FreBIS: Frequency-Based Stratification for Neural Implicit Surface Representations

Naoko Sawada[1, 2], Pedro Miraldo[1], Suhas Lohit[1], Tim K. Marks[1], Moitreya Chatterjee[1]

[1] Mitsubishi Electric Research Laboratories (MERL)

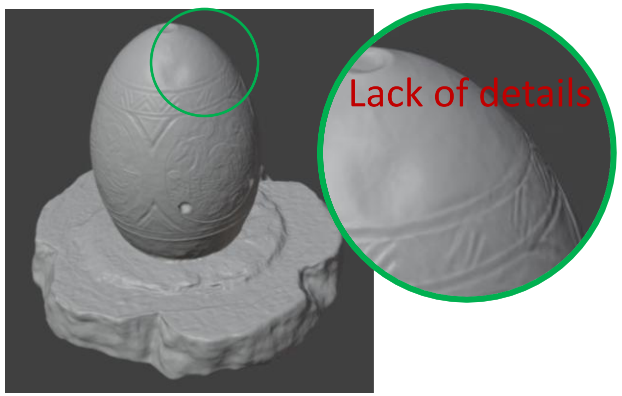
[2] Information Technology R&D Center, Mitsubishi Electric Corporation

Motivation & Problem Statement

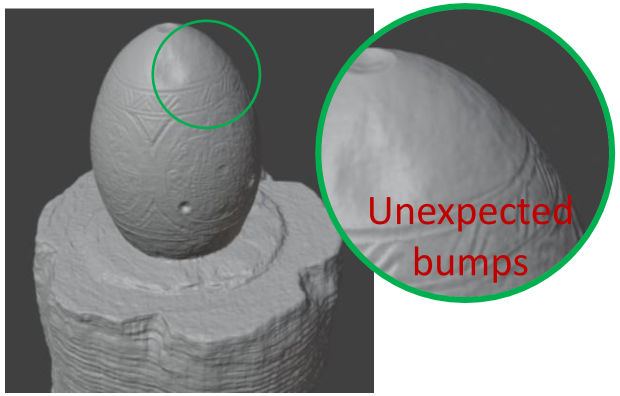
- Neural implicit surface representation enables continuous high-resolution and accurate 3D surface reconstruction.
- Existing methods use a single encoder to capture all surface frequencies.

→ **There is a tradeoff between accurate shape recovery and reconstructing the fine details.**

Reference image



Positional encoding level 6
[VolSDF, 2021]



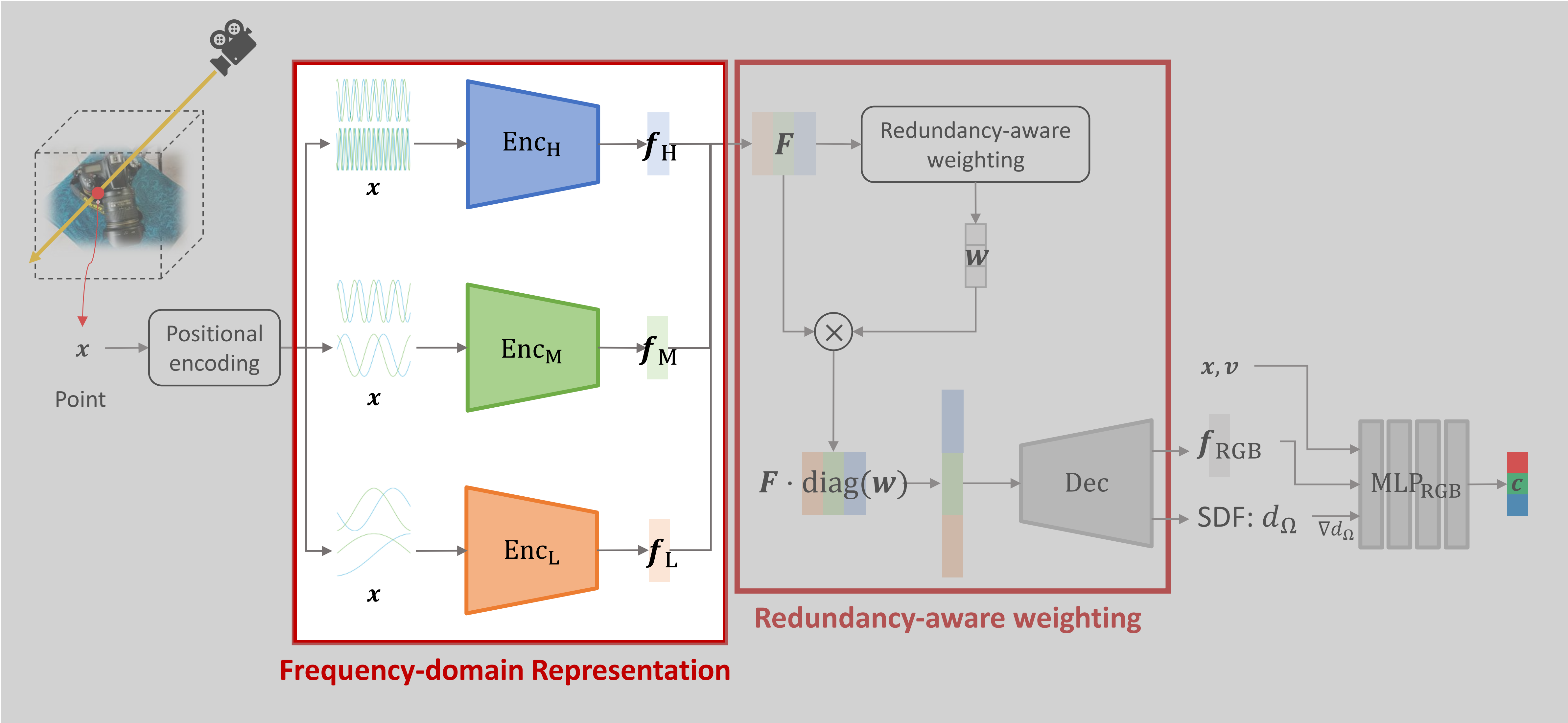
Positional encoding level 9
[VolSDF, 2021]

- **Goal**:
  Recover high-quality surfaces of a 3D scene that contains a wide variety of frequency levels.

