ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-6874-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICASSP49660.2025.10888952

Generative Sensing: Pre-training LiDAR with Masked Autoencoders for Ultra-Frugal Perception

Sina Tayebati, Theja Tulabandhula, and Amit Ranjan Trivedi
University of Illinois at Chicago (UIC), Chicago, USA

***Abstract*—We propose a disruptively frugal *generative sensing* approach for LiDAR that *generates*, rather than *senses*, parts of the environment that are either predictable based on extensive training or have limited impact on overall prediction accuracy. Our generative pre-training strategy for this purpose, radially masked autoencoding (R-MAE), also allows focusing on radial segments of the data, which captures spatial relationships and distances between objects more effectively than conventional procedures. As a result, the proposed methodology not only reduces sensing energy but also improves prediction accuracy. Our evaluations on Waymo and nuScenes show that our approach achieves over a 5% average precision improvement in detection tasks across datasets and over a 4% accuracy improvement when transferring domains from Waymo and nuScenes to KITTI. Our method achieves up to 3.17% and 2.31% improvements in mean average precision (mAP) and NDS, respectively, in nuScenes. Even with 90% radial masking, it surpasses baseline models by up to 5.59% in mAP and mean average precision with heading (mAPH) across all object classes in the Waymo dataset. Resultantly, R-MAE reduces the sensing energy by $\sim$9.1$\times$ than the conventional LiDAR processing despite computational overheads. Codes are available: *https://github.com/sinatayebati/Radial_MAE*.**

***Index Terms*—LiDAR Pre-training, Masked Autoencoder, Ultra-Efficient 3D Sensing.**

I. Introduction

Multispectral sensors like LiDARs (Light Detection and Ranging) excel in depth perception and object detection under various lighting conditions, including complete darkness and bright sunlight. Unlike cameras, LiDAR outputs are unaffected by optical illusions or ambient light variations. However, due to their *active sensing*–where they emit signals and measure reflections–LiDARs are significantly more energy-intensive than cameras. For example, Luminar's LiDAR can consume up to 25 watts [1], compared to the 1-2 watts required by modern digital cameras [2], resulting in significant adoption barriers for LiDAR sensing in many low-power applications.

We propose a generative sensing LiDAR approach that *generates* rather than *senses* parts of the environment that are predictable based on extensive training or have limited impact on overall prediction accuracy. By minimally sensing the environment and using generative models to ***fill in the blanks***, our approach can dramatically reduce LiDAR energy consumption. While previous works [3]–[7] have used generative AI to compress LiDAR point cloud data for more efficient processing, they miss upon the significant opportunity to minimize sensor power itself. By rather focusing on minimizing sensing energy than feature dimensions alone, our approach aligns with a notable trend in semiconductor technology advancements, where computing energy decreases at a much faster rate than sensing energy, such as with emerging technologies [8]–[13]. The computing energy of a digital technology is determined by how it represents binary bits '0' and '1,' and with each new generation of transistors and emerging technologies, the energy of binary representations continues to dramatically reduce. Meanwhile, sensing energy, especially for active sensing, is more fundamentally constrained by environmental factors such as atmospheric absorption, signal scattering, and reflection characteristics [14], [15]. Therefore, by leveraging generative models to overcome these fundamental energy reduction barriers, our approach provides a more scalable pathway of energy reduction.

Leveraging generative modeling for LiDAR sensing efficiency, we introduce R-MAE, a pre-training paradigm employing a masked autoencoder with a novel range-aware radial masking strategy. This approach reduces the need for extensive sensing by combining partial observations with a pre-trained generative model to reconstruct the 3D scene. Extensive experiments on Waymo [16] and nuScenes [17] show that R-MAE preserves spatial continuity and ensures viewpoint invariance even with 90% masking. Training with R-MAE also allows the model to focus on radial aspects of the data, better capturing spatial relationships. Thus, R-MAE achieved >5% average precision improvement in detection tasks across Waymo and other datasets, with >4% improvement in domain transfers from Waymo and nuScenes to KITTI.

