

# NeuRAD: Neural Rendering for Autonomous Driving

Adam Tonderski†,1,4 Carl Lindstrom†,1,2 Georg Hess†,1,2 William Ljungbergh1,3 Lennart Svensson2 Christoffer Petersson1,2

- Problem/Objective

- 자율주행에 사용될 데이터 셋 생성
- 새로운 관점에서의 장면 생성

- Contribution/Key Idea

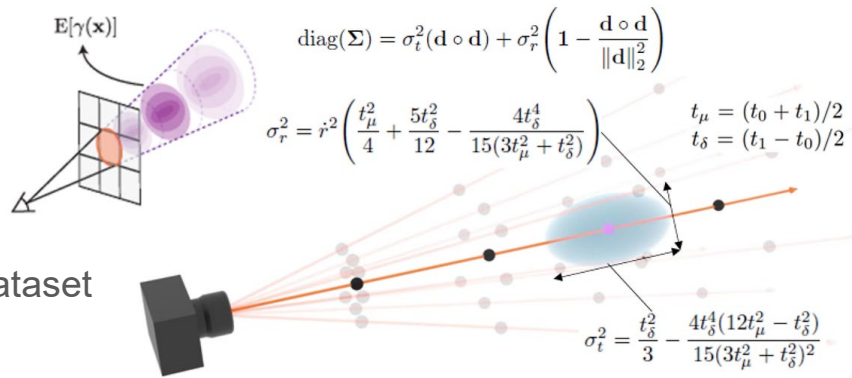
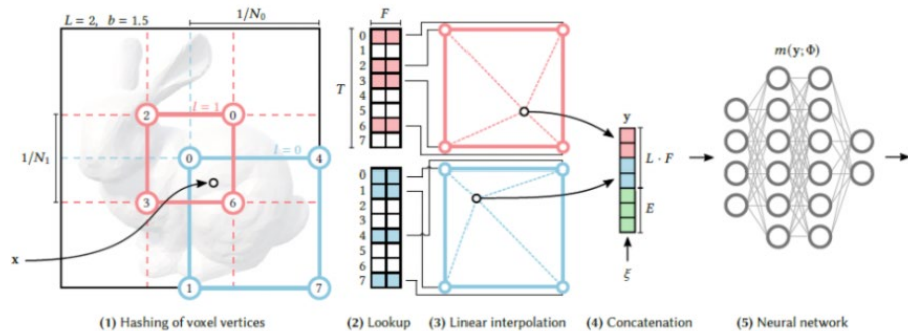
- LiDAR 모델링과 360° 카메라 통합 처리
- Single network를 통한 dynamic/static scene 처리 (only by positional embedding)
- key sensor characteristic(rolling shutter, ray drop)을 살리면서 모델링
- 5가지 dataset에서 검증

# NeuRAD: Neural Rendering for Autonomous Driving

Adam Tonderski†,1,4 Carl Lindstrom†,1,2 Georg Hess†,1,2 William Ljungbergh1,3 Lennart Svensson2 Christoffer Petersson1,2

CVPR 2024

- I-NGP: grid level별로 주변 4개의 좌표를 선택하고 hashtable을 통해 encoding해서 embedding값을 추출
- Mip-NeRF는 multi-scale을 고려, positional encoding 대신 conical frustum을 approximation 하는 Gaussian으로 encoding




Nerf 분야는 계속 발전하고, 기존의 방법들을 이용 training dataset augmentation, test dataset create 등으로 AD 분야에 기여

But) 긴 Training, 힘든 generalization, dense semantic 정보 필요

opacity times the accumulated transmittance

- NeRF와 유사한 NFF(neural feature field)를 이용

$$\mathbf{f}(\mathbf{r}) = \sum_{i=1}^{N_r} w_i \mathbf{f}_i, \quad w_i = \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j).$$



- 차이점 position, view direction이 input
  - NeRF : density, RGB 출력
  - 논문: implicit geometry(SDF), feature vector 출력




s



f

$$\alpha_i = 1 / (1 + e^{\beta s_i})$$


Opacity



SDF value at xi

# NeuRAD: Neural Rendering for Autonomous Driving

Adam Tonderski†,1,4 Carl Lindstrom†,1,2 Georg Hess†,1,2 William Ljungbergh1,3 Lennart Svensson2 Christoffer Petersson1,2

