
DTU

Enhancing Sparse Radar Point Clouds using Diffusion Models For Radar-Only 3D Detection

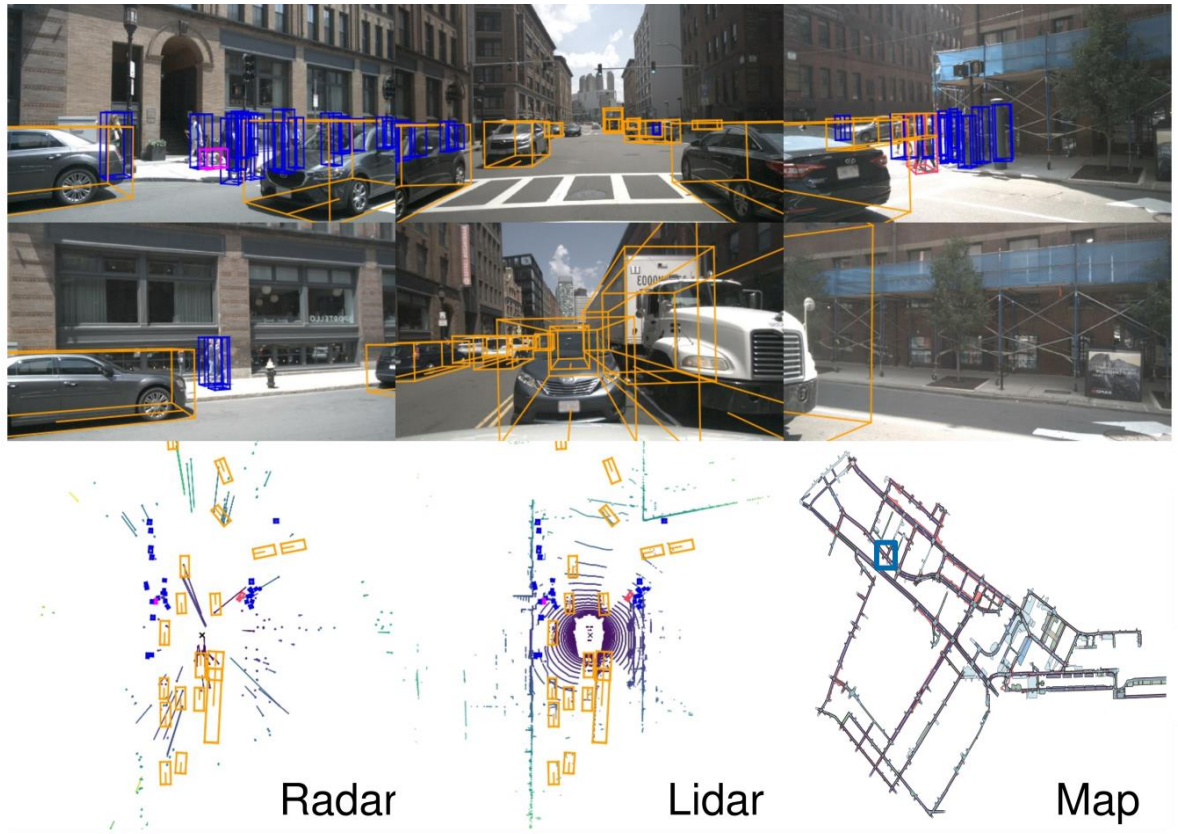
Hassan Hotait

Msc Thesis

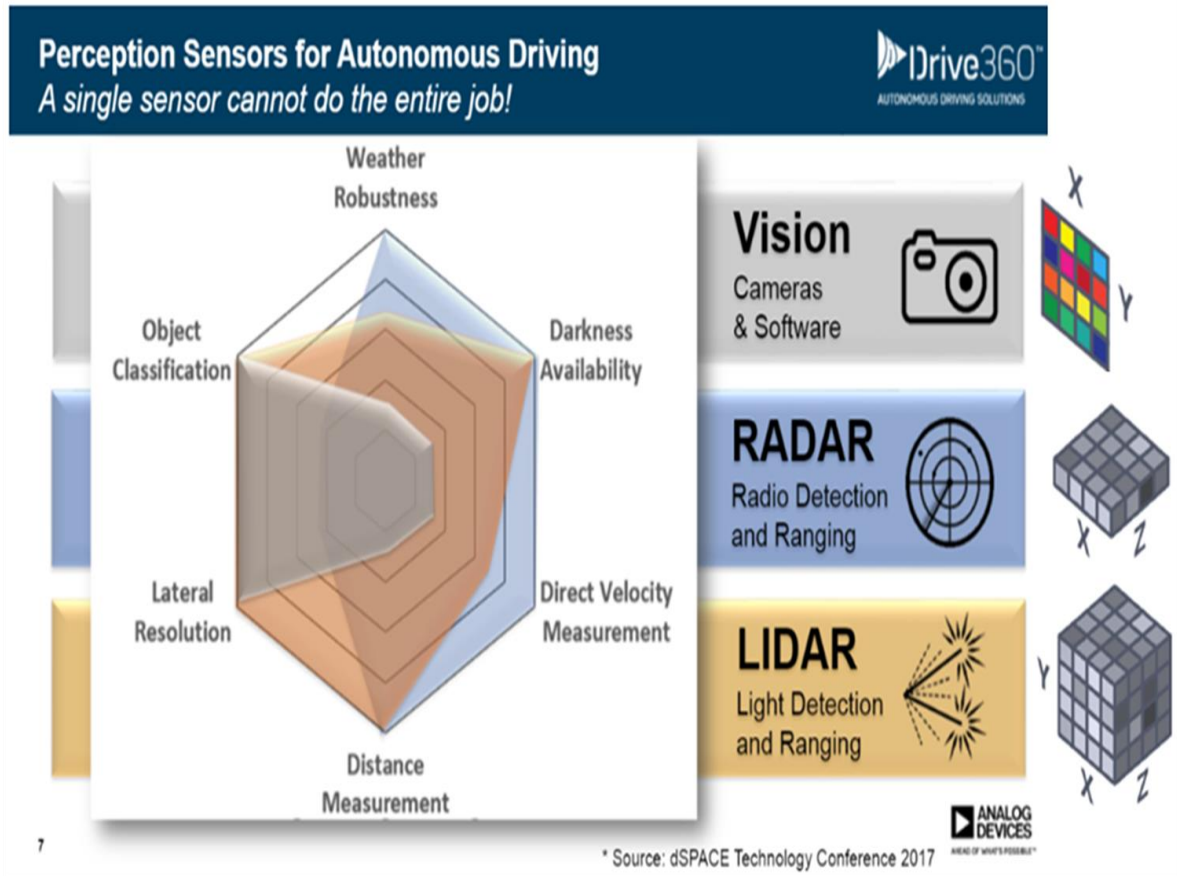
Outline

- 3D Detection in Autonomous Driving
- Thesis Scope
- Recap on Image Based Diffusion & Design Choices
- Our Method
- Qualitative & Quantitative Results
- Future Work

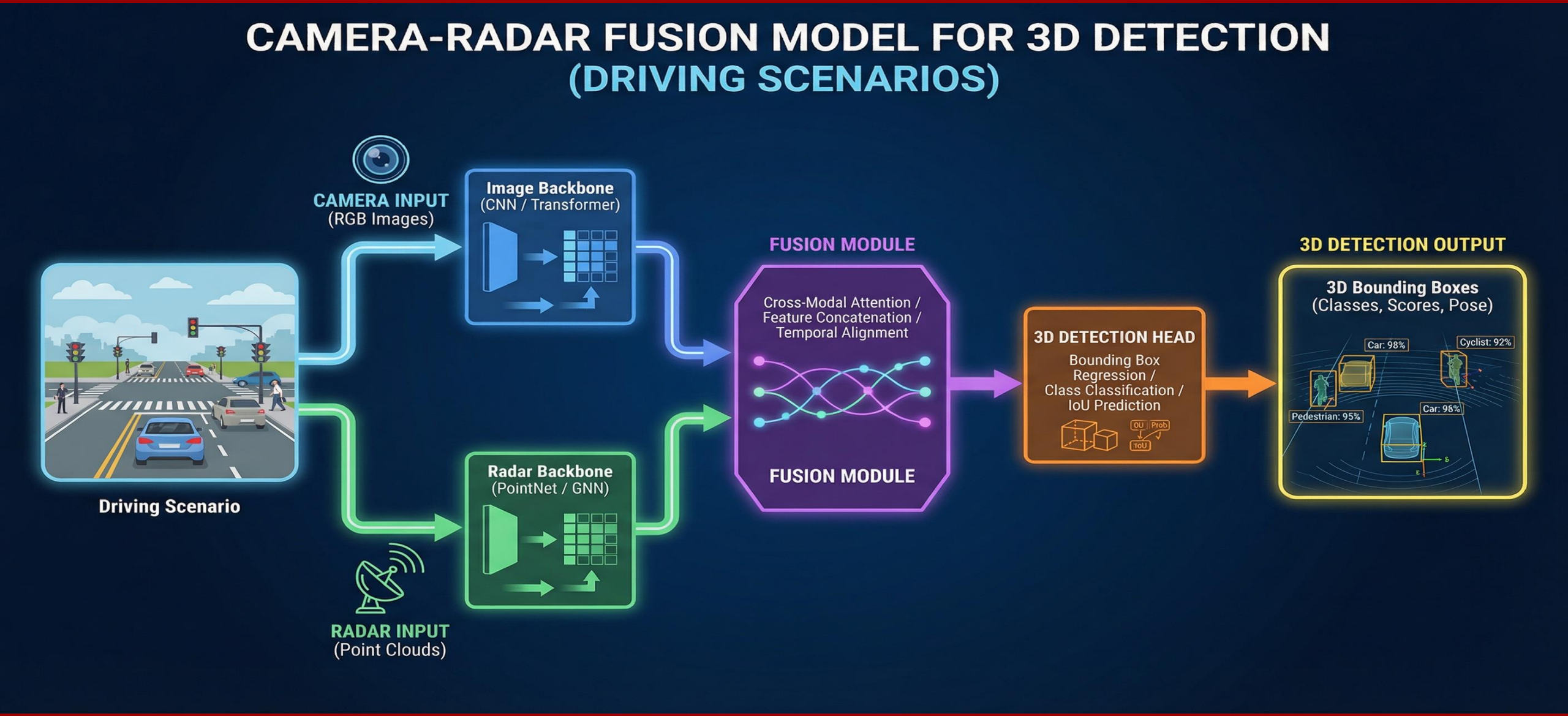
3D Detection



"Ped with pet, bicycle, car makes a u-turn, lane change, peds crossing crosswalk"



