FFAM: Feature Factorization Activation Map for Explanation of 3D Detectors

Shuai Liu, Boyang Li, Zhiyu Fang, Mingyue Cui and Kai Huang*

School of Computer Science and Engineering, Sun Yat-sen University

Github: https://github.com/Say2L/FFAM

Introduction

NEURAL INFORMATION PROCESSING SYSTEMS

Background

- ☐ LiDAR-based 3D object detection has made impressive progress recently, yet most existing models are black-box, lacking interpretability.

- ☐ Previous explanation approaches primarily focus on analyzing image-based models and are not readily applicable to LiDAR-based 3D detectors.

Introduction

NEURAL INFORMATION PROCESSING SYSTEMS

Challenges

- ☐ Point clouds are inherently three dimensional.
- ☐ Explanation for specific 3D object in point cloud.
- ☐ Point clouds are sparsely distributed in 3D space.

Introduction

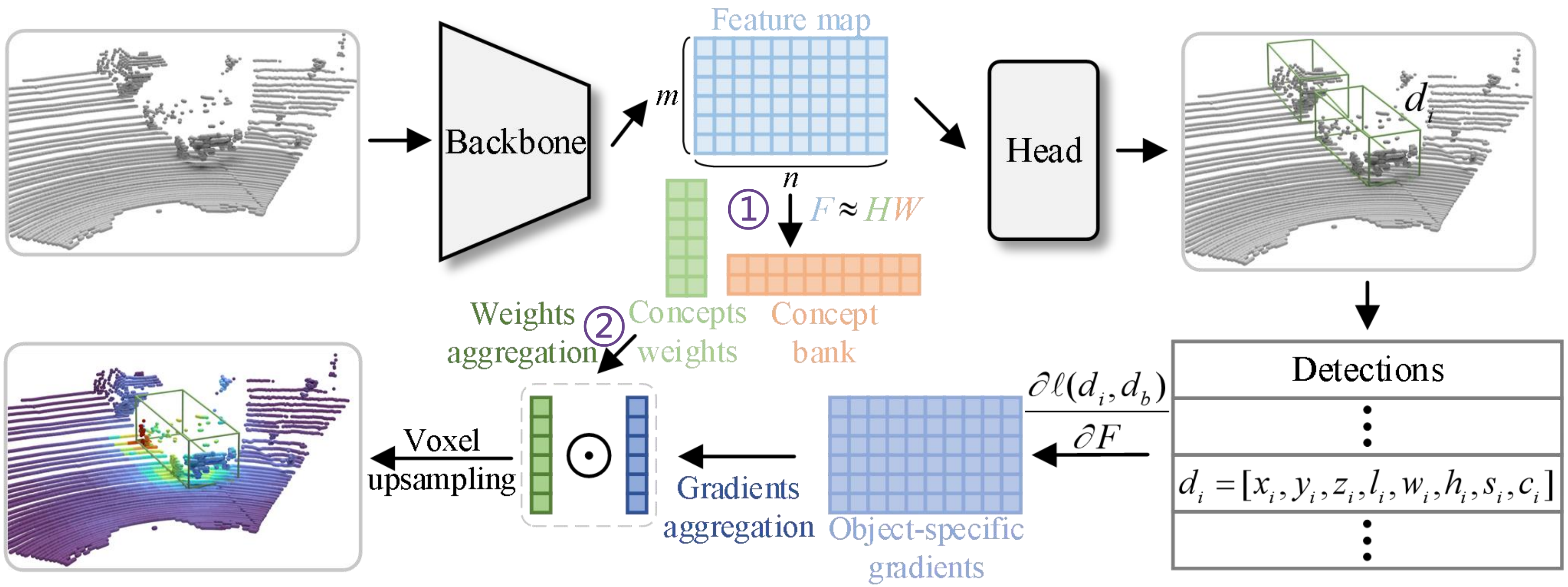
NEURAL INFORMATION PROCESSING SYSTEMS

Contribution

- ☐ We propose a feature factorization activation map (FFAM) method to obtain high-quality visual explanations for 3D detectors.
- ☐ We introduce non-negative matrix factorization (NMF) in explaining point cloud detectors.
- ☐ We utilize feature gradients of an object-specific loss to refine the global concept activation map.
- ☐ A voxel upsampling strategy is proposed to upsample sparse voxels.

