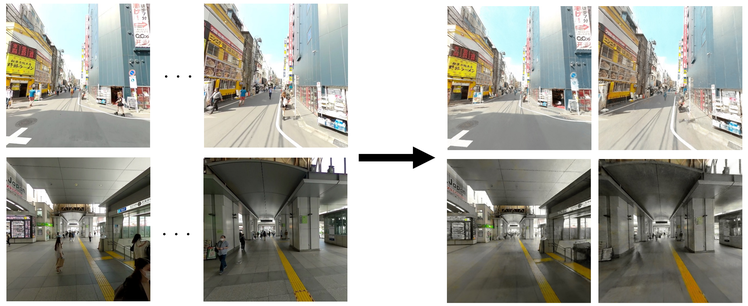
Entity-NeRF: Detecting and Removing Moving Entities in Urban Scenes

Takashi Otonari, Satoshi Ikehata, Kiyoharu Aizawa
[1]The University of Tokyo, [2]National Institute of Informatics (NII), [3]Tokyo Institute of Technology

CVPR SEATTLE, WA JUNE 17-21, 2024

Introduction

Build a static NeRF removing moving objects for urban scenes



Images in urban scenes including moving objects

Static NeRF removing moving objects

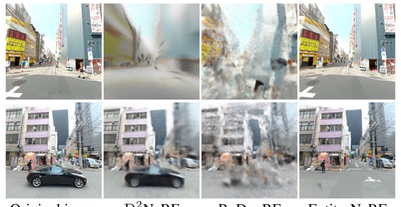
Challenges in Urban Scenes

In urban scenes, A multitude of moving objects of various categories and scales coexist.

Explicit modeling of scene dynamics

- Flow (NSFF, ⋯)
- Deformation (HyperNeRF, ⋯)
- 3D bounding box (NSG, ⋯)
- Self-supervised (D^{2}NeRF, ⋯)

→ fails to train scene dynamics



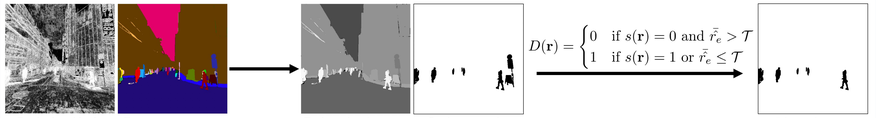
Original image D^{2}NeRF RoDynRF Entity-NeRF

Remove scene dynamic

→ excludes static backgrounds and inconsistently removes moving objects for moving objects

RGB Residuals Ground truth RobustNeRF weights

Method: Entity-NeRF



Ranked residuals Entity segmentation Entity-wise Average of Residual Ranks (EARR) $\bar{r}_e$ Stationary entity classification $s(\mathbf{r})$

$$D(\mathbf{r}) = \begin{cases} 0 & \text{if } s(\mathbf{r}) = 0 \text{ and } \bar{r}_e > \mathcal{T} \\ 1 & \text{if } s(\mathbf{r}) = 1 \text{ or } \bar{r}_e \leq \mathcal{T} \end{cases}$$

Distractiveness $D(\mathbf{r})$

Entity-wise Average of Residual Ranks (EARR)

(1) Ranking of residuals

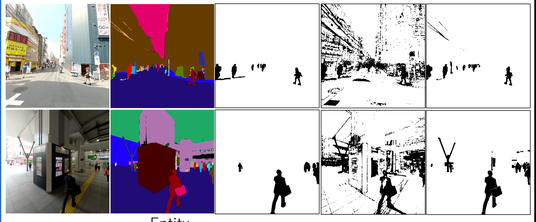
Residuals Ranked residuals

(2) Averaging of residuals per entity

Entity segmentation

(3) Thresholding

Loss weight



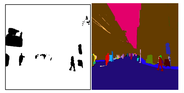
Input Entity segmentation Ground truth RobustNeRF EARR

Stationary Entity Classification

Thing: objects with a well-defined shape, e.g. car, person

Stuff: amorphous background regions, e.g. road, sky

(1) Pixel-wise thing/stuff segmentation



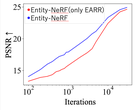
Pixel-wise thing/stuff segmentation Entity segmentation

(2) Entity-wise thing/stuff segmentation

Entity-wise thing/stuff segmentation

(3) Thing → EARR-based loss weight calculation

Stuff → Set the weight to 1



Stationary entity classification
- boosts training efficiency
- enhances final PSNR.

MovieMap Dataset

We created a new dataset to evaluate whether only the background was reconstructed in an urban scenes.

1. Mask annotation of moving objects
2. NeRF training exluding moving objects
3. Rendering background-only images
4. Use rendered background-only images as ground-truth images

Images Masks

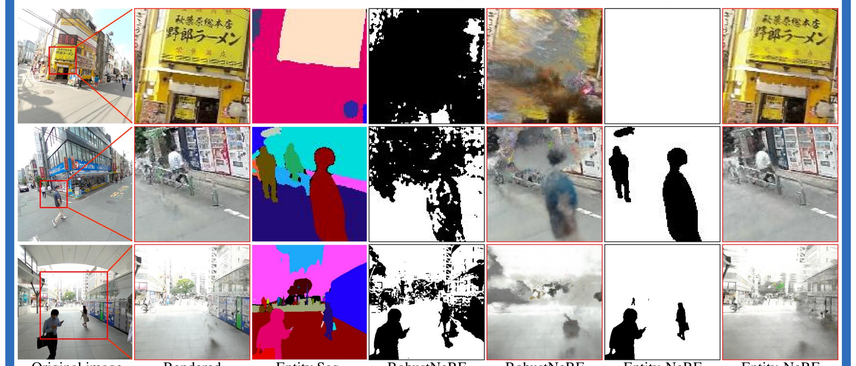
Static NeRF Training & Rendering

Background-only images

Experiments

Qualitative Evaluation on MovieMap Dataset

Entity-NeRF outperforms RobustNeRF in removing moving objects and reconstructing static backgrounds.



Original image Rendered ground-truth Entity Seg. RobustNeRF loss weights RobustNeRF Entity-NeRF loss weights Entity-NeRF

Quantitative Evaluation on MovieMap Dataset

Entity-NeRF reached a background PSNR close to MSE and surpassed existing methods in foreground one.

Model	Loss	foreground PSNR↑	background PSNR↑	PSNR↑	SSIM↑	LPIPS↓
Nerfacto	Mean-squared error (MSE)	12.10	**25.07**	**24.96**	**0.87**	**0.10**
	RobustNeRF	17.63	21.74	23.19	0.84	0.12
	Entity-NeRF (Ours)	**19.82**	24.00	24.93	0.85	0.12
Mip-NeRF 360	Mean-squared error (MSE)	11.40	**27.36**	24.22	**0.88**	**0.13**
	RobustNeRF	20.15	22.52	22.87	0.83	0.18
	Entity-NeRF (Ours)	**20.74**	25.50	**25.23**	0.84	0.15

IoU: annotated mask vs. weights

Entity-NeRF can give closer loss weights to the annotated mask.

	IoU $D(\mathbf{r}) = 1 \uparrow$ (static backgrounds)	IoU $D(\mathbf{r}) = 0 \uparrow$ (moving objects)
RobustNeRF	0.84	0.14
Entity-NeRF	**0.98**	**0.59**

Sensitivity to Hyperparameters