

Models	Scaling	Dataset	$\delta < 1.25 \uparrow$	$\delta < 1.25^2 \uparrow$	$\delta < 1.25^3 \uparrow$	Abs Rel $\downarrow$	RMSE _{log} $\downarrow$	RMSE $\downarrow$
DPT	Linear Fit	DDAD	0.802	0.954	0.990	0.163	0.254	10.342
	Global	KITTI	0.752	0.925	0.969	0.183	0.312	15.967
	Image	NYUv2,KITTI	0.763	0.931	0.975	0.179	0.308	14.468
	RSA (Ours)	NYUv2,KITTI	0.777	0.938	0.981	0.171	0.284	13.539
MiDas	Linear Fit	DDAD	0.664	0.912	0.973	0.209	0.301	18.341
	Global	KITTI	0.603	0.864	0.925	0.253	0.336	20.594
	Image	NYUv2,KITTI	0.616	0.883	0.934	0.231	0.331	20.034
	RSA (Ours)	NYUv2,KITTI	0.631	0.903	0.966	0.223	0.325	19.342
DepthAnything	Linear Fit	DDAD	0.673	0.932	0.983	0.182	0.286	18.423
	Global	KITTI	0.612	0.883	0.963	0.221	0.323	21.345
	Image	NYUv2,KITTI	0.623	0.890	0.968	0.217	0.316	20.834
	RSA (Ours)	NYUv2,KITTI	0.642	0.903	0.976	0.207	0.303	19.715

Table 1: **Zero-shot generalization to DDAD.** Evaluate on the DDAD evaluation set without any fine-tuning.

Models	Scaling	Dataset	$\delta < 1.25 \uparrow$	$\delta < 1.25^2 \uparrow$	$\delta < 1.25^3 \uparrow$	Abs Rel $\downarrow$	log ₁₀ $\downarrow$	RMSE $\downarrow$
DPT	Linear Fit	VOID	0.782	0.962	0.990	0.150	0.064	0.340
	Global	NYUv2	0.456	0.743	0.912	0.312	0.136	0.896
	Image	NYUv2,KITTI	0.516	0.812	0.936	0.289	0.112	0.634
	RSA (Ours)	NYUv2,KITTI	0.601	0.886	0.970	0.254	0.096	0.444
MiDas	Linear Fit	VOID	0.500	0.781	0.899	0.347	0.130	0.829
	Global	NYUv2	0.268	0.597	0.735	0.512	0.193	1.346
	Image	NYUv2,KITTI	0.304	0.626	0.812	0.487	0.159	0.913
	RSA (Ours)	NYUv2,KITTI	0.392	0.696	0.892	0.448	0.148	0.660
DepthAnything	Linear Fit	VOID	0.249	0.465	0.643	0.682	0.254	1.251
	Global	NYUv2	0.084	0.194	0.376	1.674	0.389	2.046
	Image	NYUv2,KITTI	0.093	0.215	0.412	1.497	0.345	1.963
	RSA (Ours)	NYUv2,KITTI	0.104	0.262	0.450	1.287	0.323	1.716

Table 2: **Zero-shot generalization to VOID.** Evaluate on the VOID evaluation set without any fine-tuning.

Models	Scaling	Dataset	$\delta < 1.25 \uparrow$	$\delta < 1.25^2 \uparrow$	$\delta < 1.25^3 \uparrow$	Abs Rel $\downarrow$	log ₁₀ $\downarrow$	RMSE $\downarrow$
ZeroDepth (from main paper)	DA	NYUv2	0.901	0.961	-	0.100	-	0.380
ZeroDepth (from main paper)	DA,Median	NYUv2	0.926	0.986	-	0.081	-	0.338
3D-2D Object Sizes	Mean	NYUv2,KITTI	0.559	0.865	0.964	0.229	0.105	0.723
3D-2D Object Sizes	Median	NYUv2,KITTI	0.588	0.893	0.974	0.227	0.097	0.688
DPT	RSA (Ours)	NYUv2	0.916	0.990	0.998	0.097	0.042	0.347
	RSA (Ours)	NYUv2,KITTI	0.913	0.988	0.998	0.099	0.042	0.352
MiDas	RSA (Ours)	NYUv2	0.731	0.955	0.993	0.171	0.072	0.569
	RSA (Ours)	NYUv2,KITTI	0.737	0.959	0.993	0.168	0.071	0.561
DepthAnything	RSA (Ours)	NYUv2	0.775	0.975	0.997	0.147	0.065	0.484
	RSA (Ours)	NYUv2,KITTI	0.776	0.974	0.996	0.148	0.065	0.498