Yariv et al., 2021, “Volume rendering of neural implicit surfaces,” in *Proceedings of the 35th International Conference on Neural Information Processing Systems*

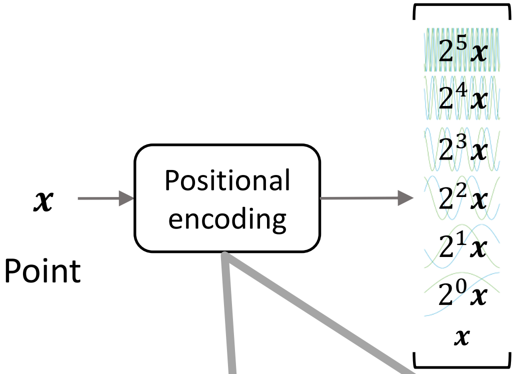
Method: FreBIS

The FreBIS framework consists of two modules:



Frequency-Domain Representation

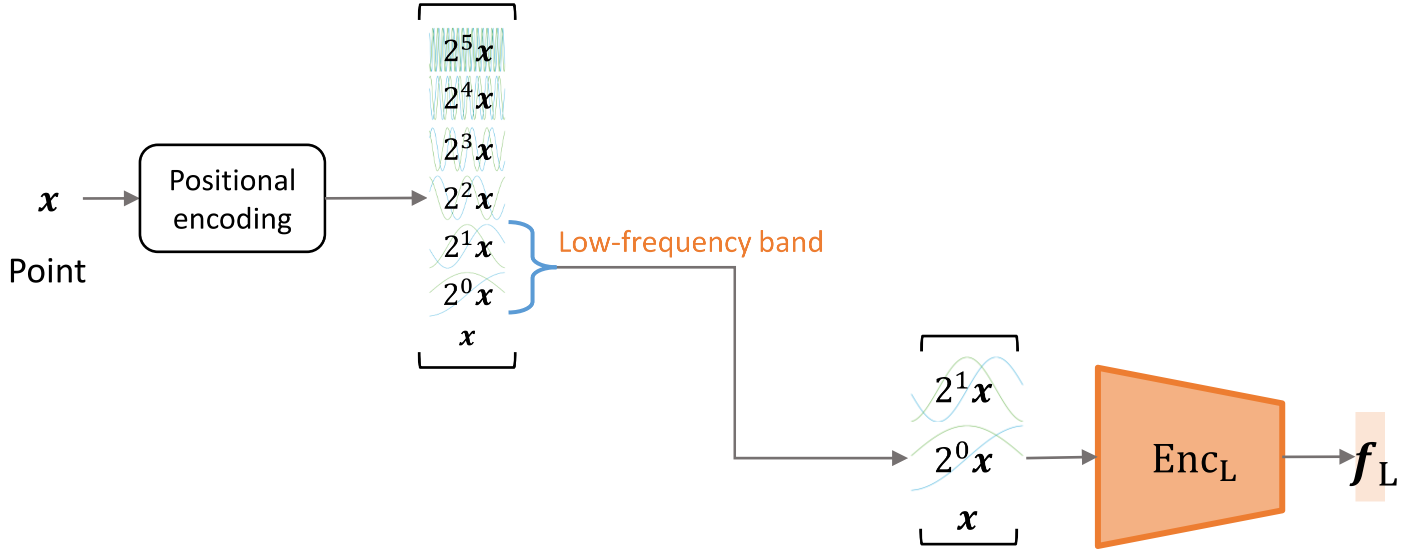
- Three encoders convert the input to features corresponding to different frequency bands (low, middle, high).
- FreBIS uses positional encoding to transform the input coordinate into frequency domains.
  - e.g., frequency level $L = 6$



$$\mathit{Embed}(\boldsymbol{x}) = \left(\boldsymbol{x}, \sin(\boldsymbol{x} \times 2^0), \cos(\boldsymbol{x} \times 2^0), \cdots, \sin(\boldsymbol{x} \times 2^{L-1}), \cos(\boldsymbol{x} \times 2^{L-1})\right)$$

