PhD Open Days

NeRF View Synthesis: Subjective Quality Assessment and Objective Metrics Evaluation

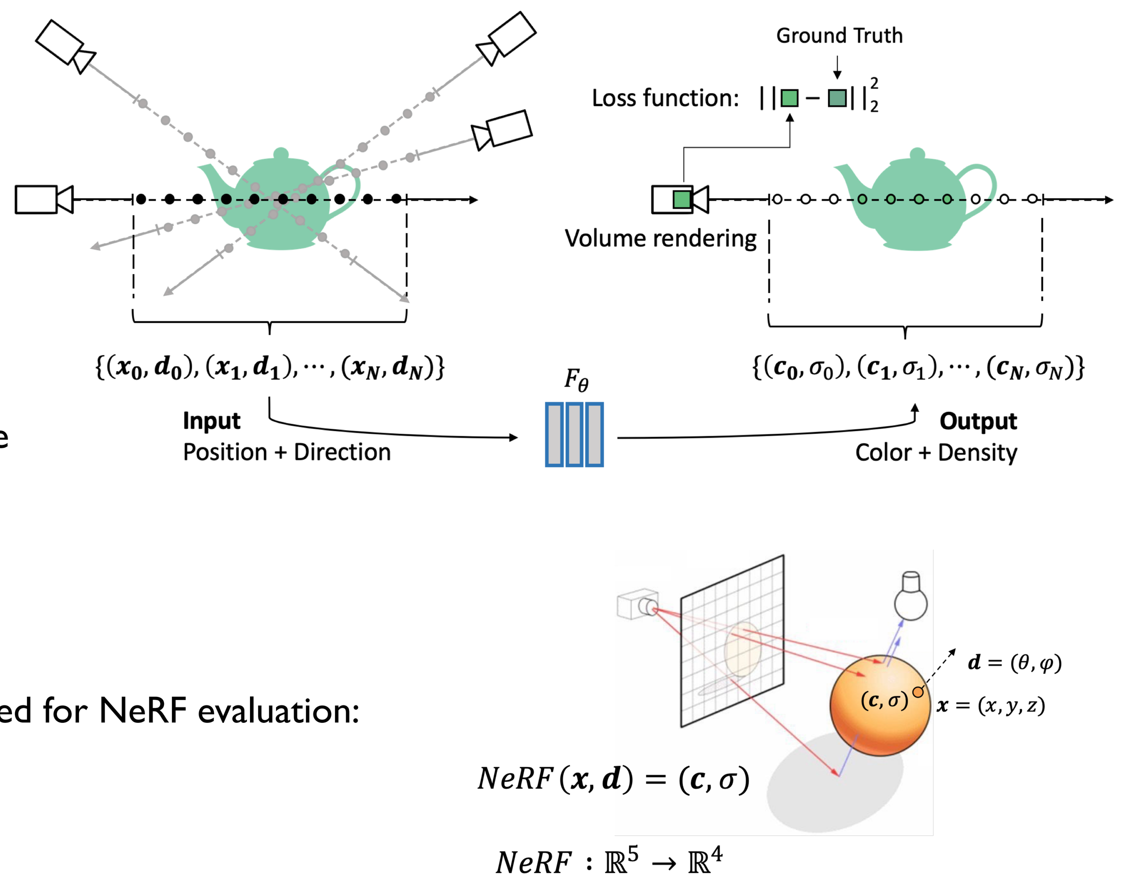
Pedro Martin

1. Introduction

- Neural Radiance Fields (NeRF) methods are a powerful technique for synthesizing novel views of a visual scene from a set of input views.
- NeRF represent 3D scenes as a radiance field typically modelled by a Multi-Layer Perceptron (MLP).
- Neural-network-free methods have also been proposed seeking the rendering time reduction.

