

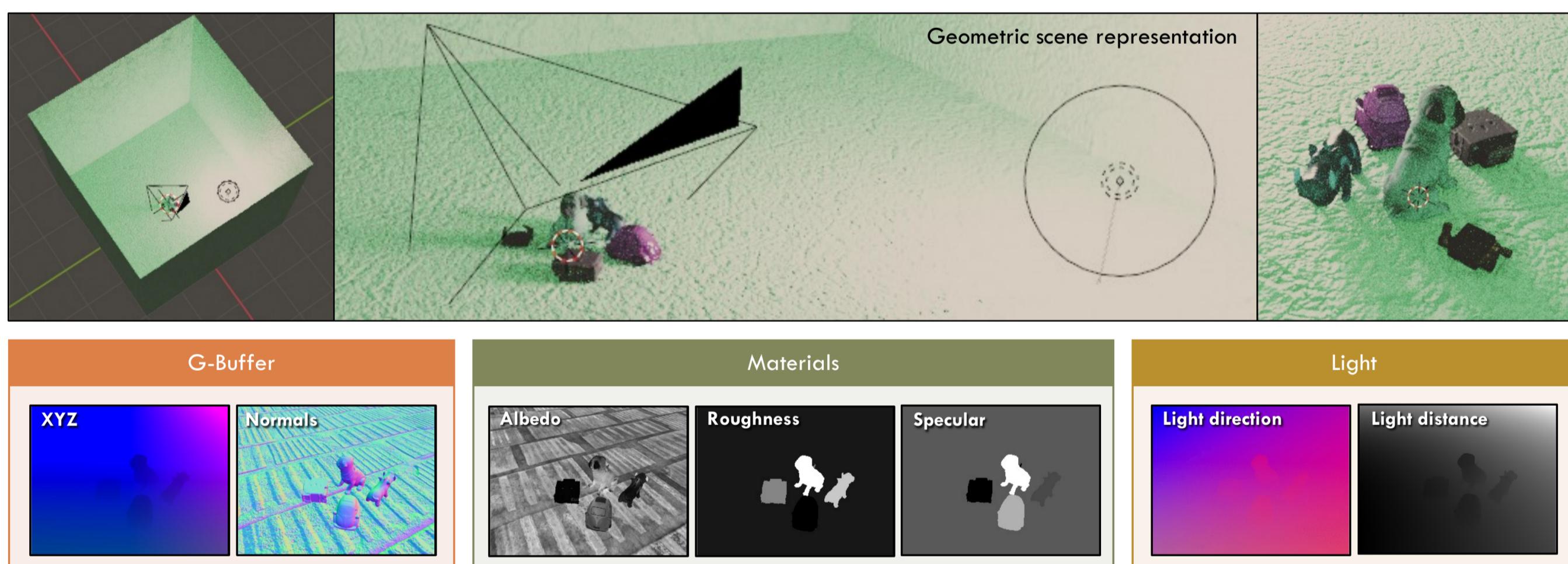
# Photo-realistic Neural Domain Randomization