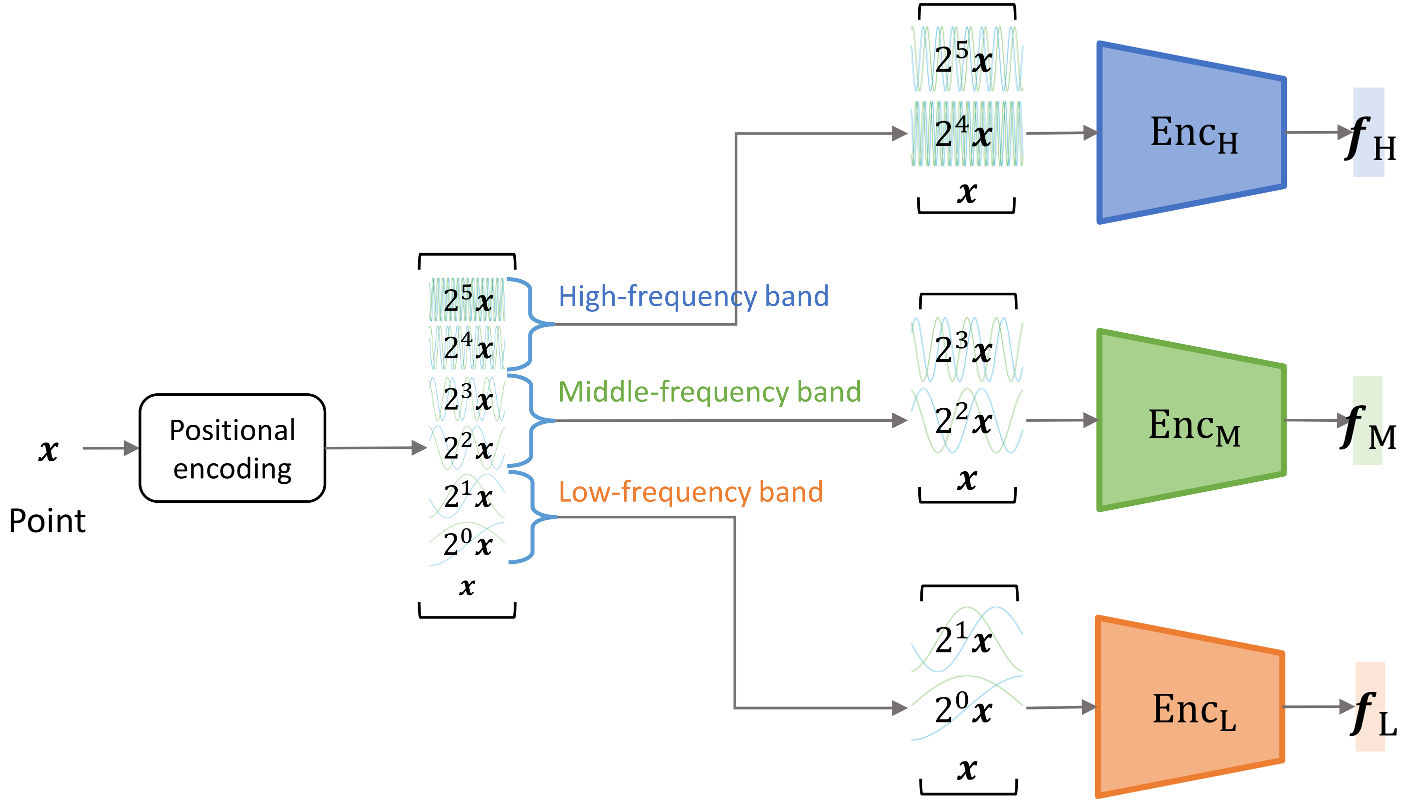
Frequency-Domain Representation

- Three encoders convert the input to features corresponding to different frequency bands (low, middle, high).
- FreBIS uses positional encoding to transform the input coordinate into frequency domains.
  - e.g., frequency level $L = 6$



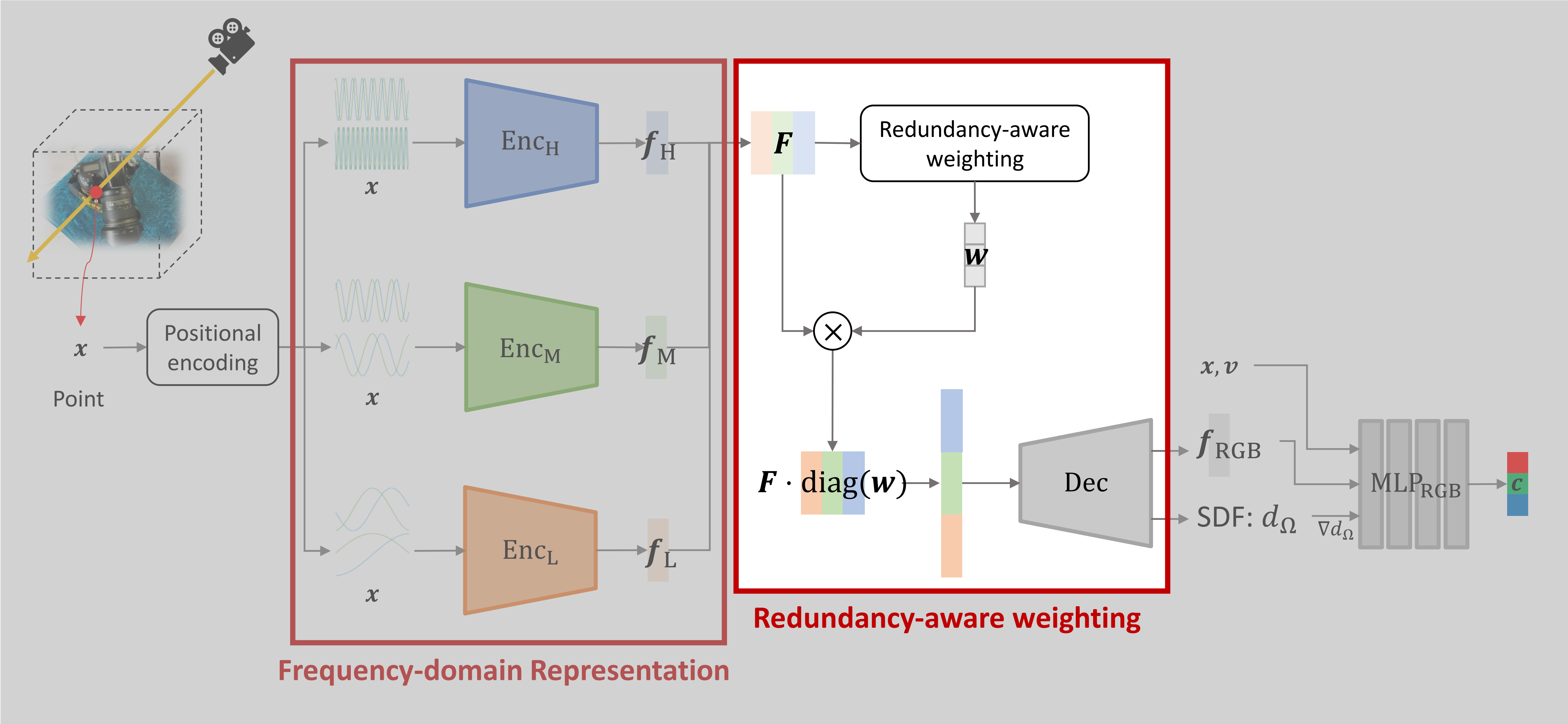
Frequency-Domain Representation

- Three encoders convert the input to features corresponding to different frequency bands (low, middle, high).
- FreBIS uses positional encoding to transform the input coordinate into frequency domains.
  - e.g., frequency level $L = 6$



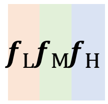
Method: FreBIS

The FreBIS framework consists of two modules:



Redundancy-Aware Weighting

- The redundancy-aware weighting module **maximally utilizes the encoder capacity** and effectively combines the learned complementary information by encouraging **dissimilarity between the learned representations**.
  - i.e., the higher weights are on features that are dissimilar to other features, while the lower weights are on features that are similar to other features.