II. Physical Sensing vs. Generative Sensing

LiDAR's power consumption includes laser emission, scanning, signal processing, and data acquisition, requiring $P_{\text{total}} = P_{\text{laser}} + P_{\text{scan}} + P_{\text{signal}} + P_{\text{control}}$. The laser emitter's power, P_{laser}, depends on energy per pulse, E_{pulse}, pulse repetition frequency, f_{pulse}, and laser efficiency, η_{laser}: $P_{\text{laser}} = \frac{E_{\text{pulse}} \times f_{\text{pulse}}}{\eta_{\text{laser}}}$. Control and data acquisition power, P_{control}, includes data handling, system control, and communication: $P_{\text{control}} = P_{\text{ADC}} + P_{\text{MCU}}$, where P_{ADC} is the power consumption of analog-to-digital converters and P_{MCU} is microcontroller's power consumption.

Importantly, these energy components face fundamental *energy-accuracy-range trade-offs*. For instance, as range R increases, E_{pulse} increases: $E_{\text{pulse}} = \frac{P_r \cdot (4\pi R^2)^2 \cdot \tau}{A_r \cdot \rho \cdot \eta}$, where A_r is receiver aperture area, ρ is target reflectivity, η is system efficiency, and τ is laser pulse width. Since minimum received signal strength P_r cannot be arbitrarily small, transmission energy increases as R^4. Similarly, angular precision, $\Delta\theta$, measures the accuracy of angle distinction between objects,

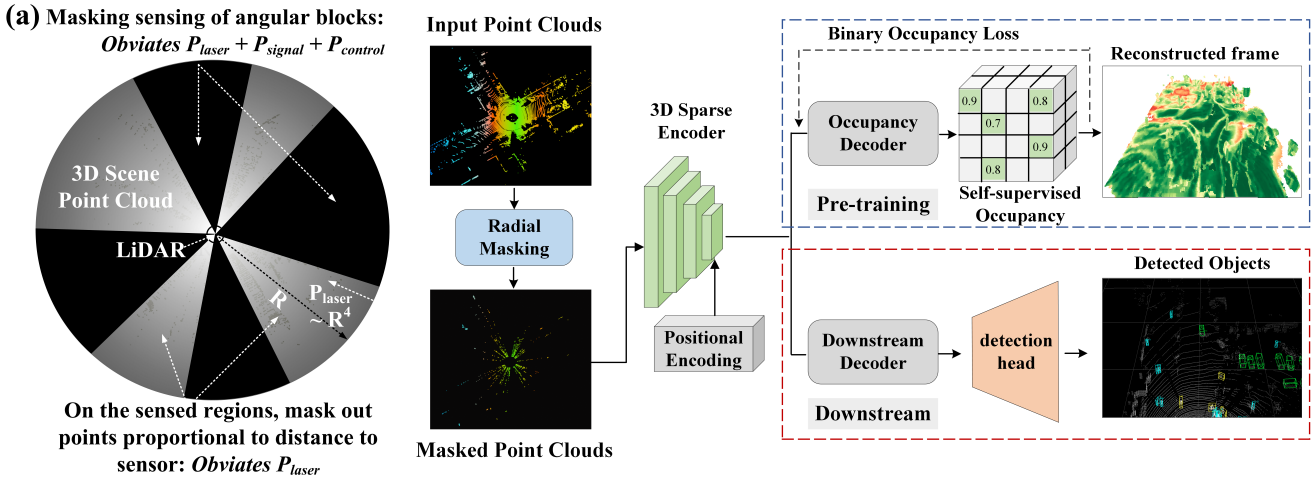

Fig. 1: **(a) Radially masked autoencoding (R-MAE) strategy:** Angular regions are completely masked out as shown in the black by turning off laser emissions. Even on the sensed regions, points are probabilistically dropped-off proportional to their distance R. Notably, $P_{\text{laser}} \sim R^4$, therefore the pretraining encourages models to predict accurately with a low power laser. **(b) R-MAE processing flow:** The input point cloud is voxelized and radially masked based on voxel distance from the sensor. A 3D spatially sparse convolutional encoder extracts latent features, while a decoder reconstructs the 3D scene.

determined by beam divergence and laser aperture diameter D: $\Delta\theta = \frac{\lambda}{D}$, where λ is laser wavelength. Finer angular precision (smaller $\Delta\theta$) requires larger aperture D, increasing LiDAR's footprint, while lower λ for finer precision is constrained by eye safety and atmospheric absorption issues due to higher transmitted wave energy [18]. Due to such intricate interactions among various LiDAR performance metrics, simultaneously achieving high precision, extended range, low footprint, energy efficiency, and cost-effectiveness is challenging for most LiDAR systems. We address these challenges through a novel generative sensing paradigm where *training data* rather than *sensor design and energy* are relied upon. As a result, our approach reduces the sensing energy by $\sim 9.1\times$ than the conventional LiDAR sensing.