Sergey Zakharov\*, Rareş Ambrus\*, Vitor Guizilini, Wadim Kehl, Adrien Gaidon

## Motivation

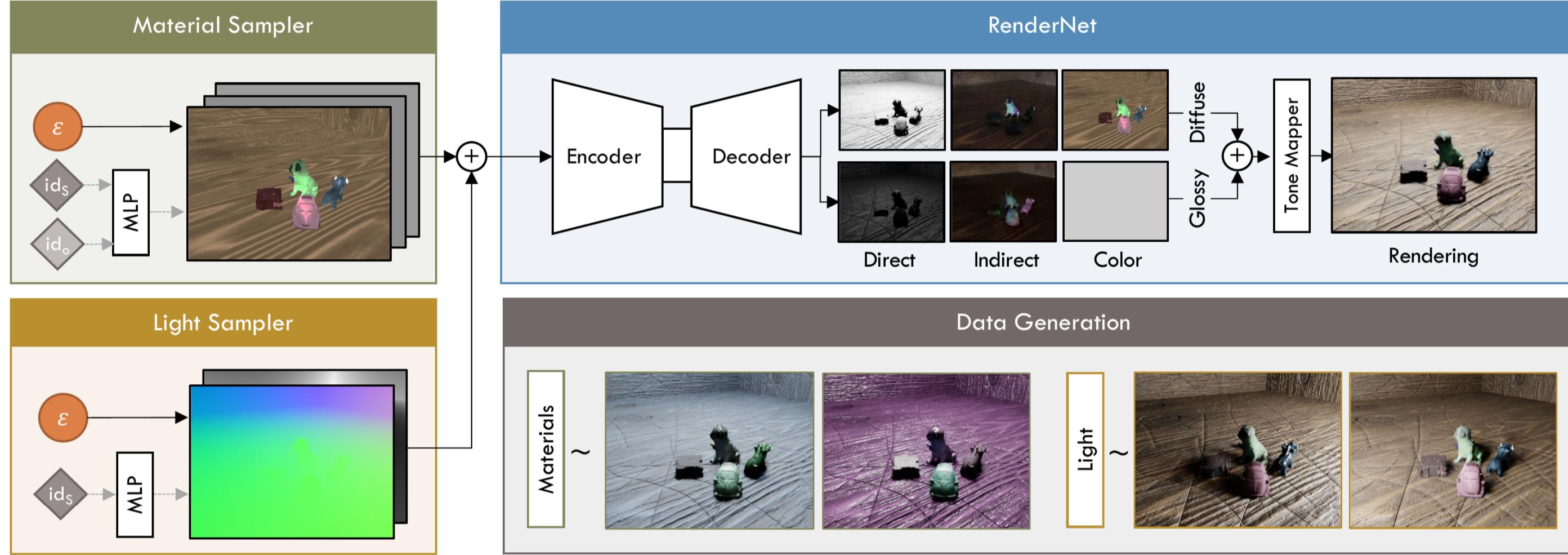
- We unify **photo-realistic rendering** and **domain randomization** for synthetic data generation
- Our learned deferred renderer, **RenderNet**, allows **flexible randomization of physical parameters** while being **1,600x** faster than comparable ray-tracers
- Our approach yields **state-of-the-art zero-shot sim-to-real transfer** for **6D object detection** and **monocular depth estimation**

## Geometric Representation



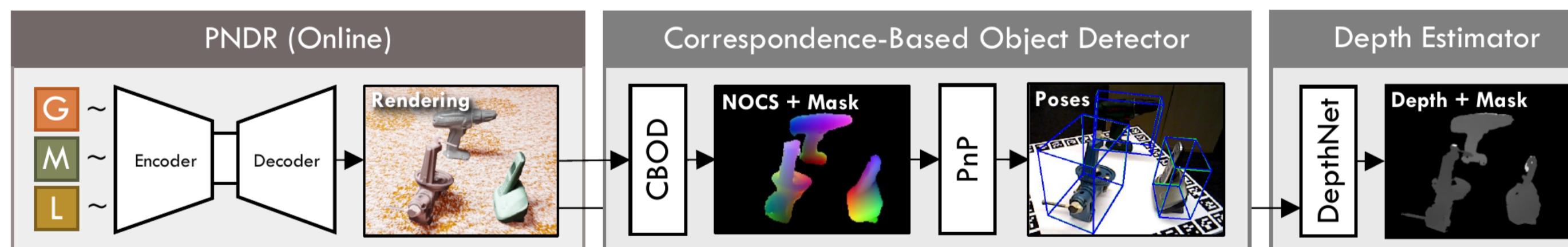
- We place 3D objects inside an empty room ensuring no collisions. **Random materials** are assigned to objects and walls and a randomly positioned light source illuminates the scene.
- Output buffers consist of **G-Buffer** (scene coordinates in camera space  $X$ , surface normals map  $N$ ), **material properties** (albedo  $A$ , roughness  $R$ , specularity  $S$ ), and **lighting** (light direction map  $L_{dir}$ , and light distance map  $L_{dist}$ )

## Method



- The main component of our domain randomization method is the ray tracer approximator (**RenderNet**). It takes a **G-buffer** as well as random **material maps** and **light maps** produced by corresponding samplers and generates intermediate light outputs. These outputs are then combined using a **tone mapper** to generate a final rendering.
- The lower-right row shows different material and light samples (e.g., roughness, specularity, light position).

