PINs: Progressive Implicit Networks for Multi-Scale Neural Representations

Zoe Landgraf [*+] Alexander Sorkine−Hornung[+], Ricardo Silveira Cabral [+]

Imperial College London* , Meta [+]

ICML 2022, Spotlight presentation

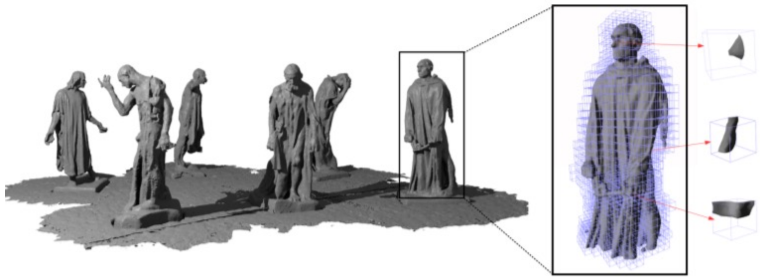
Introduction

What are Neural Implicit Representations

x, y, z → Positional encoding → MLP → Scene element (pixel, voxel)

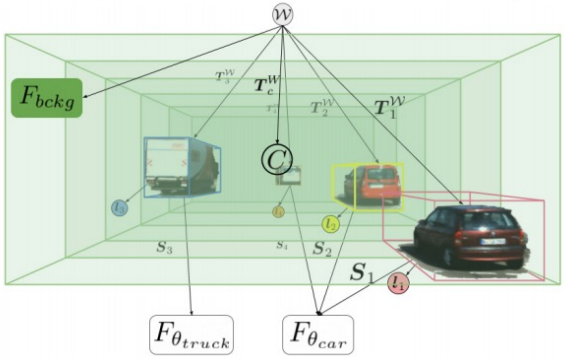
Introduction

Popular as effective scene encoders



Deep Local Shapes[1]

NeRF[2]

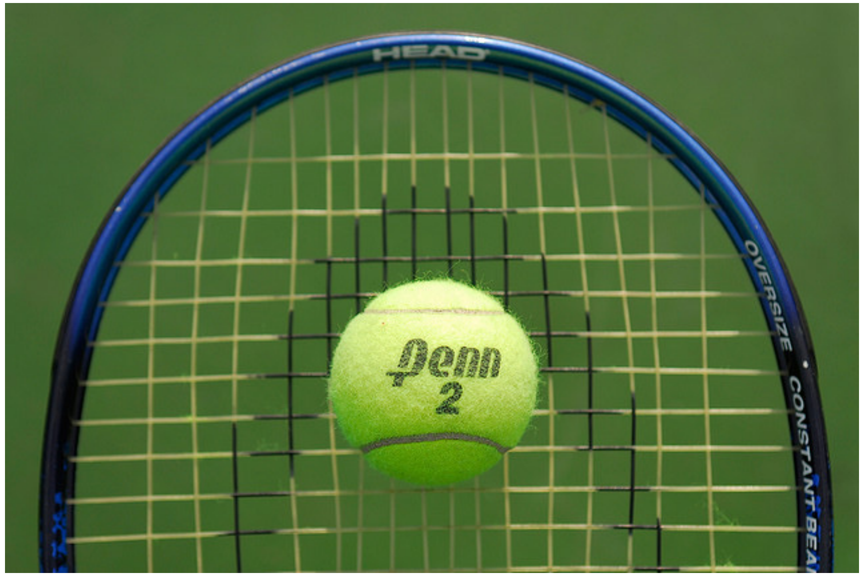


Neural Scene Graphs[3]

1) Deep Local Shapes: Learning Local SDF Priors for Detailed 3D Reconstruction, R. Chabra et. al., *ECCV 2020*
2) NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, B. Mildenhall, ECCV 2020
3) Neural Scene Graphs for Dynamic Scenes, J. Ost et. al. CVPR 2021

Problem statement

Can't reconstruct scenes with a wide frequency spectrum well:



Original

Reconstruction
(Implementation of Fourier Feature Networks[1])

1) Fourier features let networks learn high frequency functions in low dimensional domains, M. Tancik et. al., *NeurIPS 2020*

Problem statement

Noise is introduced into smooth regions

Original

Reconstruction
(Implementation of Fourier Feature Networks[1])

Why? High frequency components in the positional encoding

1) Fourier features let networks learn high frequency functions in low dimensional domains, M. Tancik et. al., *NeurIPS 2020*

Problem statement

Lower frequencies in positional encoding yields blurry reconstructions

Original

Reconstruction
(Implementation of Fourier Feature Networks[1])