III. Radially Masked Autoencoding (R-MAE)

While random masking has proven effective in pre-training models across various modalities [19]–[21], its direct application to large-scale LiDAR point clouds is challenging. LiDAR data is inherently irregular and sparse, making conventional block-wise masking less effective. To address this, we propose Radially Masked Autoencoding (R-MAE), which masks random angular portions of a LiDAR scan and uses an autoencoder to predict occupancy in these unsensed regions. Pre-training R-MAE on unlabeled point clouds enables it to capture underlying geometric and semantic structures. By generating, rather than sensing, significant portions of the environment, R-MAE reduces LiDAR scan requirements. Additionally, R-MAE is easily implementable on modern LiDAR systems by allowing for selective laser activation during inference. Key components of R-MAE are detailed below:

A. Radial Masking Strategy

To process large-scale LiDAR point clouds while mimicking the sensor's radial scanning mechanism, we employ a voxel-based radial masking strategy. The point cloud is voxelized into a set of non-empty voxels V, where each voxel $v_i \in V$ is characterized by a feature vector $f_i \in \mathbb{R}^C$ encompassing its geometric and reflectivity properties. The radial masking function, denoted as $M : V \to \{0,1\}$, operates on the cylindrical coordinates of each voxel v_i, represented as (r_i, θ_i, z_i), where r_i is the radial distance, θ_i is the azimuth angle, and z_i is the height in two steps:

Stage 1: Angular Group Selection: Voxels are grouped based on their azimuth angle θ_i into N_g angular groups, where each group spans an angular range of $\Delta\theta = \frac{2\pi}{N_g}$. A subset of these groups is randomly selected with a selection probability $p_g = 1 - m$, where m is the desired masking ratio.

Stage 2: Range-Aware Masking within Selected Groups: Let $G_s \subset \{1, 2, ..., N_g\}$ denote the indices of the selected groups. Within each selected group $g_j \in G_s$, voxels are further divided into N_d distance subgroups based on their radial distance r_i. The distance ranges for these subgroups are defined by thresholds $r_{t_1}, r_{t_2}, ..., r_{t_{N_d}}$. For each voxel v_i in a selected group g_j, a masking decision is made based on its distance subgroup $k(v_i)$ and a range-dependent probability $p_{m_{j,k}}$:

$$M(v_i) = \begin{cases} 0, & \text{if } g(v_i) \in G_s \text{ and } \text{Bernoulli}(p_{m_{g(v_i),k(v_i)}}) = 1 \\ 1, & \text{otherwise} \end{cases}$$

where $g(v_i)$ denotes the group index to which voxel v_i belongs, $k(v_i)$ denotes the distance subgroup index to which voxel v_i belongs within its group, $\text{Bernoulli}(p)$ represents a Bernoulli random variable with success probability p.

B. Convolutional Encoder and Decoder

We employ 3D sparse convolutions [31] for encoder, operating only on non-empty voxels, to reduce memory usage. Specifically, unlike Transformer-based methods that flatten 3D point clouds into 2D pillars [5], [32], sparse convolutions

TABLE I: Accuracy of R-MAE against current art on Waymo Validation set.