## ● Camera/LiDAR modeling

- 카메라 rendering → volume rendering을 수행  
low resolution + CNN을 이용해 upsampling하여 고화질 렌더링
- Lidar generation
  - 이전과 차이는 논문에서는 return되지 않는 ray들도 포함
  - Ray dropping으로 return된 ray의 세기가 너무 약할때 나타나고 sim-to-real gap을 줄이기위한 중요한 요소.
  - mirror, glass or wet road surface에서 종종 발생하는 현상
  - redered ray feature(위의 식을 거쳐나온)를 MLP을 통해 ray drop을 표현할 수 있다고 언급

For a lidar point, we shoot a ray  $\mathbf{r}(\tau) = \mathbf{o} + \tau\mathbf{d}$ , where  $\mathbf{o}$  is the origin of the lidar and  $\mathbf{d}$  is the normalized direction of the beam. We then find the expected depth  $D_l$  of a ray as  $\mathbb{E}[D_l(\mathbf{r})] = \sum_{i=1}^{N_r} w_i \tau_i$ . For predicting intensity, we volume render the ray feature following (1) and pass the feature through a small MLP.

## NeuRAD: Neural Rendering for Autonomous Driving

Adam Tonderski†,1,4 Carl Lindstrom†,1,2 Georg Hess†,1,2 William Ljungbergh1,3 Lennart Svensson2 Christoffer Petersson1,2

- **Rolling shutter modeling**

- Rolling shutter modeling을 제공. 이미지 왜곡이 발생x
- 각 pixel의 시간을 개별적으로 계산

- **Differing camera settings**

- AD dataset은 보통 여러개의 카메라를 설정 → 센서 특성을 모델링하기 위한 모듈
- rendering전 sensor embedding 단계, 각 센서마다 learnable embedding → 최종 volume rendering 된 결과에 concat

and MLP together with  $g$ . However, as we know which image comes from which sensor, we instead learn a single embeddings per sensor, **minimizing the potential for overfitting**, and allowing us to use these *sensor embeddings* when generating novel views. As we render features rather than color, we apply these embeddings after the volume rendering, significantly reducing computational overhead.

# NeuRAD: Neural Rendering for Autonomous Driving

Adam Tonderski†,1,4 Carl Lindstrom†,1,2 Georg Hess†,1,2 William Ljungbergh1,3 Lennart Svensson2 Christoffer Petersson1,2