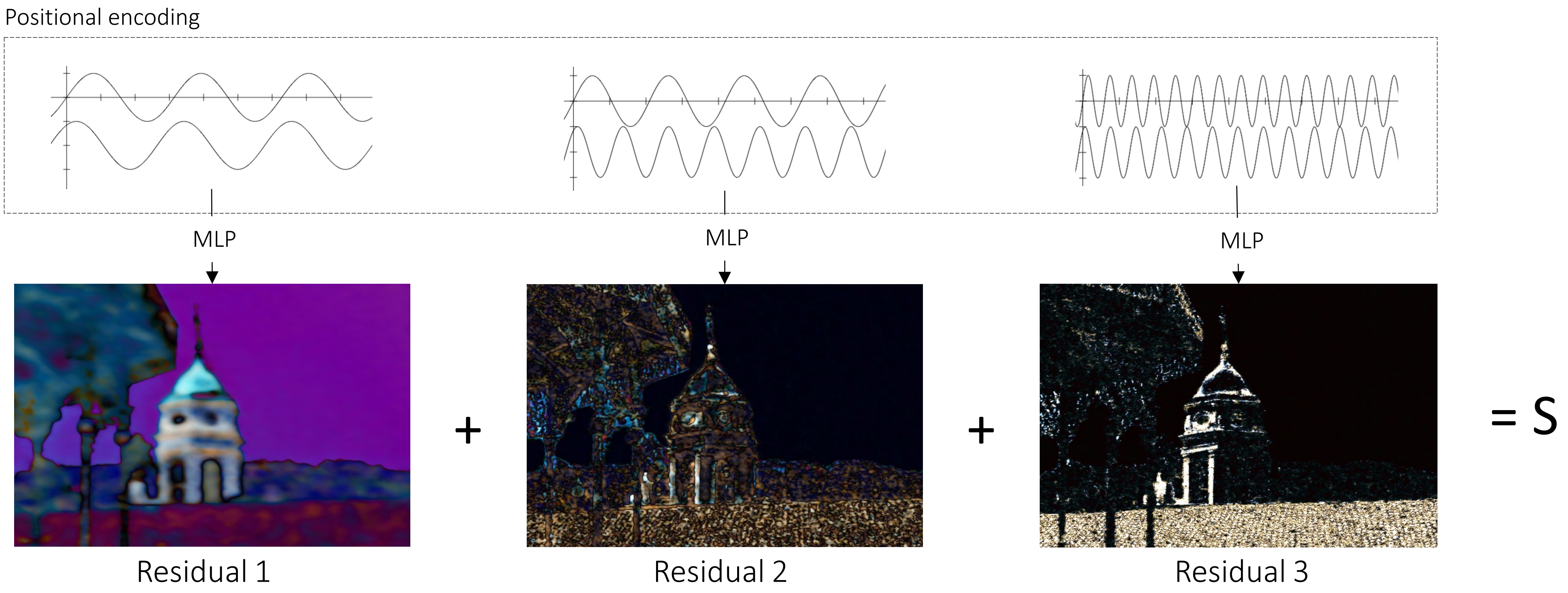
1) Fourier features let networks learn high frequency functions in low dimensional domains, M. Tancik et. al., *NeurIPS 2020*