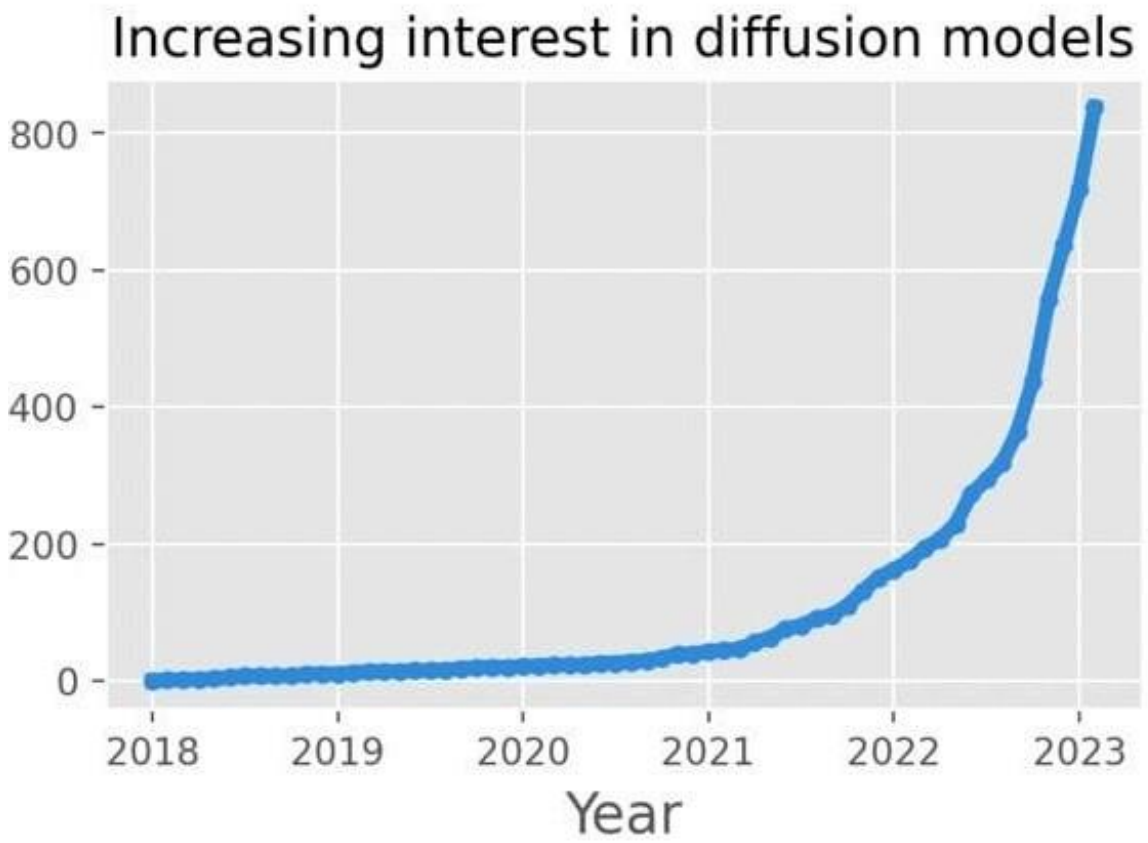
Camera Radar Fusion For 3D Detection



Research Gap & Thesis Scope

Method					
Date	Name	Modalities	Map data	External data	mAP
		Radar	All	All	
2024-03-08	RadarDistill	Radar	no	no	0.205
2022-02-03	KPConvPillars	Radar	no	no	0.049
2020-09-22	Radar-PointGNN	Radar	no	no	0.005

Method	Input	KD	Car(AP)↑	mAP↑
BEVFusion* [20]	C,R		65.9	38.3
RadarDistill-CR	C,R	✓	**67.7**	**39.6**

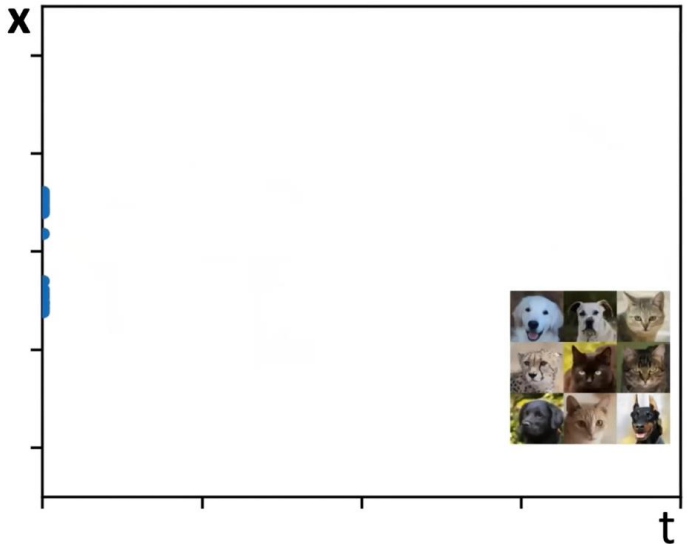


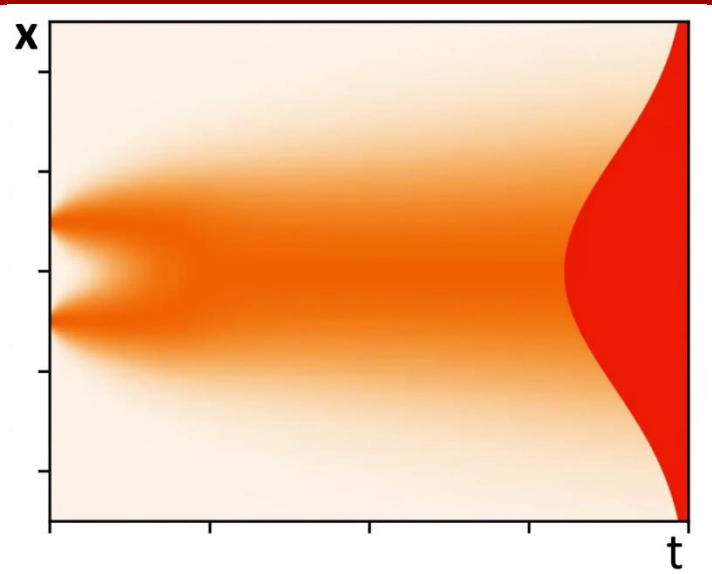
Thesis Scope: *Using diffusion models for enhancing sparsity of radar pointclouds for radar-only 3D detection*

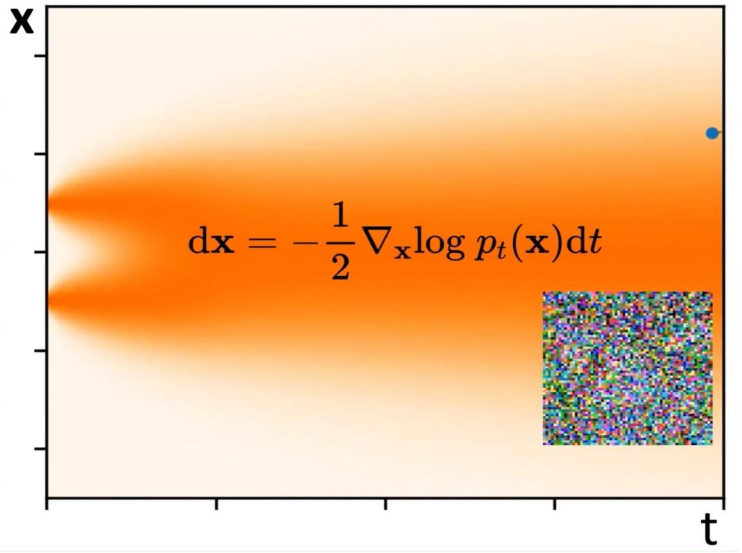
Related Work – Image Based Diffusion Generative Methods

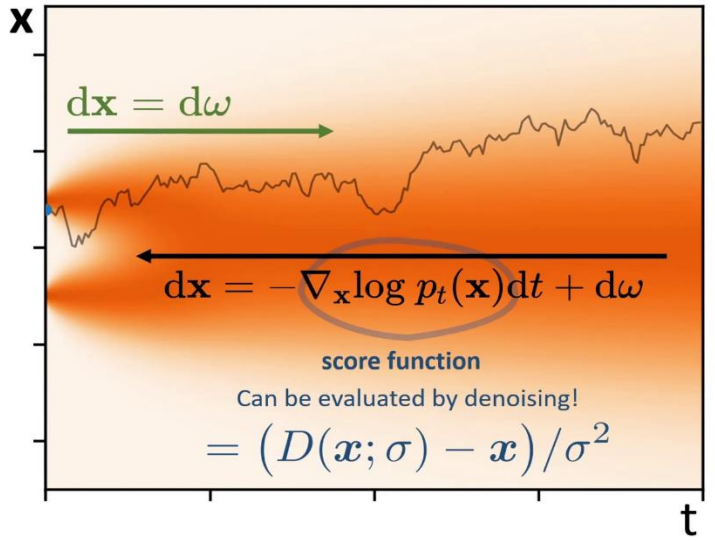
		VP	VE	iDDPM	+ DDIM
Sampling					
ODE solver		Euler	Euler	Euler	
Time steps	$t_{i<N}$	$1+\frac{i}{N-1}(\epsilon_s-1)$	$\sigma_{\max}^2\left(\sigma_{\min}^2/\sigma_{\max}^2\right)^{\frac{i}{N-1}}$	$u_{\lfloor j_0+\frac{M-1-j_0}{N-1}i+\frac{1}{2}\rfloor}$, where	
				$u_M=0$	
				$u_{j-1}=\sqrt{\frac{u_j^2+1}{\max(\alpha_{j-1}/\alpha_j,C_1)}-1}$	
Schedule	$\sigma(t)$	$\sqrt{e^{\frac{1}{2}\beta_d t^2+\beta_{\min}t}-1}$	$\sqrt{t}$	t	
Scaling	$s(t)$	$1/\sqrt{e^{\frac{1}{2}\beta_d t^2+\beta_{\min}t}}$	1	1	
Network and preconditioning					
Architecture of F_θ		DDPM++	NCSN++	DDPM	
Skip scaling	$c_{\text{skip}}(\sigma)$	1	1	1	
Output scaling	$c_{\text{out}}(\sigma)$	$-\sigma$	σ	$-\sigma$	
Input scaling	$c_{\text{in}}(\sigma)$	1		$1/\sqrt{\sigma^2+1}$	
Noise cond.	$c_{\text{noise}}(\sigma)$	(		$\min_j\lvert u_j-\sigma\rvert$	
Training					
Noise distribution				$\mathcal{U}\{0,M-1\}$	
Loss weighting	$\lambda(\sigma)$	$1/\sigma^2$		: *)	
Parameters		$\beta_d=19.9, \beta_{\min}=0.1$	$\sigma_{\min}=0.02$	$\frac{j}{M(C_2+1)}$)	
		$\epsilon_s=10^{-3}, \epsilon_t=10^{-5}$	$\sigma_{\max}=100$	$C_1=$, $C_2=0.008$	
		$M=1000$		$M=1000, j_0=8^\dagger$	