- The NeRF methods inputs are:
  - 3D spatial location
  - 2D viewing direction
- The NeRF methods outputs are:
  - 3D RGB color
  - Opacity level
- NeRF view synthesis may generate artifacts:
  - Floaters
  - Flawed geometry
  - Flickering object edges

Objective quality metrics typically used for NeRF evaluation:
- PSNR
- SSIM
- LPIPS



2. Objectives and Contributions

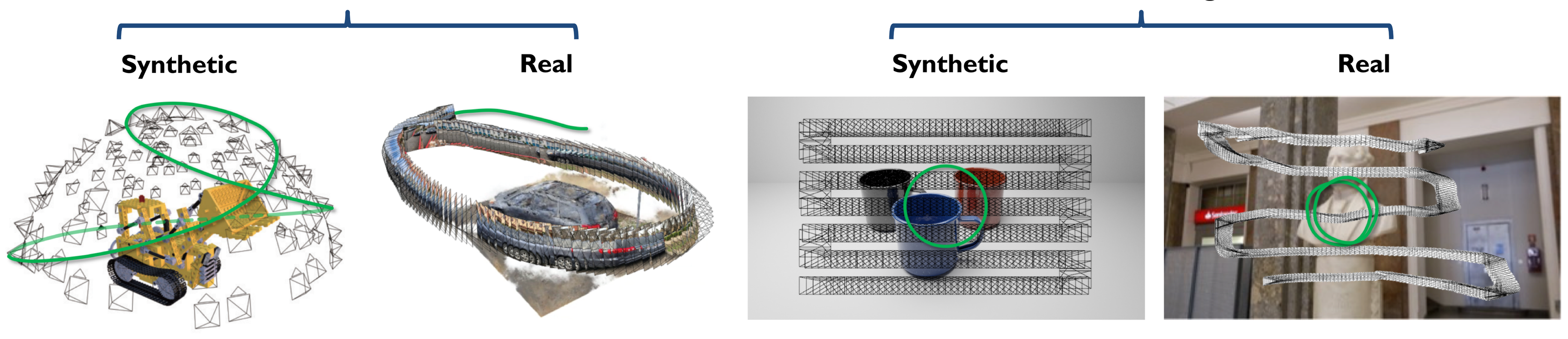
Objectives

- To study, in a subjective way, the impact of the different artifacts produced by several NeRF methods for different classes of visual scenes.
- To evaluate the performance of state-of-the-art image and video quality assessment (IQA and VQA) metrics considering the results obtained on the subjective assessment study.

Contributions

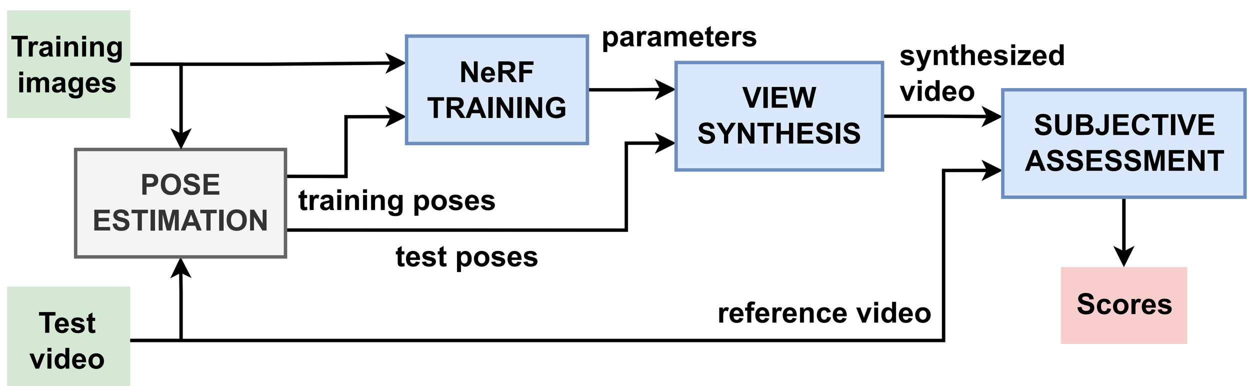
Dataset webpage

- The creation of a new dataset of FF synthetic and real visual scenes with the respective camera poses.
- The subjective evaluation of the impact on perceived quality of NeRF view synthesis.
- The evaluation of objective quality assessment metrics developed for 2D images and video, using several scene classes:



3. NeRF Creation Framework

- The NeRF creation framework is composed by three main steps:
  - NeRF training
  - View synthesis
  - Subjective assessment



- NeRF methods were selected according to:
  - Synthesis quality performance
  - Training and synthesis speed
  - Suitability to the considered scene classes

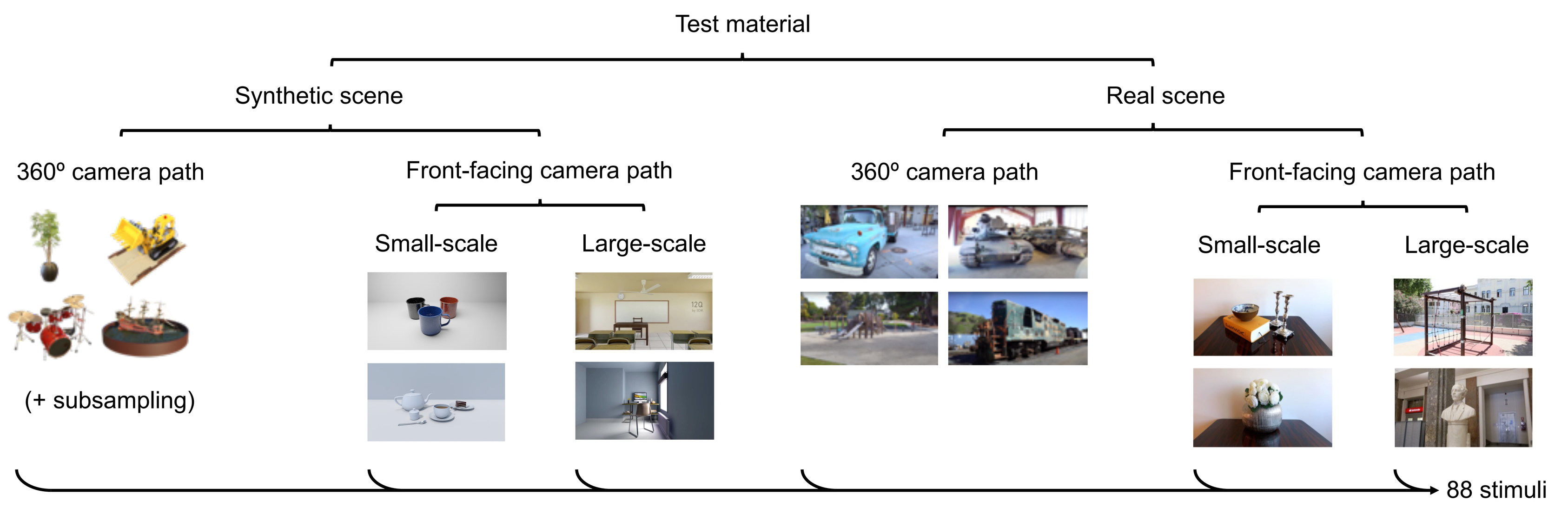
NeRF Method	Approach	Scene Application
DVGO	Grid-based	All
Instant-NGP	Grid-based	Synthetic scenes
Mip-NeRF 360	MLP-based	Unbounded scenes
Nerfacto	MLP-based	Unbounded scenes
NeRF++	MLP-based	Unbounded scenes
Plenoxels	Grid-based	360° synthetic scenes
TensoRF	Grid-based	Synthetic scenes

- Due to the high training and synthesis times of MLPs a branch of MLP-free methods has emerged based on voxel grids.