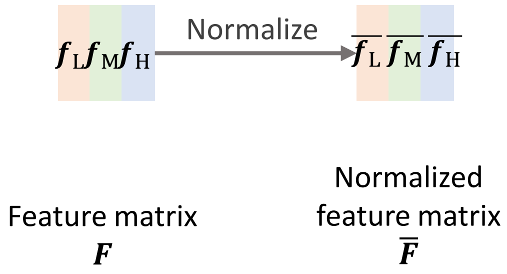


Feature matrix
$\boldsymbol{F}$

Redundancy-Aware Weighting

- The redundancy-aware weighting module **maximally utilizes the encoder capacity** and effectively combines the learned complementary information by encouraging **dissimilarity between the learned representations**.
  - i.e., the higher weights are on features that are dissimilar to other features, while the lower weights are on features that are similar to other features.

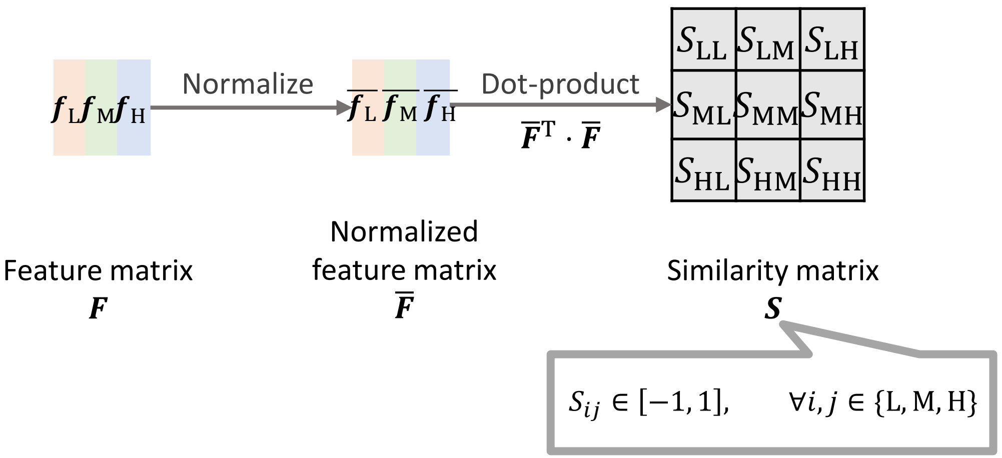
Feature normalization



Redundancy-Aware Weighting

- The redundancy-aware weighting module **maximally utilizes the encoder capacity** and effectively combines the learned complementary information by encouraging **dissimilarity between the learned representations**.
  - i.e., the higher weights are on features that are dissimilar to other features, while the lower weights are on features that are similar to other features.

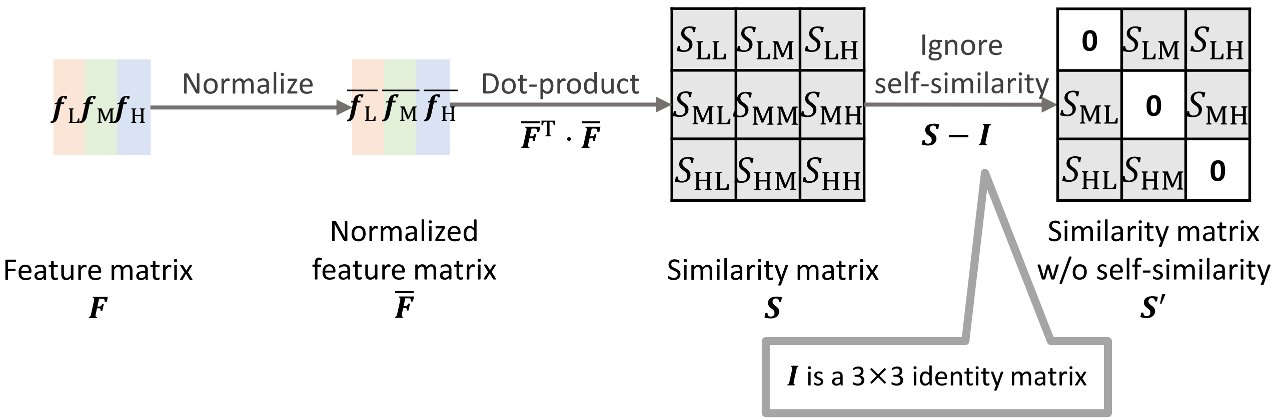
Similarity between features



Redundancy-Aware Weighting

- The redundancy-aware weighting module **maximally utilizes the encoder capacity** and effectively combines the learned complementary information by encouraging **dissimilarity between the learned representations**.
  - i.e., the higher weights are on features that are dissimilar to other features, while the lower weights are on features that are similar to other features.