Related Work – Diffusion Based Generative Methods

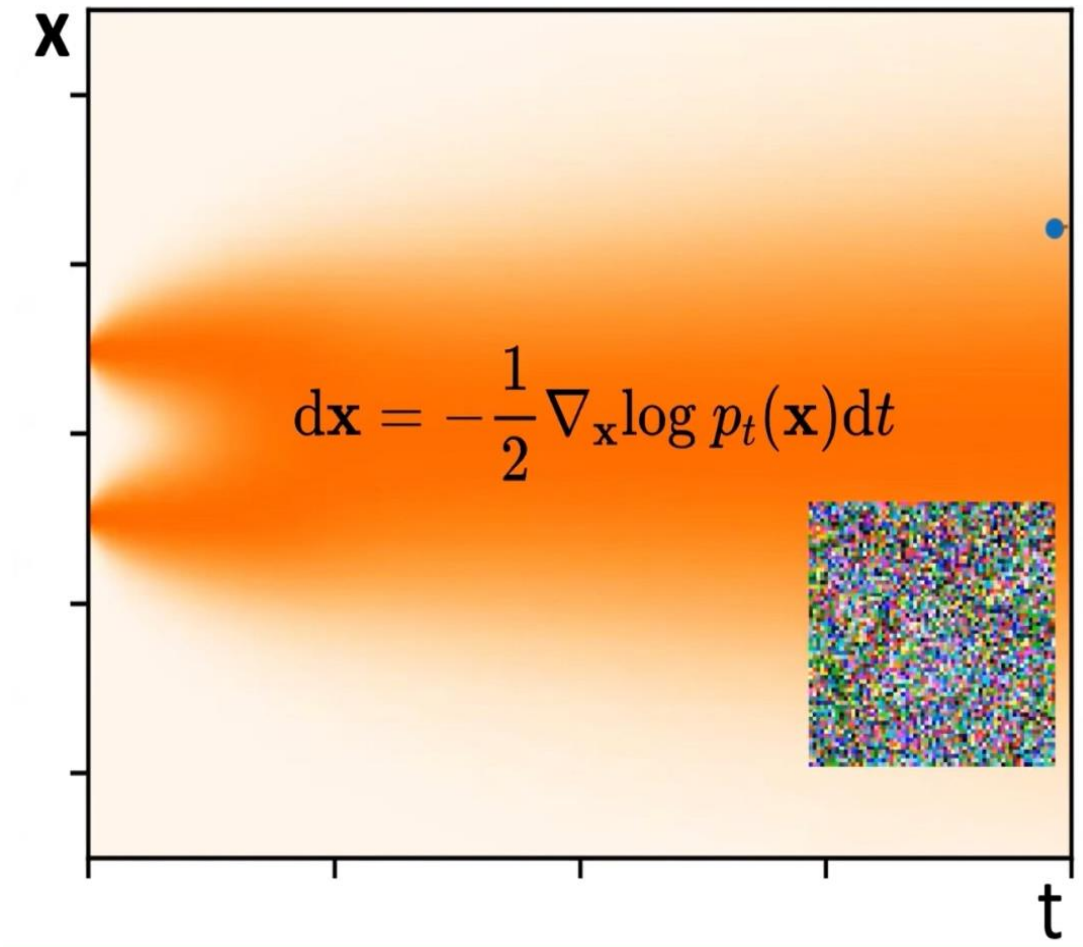




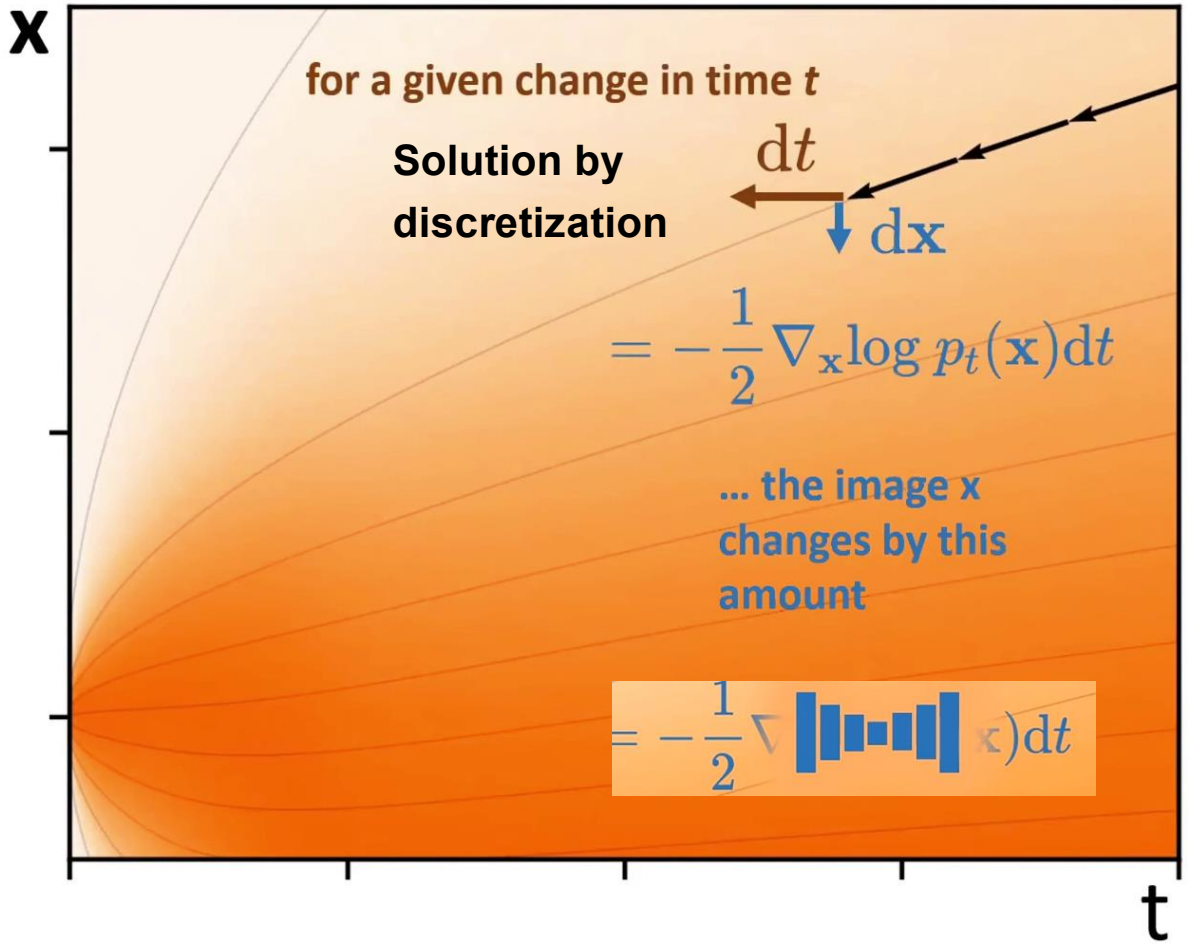




Related Work – Diffusion Based Generative Methods



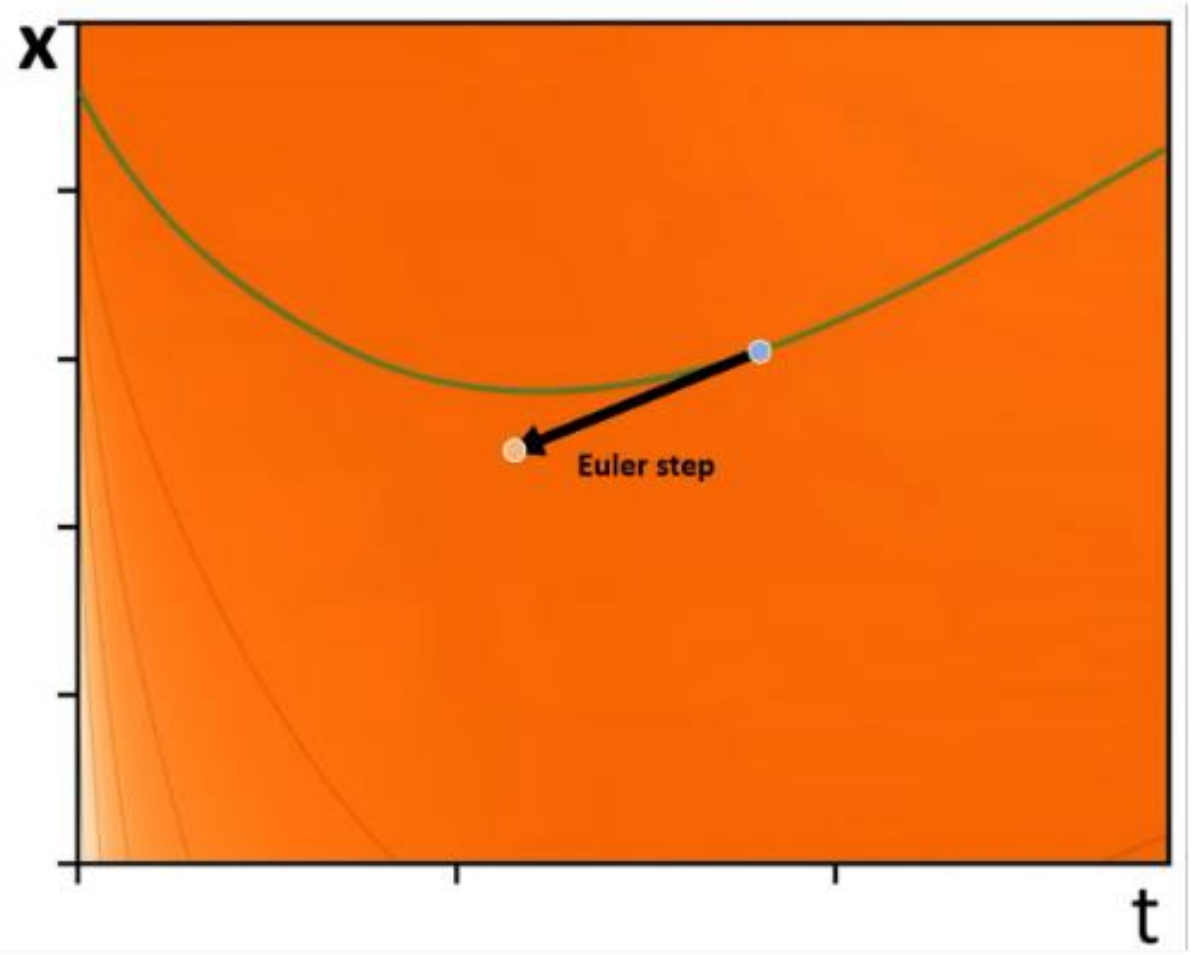
Forward & Reverse Process Loop

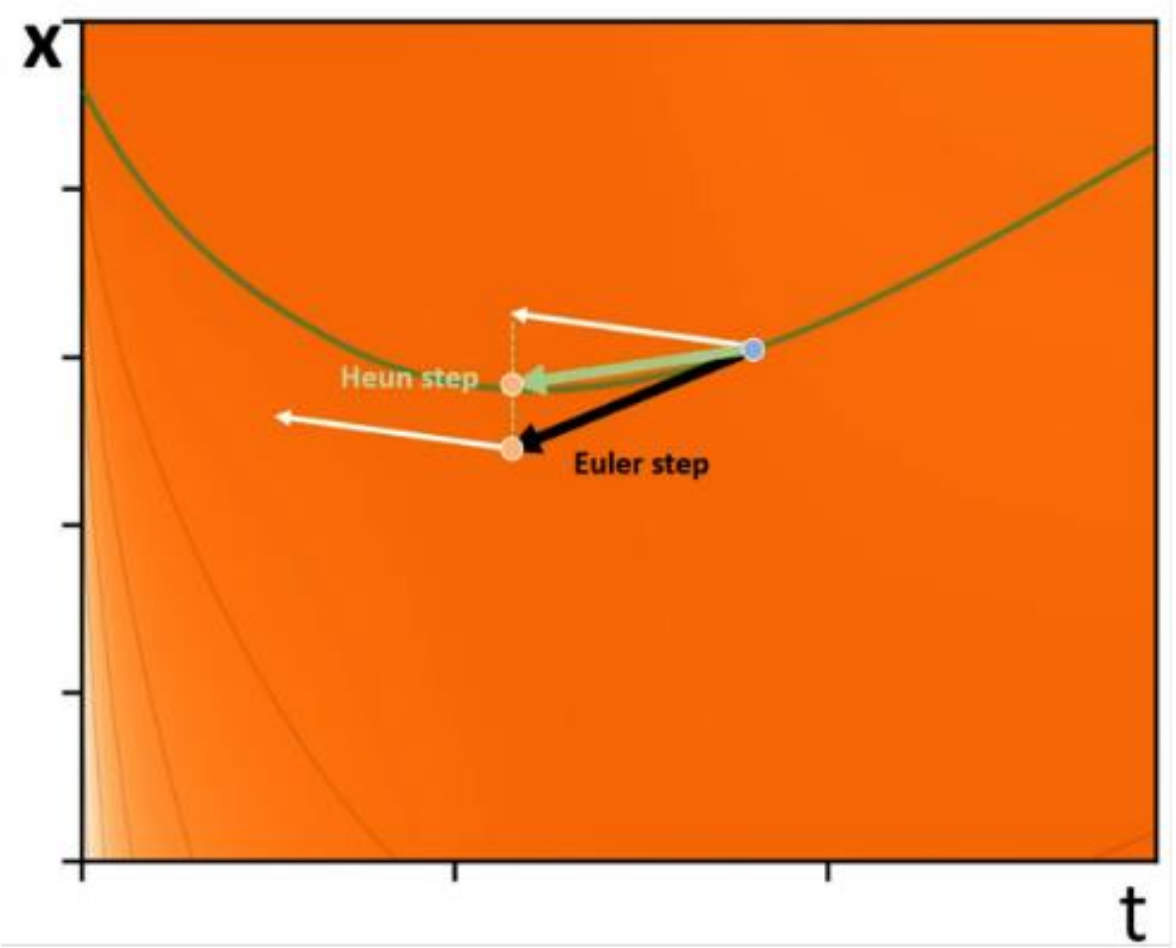


Score Function Predicted by NN

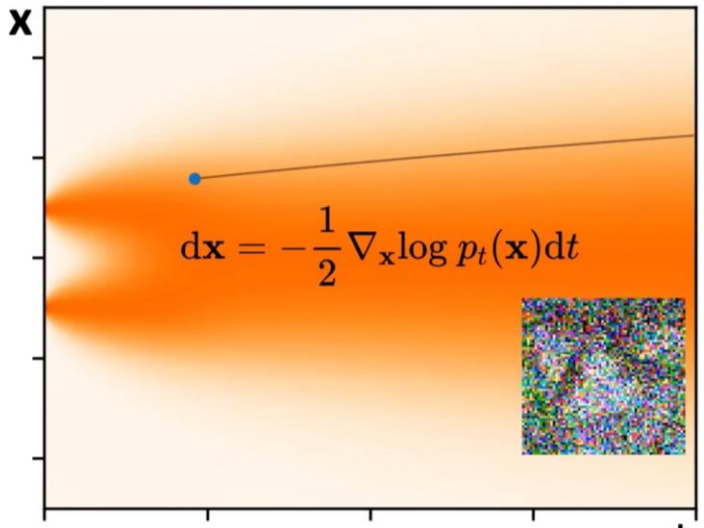
Related Work – Diffusion Based Generative Methods