4. Experimental Setup

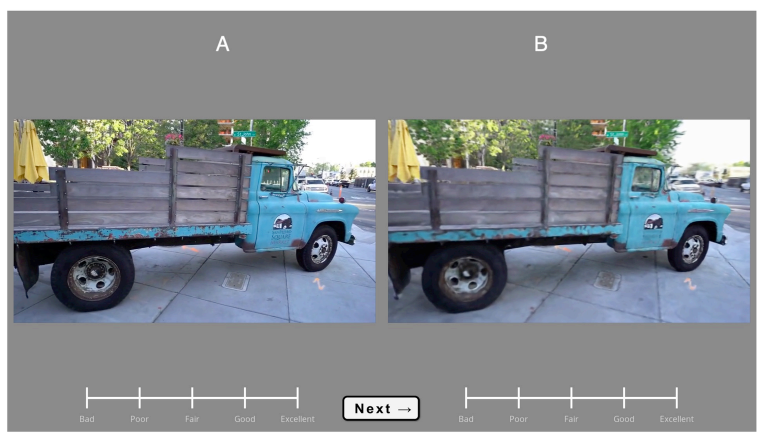
Test Material

- For 360° scenes:
  - **360° synthetic scenes:** Realistic Synthetic 360° (4 scenes, 100 training images, 800×800 pixels)
  - **360° real scenes:** Tanks and Temples (4 scenes, 226-277 training images, 1077×546, 1008×548, 982×546, and 980×546 pixels, respectively)
- For FF scenes:
  - **FF synthetic scenes:** IST/IT dataset (4 scenes, 228-377 training images, 960×540 pixels)
  - **FF real scenes:** IST/IT dataset (4 scenes, 300 training images, 960×540 pixels)



Subjective Assessment Methodology

- **DSCQS** subjective assessment method:
  - Score: 0 - 100 (with 5 quality labels: bad, poor, fair, good, excellent)
  - Synthesized and reference videos
  - Side-by-side with random order
- **Total number of stimuli:** 88 pairs
- **Test session duration:** around 30 minutes
- **Monitor:** computer display with 1920×1080 pixels
- **Total number of participants:** 22 non-expert
- **Final scores** were converted to **DMOS** values