Problem statement

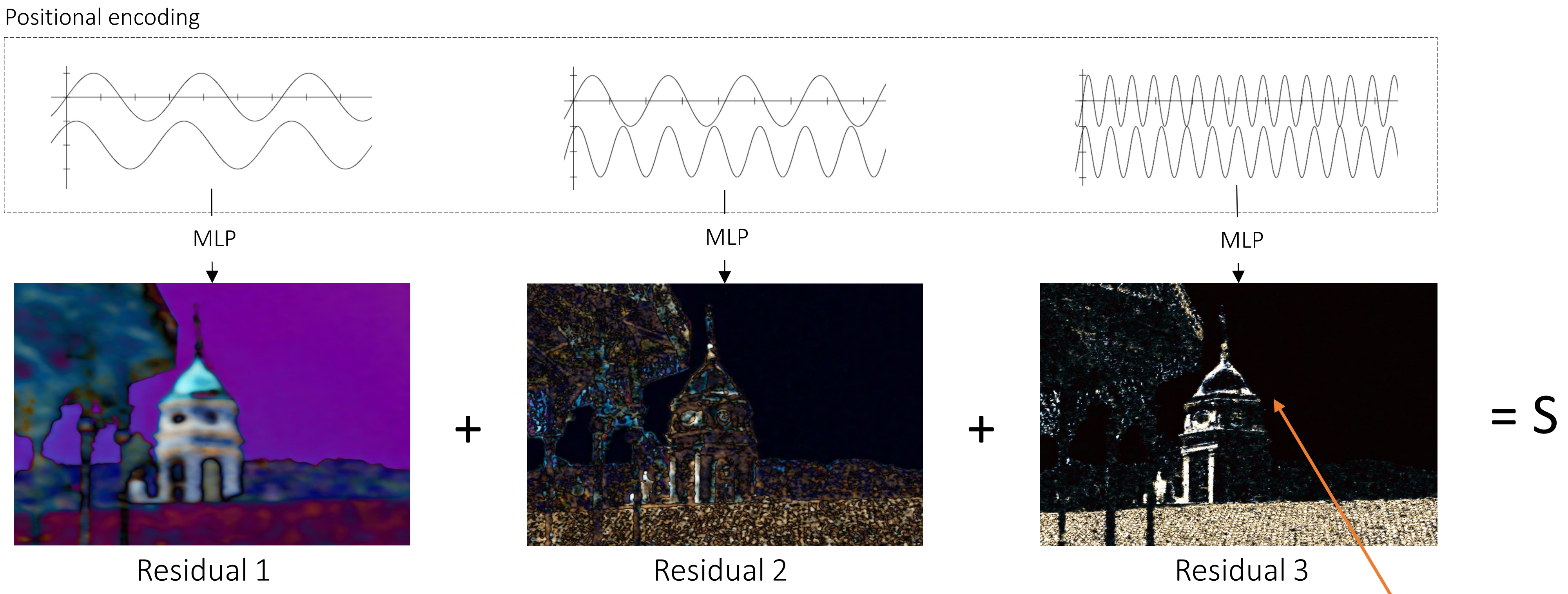
How do we remove noise in smooth regions while keeping the high frequency detail of the reconstruction?

1) Fourier features let networks learn high frequency functions in low dimensional domains, M. Tancik et. al., *NeurIPS 2020*