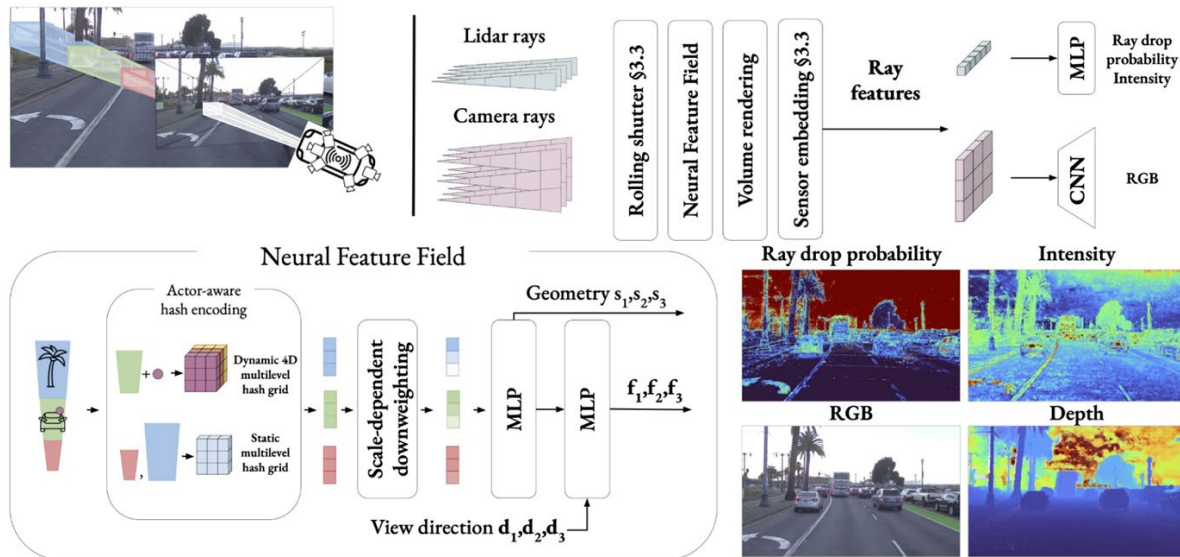


Figure 2. Overview of our approach. We learn a joint neural feature field for the statics and dynamics of an automotive scene, where the two are discerned only by our actor-aware hash encoding. Points that fall inside actor bounding boxes are transformed to actor-local coordinates and, together with actor index, used to query the 4D hash grid. We decode the volume rendered ray-level features to RGB values using an upsampling CNN, and to ray drop probability and intensity using MLPs.

1. Time 계산: Rolling shutter modeling
1. NFF 입력: 3D 위치 및 시간 전달
1. 샘플링: Efficient sampling
1. Dynamic/Static 분류: Multi-resolution hash grid
1. 임베딩 추출: Embedding vector
1. 가중치 조정: Gaussian weighting
1. SDF 계산: MLP 첫 번째 layer
1. Feature map 생성: MLP layer
1. Sensor 결합: Embedding 결합
1. 최종 렌더링: CNN 렌더링

# NeuRAD: Neural Rendering for Autonomous Driving

Adam Tonderski†,1,4 Carl Lindstrom†,1,2 Georq Hess†,1,2 William Ljunqbergh1,3 Lennart Svensson2 Christoffer Petersson1,2

Table 1. Image novel view synthesis performance comparison to state-of-the-art methods across five datasets. \*our reimplementation. †baselines from [44, 46, 47]. §partial results due to training instability. **Bold/underline** for best/second-best.

		PSNR ↑	SSIM ↑	LPIPS ↓
Panda FC	Instant-NGP† [27, 47]	24.03	0.708	0.451
	UniSim [47]	25.63	0.745	0.288
	UniSim*	25.44	0.732	0.228
	NeuRAD (ours)	26.58	0.778	0.190
	NeuRAD-2x (ours)	<b>26.84</b>	<b>0.801</b>	<b>0.148</b>
Panda 360	UniSim*	23.50	0.692	0.330
	NeuRAD (ours)	25.97	0.758	0.242
	NeuRAD-2x (ours)	<b>26.47</b>	<b>0.779</b>	<b>0.196</b>
nuScenes	Mip360† [3, 46]	24.37	0.795	0.240
	S-NeRF [46]	26.21	<b>0.831</b>	0.228
	NeuRAD (ours)	26.99	0.815	0.225
	NeuRAD-2x (ours)	<b>27.13</b>	0.820	<b>0.205</b>
KITTI MOT	SUDS† [34, 44]	23.12	0.821	0.135
	MARS [44]	24.00	0.801	0.164
	NeuRAD (ours)	27.00	0.795	0.082
	NeuRAD-2x (ours)	<b>27.91</b>	<b>0.822</b>	<b>0.066</b>
Argo2	UniSim*	23.22§	0.661§	0.412§
	NeuRAD (ours)	26.22	0.717	0.315
	NeuRAD-2x (ours)	<b>27.73</b>	<b>0.756</b>	<b>0.233</b>
ZOD	UniSim*	27.97	0.777	0.239
	NeuRAD (ours)	29.49	0.809	0.226
	NeuRAD-2x (ours)	<b>30.59</b>	<b>0.857</b>	<b>0.210</b>

Table 2. Lidar novel view synthesis performance comparison to state-of-the-art methods. Depth is median L2 error [m]. Intensity is RMSE. Drop acc. denotes ray drop accuracy. Chamfer denotes chamfer distance, normalized with num. ground truth points [m].