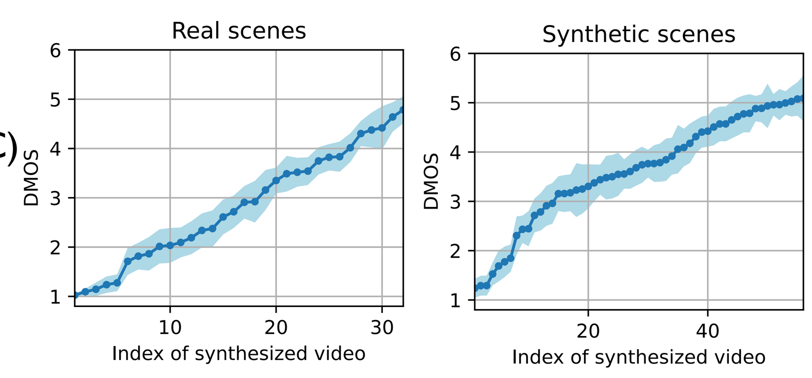
$$DMOS = MOS_{syn} - MOS_{ref} + 5$$

Objective Metrics Performance Assessment

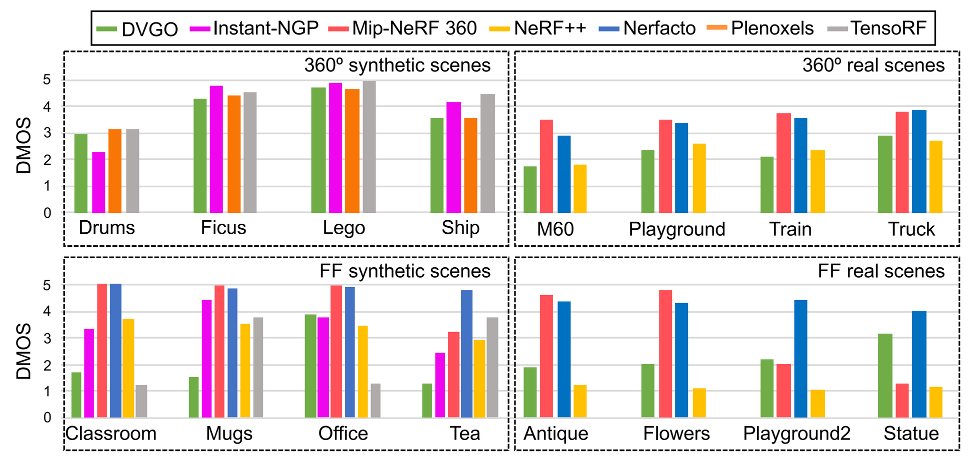
- Objective quality metrics selection was based on several criteria:
  - **NVS literature:** PSNR-Y, PSNR-YUV, SSIM, MS-SSIM, LPIPS, and FovVideoVDP
  - **Best performances:** PSNR-HVS, IW-SSIM, VIF, VIFp, FSIM, VSI, MAD, GMSD, and NLPD
  - **Learning-based metrics:** LPIPS, ST-LPIPS, DISTS, and VMAF.

5. Experimental Results and Conclusions

- **Metrics' performance evaluation:**
  - Pearson Linear Correlation Coefficient (PLCC)
  - Spearman Rank Order Correlation Coefficient (SROCC)
- **Disparity compensation:**
  - Block-matching algorithm
  - Macroblock size of 32×32 pixels
  - Search window of 12 pixels



- The coverage of the total span of qualities is verified for both real and synthetic scenes.



NVS Method	Real		Synthetic	
	FF	360°	FF	360°
DVGO	2.31	2.28	2.11	3.90
Instant-NGP	N/A	N/A	3.50	4.04
Mip-NeRF 360	3.18	**3.64**	4.54	N/A
NeRF++	1.13	2.37	3.40	N/A
Nerfacto	**4.28**	3.41	**4.92**	N/A
Plenoxels	N/A	N/A	N/A	3.94
TensoRF	N/A	N/A	2.52	**4.31**
Average	2.72	2.92	3.50	4.05

- 360° synthetic scenes are the only class of scenes with an average DMOS rating greater than 4.
- 360° real scenes have, on average, higher DMOS values than FF real scenes.

PLCC

Metric	Synthetic			Real			All	Real*			All*
	360	FF	All	360	FF	All		360*	FF*	All*	
MSE-RGB	0.94	0.89	0.87	0.59	0.78	0.57	0.65	0.70	0.66	0.59	0.73
PSNR-Y	0.94	0.84	0.86	0.57	0.66	0.51	0.67	0.65	0.70	0.56	0.70
PSNR-YUV	0.94	0.85	0.86	0.63	0.63	0.49	0.64	0.76	0.64	0.52	0.67
PSNR-HVS	0.93	0.85	0.86	0.58	0.69	0.53	0.69	0.60	0.81	0.64	0.73
SSIM	0.90	0.80	0.69	0.70	0.70	0.47	0.60	0.82	0.72	0.57	0.62
MS-SSIM	0.95	0.88	0.77	0.71	0.72	0.53	0.65	0.78	0.80	0.63	0.69
IW-SSIM	0.93	0.83	0.80	0.66	0.67	0.53	0.68	0.82	0.87	0.77	0.78
VIF	0.91	0.74	0.77	0.67	0.61	0.51	0.66	0.77	0.80	0.73	0.75
VIFp	0.90	0.75	0.75	0.77	0.63	0.51	0.63	0.86	0.78	0.70	0.71
FSIM	0.93	0.87	0.73	0.69	0.72	0.48	0.66	0.83	0.81	0.65	0.63
VSI	0.93	0.77	0.78	0.65	0.75	0.60	0.68	0.77	0.83	0.78	0.77
MAD	0.87	0.63	0.71	0.76	0.67	0.56	0.67	0.68	0.78	0.67	0.71
LPIPS	0.87	0.90	0.55	0.75	0.75	0.67	0.60	0.90	0.87	0.76	0.64
ST-LPIPS	0.92	0.92	0.77	0.88	0.82	0.68	0.66	0.87	0.90	0.71	0.68
DISTS	0.94	0.83	0.81	0.61	0.71	0.67	0.76	0.70	0.78	0.72	0.76
GMSD	0.95	0.81	0.82	0.74	0.74	0.52	0.67	0.84	0.81	0.68	0.72
NLPD	0.95	0.87	0.86	0.65	0.73	0.52	0.69	0.70	0.79	0.61	0.72
VMAF	0.87	0.76	0.79	0.81	0.64	0.61	0.64	0.78	0.92	0.80	0.77
FovVideoVDP	0.96	0.90	0.80	0.49	0.76	0.65	0.70	0.41	0.83	0.65	0.73