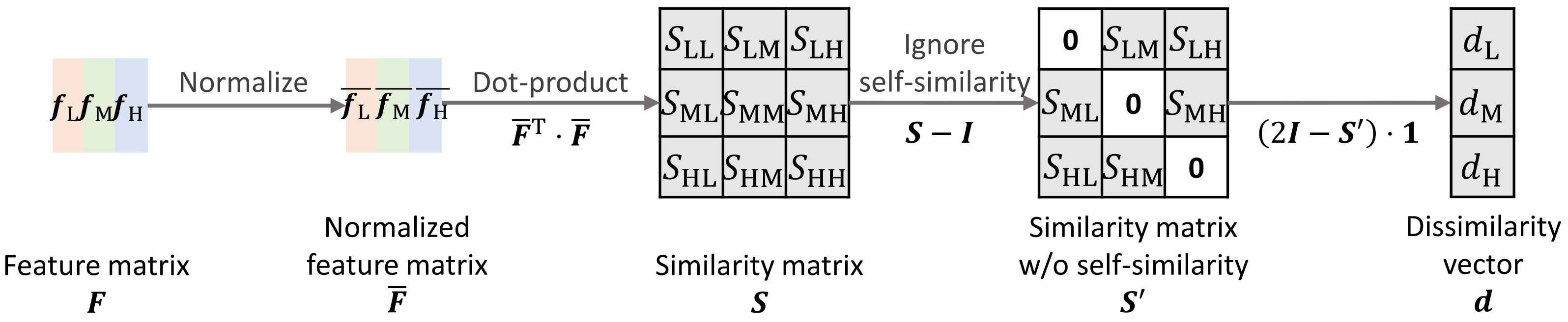
Off-diagonal similarity



Redundancy-Aware Weighting

- The redundancy-aware weighting module **maximally utilizes the encoder capacity** and effectively combines the learned complementary information by encouraging **dissimilarity between the learned representations**.
  - i.e., the higher weights are on features that are dissimilar to other features, while the lower weights are on features that are similar to other features.

