Gumbel-NeRF: Representing Unseen Objects as Part-Compositional Neural Radiance Fields

C. Hsu[1], Y. Sekikawa[2], S. Ikehata[1], R. Kawakami[1], I. Sato[1,2]

[1]Institute of Science Tokyo, Japan; [2]Denso IT Laboratory, Inc., Japan

Recognition and Learning Algorithm Laboratory

Institute of SCIENCE TOKYO

DENSO IT LAB

1. Introduction

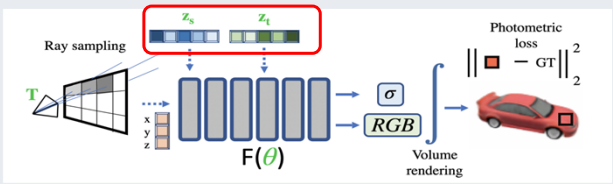
Background: Birds' eye-view vision systems provide intuitive understanding of surrounding objects for driver's assistance.

BEV vision system

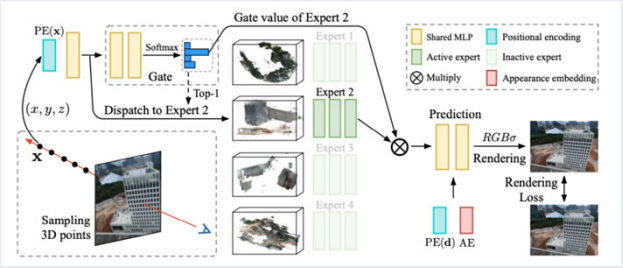
Goal: To develop NeRF models **generalizable to test instances** with enhanced expressivity by **part-based expert modules**.

2. Existing Method

- CodeNeRF[Wonbong+,ICCV2021] models multiple objects for **generalization** by conditioning on learnable latent codes.



- Switch-NeRF[Zhenxing+,ICLR2023] decomposes single scene using an input-based gating network and separate sub-models to enhance **expressivity**.



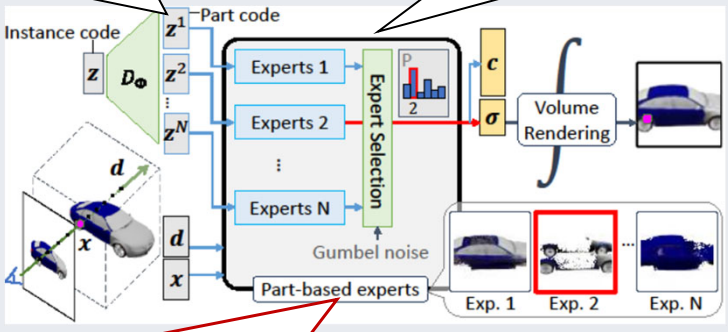
⇒ Combine two methods may achieve our goal.

3. Proposed Method

- **Gumbel-NeRF (GN):** conditional NeRF with *gate-free, density-based* selection of *part-aware* sub-models.

Expert-Specific Codes: Encode properties of a part that ties with particular expert.

Part-Based Experts: Learns to decompose scene into semantic parts (e.g., wheels).



Gate-free Density-Based Selection:

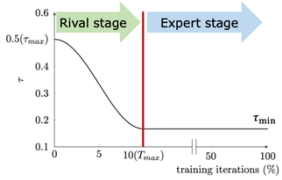
Avoids gating bottleneck by exploiting more info from experts.

$$\sigma = \mathbb{1}_{\max}\left(\frac{\sigma^i}{\tau} + g_i\right), \quad \text{where } \tau\text{: Temperature},\ g\text{: Gumbel noise}$$

Rival-to-Expert Training:

By proper noising with scheduled τ, Experts compete until one stands out and gradually becomes an expert.

$$\min_{\Theta,\Phi,\boldsymbol{z}} \sum |I - \hat{I}|^2$$



One-shot test-time optimization:

Optimizes $\boldsymbol{z}$ to adapt to novel cars with as few as one input image.

- **Baseline:** "Coded Switch-NeRF", a naïve model that combines CodeNeRF and Switch-NeRF.

4. Evaluation

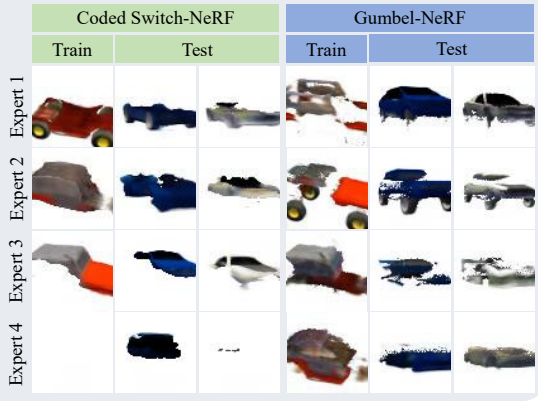
Exp-1: Synthesized BEV Image Quality on SRN Benchmark

- Methods: CodeNeRF (CN), Coded Switch-NeRF (CSN), Gumbel-NeRF (GN).
- Results: Gumbel-NeRF outperforms on several image quality metrics.

Example Input	Method	PSNR	SSIM	LPIPS			CN	CSN	GN	Ground Truth
				VGG	Alex	Squeeze				
	CN	**12.84**	0.475							
	CSN	12.19	0.546	0.374	0.192	0.210				
	GN	12.61	**0.576**	**0.346**	**0.177**	**0.193**				
	CN	14.85	0.562							
	CSN	14.16	0.625	0.317	0.159	0.171				
	GN	**15.20**	**0.659**	**0.286**	**0.141**	**0.154**				
	CN	18.22	0.665							
	CSN	17.23	0.717	0.256	0.120	0.129				
	GN	**18.66**	**0.749**	**0.230**	**0.105**	**0.116**				

Exp-2: Decomposition

- Both are consistent across objects.
- Gumbel-NeRF utilize experts more evenly.



This work is supported by DENSO IT LAB Recognition and Learning Algorithm Collaborative Research Chair (Tokyo Tech.).