Sampling Design Choices – ODE Solver

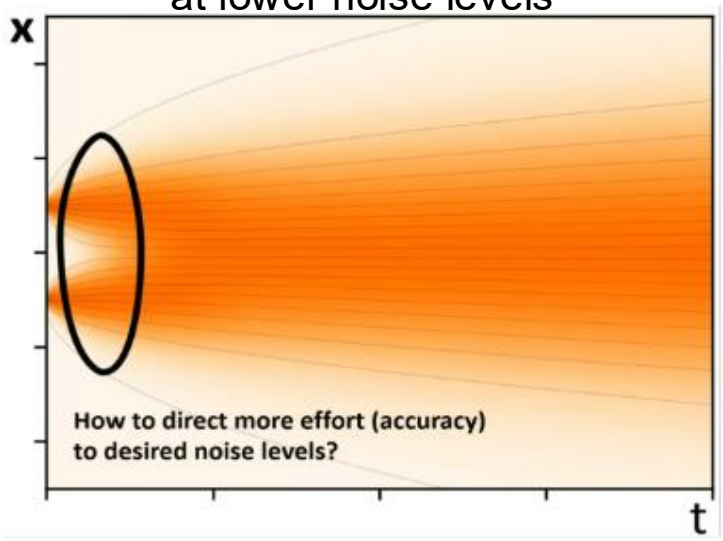


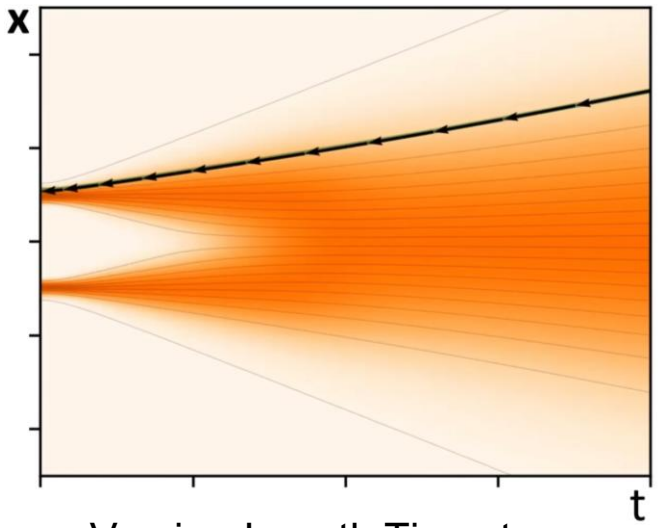


Related Work – Diffusion Based Generative Methods

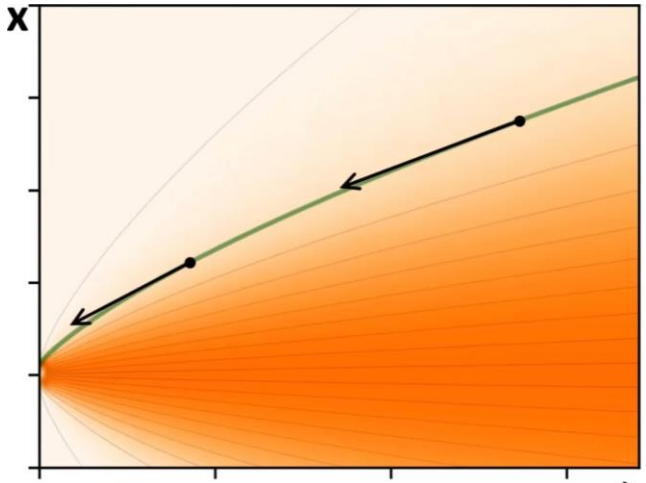


Most Details are constructed at lower noise levels



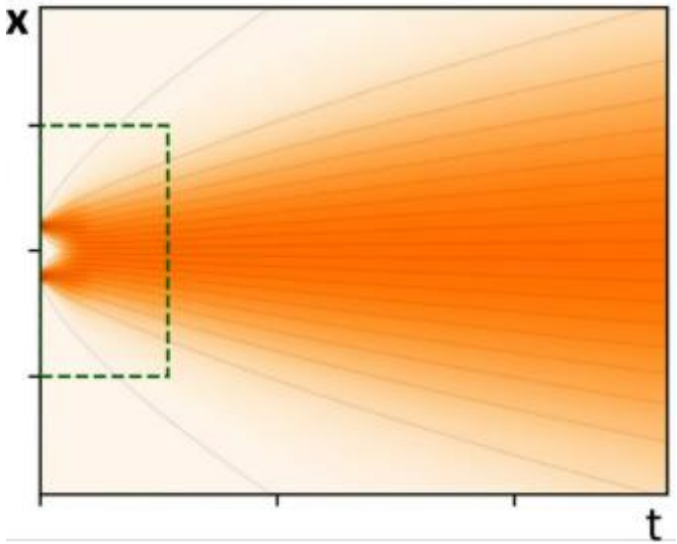


Varying Length Timesteps

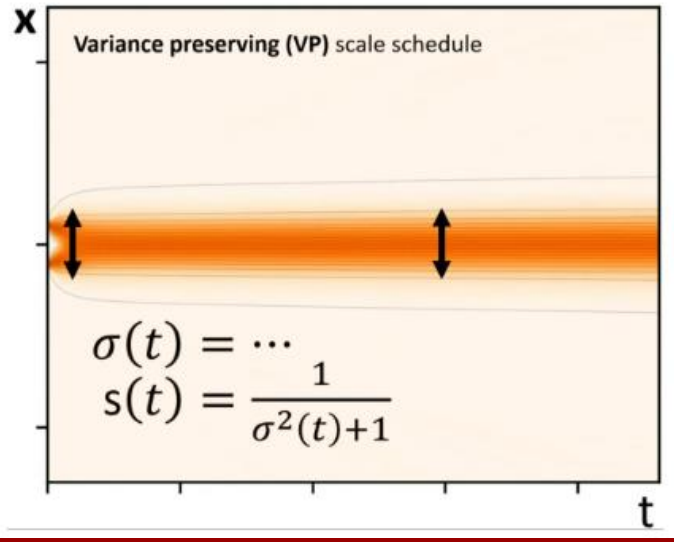


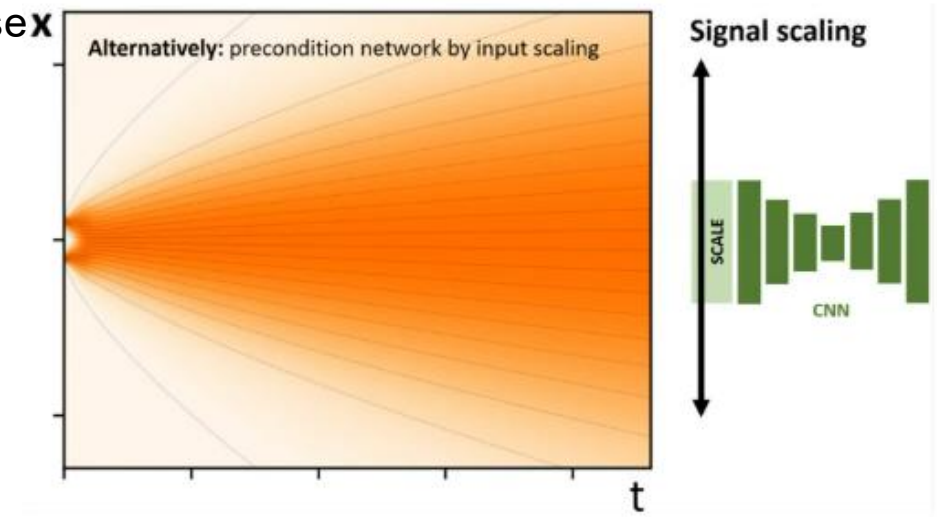
Family of Noise Schedules

Sampling Design Choices – Scale Schedule & Signal Scaling

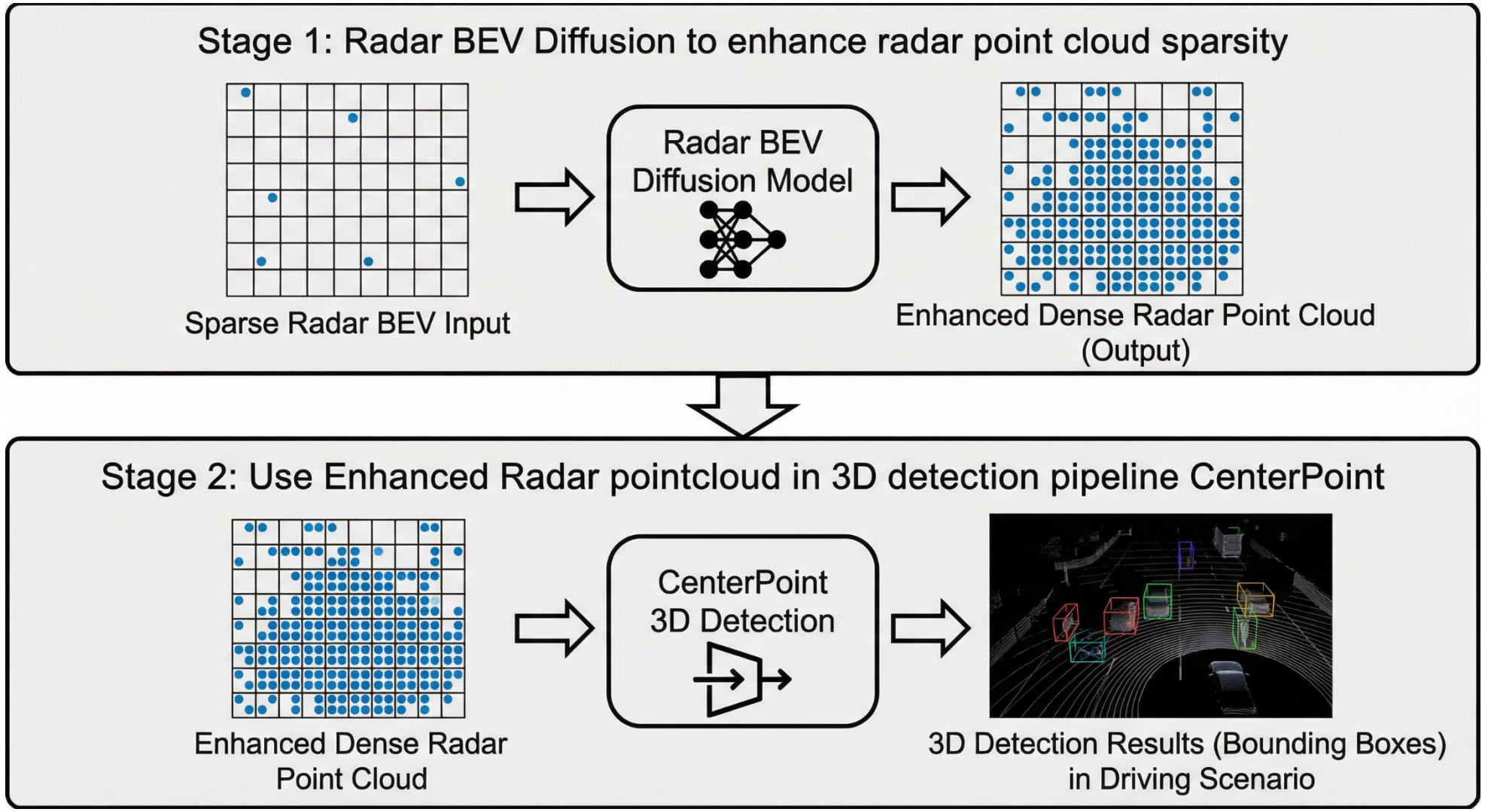


Noise magnitude grows unbounded at higher noise levels

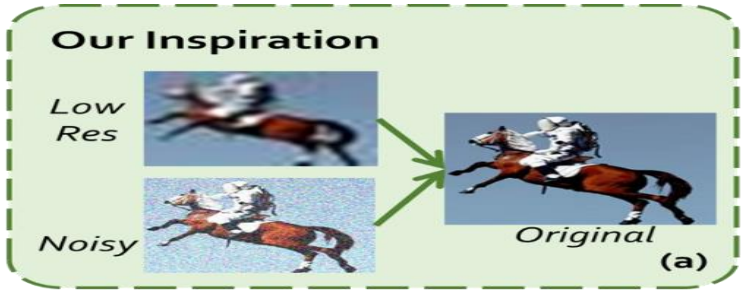


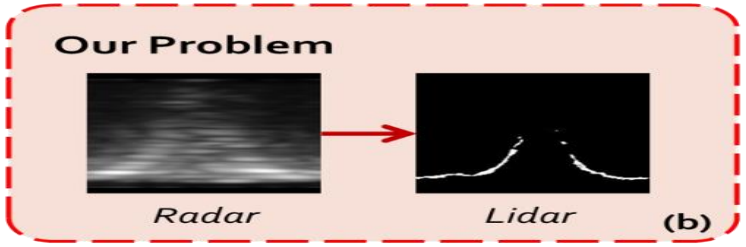


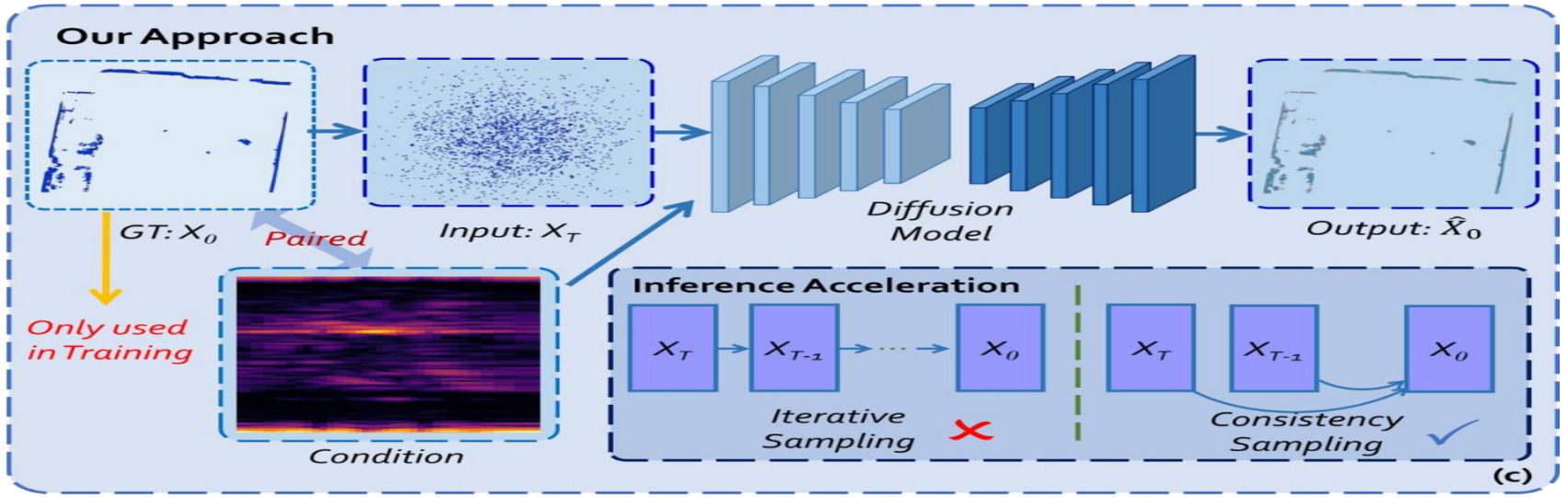
2 Stage Approach For Radar-Only 3D Detection

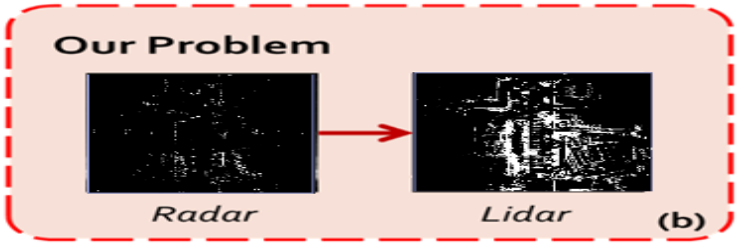


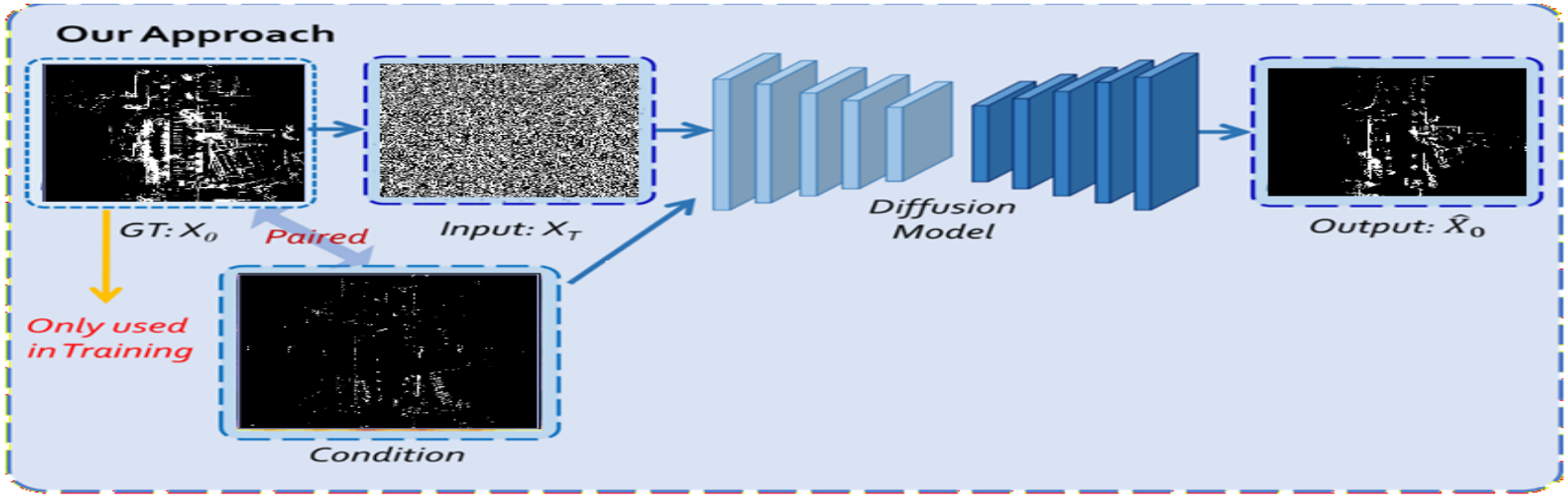
Methodology – Starting Point (Related Work)