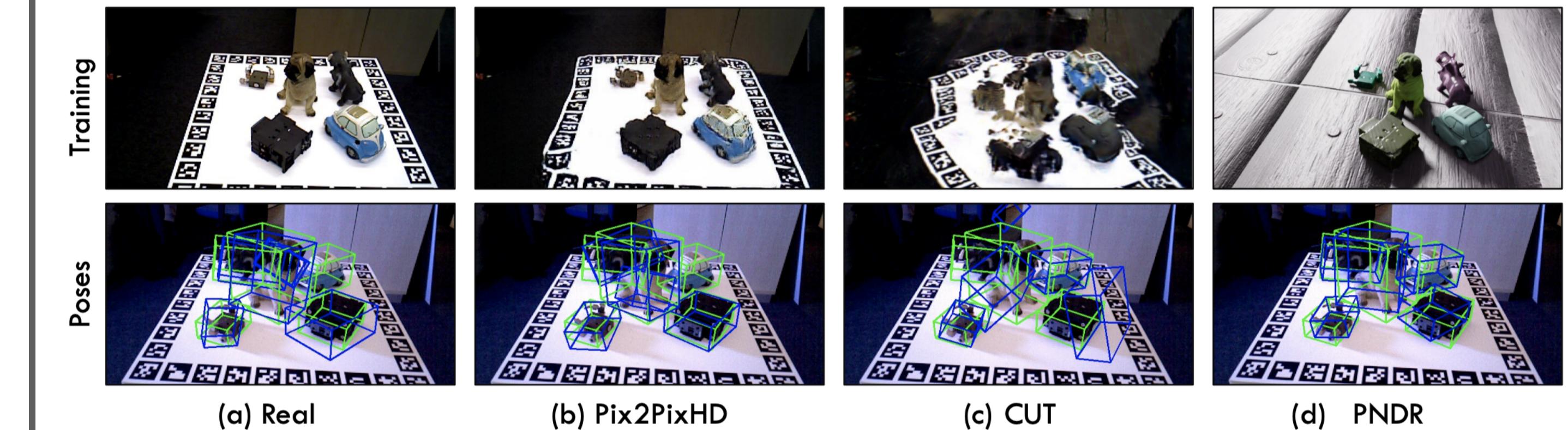
## Downstream Tasks



- During training, downstream tasks take PNDR renderings generated **online**, providing new realistic augmentations **at each iteration**.

## Results

**Dynamic Lighting Benchmark:** Training on photorealistic synthetic data is competitive with real data training and generalizes better to new domains.



Train	Method	HB Scene 5				HB Scene 10 (Lighting)							
		Car	P12	P15	Pumba	Dog	Mean	Car	P12	P15	Pumba	Dog	Mean
Real		91.76	94.12	79.34	94.41	95.00	90.93	26.69	35.88	8.75	25.22	14.63	22.24
Real	Pix2Pix [22]	81.84	76.47	38.53	76.76	94.12	73.54	37.13	35.59	13.75	24.78	33.82	29.01
+ CAD†	Pix2PixHD [56]	90.96	92.21	73.09	91.91	95.00	89.59	37.21	40.96	21.25	29.26	15.59	32.33
Real	CycleGAN [66]	49.41	31.10	24.12	40.88	73.24	43.75	27.72	2.28	6.91	7.06	11.47	11.09
+ CAD†	CUT [40]	56.10	28.97	29.34	41.47	85.29	66.49	27.35	4.63	7.35	8.90	12.13	25.36
RayTraced - 1088	85.59	86.76	61.18	89.71	94.85	83.62	47.28	36.84	9.12	36.25	23.38	30.57	
RayTraced - 2176	86.99	90.00	63.01	91.47	95.00	85.29	50.88	38.82	10.00	35.29	30.96	33.19	
RayTraced - 4352	89.71	88.97	66.91	92.35	95.00	86.59	52.43	38.75	10.29	41.47	43.82	37.35	
CAD	Ours - 1088	89.93	91.62	71.99	92.35	95.00	88.18	58.01	42.50	10.59	46.18	44.93	40.44

6D Object Detection

Method	Training	HB5			HB10		
		AbsRel↓	RMSE↓	a1↑	AbsRel↓	RMSE↓	a1↑
Monodepth2	Raytraced	0.082	0.09	0.951	0.162	0.148	0.805
	PNDR	<b>0.075</b>	<b>0.083</b>	<b>0.966</b>	<b>0.154</b>	<b>0.14</b>	<b>0.83</b>

Monocular Depth Estimation

## Generalization & Ablation

Train	Test	PSNR↑	SSIM↑	LPIPS↓
$HB_2$ - Train	$HB_2$ - Train	30.36	0.96	0.03
	$HB_2$ - Test	26.14	0.94	0.05
	$HB_5$ - Test	24.14	0.92	0.06
$HB_5$ - Train	$HB_5$ - Train	30.39	0.96	0.03
	$HB_5$ - Test	26.40	0.94	0.06
	$HB_2$ - Test	23.72	0.92	0.07

We achieve strong generalization performance when PNDR is applied on a completely different scene containing new objects and materials.

Modes	ADD <sub>AUC</sub> ↑	IoU ↑	Corr (mm) ↓
A	85.43	83.29	38.56
+ S + R	3%	3%	11%
Fixed	73.04	72.24	76.94
Dynamic	21%	19%	55%
$D_{dir}$ , $G_{dir}$	87.09	85.55	38.42
+ $D_{ind}$ , $G_{ind}$	1%	0%	11%
Full	88.18	85.97	34.32

We analyze the effect of different augmentations on downstream task performance. Lighting is by far the most important randomization parameter.