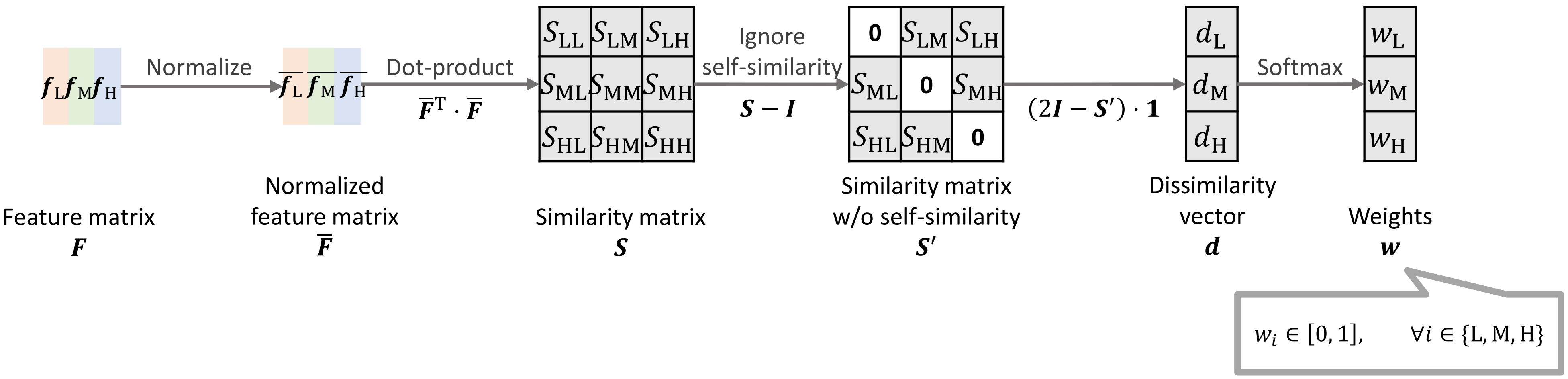
Dissimilarity between features



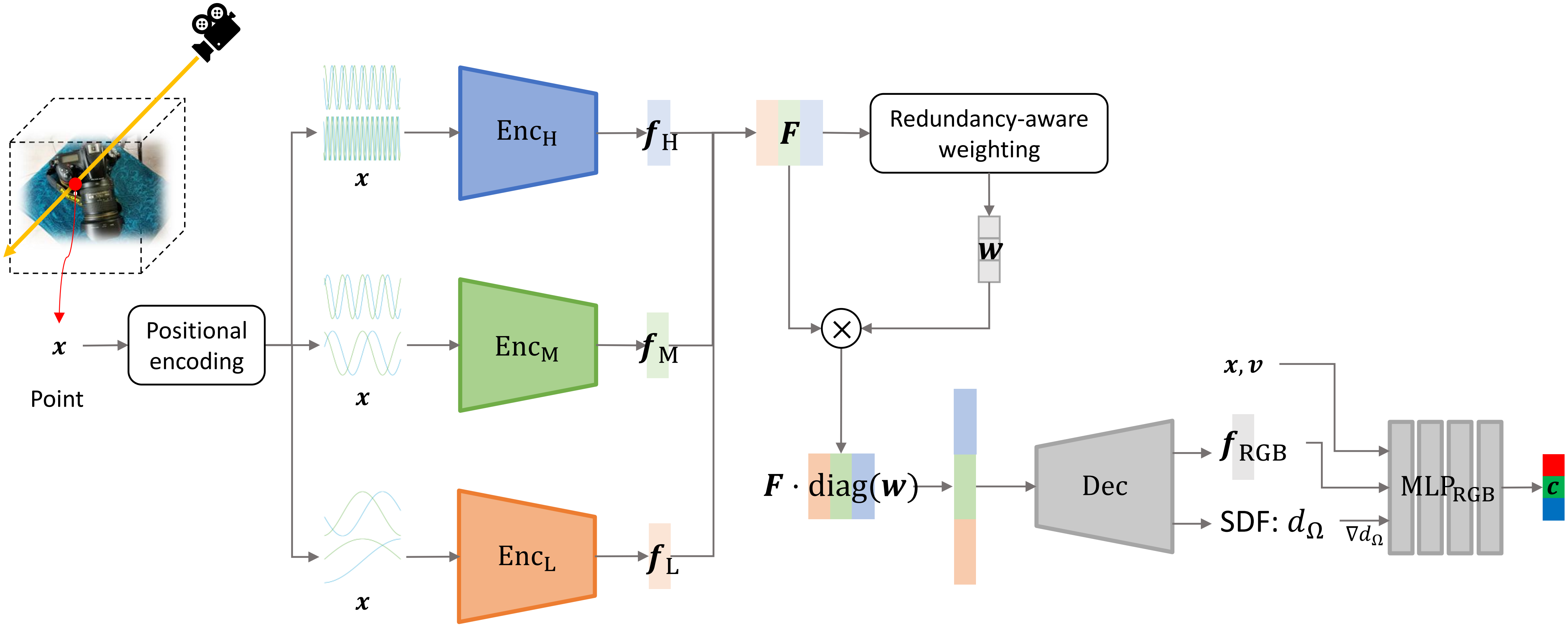
Redundancy-Aware Weighting

- The redundancy-aware weighting module **maximally utilizes the encoder capacity** and effectively combines the learned complementary information by encouraging **dissimilarity between the learned representations**.
  - i.e., the higher weights are on features that are dissimilar to other features, while the lower weights are on features that are similar to other features.

Rescaled dissimilarity



Overall FreBIS Framework



Experimental Setup

Dataset

- **BlendedMVS** : Object-centric real-world scenes with complex backgrounds
  - **Number of Scenes** : 9 scenes
  - **Number of views** : 31~144 views
  - **Resolution** : 768 × 576

Baselines

- VolSDF [Yariv et al., 2021] : 0.5M parameters
- Scaled-up VolSDF : 1.4M parameters (roughly the same as Ours)

*Scaled-up VolSDF: An adaptation of VolSDF, where the number of parameters is increased to be roughly the same as Ours

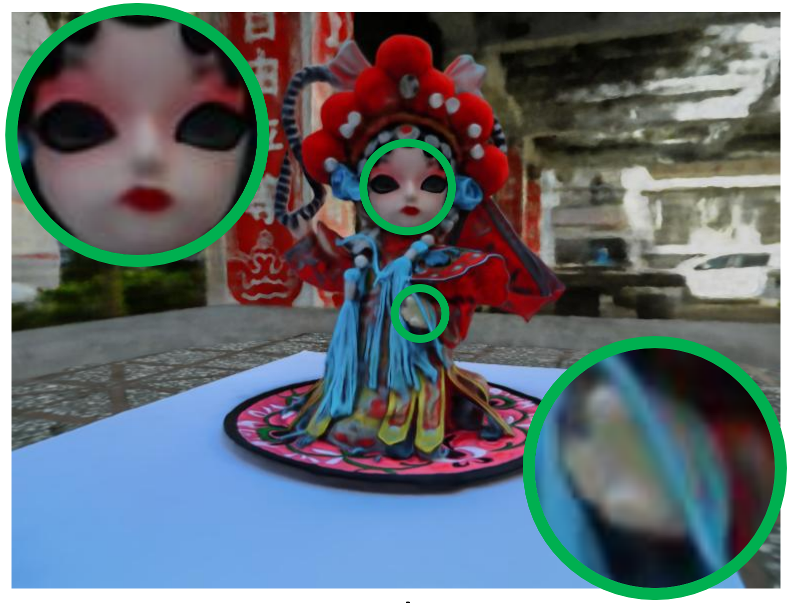
Quantitative Results: BlendedMVS

	Method (no. of parameters)	Doll	Egg	Head	Angel	Bull	Robot	Dog	Bread	Camera	**Mean**
	VolSDF [52] (0.5M)	25.43	27.23	26.94	30.28	26.18	26.39	28.44	**31.18**	22.96	27.23
PSNR(↑)	Scaled-up VolSDF (1.4M)	26.07	27.15	26.62	30.37	26.08	25.07	28.32	29.44	23.02	26.90
	Ours (1.4M)	**26.22**	**27.48**	**27.29**	**30.52**	**26.33**	**26.69**	**28.56**	30.22	**23.08**	**27.38**
	VolSDF [52] (0.5M)	0.911	0.943	0.959	0.989	0.970	0.957	0.950	**0.988**	0.928	0.955
SSIM(↑)	Scaled-up VolSDF (1.4M)	0.925	0.943	0.956	**0.990**	0.970	0.946	0.949	0.980	0.929	0.954
	Ours (1.4M)	**0.928**	**0.946**	**0.961**	**0.990**	**0.971**	**0.962**	**0.952**	0.983	**0.930**	**0.958**
	VolSDF [52] (0.5M)	0.041	0.032	0.017	0.007	0.021	0.032	0.027	**0.006**	0.045	0.025
LPIPS(↓)	Scaled-up VolSDF (1.4M)	**0.035**	0.032	0.018	**0.006**	0.021	0.043	0.028	0.011	0.045	0.027
	Ours (1.4M)	**0.035**	**0.030**	**0.015**	**0.006**	**0.020**	**0.030**	**0.026**	0.009	**0.044**	**0.024**