Method	mAP/mAPH L2	Vec_L1 AP/APH	Vec_L2 AP/APH	Ped_L1 AP/APH	Ped_L2 AP/APH	Cyc_L1 AP/APH	Cyc_L2 AP/APH
CenterPoint [22]	64.51 / 61.92	71.33 / 70.76	63.16 / 62.65	72.09 / 65.49	64.27 / 58.23	68.68 / 67.39	66.11 / 64.87
+ GCC-3D [23]	$65.29^{+0.78}$ / $62.79^{+0.87}$	-	63.97 / 63.47	-	64.23 / 58.47	-	67.68 / 66.44
+ PropCount [24]	$66.42^{+1.91}$ / $63.85^{+1.93}$	-	64.94 / 64.42	-	66.13 / 60.11	-	68.19 / 67.01
+ Occ-MAE [4]	$65.86^{+1.35}$ / $63.23^{+1.31}$	71.89 / 71.33	64.05 / 63.53	73.85 / 67.12	65.78 / 59.62	70.29 / 69.03	67.76 / 66.53
+ R-MAE (Ours)	**67.20**$^{+2.69}$ **/ 64.76**$^{+2.84}$	**73.38 / 72.85**	**65.28 / 64.79**	**74.84 / 68.68**	**66.90 / 61.24**	**72.05 / 70.84**	**69.43 / 68.26**
PV-RCNN [25]	64.84 / 60.86	75.41 / 74.74	67.44 / 66.80	71.98 / 61.24	63.70 / 53.95	65.88 / 64.25	63.39 / 61.82
+ GCC-3D [4]	$61.30^{-3.54}$ / $58.18^{-2.68}$	-	65.65 / 65.10	-	55.54 / 48.02	-	62.72 / 61.43
+ PropCount [24]	$62.62^{-2.22}$ / $59.28^{-1.58}$	-	66.04 / 65.47	-	57.58 / 49.51	-	64.23 / 62.86
+ MAELi [6]	- / $62.14^{+1.28}$	-	- / 67.34	-	- / 56.32	-	- / 62.76
+ Occ-MAE [4]	$66.17^{+1.33}$ / $61.98^{+1.12}$	75.94 / 75.28	67.94 / 67.34	74.02 / 63.48	64.94 / 55.57	67.21 / 66.49	65.62 / 63.02
+ R-MAE (Ours)	**68.95**$^{+4.11}$ **/ 66.45**$^{+5.59}$	**76.72 / 76.22**	**68.38 / 67.92**	**78.19 / 71.74**	**69.63 / 63.68**	**72.44 / 70.32**	**68.84 / 67.76**

TABLE II: Accuracy of R-MAE against current art on nuScenes validation set.

Method	**Data Modality**	**mAP↑**	**NDS↑**	**mATE↓**	**mASE↓**	**mAOE↓**	**mAVE↓**	**mAAE↓**
CenterPoint [22]	LiDAR	56.03	64.54	30.11	25.55	38.28	21.94	18.87
+ GCC-3D [23]	LiDAR	$57.3^{+1.3}$	$65.0^{+0.5}$	-	-	-	-	-
+ BEV-MAE [26]	LiDAR	$57.2^{+1.2}$	$65.1^{+0.6}$	-	-	-	-	-
+ Occupancy-MAE [4]	LiDAR	$57.8^{+1.8}$	$65.9^{+1.4}$	28.8	24.4	37.5	20.6	18.0
+ R-MAE (Ours)	LiDAR	**59.20**$^{+3.17}$	**66.85**$^{+2.31}$	29.73	25.71	34.16	20.02	17.91
Transfusion [27]	LiDAR	64.58	69.43	27.96	25.37	29.35	27.31	18.55
+ R-MAE (Ours)	LiDAR	**65.01**$^{+0.43}$	**70.17**$^{+0.74}$	28.19	25.20	26.92	24.27	18.71
BEVFusion [28]	LiDAR + Camera	65.91	70.20	28.26	25.43	28.88	26.80	18.67
+ R-MAE (Ours)	LiDAR + Camera	**66.40**$^{+0.49}$	**70.41**$^{+0.21}$	28.31	25.54	29.57	25.87	18.60

TABLE III: Average Precision (AP) of R-MAE against current art on KITTI validation set († results are reproduced by us).

Model	**Car**	**Pedestrian**	**Cyclist**	**Model**	**Car**	**Pedestrian**	**Cyclist**
SECOND† [29]	79.08	44.52	64.49	PV-RCNN [25]	82.28	51.51	69.45
+ Occupancy-MAE† [4]	$79.12^{+0.04}$	$45.35^{+0.83}$	$63.27^{-1.22}$	+ Occ.-MAE† [4]	$82.43^{+0.15}$	$48.13^{-3.38}$	$71.51^{+2.06}$
+ ALSO† [30]	$78.98^{-0.10}$	$45.33^{+0.81}$	$66.53^{+2.04}$	+ ALSO† [30]	$82.52^{+0.24}$	$52.63^{+1.12}$	$70.20^{+0.75}$
+ R-MAE (Ours)	**79.10**$^{+0.02}$	**46.93**$^{+2.41}$	**67.75**$^{+3.26}$	+ R-MAE (Ours)	**82.82**$^{+0.54}$	**51.61**$^{+0.10}$	**73.82**$^{+4.37}$