Data Augmentation

CUT MIX AUGMENTATION

Cut Mix Augmentation increases the **SPATIAL DISTORTION** in the training images, making the Student Network learn more general and robust representations!

STEP 1: Take two pairs of images and their pseudo-labelled depth maps

STEP 2: Pasting a rectangular portion of Image A on top of Image B

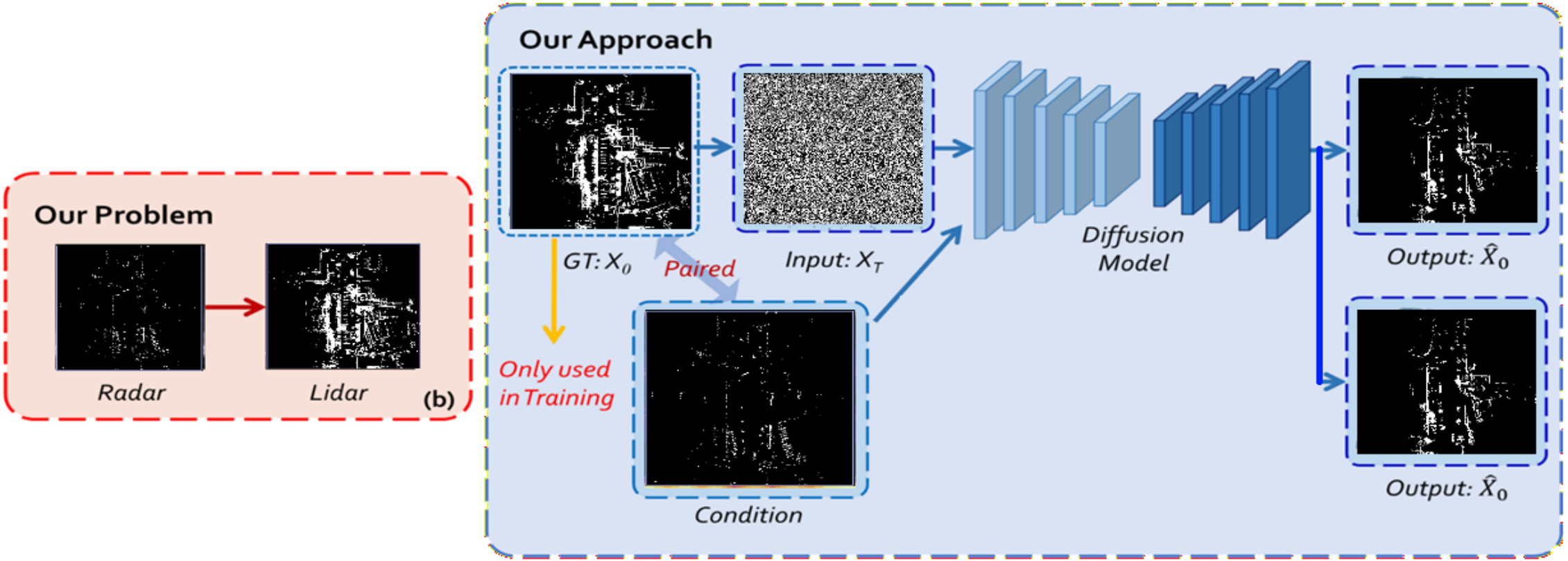
Methodology – Data Representation

Model	Preprocessing Pipeline	mAP
CenterPoint (Benchmark)	• Points From LiDAR MultiSweeps • No Motion Compensation • Temporal Encoding	56%
RadarDistill (Target)	-	20%
CenterPoint (UpperBound)	• Points From LiDAR MultiSweeps • Motion Compensation • Ground Point Removal • Points to BEV 1. Occupancy BEV Map 2. Constant Zero Height Map • BEV To Points	13%
RadarBEVDiff(Cut&Mix) + CenterPoint (Our 2 Stage Method)	• Points from Denoised BEV Occupancy Map • BEV To Points	6%