Proposed Method

Framework of FFAM



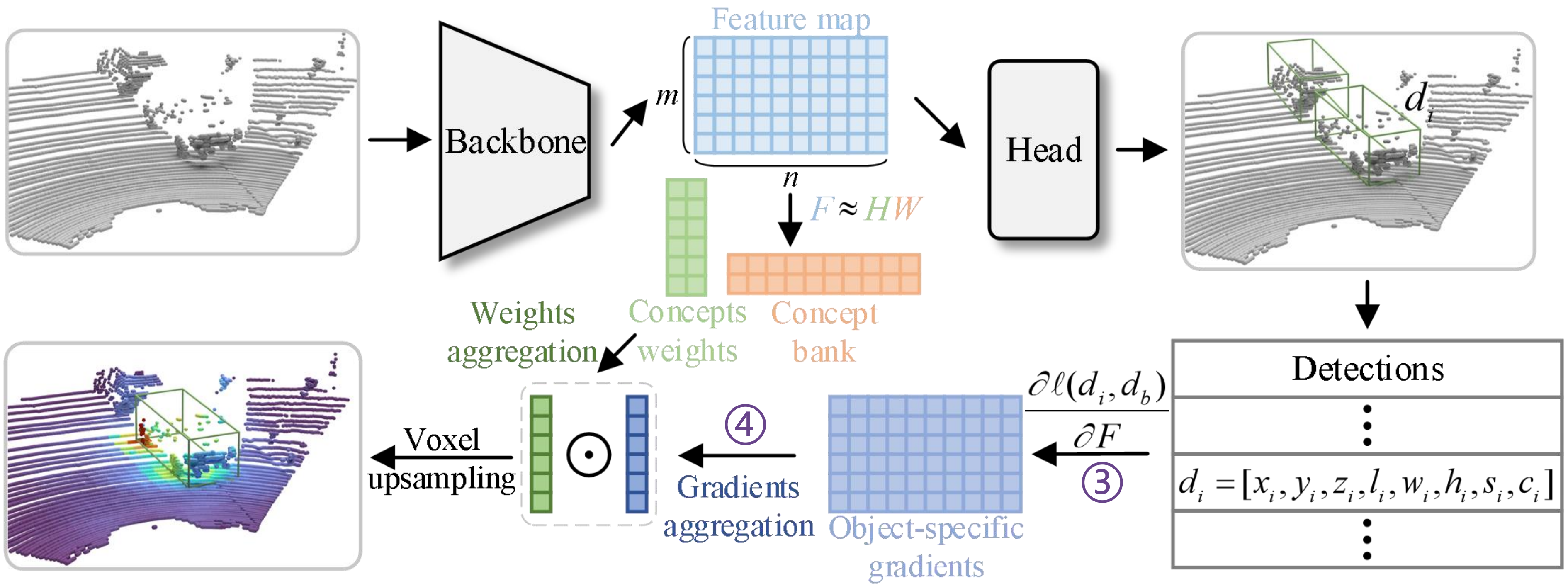
- Calculate the global activation map V:

① $$\mathrm{NMF}(A) = \underset{\hat{A}}{\operatorname{argmin}} \left\| A - \hat{A} \right\|_F^2$$
$$\text{s.t. } \hat{A} = HW, \forall ij, H_{ij}, W_{ij} \geq 0$$

② $$V = \sum_{j=1}^{r} H_{\cdot j}$$

Proposed Method

Framework of FFAM



- Obtain the object-specific activation map M:

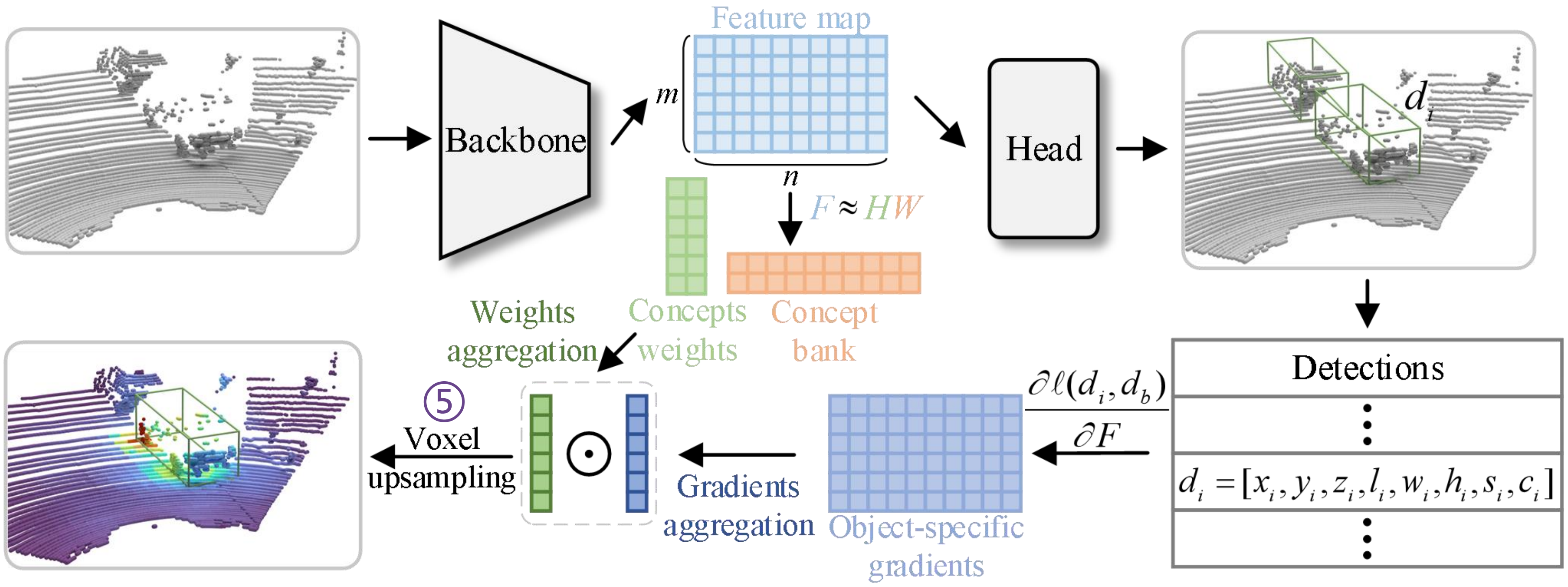
③ $G = \dfrac{\partial \ell}{\partial F}$