SROCC

Metric	Synthetic			Real			All	Real*			All*
	360	FF	All	360	FF	All		360*	FF*	All*	
MSE-RGB	0.95	0.72	0.82	0.57	0.64	0.51	0.69	0.49	0.60	0.54	0.71
PSNR-Y	0.95	0.77	0.82	0.57	0.64	0.50	0.68	0.50	0.64	0.48	0.70
PSNR-YUV	0.95	0.74	0.83	0.59	0.61	0.48	0.66	0.54	0.55	0.46	0.68
PSNR-HVS	0.95	0.77	0.81	0.57	0.66	0.48	0.69	0.51	0.70	0.50	0.73
SSIM	0.87	0.67	0.65	0.64	0.52	0.31	0.59	0.79	0.64	0.49	0.62
MS-SSIM	0.96	0.68	0.74	0.74	0.66	0.44	0.67	0.74	0.70	0.51	0.70
IW-SSIM	0.93	0.68	0.74	0.68	0.67	0.48	0.68	0.81	0.87	0.73	0.75
VIF	0.91	0.64	0.72	0.67	0.61	0.43	0.65	0.77	0.78	0.65	0.71
VIFp	0.92	0.67	0.74	0.81	0.63	0.45	0.65	0.89	0.77	0.68	0.70
FSIM	0.92	0.67	0.62	0.77	0.58	0.39	0.64	0.81	0.80	0.63	0.61
VSI	0.94	0.58	0.71	0.69	0.62	0.56	0.67	0.71	0.81	0.77	0.73
MAD	0.83	0.58	0.69	0.71	0.64	0.47	0.67	0.54	0.77	0.62	0.70
LPIPS	0.74	0.86	0.39	0.75	0.78	0.59	0.55	0.89	0.87	0.72	0.57
ST-LPIPS	0.90	0.87	0.77	0.86	0.80	0.61	0.70	0.87	0.85	0.65	0.71
DISTS	0.94	0.69	0.77	0.64	0.70	0.64	0.72	0.70	0.81	0.70	0.72
GMSD	0.95	0.67	0.75	0.80	0.54	0.41	0.67	0.77	0.74	0.61	0.72
NLPD	0.95	0.70	0.81	0.61	0.66	0.43	0.68	0.71	0.70	0.53	0.71
VMAF	0.89	0.57	0.74	0.75	0.76	0.50	0.67	0.61	0.88	0.70	0.73
FovVideoVDP	0.96	0.70	0.75	0.45	0.71	0.44	0.68	0.39	0.81	0.51	0.71

Max
Min

- Learning-based and VQA metrics stand out.
- Metrics based on pixel-wise differences are the most effective for synthetic scenes.
- Metrics' performance benefits greatly from disparity compensation

Maria Paula Queluz, João Ascenso, António Rodrigues

TÉCNICO LISBOA
instituto de telecomunicações
PDEEC

phdopendays.tecnico.ulisboa.pt

This work was supported by the Fundação para a Ciencia e a Tecnologia (FCT) through national funds, and when applicable, co-funded by EU funds under the projects UIDB/50008/2020 and PTDC/EEI-COM/7775/2020.