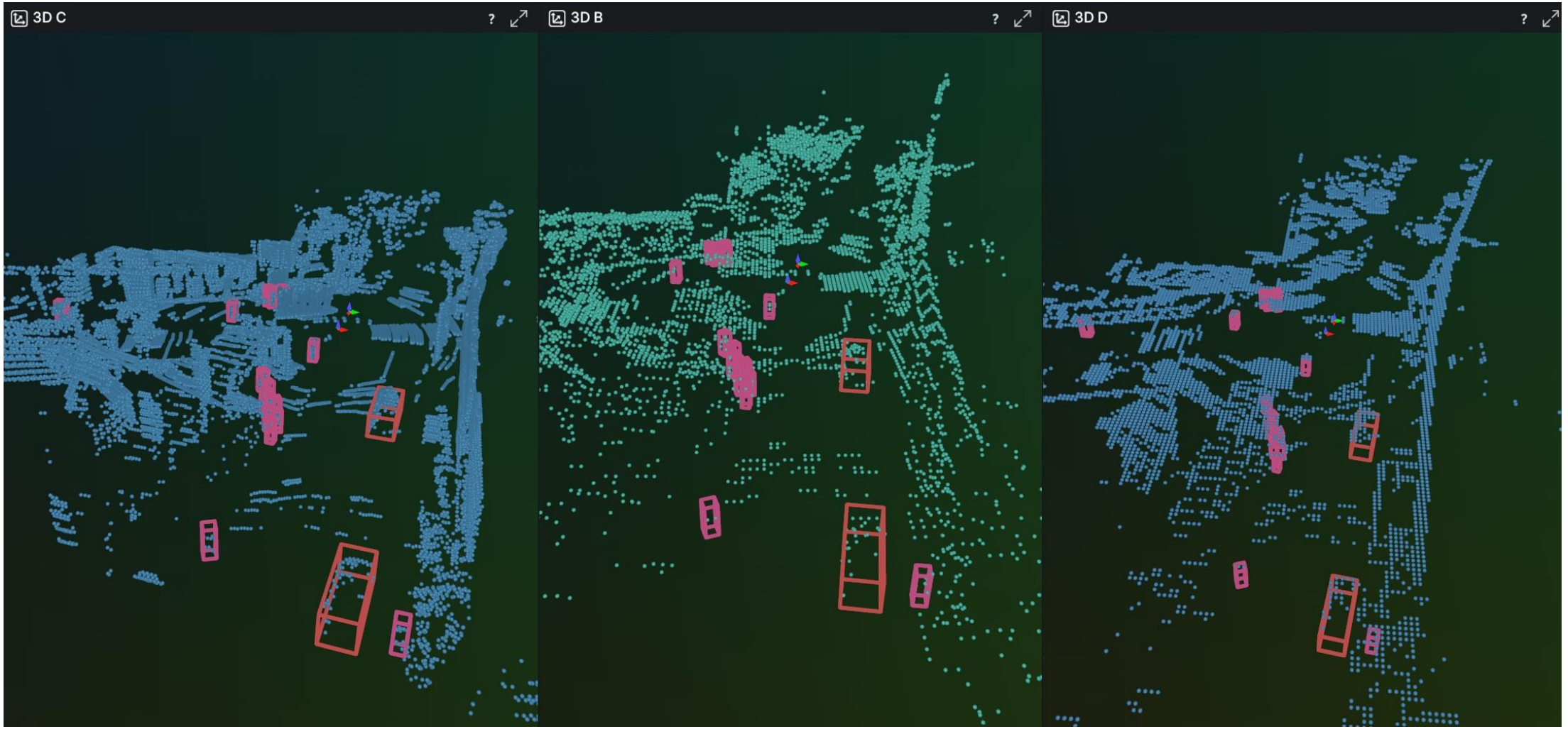
Question:

How to preserve height information in BEV representation?

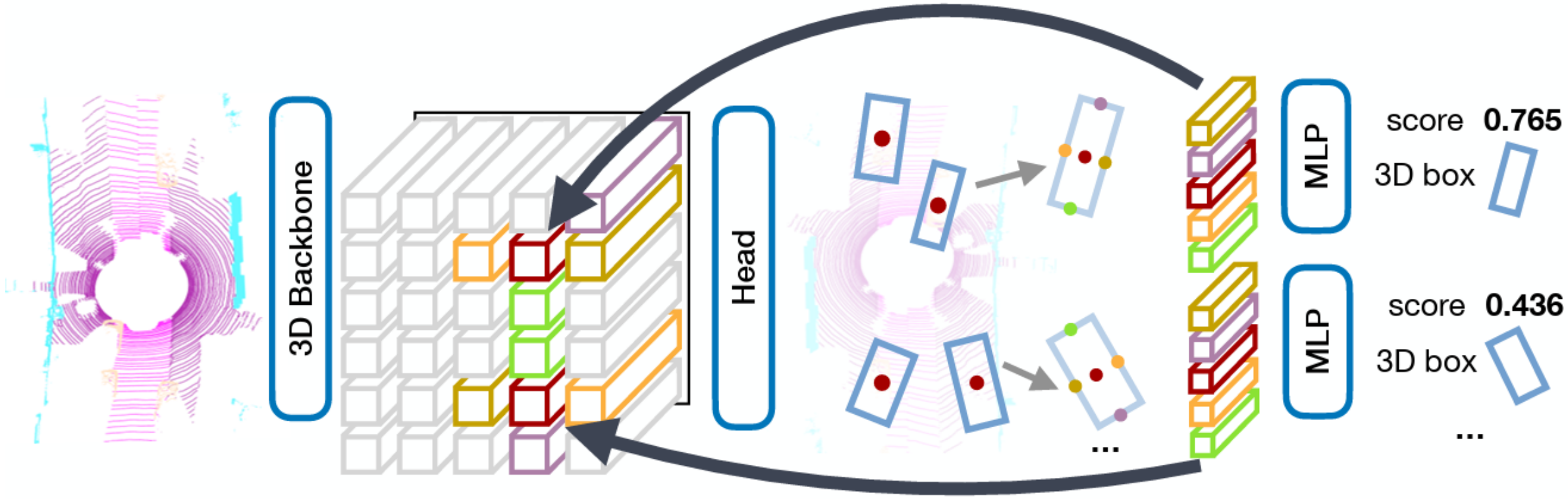
BEV Height Map Encoding



Data Representation

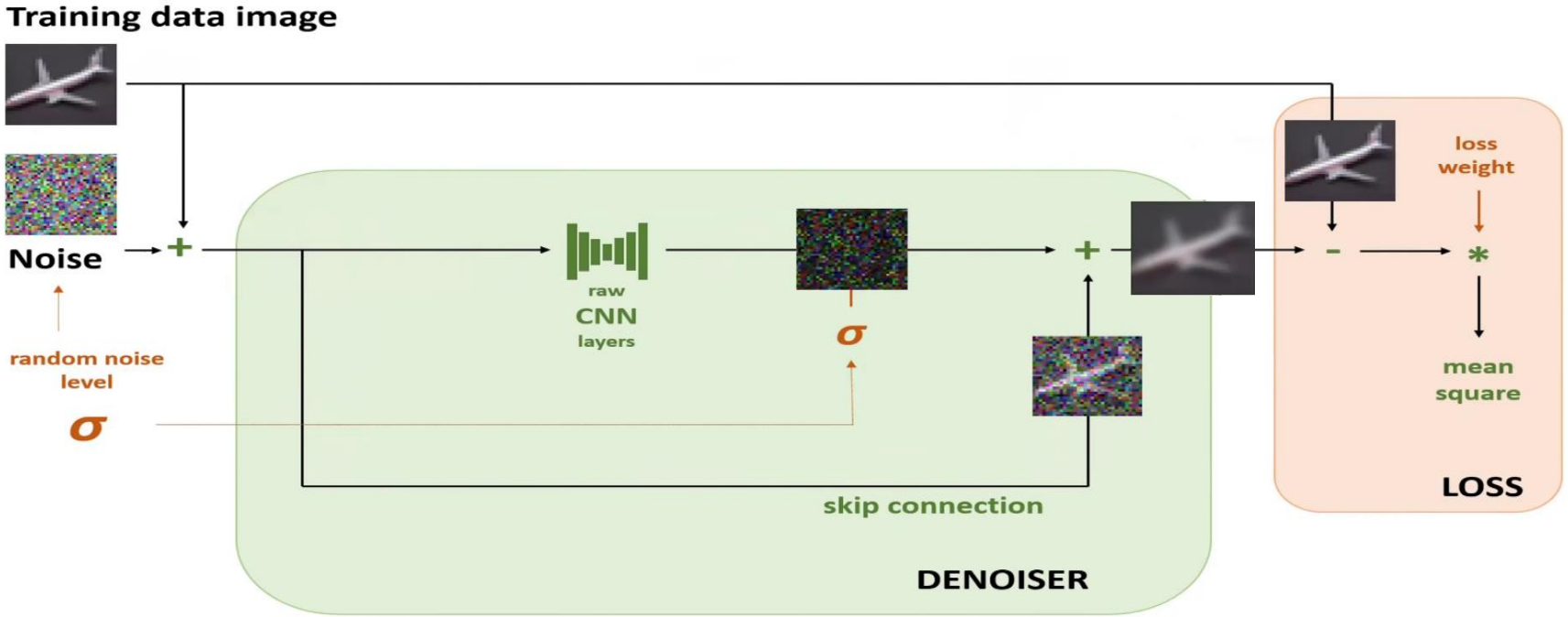


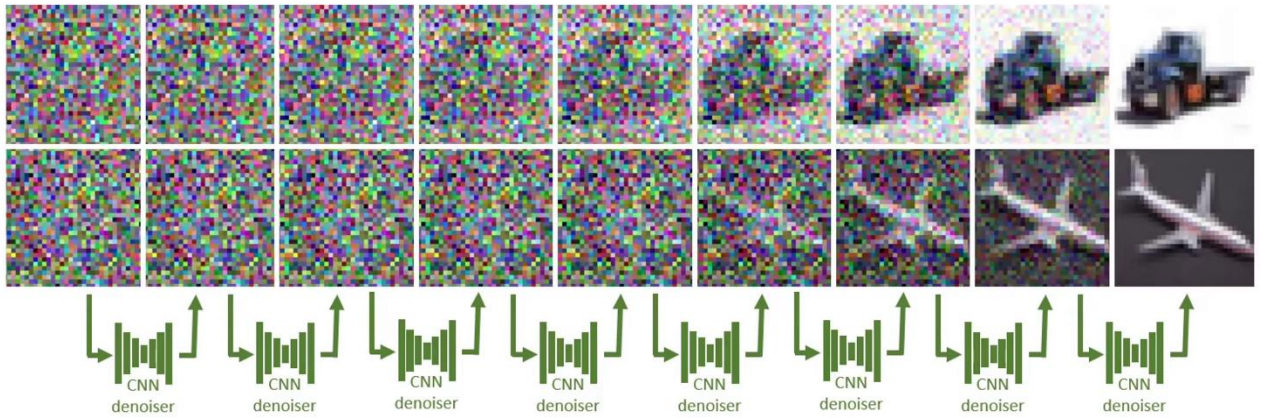
CenterPoint – 3D Detection Benchmark



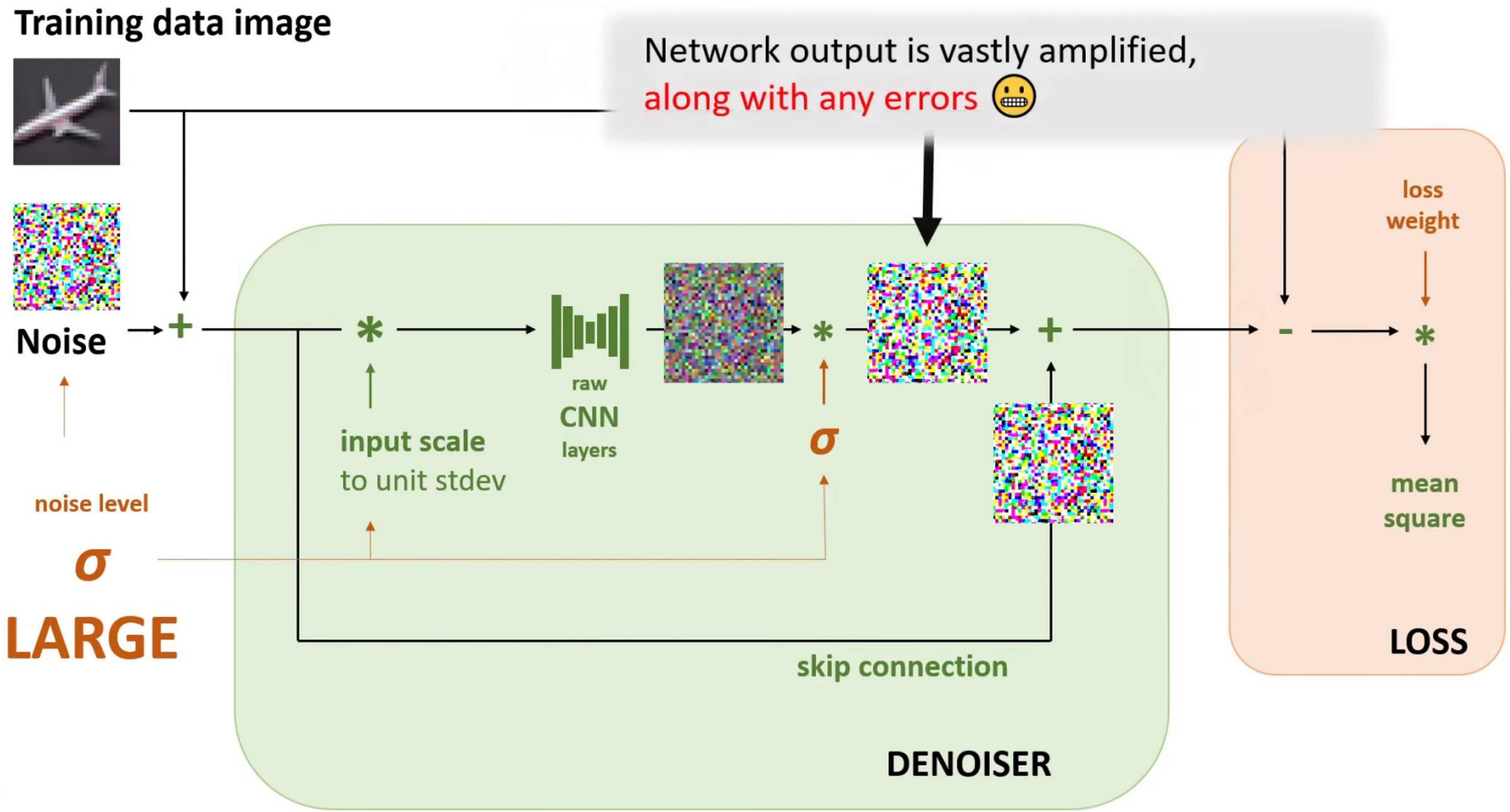
(a) Point. cloud (b) Map-view features (c) First stage: Centers and 3D boxes (d) Second stage: Score and 3D boxes