PINs

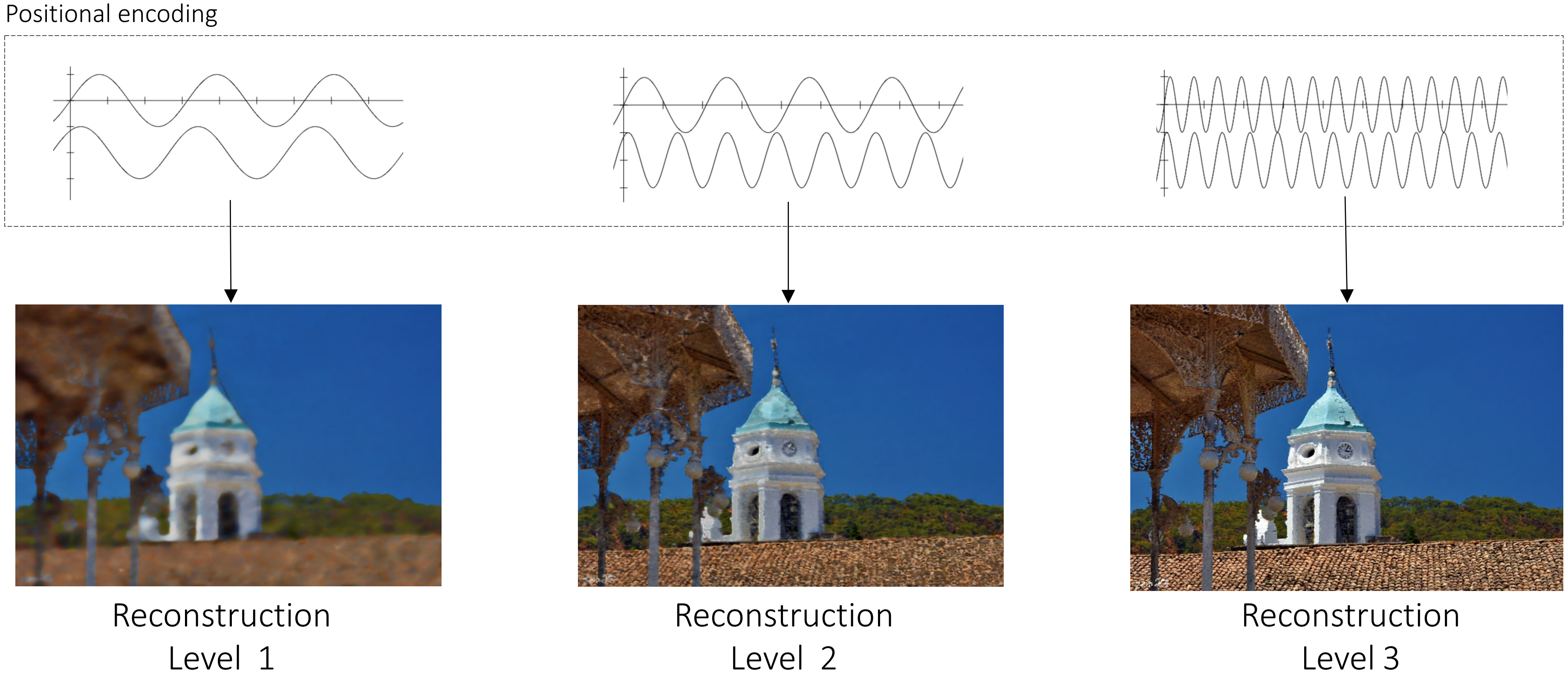

Positional encoding
MLP
MLP
MLP
Residual 1
+
Residual 2
+
Residual 3
= S

PINs



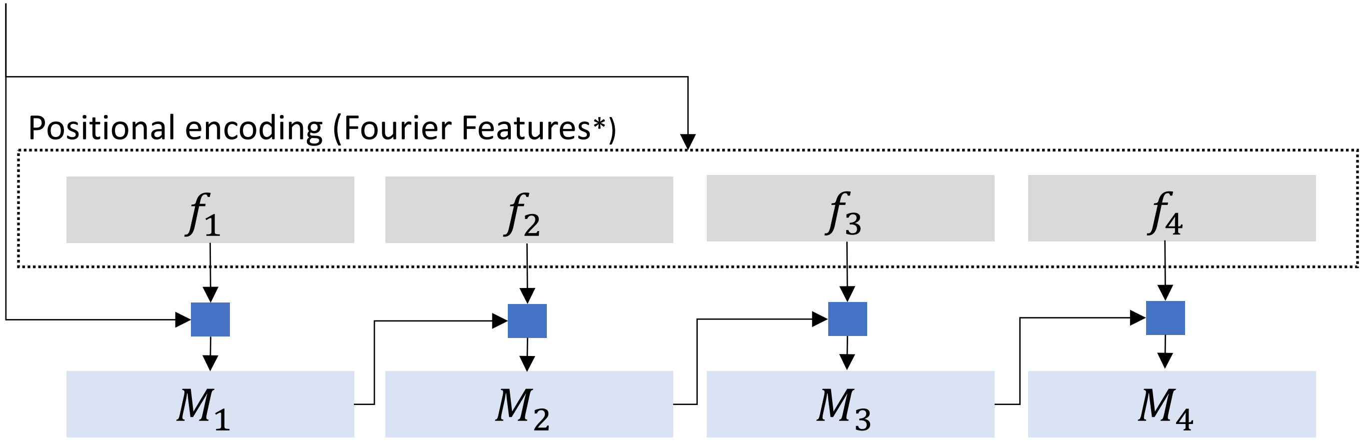
Later layers learn high frequency detail locally

