.5	82.6	85.7	761	781	718	756	0.62	0.62	332	363	0.0	0.0	0.0
PDN [6]	239	324	96.0	91.0	81.5	82.9	739	759	751	807	0.62	0.76	279	342	16.9	26.5	43.4
DG [1]	307	503	8.8	16.8	82.0	82.7	743	762	766	812	0.74	0.87	313	337	0.0	1.5	1.5
DIP [7]	96	145	17.0	22.4	94.4	94.6	956	952	789	751	1.94	2.12	540	557	0.0	1.5	1.5
MSG [5]	297	408	7.3	10.0	81.2	83.8	723	810	626	604	0.90	1.01	351	391	0.0	0.7	0.7
DIP-v	**69**	**117**	**4.1**	9.3	81.0	86.0	700	789	**516**	**537**	0.64	0.86	**266**	327	**64.0**	18.4	**82.4**
MSG-v	127	203	5.4	**8.4**	76.9	79.4	678	735	739	803	0.60	0.64	283	330	2.2	11.8	14.0

Table 8: Quantitative evaluation on samples with small number of missing measurements from Middlebury dataset. The best result is in bold, the second best is underlined.

	$\mathrm{RMSE_d}$		$\mathrm{BadPix_d(5cm)}$		$\mathrm{BadPix_v(5)}$		$\mathrm{DSSIM_v}$		$\mathrm{LPIPS_v}$		$\mathrm{Bumpiness_d}$		$\mathrm{RMSE_v}$		User, 1st	User, 2nd	Top 2
	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x8	x4	x4	x4
Playroom																	
Bicubic	1263	1744	14.4	20.4	**72.0**	**76.9**	**684**	**783**	509	675	**0.80**	**0.52**	386	441	0.0	2.2	2.2
SRfS [2]	97	151	26.9	42.1	88.1	91.2	802	829	663	715	1.24	1.08	493	540	0.0	0.0	0.0
PDN [6]	181	253	85.0	69.7	86.3	89.2	820	862	583	656	1.54	1.88	472	543	25.7	**61.8**	87.5
DG [1]	425	133	22.4	25.0	85.4	86.0	845	826	779	691	1.96	1.63	519	469	1.5	2.2	3.7
DIP [7]	58	91	18.4	20.0	93.0	93.2	941	937	647	612	3.09	2.86	602	592	1.5	2.9	4.4
MSG [5]	433	349	16.1	22.3	85.8	89.9	855	911	685	705	2.51	2.74	576	616	0.0	0.0	0.0
DIP-v	**49**	**83**	**5.4**	**12.2**	83.8	88.5	728	847	**459**	**530**	1.29	1.52	**357**	**433**	**70.6**	27.2	**97.8**
MSG-v	112	166	9.4	15.5	75.2	80.1	721	810	565	615	1.46	1.67	453	510	0.0	3.7	3.7
Backpack																	
Bicubic	985	1078	14.3	11.5	**62.7**	**69.4**	**639**	**730**	564	692	**0.60**	**0.45**	**392**	**424**	2.2	34.6	36.8
SRfS [2]	69	83	18.9	25.5	89.9	89.9	831	847	630	651	1.37	1.26	500	505	0.0	0.0	0.0
PDN [6]	173	207	81.6	65.2	80.4	85.4	770	820	609	719	1.59	1.96	519	553	3.7	**37.5**	41.2
DG [1]	325	465	26.5	39.9	77.0	82.0	765	808	650	696	1.62	2.07	529	545	0.7	2.9	3.7
DIP [7]	41	67	8.2	17.9	93.2	94.5	943	984	766	692	3.36	2.95	639	645	1.5	11.8	13.2
MSG [5]	211	170	15.1	10.4	76.5	86.9	762	856	671	723	2.11	2.31	577	609	0.7	0.0	0.7
DIP-v	**38**	**62**	**6.2**	12.9	82.5	88.7	677	768	**457**	**496**	1.31	1.43	409	448	**90.4**	5.1	**95.6**
MSG-v	113	89	10.8	**5.9**	65.3	72.5	663	752	577	635	1.07	1.15	462	480	0.0	5.9	5.9
Jadeplant																	
Bicubic	1017	1297	19.3	18.4	**68.8**	**75.7**	695	788	545	696	**0.97**	**0.62**	**449**	**464**	2.3	27.7	30.0
SRfS [2]	105	143	39.5	48.9	87.2	92.7	787	839	637	719	1.96	1.70	551	583	0.0	0.0	0.0
PDN [6]	161	205	81.8	62.0	82.4	88.0	778	849	551	625	1.95	2.20	512	572	19.1	**41.4**	60.5
DG [1]	326	512	27.8	47.6	82.6	86.8	791	823	718	670	2.28	2.66	567	601	0.0	0.5	0.5
DIP [7]	**70**	121	16.7	32.8	91.2	92.3	913	911	735	764	3.19	3.21	615	638	0.0	0.5	0.5
MSG [5]	216	263	21.1	17.8	81.5	87.6	796	880	751	783	2.73	2.90	614	649	0.0	0.0	0.0
DIP-v	84	121	21.8	24.0	86.6	89.5	820	870	**542**	654	1.92	1.99	503	535	**78.2**	20.5	**98.6**
MSG-v	109	**117**	**13.9**	**11.2**	71.0	79.0	**688**	**781**	605	**622**	1.61	1.66	507	529	0.5	8.2	8.6

Table 9: Quantitative evaluation on samples with large number of missing measurements from Middlebury dataset. The best result is in bold, the second best is underlined.

References

[1] Shuhang Gu, Wangmeng Zuo, Shi Guo, Yunjin Chen, Chongyu Chen, and Lei Zhang. Learning dynamic guidance for depth image enhancement. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3769–3778, 2017. 3, 12, 13, 14, 15

[2] Bjoern Haefner, Yvain Quéau, Thomas Möllenhoff, and Daniel Cremers. Fight ill-posedness with ill-posedness: Single-shot variational depth super-resolution from shading. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 164–174, 2018. 3, 12, 13, 14, 15

[3] Katrin Honauer, Ole Johannsen, Daniel Kondermann, and Bastian Goldluecke. A dataset and evaluation methodology for depth estimation on 4d light fields. In *Asian Conference on Computer Vision*, pages 19–34. Springer, 2016. 1

[4] Katrin Honauer, Lena Maier-Hein, and Daniel Kondermann. The hci stereo metrics: Geometry-aware performance analysis of stereo algorithms. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2120–2128, 2015. 1

[5] Tak-Wai Hui, Chen Change Loy, and Xiaoou Tang. Depth map super-resolution by deep multi-scale guidance. In *European Conference on Computer Vision*, pages 353–369. Springer, 2016. 3, 12, 13, 14, 15

[6] David Riegler, Gernot aand Ferstl, Matthias Rüther, and Horst Bischof. A deep primal-dual network for guided depth super-resolution. In *British Machine Vision Conference*. The British Machine Vision Association, 2016. 3, 12, 13, 14, 15

[7] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 3, 12, 13, 14, 15

[8] Stefan Van der Walt, Johannes L Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. scikit-image: image processing in python. *PeerJ*, 2:e453, 2014. 1

[9] Jun Xie, Rogerio Schmidt Feris, and Ming-Ting Sun. Edge-guided single depth image super resolution. *IEEE Transactions on Image Processing*, 25(1):428–438, 2016. 3, 12, 13, 14

[10] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 1