End-To-End Approach

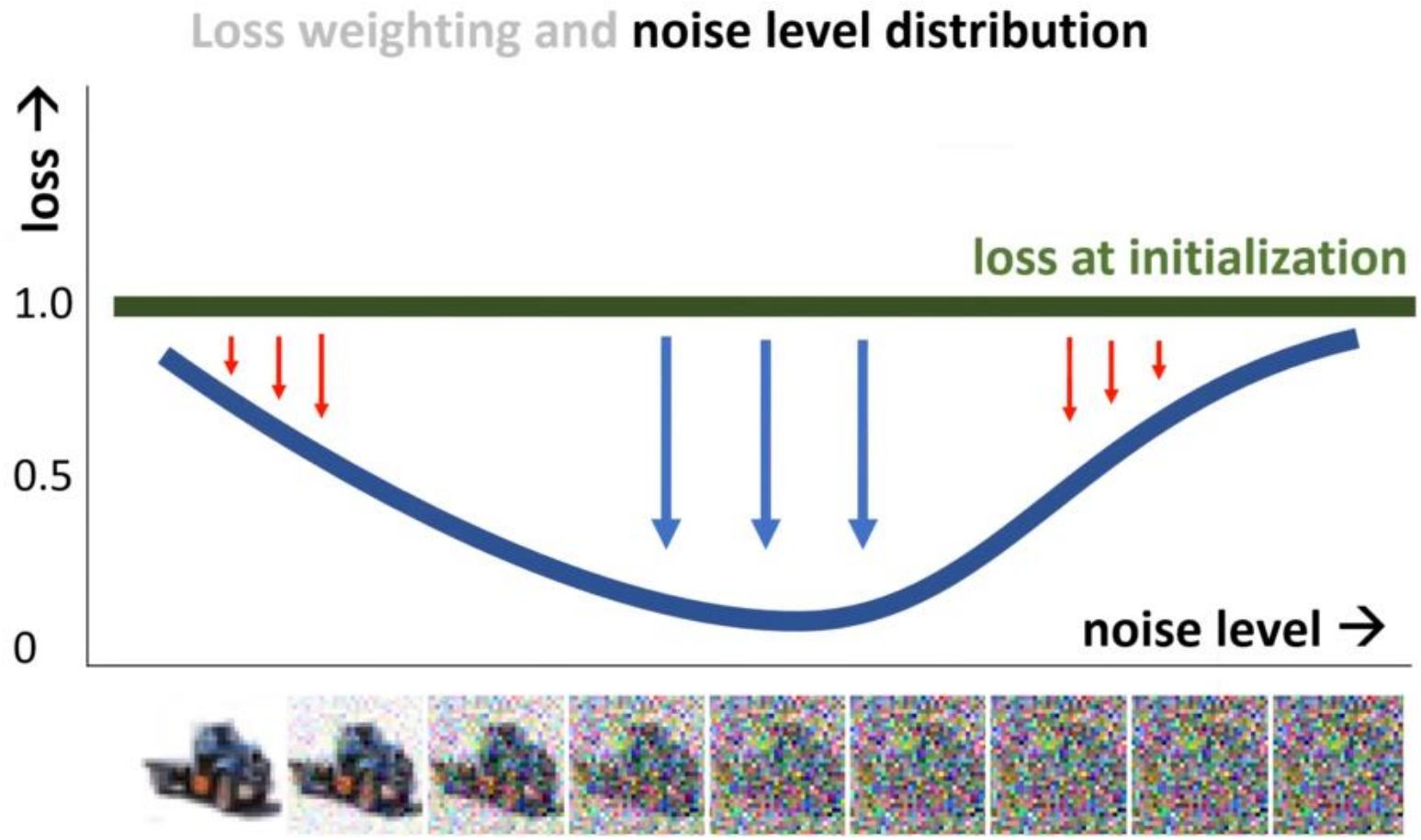




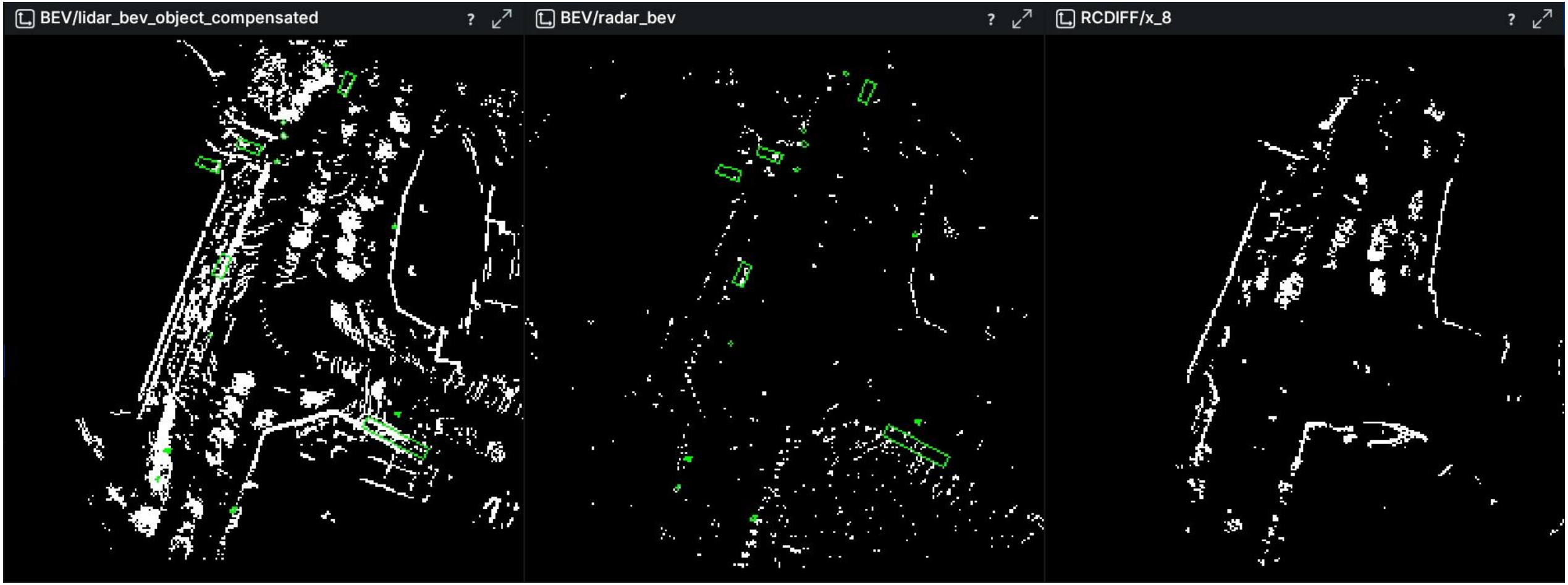
CenterPoint – 3D Detection Benchmark



End-To-End Approach



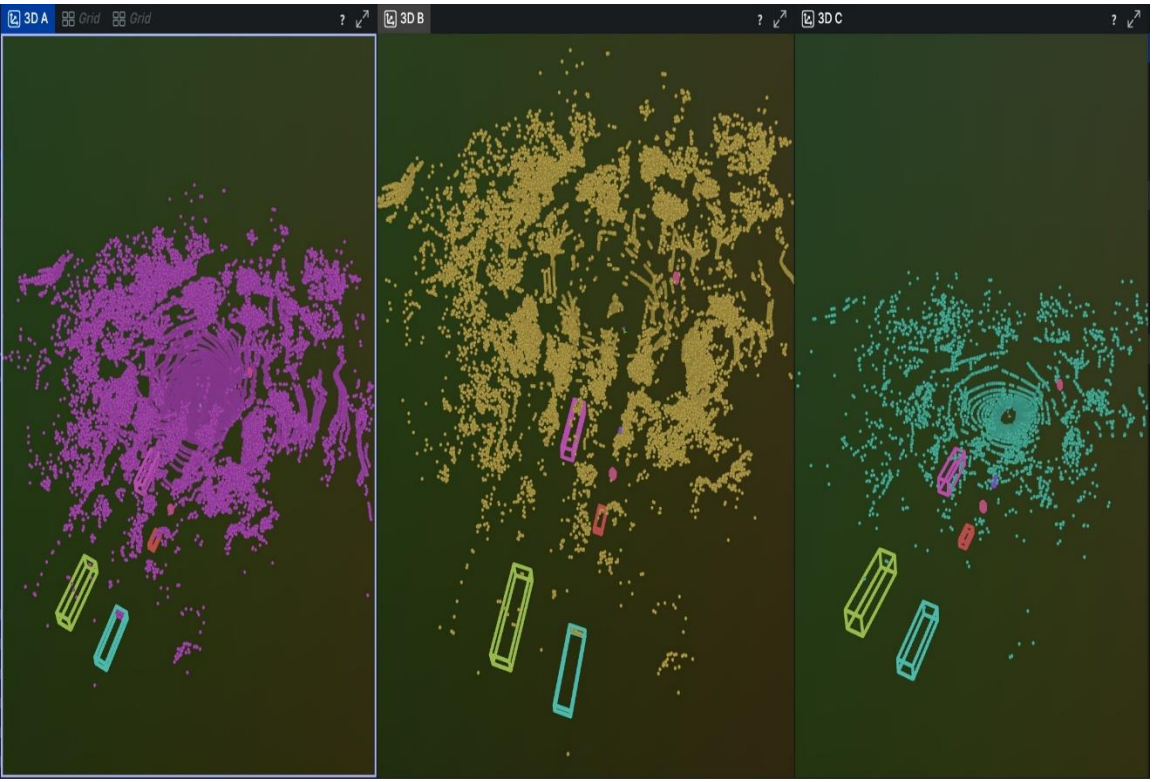
Qualitative Results

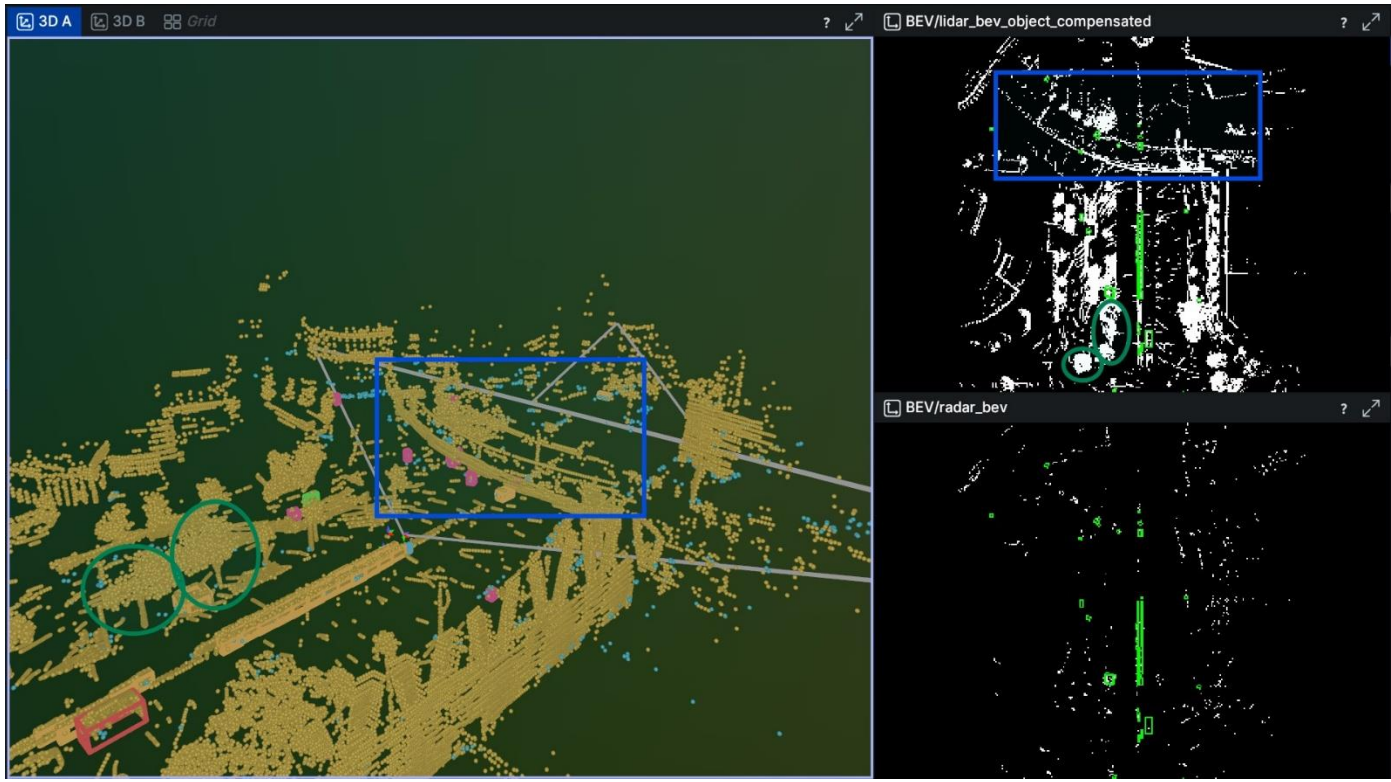


Quantitative Results

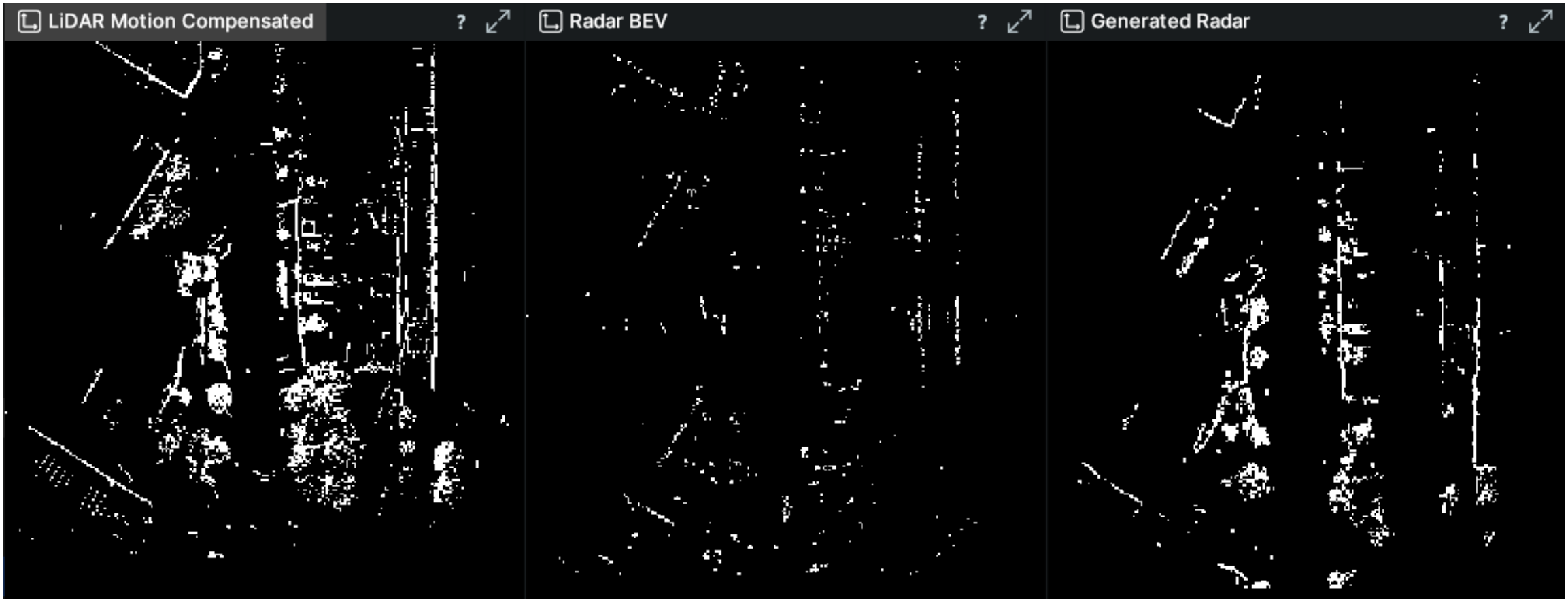
Approach	Occupancy IoU (%) Train-Val	Height MSE Train-Val	mAP (%)
• End-To-End: Train Occupancy & Height Map Denoise + CenterPoint – Loss: MSE + L_{cls} + L_{reg} – Conditioning Signal: Radar Occupancy + Height Map – Augmentation: None	19-12	15-11	2
• Stage 1: Train Occupancy & Height Denoiser – Loss: MSE + BCE – Conditioning Signal: Radar Occupancy + **Height Map** – Augmentation: Cut & Mix • Stage 2: Train CenterPoint on Generated Points	37-18	6-8	5
• Stage 1: Train Occupancy & Height Denoiser – Loss: MSE – Conditioning Signal: Radar Occupancy Map – Augmentation: Cut & Mix • Stage 2: Train CenterPoint on Generated Points	38-21	6-9	5