		Depth ↓	Intensity ↓	Drop acc. ↑	Chamfer ↓
Panda FC	UniSim	0.10	0.065	-	-
	UniSim*	0.07	0.085	91.0	11.2
	NeuRAD (ours)	<b>0.01</b>	<b>0.062</b>	<b>96.2</b>	<b>1.6</b>
Panda 360	UniSim*	0.07	0.087	91.9	10.3
	NeuRAD (ours)	<b>0.01</b>	<b>0.061</b>	<b>96.1</b>	<b>1.9</b>



# NeuRAD: Neural Rendering for Autonomous Driving

CVPR 2024

Adam Tonderski†,1,4 Carl Lindstrom†,1,2 Georg Hess†,1,2 William Ljungbergh1,3 Lennart Svensson2 Christoffer Petersson1,2

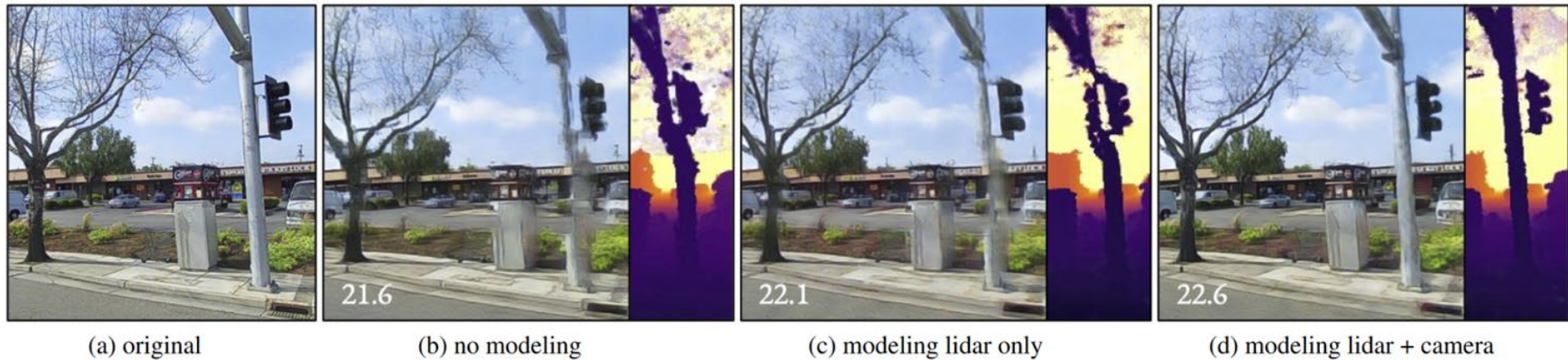


Figure 3. Impact of modeling rolling shutter in a high-speed scenario (with inset PSNR). (a) original side-camera image. Omitting the rolling shutter entirely (b) results in extremely blurry renderings and unrealistic geometry, especially for the pole. Modeling the lidar rolling shutter (c) improves the quality, but it is only when both sensors are modeled correctly (d) that we get realistic renderings.

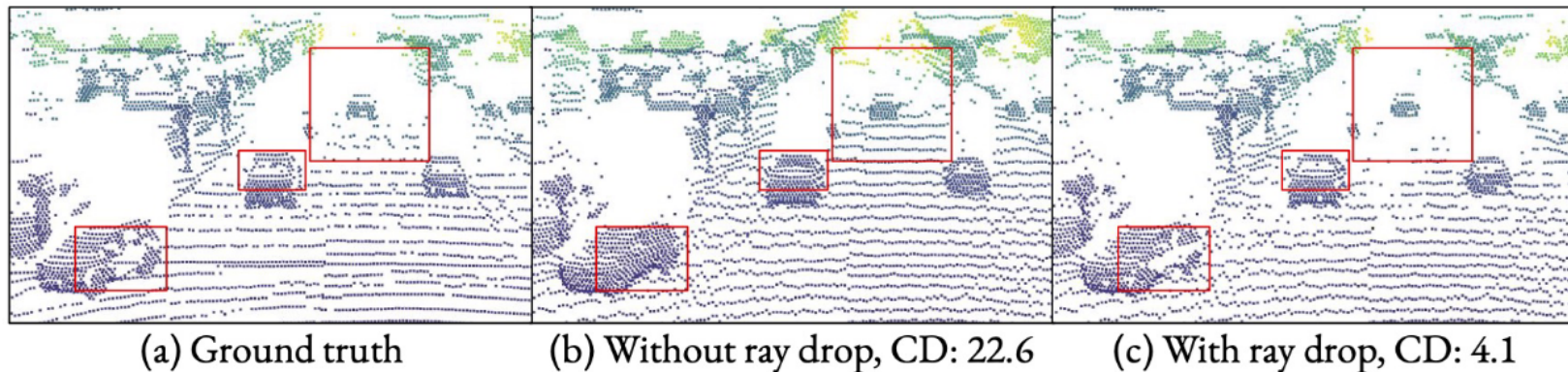


Figure 4. Visualization of ray drop effects for lidar simulation. Highlighted parts show areas where ray dropping effects are important to consider in order to simulate realistic point clouds. CD denotes Chamfer distance normalized by num. GT points.