directly operate in 3D space, preserving the scene's geometric structure and enabling the model to learn more nuanced spatial relationships. The encoder, $E : V_s \times \mathbb{R}^C \to \mathbb{R}^L$, transforms the input features $f_i \in \mathbb{R}^C$ of the unmasked voxels $v_i \in V_s$ into a lower-dimensional latent representation $z_i \in \mathbb{R}^L$ through a series of sparse convolutional blocks with 3D convolutions, batch normalization, and ReLU activation. Residual connections are employed to improve gradient flow. The resulting latent representation z_i encapsulates learned geometric and semantic features, which are then passed to the decoder.

The decoder, $D : \mathbb{R}^L \to \mathbb{R}^{|V|}$, reconstructs the 3D scene by predicting the occupancy probability $\hat{o}_i$ for masked voxels v_i. Each deconvolution layer follows batch normalization and ReLU activation to upsample the feature maps, thereby increasing spatial resolution until the original voxel grid is reconstructed. The final layer outputs the predicted occupancy probability $\hat{o}_i$ for each voxel, which is then compared to the ground truth occupancy o_i using a binary cross-entropy loss.

C. Occupancy Loss

We adopt occupancy prediction as a pretext task to encourage the model to learn global features representative of different object categories. Occupancy prediction is framed as a binary classification problem due to the prevalence of empty voxels in outdoor scenes and our deliberate partial sensing to conserve energy. The binary cross-entropy loss with logits L_{occup} is used to supervise the reconstruction:

$$-\frac{1}{|B|}\sum_{i\in B}\frac{1}{|Q_s|}\sum_{q\in Q_s}\left[o_q^i \log(\sigma(\hat{o}_q^i)) + (1-o_q^i)\log(1-\sigma(\hat{o}_q^i))\right]$$

where $\hat{o}_q^i$ is the estimated probability of query voxel q of the i-th training sample, o_q^i is the corresponding ground truth occupancy (1 for occupied, 0 for empty), and σ is the sigmoid function. $|B|$ is the batch size, and $|Q_s|$ is the number of query voxels in the sphere centered on S.

IV. Accuracy-Energy of Generative Sensing

A. 3D Object Detection

In Table I, we evaluated R-MAE on the Waymo validation set, where it surpassed the accuracy of other pre-training methods. Notably, R-MAE consistently achieved the highest mAP and mAPH scores across both CenterPoint and PV-RCNN, with gains of +2.69/+2.84 in CenterPoint and +4.11/+5.59 in PV-RCNN. It also shows significant improvements in detecting various object types, such as pedestrians (Ped_L1, Ped_L2),

TABLE IV: Accuracy comparisons under domain transfer: We pre-trained R-MAE with Waymo and nuScenes and fine-tuned with KITTI. Evaluation results are presented at moderate difficulty level and 40 recall positions.

Task	method	Car	Pedestrian	Cyclist
	SECOND [29]	79.08	44.52	64.49
Waymo → KITTI	+ R-MAE (Ours)	**79.30**$^{+0.22}$	**48.61**$^{+4.09}$	**66.62**$^{+2.13}$
nuScenes → KITTI	+ R-MAE (Ours)	**79.32**$^{+0.24}$	**46.05**$^{+1.53}$	**68.27**$^{+3.78}$

vehicles (Vec_L1, Vec_L2), and cyclists (Cyc_L1, Cyc_L2), especially with better accuracy for small and dynamic objects.

Table II shows R-MAE's performance on nuScenes. Pre-trained on nuScenes LiDAR data, R-MAE improved LiDAR-only models CenterPoint [22] and Transfusion [27] by 2.31% to 3.17% in NDS (NuScenes detection score combining mAP with localization, size, orientation, velocity, and attribute accuracy) and mAP. For multi-modal BEVFusion [28], pre-training only the LiDAR component yielded 0.49% and 0.21% gains. R-MAE outperformed other pre-training methods when fine-tuned with CenterPoint, showing the advantage of our method.