PINs

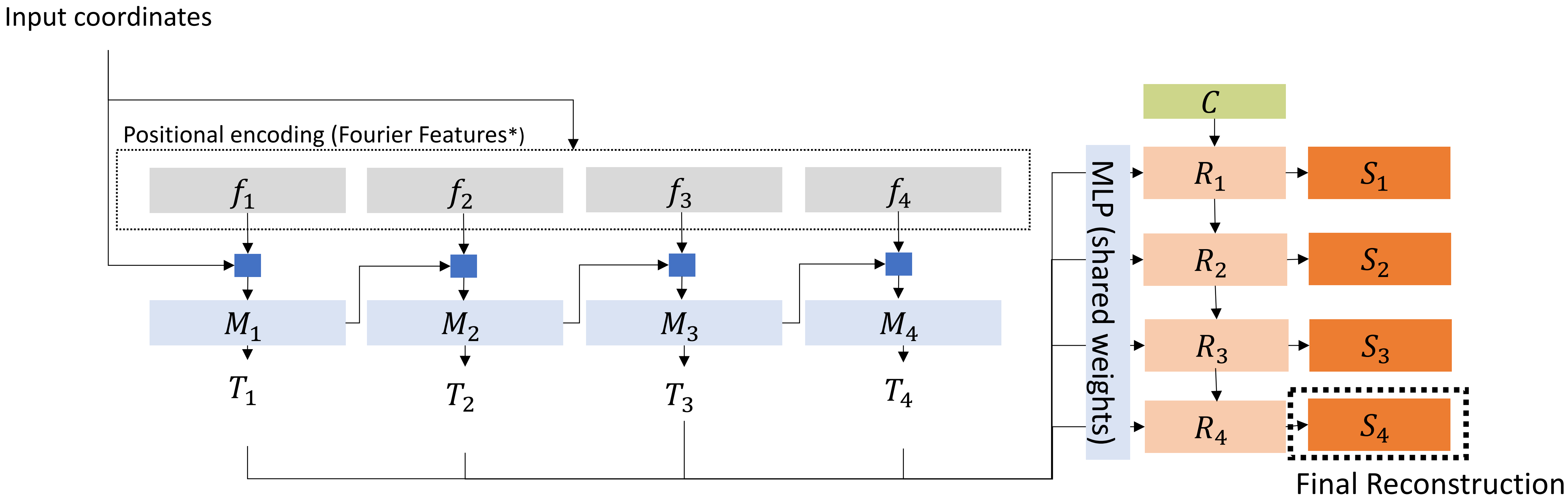

Positional encoding
Reconstruction
Level 1
Reconstruction
Level 2
Reconstruction
Level 3