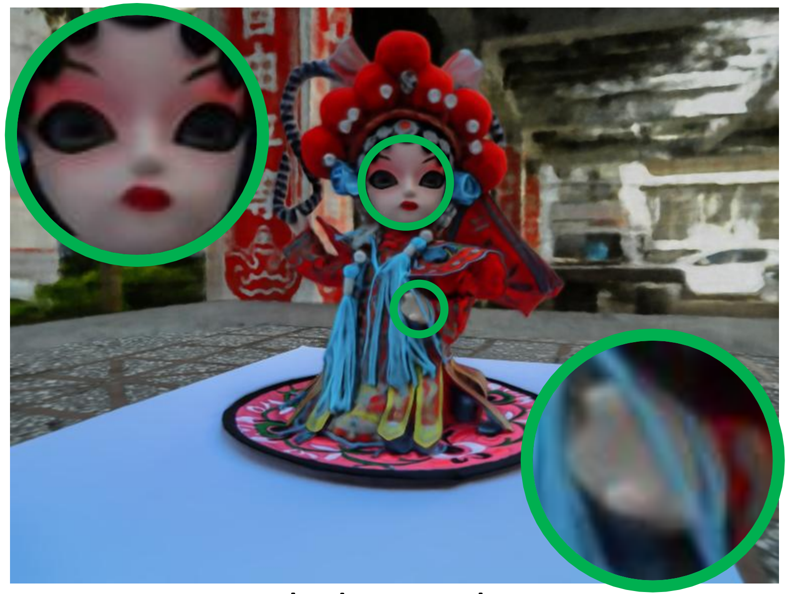
The proposed method improves the rendering quality in terms of PSNR, SSIM, and LPIPS.

Qualitative Results: BlendedMVS (Doll)

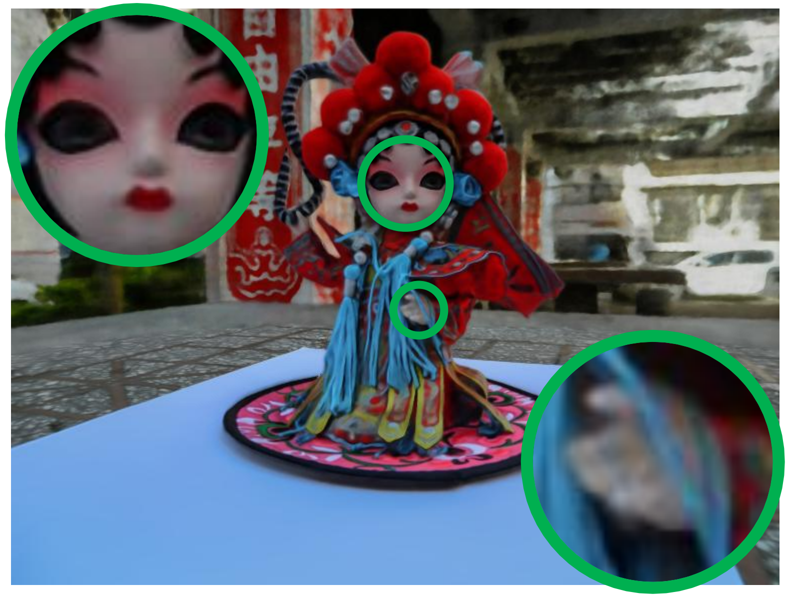
Ground truth



VolSDF



Scaled-up VolSDF

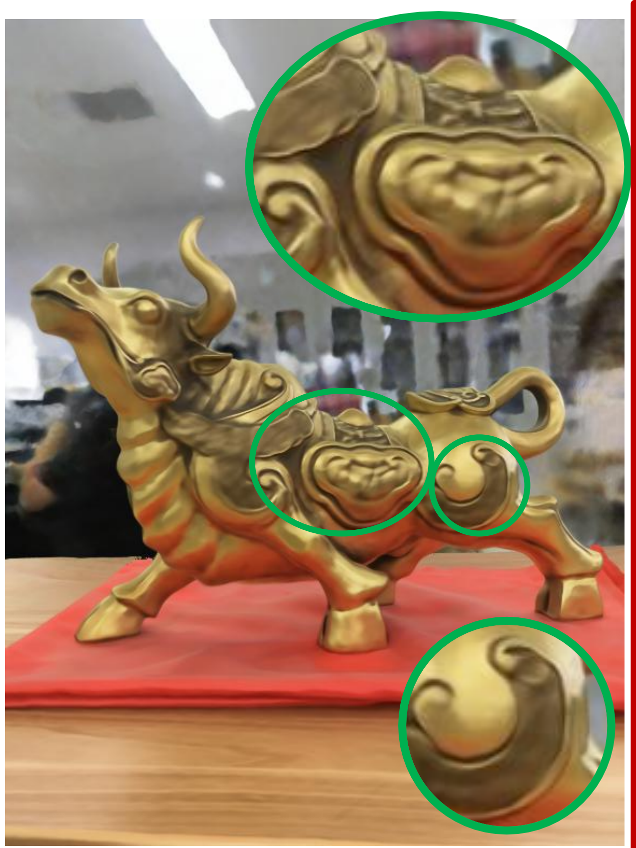


Ours

Qualitative Results: BlendedMVS (Bull)