Table 3: **Comparing with works using object scales on NYU-Depth-v2.**

Models	Scaling	Dataset	$\delta < 1.25 \uparrow$	$\delta < 1.25^2 \uparrow$	$\delta < 1.25^3 \uparrow$	Abs Rel $\downarrow$	RMSE _{log} $\downarrow$	RMSE $\downarrow$
ZeroDepth (from main paper)	DA	KITTI	0.892	0.961	0.977	0.102	0.196	4.378
ZeroDepth (from main paper)	DA,Median	KITTI	0.886	0.965	0.984	0.105	0.178	4.194
3D-2D Object Sizes	Mean	NYUv2,KITTI	0.537	0.837	0.943	0.255	0.312	6.951
3D-2D Object Sizes	Median	NYUv2,KITTI	0.566	0.864	0.968	0.253	0.273	5.600
DPT	RSA (Ours)	KITTI	0.963	0.995	0.999	0.061	0.090	2.354
	RSA (Ours)	NYUv2,KITTI	0.962	0.994	0.998	0.060	0.089	2.342
MiDas	RSA (Ours)	KITTI	0.798	0.948	0.981	0.163	0.185	4.082
	RSA (Ours)	NYUv2,KITTI	0.782	0.946	0.980	0.160	0.194	4.232
DepthAnything	RSA (Ours)	KITTI	0.780	0.958	0.988	0.160	0.189	4.437
	RSA (Ours)	NYUv2,KITTI	0.756	0.956	0.987	0.158	0.191	4.457

Table 4: **Comparing with works using object scales on KITTI Eigen Split.**

Models	Scaling	Dataset	$\delta < 1.25 \uparrow$	$\delta < 1.25^2 \uparrow$	$\delta < 1.25^3 \uparrow$	Abs Rel $\downarrow$	log ₁₀ $\downarrow$	RMSE $\downarrow$
DPT	Objects + Background	NYUv2,KITTI	0.911	0.989	0.998	0.102	0.043	0.350
	Detector feature	NYUv2,KITTI	0.903	0.989	0.998	0.101	0.045	0.366
	RSA (Ours)	NYUv2,KITTI	0.913	0.988	0.998	0.099	0.042	0.352
MiDas	Objects + Background	NYUv2,KITTI	0.745	0.970	0.996	0.160	0.069	0.531
	Detector feature	NYUv2,KITTI	0.714	0.963	0.995	0.181	0.073	0.554
	RSA (Ours)	NYUv2,KITTI	0.737	0.959	0.993	0.168	0.071	0.561
DepthAnything	Objects + Background	NYUv2,KITTI	0.799	0.980	0.998	0.141	0.061	0.472
	Background only	NYUv2,KITTI	0.645	0.931	0.988	0.185	0.084	0.613
	Detector feature	NYUv2,KITTI	0.705	0.961	0.995	0.161	0.074	0.533
	Coordinates of objects	NYUv2,KITTI	0.722	0.959	0.996	0.180	0.073	0.560
	Intrinsics of the camera	NYUv2,KITTI	0.722	0.965	0.995	0.162	0.072	0.541
	Configuration of the objects	NYUv2,KITTI	0.761	0.965	0.995	0.153	0.067	0.518
	RSA with 3-layer Transformers	NYUv2,KITTI	0.732	0.964	0.997	0.177	0.071	0.553
	RSA with 5-layer Transformers	NYUv2,KITTI	0.740	0.965	0.994	0.150	0.069	0.510
	RSA with 5-layer MLP	NYUv2,KITTI	0.764	0.970	0.995	0.150	0.066	0.503
	RSA with 3-layer MLP (Ours)	NYUv2,KITTI	0.776	0.974	0.996	0.148	0.065	0.498

Table 5: **Additional experiments on NYU-Depth-v2.**

Models	Scaling	Dataset	$\delta < 1.25 \uparrow$	$\delta < 1.25^2 \uparrow$	$\delta < 1.25^3 \uparrow$	Abs Rel $\downarrow$	RMSE _{log} $\downarrow$	RMSE $\downarrow$
DPT	Objects + Background	NYUv2,KITTI	0.957	0.995	0.999	0.069	0.095	2.397
	Detector feature	NYUv2,KITTI	0.918	0.989	0.998	0.108	0.135	2.997
	RSA (Ours)	NYUv2,KITTI	0.962	0.994	0.998	0.060	0.089	2.342
MiDas	Objects + Background	NYUv2,KITTI	0.771	0.946	0.989	0.159	0.191	4.198
	Detector feature	NYUv2,KITTI	0.735	0.934	0.985	0.168	0.202	4.468
	RSA (Ours)	NYUv2,KITTI	0.782	0.946	0.980	0.160	0.194	4.232
DepthAnything	Objects + Background	NYUv2,KITTI	0.792	0.951	0.990	0.156	0.181	4.097
	Background only	NYUv2,KITTI	0.673	0.935	0.982	0.184	0.219	5.134
	Detector feature	NYUv2,KITTI	0.708	0.934	0.989	0.173	0.210	4.660
	Coordinates of objects	NYUv2,KITTI	0.720	0.927	0.979	0.175	0.209	4.596
	Intrinsics of the camera	NYUv2,KITTI	0.734	0.946	0.990	0.163	0.198	4.459
	Configuration of the objects	NYUv2,KITTI	0.756	0.939	0.988	0.168	0.194	4.331
	RSA with 3-layer Transformers	NYUv2,KITTI	0.734	0.935	0.987	0.168	0.201	4.490
	RSA with 5-layer Transformers	NYUv2,KITTI	0.735	0.934	0.985	0.168	0.202	4.467
	RSA with 5-layer MLP	NYUv2,KITTI	0.757	0.936	0.986	0.174	0.195	4.275
	RSA with 3-layer MLP(Ours)	NYUv2,KITTI	0.756	0.956	0.987	0.158	0.191	4.457

Table 6: **Additional experiments on KITTI Eigen Split.**