PINs - architecture

Input coordinates

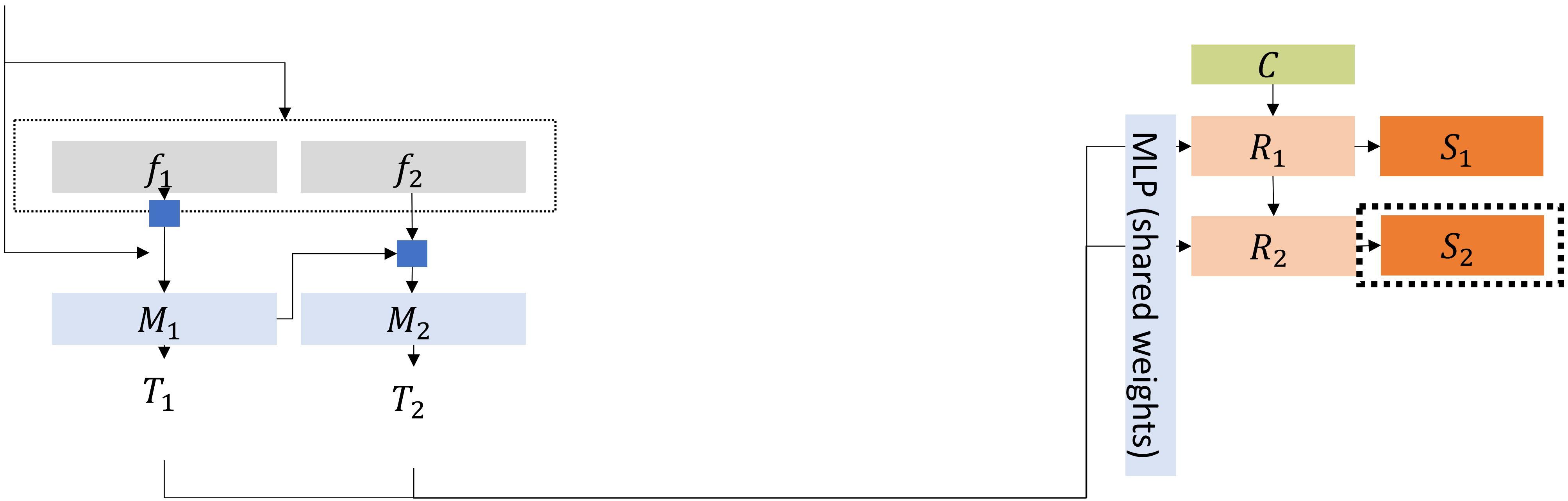


PINs - architecture

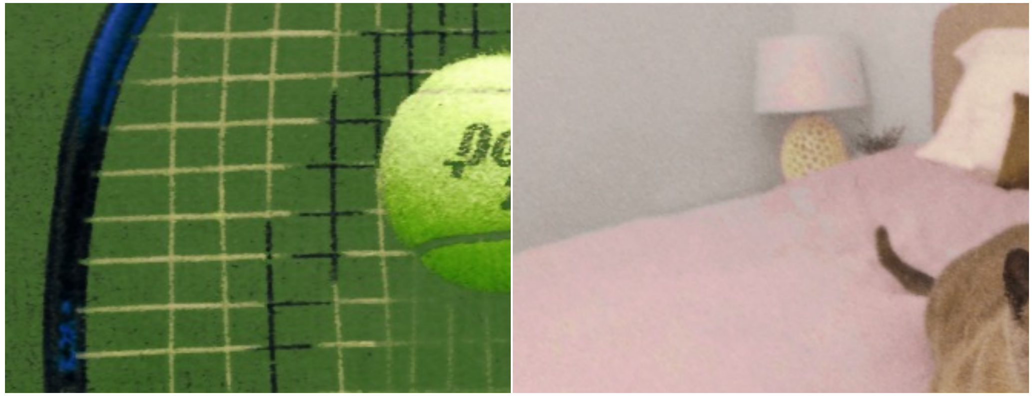


PINs - modularity

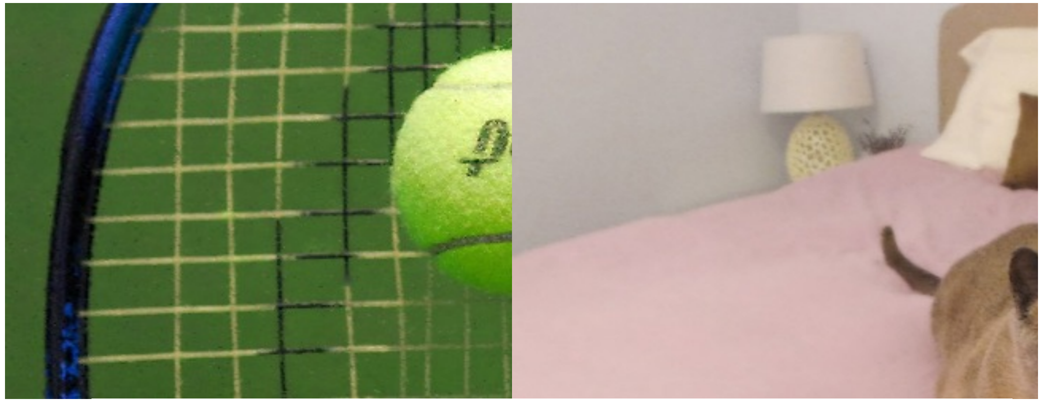
Input coordinates



Results – improved reconstruction quality



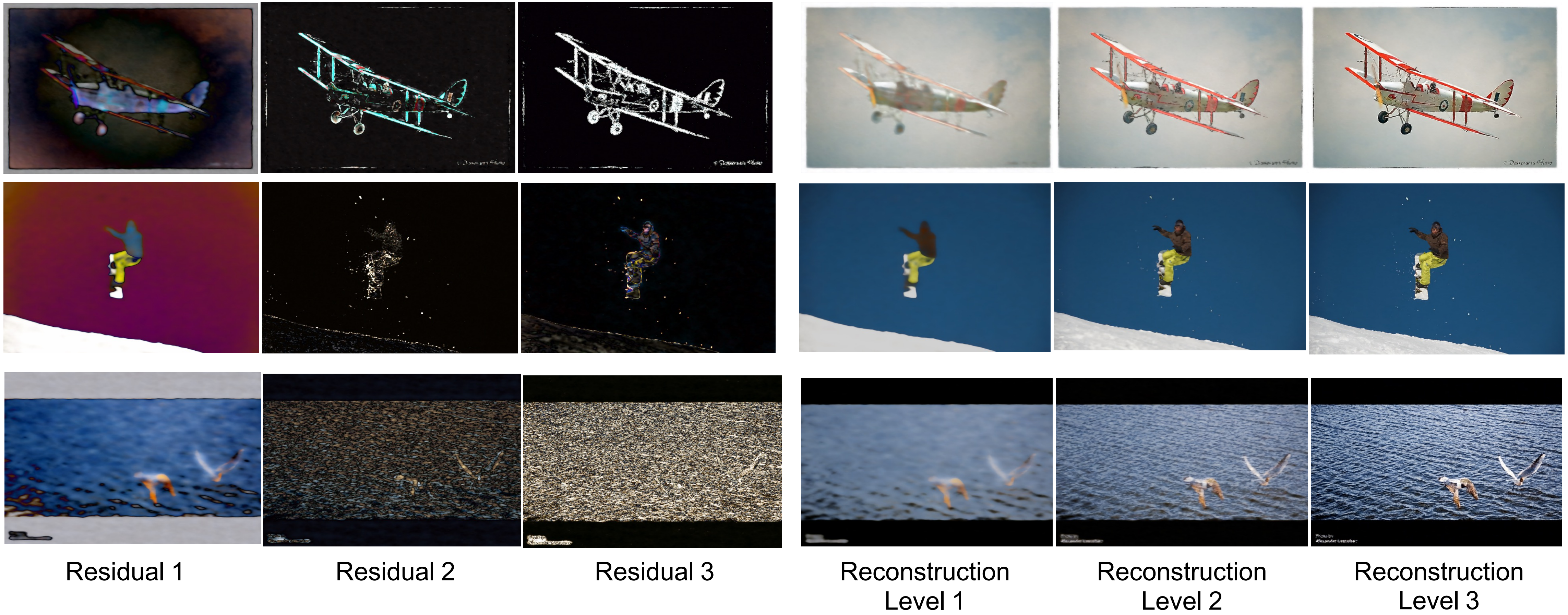
Reconstruction
(Implementation of Fourier Feature Networks[1])