Finally, Table III shows R-MAE's performance on KITTI dataset. R-MAE improves performance by 0.1% to 4.3% over SECOND [29] and PV-RCNN [25], especially for smaller objects. Compared to other pre-trained models, R-MAE shows similar avergae precision (AP) for cars and 1.2% to 1.6% improvement for pedestrians and cyclists. R-MAE enhances ALSO's [30] occupancy prediction approach using masked point clouds and a 3D MAE backbone for better semantic understanding. It also improves upon Occupancy-MAE [4] by employing radial masking, enhancing generative tasks and edge device efficiency.

B. Transferring Domain

To evaluate the transferability of the learned representation, we fine-tuned R-MAE models on the KITTI dataset using SECOND [29] and PVRCNN [25] as detection bases. As shown in Table IV, the performance gains from R-MAE pre-training are evident in the KITTI domain, indicating that the model acquires a robust, generic representation. Despite the KITTI training samples being smaller compared to Waymo and nuScenes, the R-MAE pre-trained models show significant improvements across different classes. However, the relative improvement is smaller when transferring to KITTI, likely due to the domain gap. This suggests that while R-MAE effectively learns generalizable features, variations in data domains can impact performance improvements.

C. Energy Savings and R-MAE's Overhead

In Fig. 2, we explored R-MAE's accuracy limits by varying the masking ratio and contiguous angular segment size. Fig. 2 shows the experiments using a pre-trained R-MAE fine-tuned on PointPillar [33]. In Fig. 2(a), accuracy degrades only beyond 92% masking, indicating that only 8% of LiDAR's BEV (bird's eye view) needs sensing. Fig. 2(b) highlights R-MAE's near-invariance to angular grouping range with 80% masking.

While generative sensing reduces the need for direct environmental sensing, it introduces computational overhead. Table V outlines these trade-offs. The energy for a single LiDAR pulse varies with distance: ∼500 nJ for 10 meters and ∼50 µJ for 200 meters as in [34]. Our range-aware masking (*Stage 2*, Sec. III-A) accounts for this by increasing the masking probability for farther segments of point clouds. In our scheme, up to 30 meters, masking matches the baseline considered in Fig. 2, while from 30–50 meters, a 70% masking probability is multiplied, and beyond 50 meters, a 90% masking probability is multiplied on the baseline ratio. Given that pulse energy follows $E_{pulse} \sim R^4$, this reduces the average pulse energy to ∼5.5 µJ. R-MAE requires ∼830K parameters and ∼335M FLOPs per 360° scan to reconstruct unmasked regions. On Jetson Nano (472 GFLOPs/s at 10W [35]), this computational overhead amounts to ∼7.1 mJ, while sensing energy drops to 0.79 mJ with a 0.25° horizontal resolution. Therefore, in Table V, the total sensing + reconstruction energy of R-MAE is 9.11× lower than the conventional LiDAR processing, showing the sensor energy savings far exceed the computational cost.

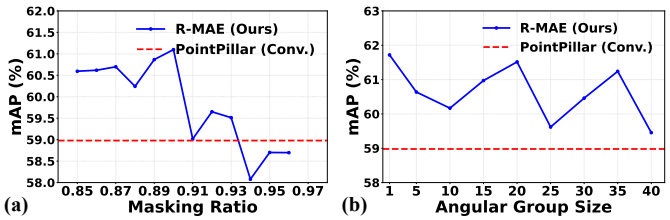

Fig. 2: Accuracy under varying pretraining conditions, **(a)** at varying masking ratios with a fixed angular range of 1 degree, and **(b)** at different angular ranges with a fixed masking ratio of 0.8. Results are compared against the state-of-the-art (SOTA) PointPillars [33] method.

TABLE V: Conventional LiDAR Processing vs. R-MAE

Parameter	Conventional	R-MAE
Masking Ratio	-	>90%
Avg. Laser Diode Energy/pulse	50 µJ	5.5 µJ
# of Model Parameters	-	830K
FLOP/360° scan	-	335M
Sensing Energy/360° scan	72 mJ	792 µJ
Reconstruction Energy/360° scan	-	7.1 mJ

V. Conclusion

R-MAE-based LiDAR processing *generates* rather than *senses* predictable or inconsequential parts of the environment, enabling ultrafrugal sensing. R-MAE improves average precision by over 5% in detection tasks and accuracy by over 4% in domain transfer on Waymo, nuScenes, and KITTI datasets. R-MAE surpasses baselines by up to 5.59% in mAP/mAPH on Waymo with 90% radial masking, thereby resulting in 9.11× lower energy than the convention LiDAR processing despite reconstruction overheads.