## ● Problem/Objective

- multimodal에서 각각 modality의 기여도를 판단

## ● Contribution/Key Idea

- sample-level 수준의 modality valuation metric 도입
- 기여도가 낮은 modality 분석 및 학습 개선
- MM-biased dataset

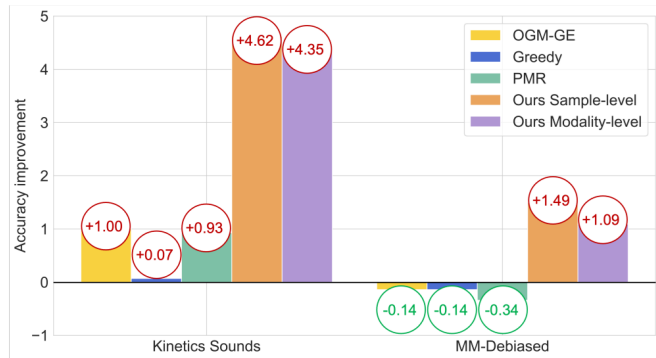
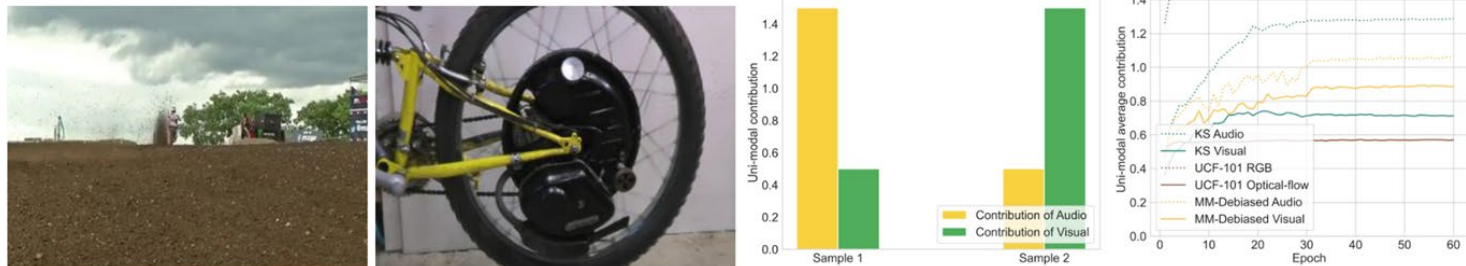


Figure 1. Accuracy improvement compared with joint training baseline of imbalanced multimodal learning methods, on Kinetics Sounds and our proposed MM-Debiased dataset. Other methods: OGM-GE [21], Greedy [33] and PMR [4].

## Enhancing Multimodal Cooperation via Sample-level Modality Valuation

Yake Wei<sup>1</sup>, Ruoxuan Feng<sup>1</sup>, Ziheng Wang<sup>1,2</sup>, Di Hu<sup>1,2</sup>,



(a) Visual of *S.1* of *motorcycling*. (b) Visual of *S.2* of *motorcycling*. (c) Valuation of *S.1* and *S.2*. (d) Avg. Contribution of dataset.

Figure 2. **(a-b):** Audio-visual samples of *motorcycling* category. **(c):** Our modality valuation of *S.1* and *S.2*. *S.1* and *S.2* denotes *Sample 1* and *Sample 2* respectively. **(d):** Uni-modal average contribution over all training samples of different dataset. Our proposed MM-Debiased dataset has less global discrepancy at dataset-level, compared with other curated dataset.

- (a), (b) 모두 motorcycling 에 대한 audio-visual 데이터셋
- (c) 를 통해 알 수 있듯 sample 마다 modality 비중 차이가 존재.  
→ 같은 네트워크로 학습하면 특정 modality에만 강인한 결과

본 논문 : Training에서 기여도가 낮은 modality를 판별하여 multi-modal cooperation을 강화 가능