④ $\omega = \sum_{k=1}^{d} |G_{\cdot k}|$

$M = \Phi(\omega) \odot \Phi(V)$

Proposed Method

Framework of FFAM

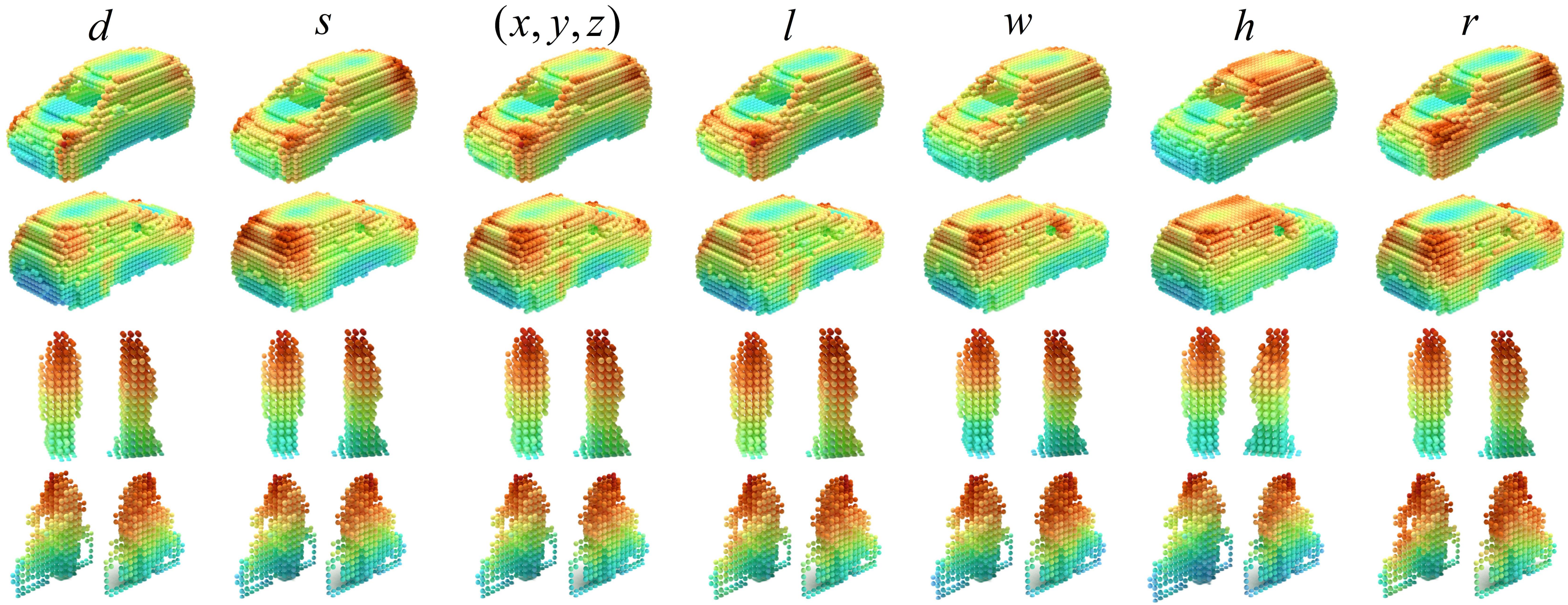


- Voxel Upsampling:

$$⑤ \quad s_p = \sum_{v \in \aleph} \frac{\Psi(d(v_p, v))}{\sum_{v \in \aleph} \Psi(d(v_p, v))} M_v$$