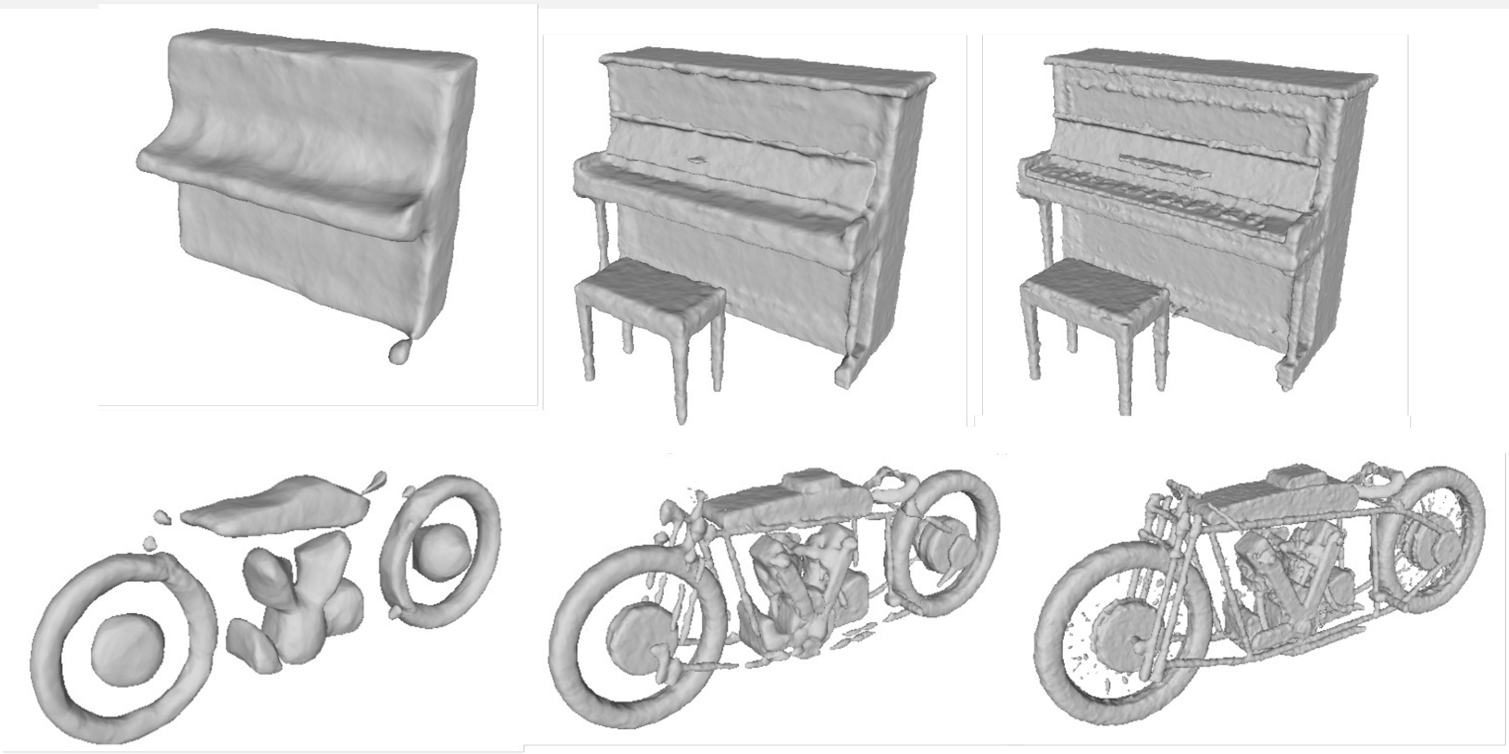
Reconstruction
PINs

1) Fourier features let networks learn high frequency functions in low dimensional domains, M. Tancik et. al., *NeurIPS 2020*