Acknowledgement: This work was supported in part by COGNISENSE, one of seven centers in JUMP 2.0, a Semiconductor Research Corporation (SRC) program sponsored by DARPA and NSF Award #2329096.

REFERENCES

[1] Corneliu Rablau, "Lidar–a new (self-driving) vehicle for introducing optics to broader engineering and non-engineering audiences," in *Education and training in optics and photonics*. Optica Publishing Group, 2019, p. 11143_138.

[2] Furkan E Sahin, "Long-range, high-resolution camera optical design for assisted and autonomous driving," in *photonics*. MDPI, 2019, vol. 6, p. 73.

[3] Honghui Yang, Tong He, Jiaheng Liu, Hua Chen, Boxi Wu, Binbin Lin, Xiaofei He, and Wanli Ouyang, "Gd-mae: generative decoder for mae pre-training on lidar point clouds," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 9403–9414.

[4] Chen Min, Liang Xiao, Dawei Zhao, Yiming Nie, and Bin Dai, "Occupancy-mae: Self-supervised pre-training large-scale lidar point clouds with masked occupancy autoencoders," *IEEE Transactions on Intelligent Vehicles*, 2023.

[5] Runsen Xu, Tai Wang, Wenwei Zhang, Runjian Chen, Jinkun Cao, Jiangmiao Pang, and Dahua Lin, "Mv-jar: Masked voxel jigsaw and reconstruction for lidar-based self-supervised pre-training," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 13445–13454.

[6] Georg Krispel, David Schinagl, Christian Fruhwirth-Reisinger, Horst Possegger, and Horst Bischof, "Maeli: Masked autoencoder for large-scale lidar point clouds," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 3383–3392.

[7] Changhyeon Lee, Leila Rahimifard, Junhwan Choi, Jeong-ik Park, Chungryeol Lee, Divake Kumar, Priyesh Shukla, Seung Min Lee, Amit Ranjan Trivedi, Hocheon Yoo, et al., "Highly parallel and ultra-low-power probabilistic reasoning with programmable gaussian-like memory transistors," *Nature Communications*, vol. 15, no. 1, pp. 2439, 2024.

[8] Samir N Ajani, Prashant Khobragade, Mrunalee Dhone, Bireshwar Ganguly, Nilesh Shelke, and Namita Parati, "Advancements in computing: Emerging trends in computational science with next-generation computing," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 7s, pp. 546–559, 2024.

[9] Mushu Li, Nan Cheng, Jie Gao, Yinlu Wang, Lian Zhao, and Xuemin Shen, "Energy-efficient uav-assisted mobile edge computing: Resource allocation and trajectory optimization," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 3, pp. 3424–3438, 2020.

[10] Giovanni Finocchio, Jean Anne C Incorvia, Joseph S Friedman, Qu Yang, Anna Giordano, Julie Grollier, Hyunsoo Yang, Florin Ciubotaru, Andrii V Chumak, Azad J Naeemi, et al., "Roadmap for unconventional computing with nanotechnology," *Nano Futures*, vol. 8, no. 1, pp. 012001, 2024.

[11] Priyesh Shukla, Ankith Muralidhar, Nick Iliev, Theja Tulabandhula, Sawyer B Fuller, and Amit Ranjan Trivedi, "Ultralow-power localization of insect-scale drones: Interplay of probabilistic filtering and compute-in-memory," *IEEE transactions on very large scale integration (VLSI) systems*, vol. 30, no. 1, pp. 68–80, 2021.

[12] Susmita Dey Manasi, Md Mamun Al-Rashid, Jayasimha Atulasimha, Supriyo Bandyopadhyay, and Amit Ranjan Trivedi, "Skewed straintronic magnetotunneling-junction-based ternary content-addressable memory—part i," *IEEE transactions on Electron Devices*, vol. 64, no. 7, pp. 2835–2841, 2017.

[13] Ahish Shylendra, Priyesh Shukla, Saibal Mukhopadhyay, Swarup Bhunia, and Amit Ranjan Trivedi, "Low power unsupervised anomaly detection by nonparametric modeling of sensor statistics," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 28, no. 8, pp. 1833–1843, 2020.

[14] Mario Bijelic, Tobias Gruber, and Werner Ritter, "A benchmark for lidar sensors in fog: Is detection breaking down?," in *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2018, pp. 760–767.