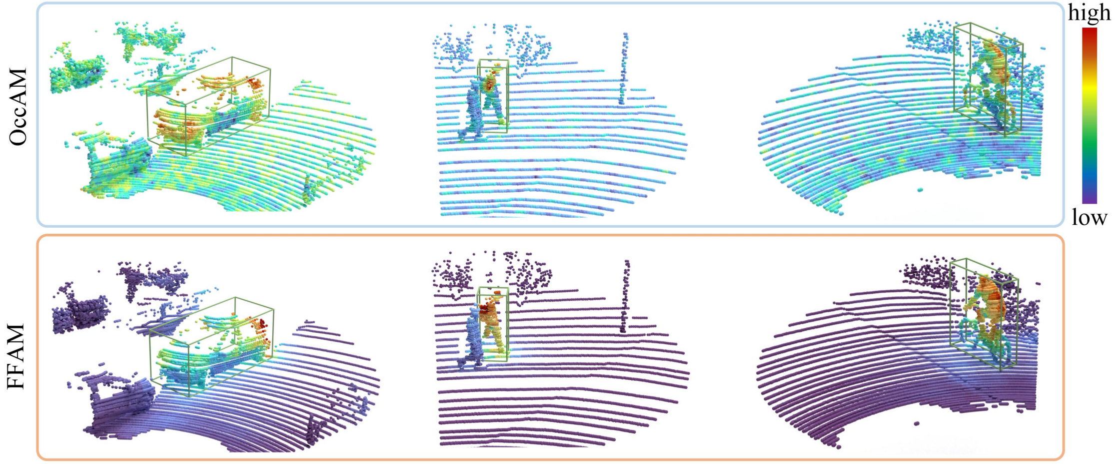
Experiments

Qualitative Results: Average saliency maps

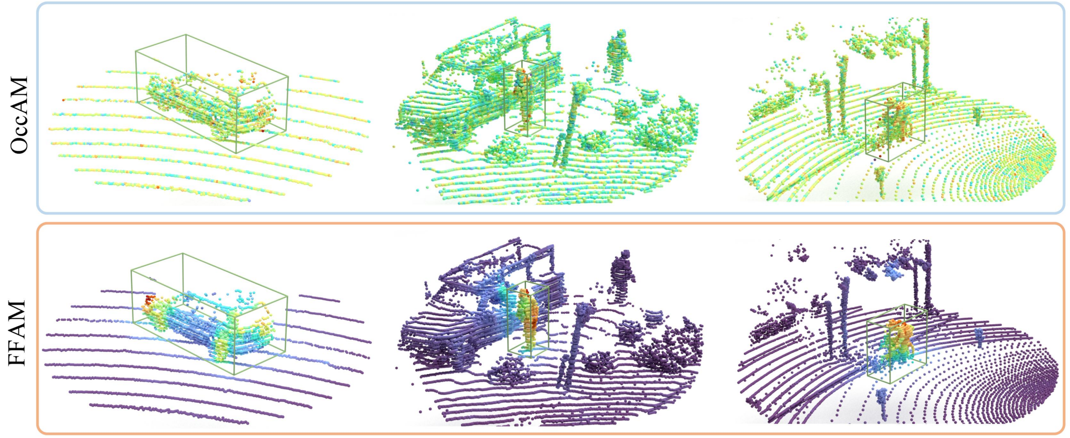


Experiments

Qualitative Results: Comparison of Saliency maps



(a) Saliency maps for SECOND on KITTI dataset.



(b) Saliency maps for CenterPoint on Waymo Open dataset.

Experiments

Quantitative Results

Method	Deletion ↓				Insertion ↑			
	All	Car	Ped.	Cyc.	All	Car	Ped.	Cyc.
Grad-CAM	0.335	0.373	0.137	0.129	0.797	0.821	0.688	0.725
ODAM	0.134	0.138	0.122	0.098	0.885	0.902	0.785	0.828
OccAM	0.286	0.311	0.146	0.167	0.863	0.880	0.761	0.790
FFAM (Ours)	**0.071**	**0.068**	**0.098**	**0.078**	**0.907**	**0.923**	**0.806**	**0.854**

Table 1: AUC for Deletion and Insertion curves. The results of different categories are reported.

Method	PG ↑				enPG ↑			
	All	Car	Ped.	Cyc.	All	Car	Ped.	Cyc.
Grad-CAM	0.093	0.080	0.166	0.163	0.021	0.022	0.014	0.011
ODAM	0.901	0.895	0.939	0.926	0.633	0.639	0.577	0.654
OccAM	0.946	0.957	0.898	0.860	0.023	0.024	0.019	0.013
FFAM (Ours)	**0.991**	**0.989**	**0.999**	**0.998**	**0.664**	**0.671**	**0.591**	**0.719**

Table 2: Comparison of Pointing game (PG) and energy-based Pointing game (enPG) metrics.