Ground truth

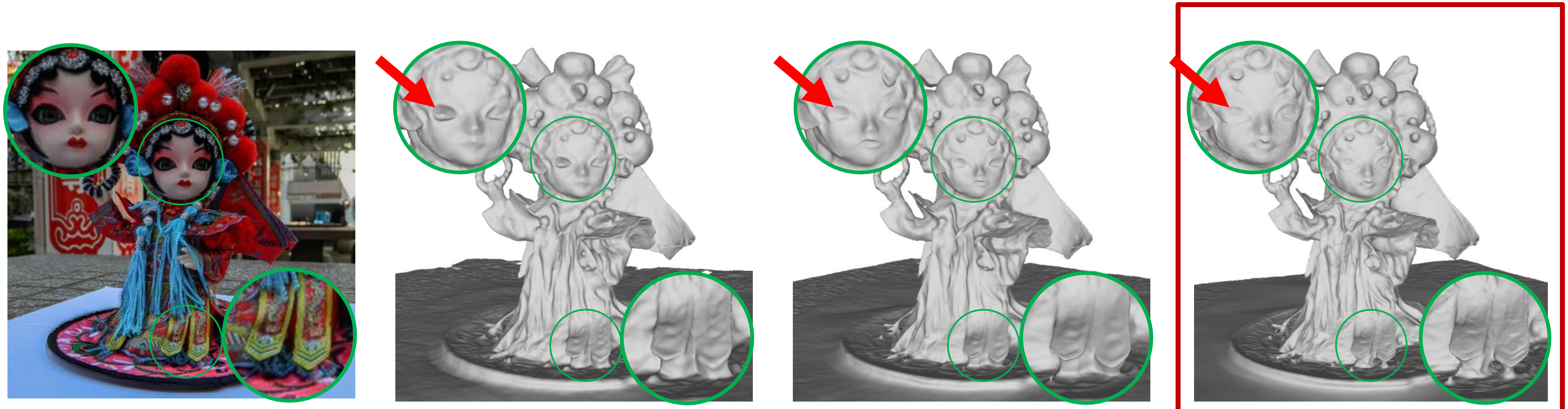

VolSDF


Scaled-up VolSDF


Ours

MITSUBISHI ELECTRIC
Changes for the Better
MITSUBISHI ELECTRIC RESEARCH LABORATORIES

Qualitative Results: BlendedMVS (Doll)

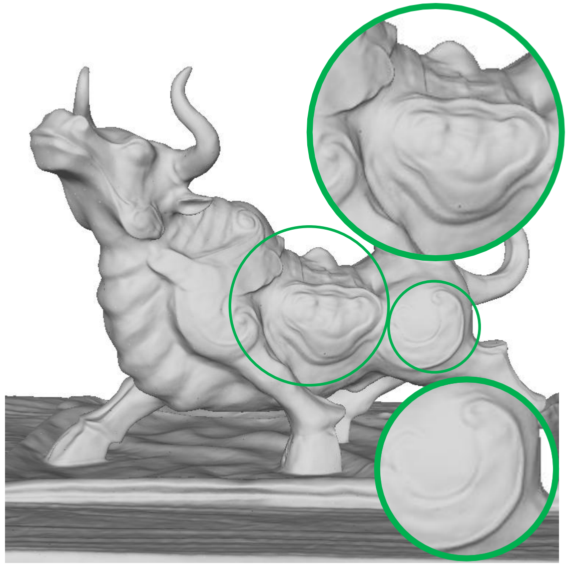


Ground truth VolSDF Scaled-up VolSDF Ours

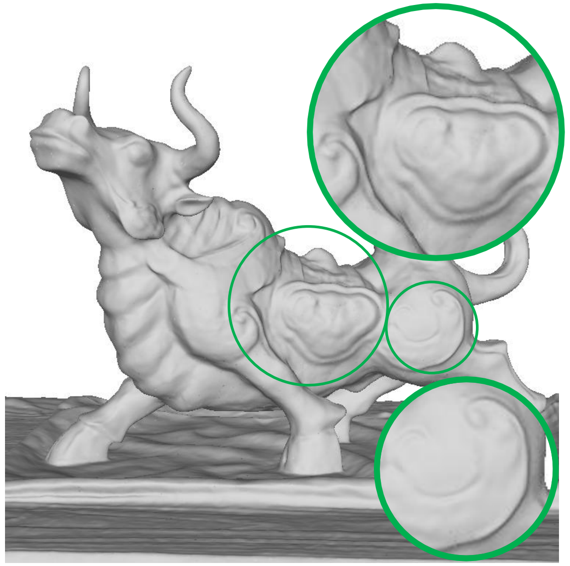
MITSUBISHI ELECTRIC
Changes for the Better
MITSUBISHI ELECTRIC RESEARCH LABORATORIES

Qualitative Results: BlendedMVS (Bull)

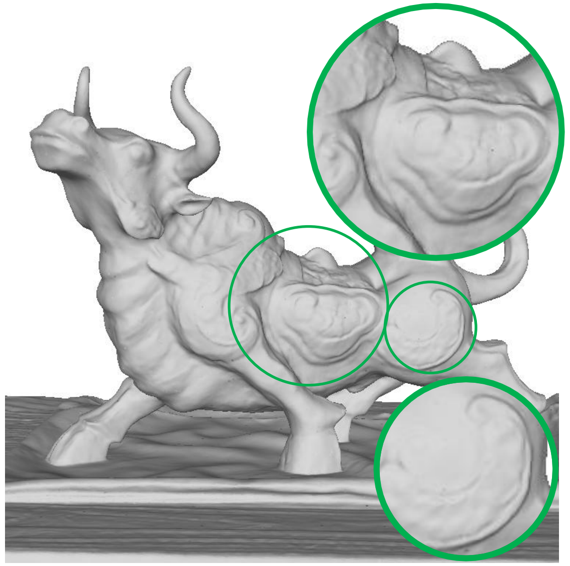
Ground truth



VolSDF



Scaled-up VolSDF



Ours

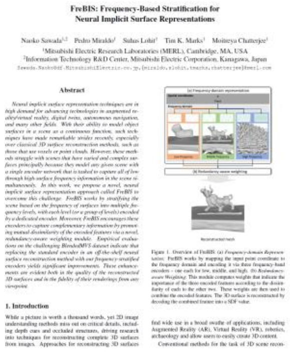
Conclusion

- We propose FreBIS, which **stratifies a scene into multiple frequency levels** according to the surface frequencies and leverages a novel **redundancy-aware weighting module,** to effectively capture complementary information.
- FreBIS improved the qualities of the reconstructed meshes, as well as rendered images.
- For future work, we plan to evaluate FreBIS on other datasets and backbones.

Thank you for your attention!

MITSUBISHI ELECTRIC
Changes for the Better

MITSUBISHI ELECTRIC RESEARCH LABORATORIES


FreBIS: Frequency-Based Stratification for Neural Implicit Surface Representations

Naoko Sawada[1,2] Pedro Miraldo[1] Suhas Lohit[1] Tim K. Marks[1] Moitreya Chatterjee[1]
[1]Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, USA
[2]Information Technology R&D Center, Mitsubishi Electric Corporation, Kanagawa, Japan

Abstract

SCAN ME

For more details, please check out our paper and poster.