[15] Joschua Schulte-Tigges, Marco Förster, Gjorgji Nikolovski, Michael Reke, Alexander Ferrein, Daniel Kaszner, Dominik Matheis, and Thomas Walter, "Benchmarking of various lidar sensors for use in self-driving vehicles in real-world environments," *Sensors*, vol. 22, no. 19, pp. 7146, 2022.

[16] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al., "Scalability in perception for autonomous driving: Waymo open dataset," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2446–2454.

[17] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom, "nuscenes: A multimodal dataset for autonomous driving," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11621–11631.

[18] Thinal Raj, Fazida Hanim Hashim, Aqilah Baseri Huddin, Mohd Faisal Ibrahim, and Aini Hussain, "A survey on lidar scanning mechanisms," *Electronics*, vol. 9, no. 5, pp. 741, 2020.

[19] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[20] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick, "Masked autoencoders are scalable vision learners," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 16000–16009.

[21] Yatian Pang, Wenxiao Wang, Francis EH Tay, Wei Liu, Yonghong Tian, and Li Yuan, "Masked autoencoders for point cloud self-supervised learning," in *European conference on computer vision*. Springer, 2022, pp. 604–621.

[22] Tianwei Yin, Xingyi Zhou, and Philipp Krahenbuhl, "Center-based 3d object detection and tracking," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 11784–11793.

[23] Hanxue Liang, Chenhan Jiang, Dapeng Feng, Xin Chen, Hang Xu, Xiaodan Liang, Wei Zhang, Zhenguo Li, and Luc Van Gool, "Exploring geometry-aware contrast and clustering harmonization for self-supervised 3d object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3293–3302.

[24] Junbo Yin, Dingfu Zhou, Liangjun Zhang, Jin Fang, Cheng-Zhong Xu, Jianbing Shen, and Wenguan Wang, "Proposalcontrast: Unsupervised pre-training for lidar-based 3d object detection," in *European conference on computer vision*. Springer, 2022, pp. 17–33.

[25] Shaoshuai Shi, Chaoxu Guo, Li Jiang, Zhe Wang, Jianping Shi, Xiaogang Wang, and Hongsheng Li, "Pv-rcnn: Point-voxel feature set abstraction for 3d object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 10529–10538.

[26] Zhiwei Lin and Yongtao Wang, "Bev-mae: Bird's eye view masked autoencoders for outdoor point cloud pre-training," *arXiv preprint arXiv:2212.05758*, 2022.

[27] Xuyang Bai, Zeyu Hu, Xinge Zhu, Qingqiu Huang, Yilun Chen, Hongbo Fu, and Chiew-Lan Tai, "Transfusion: Robust lidar-camera fusion for 3d object detection with transformers," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 1090–1099.

[28] Zhijian Liu, Haotian Tang, Alexander Amini, Xinyu Yang, Huizi Mao, Daniela L Rus, and Song Han, "Bevfusion: Multi-task multi-sensor fusion with unified bird's-eye view representation," in *2023 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2023, pp. 2774–2781.

[29] Yan Yan, Yuxing Mao, and Bo Li, "Second: Sparsely embedded convolutional detection," *Sensors*, vol. 18, no. 10, pp. 3337, 2018.

[30] Alexandre Boulch, Corentin Sautier, Björn Michele, Gilles Puy, and Renaud Marlet, "Also: Automotive lidar self-supervision by occupancy estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 13455–13465.

[31] Spconv Contributors, "Spconv: Spatially sparse convolution library," https://github.com/traveller59/spconv, 2022.

[32] Georg Hess, Johan Jaxing, Elias Svensson, David Hagerman, Christoffer Petersson, and Lennart Svensson, "Masked autoencoder for self-supervised pre-training on lidar point clouds," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2023, pp. 350–359.

[33] Alex H Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom, "Pointpillars: Fast encoders for object detection from point clouds," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 12697–12705.

[34] Filip Taneski, Tarek Al Abbas, and Robert K Henderson, "Laser power efficiency of partial histogram direct time-of-flight lidar sensors," *Journal of Lightwave Technology*, vol. 40, no. 17, pp. 5884–5893, 2022.

[35] A. Kumar Das, "Nvidia jetson nano review and benchmark - the raspberry pi killer?," 2024, Accessed: 2024-09-07.