Results – 2D regression examples



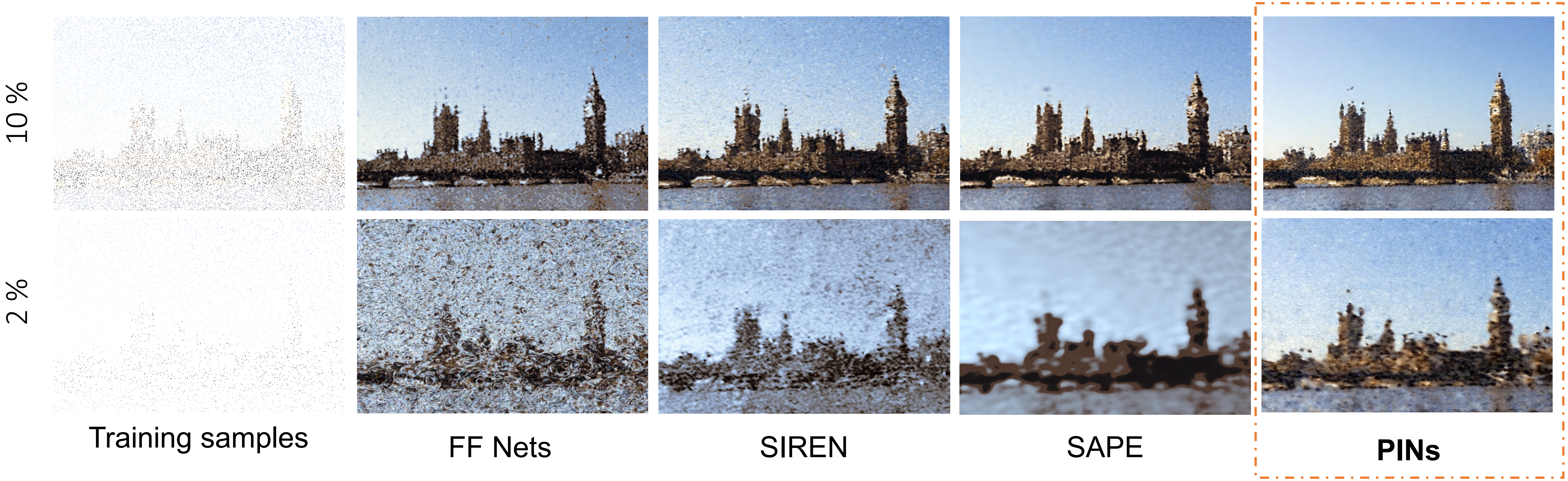
Results – 3D regression examples



Reconstruction Level 1 Reconstruction Level 2 Reconstruction Level 3

Results – robustness


Original



Training samples FF Nets SIREN SAPE **PINs**

* Please see main paper for baseline references

Results

PINs: Progressive Implicit Networks for Multi-Scale Neural Representations

Model	lamp (ChD ↓)	car (ChD ↓)	chair (ChD ↓)	sofa (ChD ↓)	motorbike (ChD ↓)	bed (ChD ↓)	camera (ChD ↓)
FF Nets	2.5±3.4e−3	2.1±4.8e−4	0.92±1.5e−4	1.5±5.6e−4	**1.0±1.5e-5**	2.83±5.2e−3	3.46±1.26e−2
SIREN	25.4±4.2e−3	2.2±5.2e−4	28.4±6.7e−3	1.6±5.5e−4	1.7±6.4e−4	3.22±6.9e−3	2.31±2.68e−3
SAPE	6.7±9.2e−2	2.2±6.6e−4	1.50±1.1e−3	6.6±0.42	2.8±3.9e-4	4.58±2.4e−2	4.38±2.7e−2
Ours	**1.5±1.0e-4**	**2.0±5.1e-4**	**0.87±1.7e-4**	**1.48±5.4e-4**	1.1±6.7e-5	**2.79±4.7e-3**	**2.05±2.7e-3**

Table 2. Evaluation on 3D models from 3D Warehouse (3DW) in terms of the bi-directional Chamfer Distance (mm)

Conclusions

- Novel, multi-scale representation with

    1) Progressive positional encoding
    2) Hierarchical MLP structure

- Improved reconstruction quality and robustness (2D & 3D scenes)

- End-to-end trainable model no explicit per-level supervision

Thank you for your attention

Poster session #2 at 6:30 pm today