• Stage 1: Train Occupancy & Height Denoiser – Loss: MSE + BCE – Conditioning Signal: Radar Occupancy Map – Augmentation: Cut & Mix • Stage 2: Train CenterPoint on Generated Points	34–15	7–11	3
• Stage 1: Train Occupancy Denoiser – Loss: MSE + LPIPS – Conditioning Signal: Radar Occupancy Map – Augmentation: Cut & Mix • Stage 2: Train CenterPoint on Generated Points	46–28	-	6
• Stage 1: Train Occupancy Denoiser – Conditioning Signal: Radar Occupancy Map – Loss: MSE + LPIPS – Augmentation: None • Stage 2: Train CenterPoint on Generated Points	39–22	-	4

Issues With Supervision Signal





Denoising Occupancy & Height Map

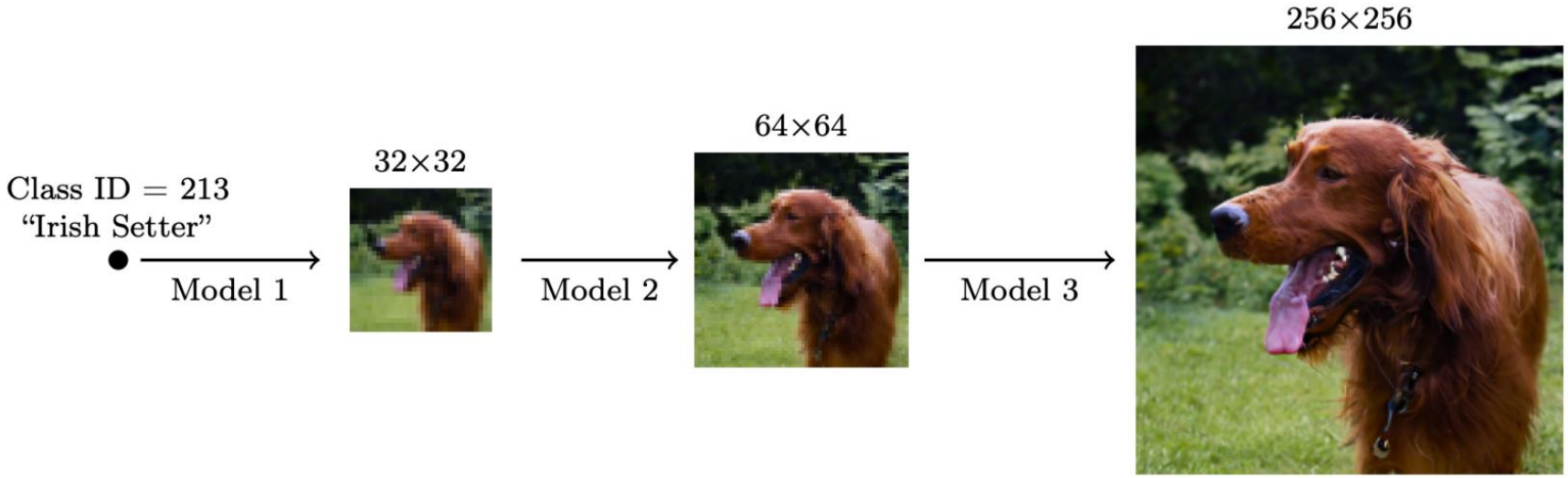


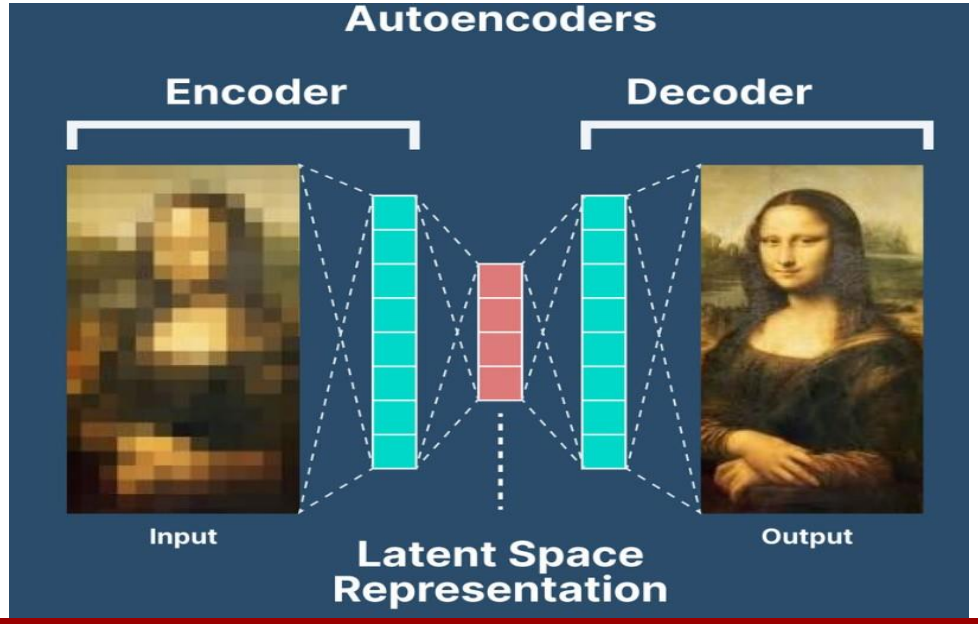
Future Work - Upsampling

Model	Preprocessing Pipeline	Pointcloud Density(Relative to Original PC)	mAP
CenterPoint	• Points From LiDAR MultiSweeps • Ego + Object Motion Compensation • Ground Point Removal • Points to BEV (Image Encoded Height) • BEV to Points	70% @ 1024×1024 32% @ 512×512	42% @ 1024×1024 23% @ 512×512

Table 7.1: Comparison of preprocessing pipelines and resulting mAP.

Future Work - Upsampling





Future Work – Diffusion in Voxel Space