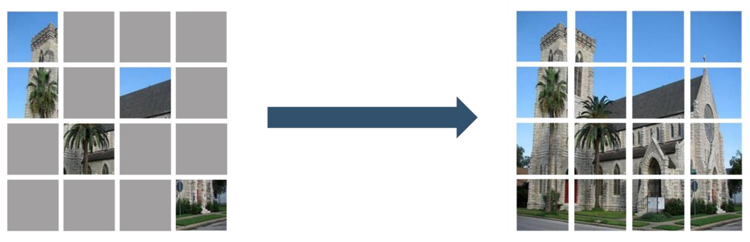
MINR: Implicit Neural Representations with Masked Image Modelling

Sua Lee, Joonhun Lee, and Myungjoo Kang
Seoul National University

Introduction

Task. **Masked Image modelling(MIM)** is notable self-supervised learning method, deliberately masking images and training models to reconstruct the hidden information for robust feature representations.



Motivation. The limitation of **Masked autoencoder(MAE)**,[1] one of the powerful MIM method, is its dependency on masking strategies.

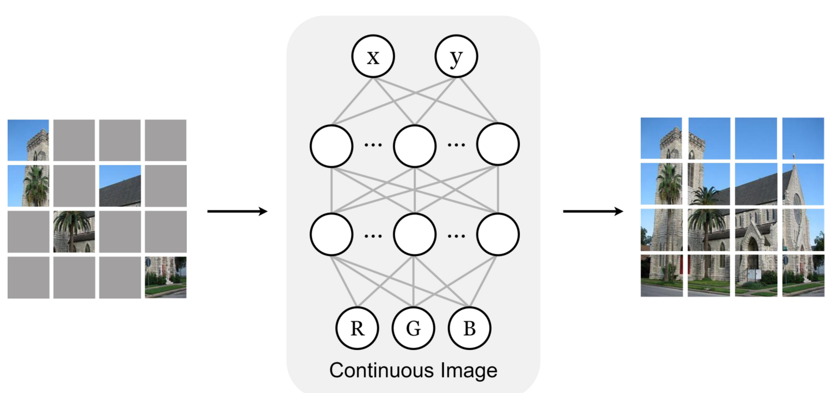
Contributions

Masked Implicit Neural Representations(MINR) framework effectively combines Implicit Neural Representations(INR) with MIM to address the limitations of MAE.

- Leverages INR to learn a continuous function, less affected by variations of visible patches information.
- Alleviates the reliance on heavy pretrained model dependencies, with considerably reduced params.
- Learned continuous function provides greater flexibility in creating embeddings for various downstream tasks.

Methodology

Integrating INRs with MIM



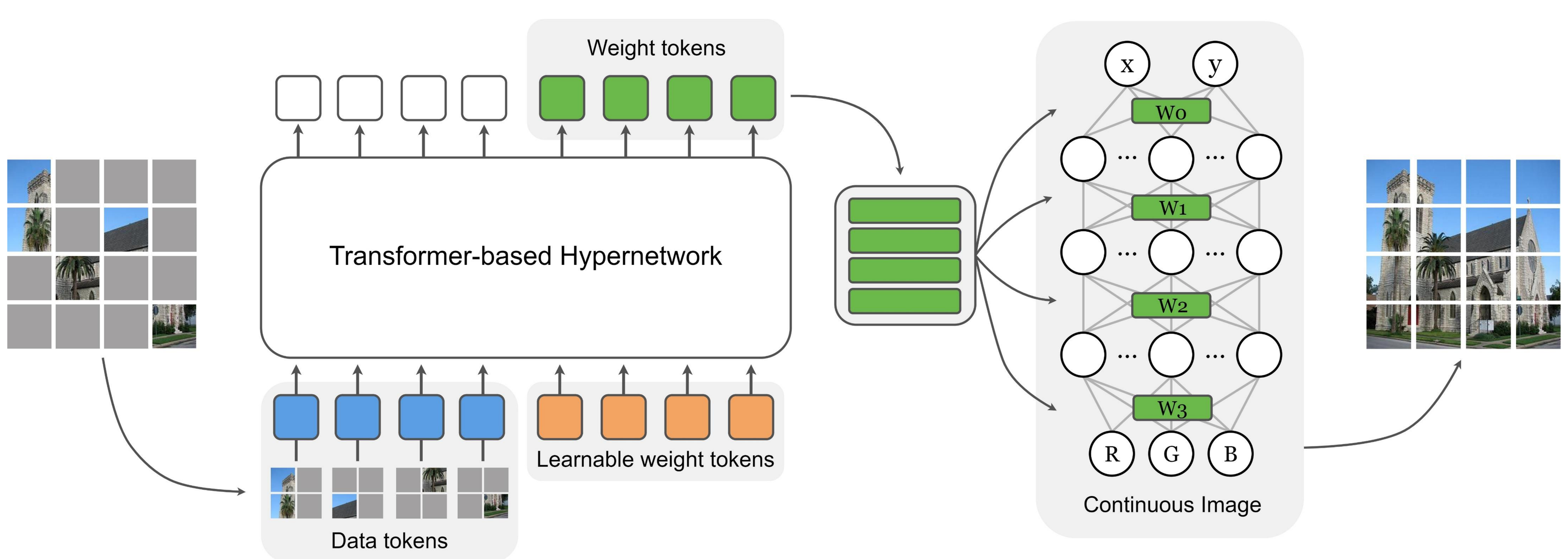
- Apply 75% random masking to a dataset $O = \{o^{(n)}\}_{n=1}^{N}$ containing N observations $o^{(n)} = \{x_i^{(n)}, y_i^{(n)}\}$:

$$M = \{m^{(n)} | m^{(n)} \coloneqq Mask^{(n)}(o^{(n)})\}_{n=1}^{N}$$

- Estimate a continuous function $f_\theta: \mathbb{R}^2 \to \mathbb{R}^3$, which maps input coordinates to corresponding properties.

$$\mathcal{L}_n\left(\theta^{(n)}; m^{(n)}\right) = \frac{1}{H \times W} \sum_{i=1}^{H \times W} \left\| y_i^{(n)} - f_{\theta^{(n)}}\left(x_i^{(n)}\right) \right\|_2^2$$

To expand independent training to a large-scale out-of-distribution(OOD) datasets, we utilizes INR approaches that utilize transformer-based hypernetwork.



TransINR[2] -based approach

- The hypernetwork predicts the entire set of INR weights $\theta^{(n)} = \{W_l\}_{l=1}^{L}$, and $\Theta = \{\theta^{(n)}\}_{n=1}^{N}$

$$\mathcal{L}(\Theta; M) = \frac{1}{N} \sum_{n=1}^{N} \mathcal{L}_n(\theta^{(n)}; m^{(n)})$$

GINR[3] -based approach

- Generalized-INR partitions the MLP hidden layers into instance-specific θ and instance-agnostic layers ϕ. It is empirically found to be effective in learning common and unique patterns across a dataset, making it ideal for OOD settings.

$$\mathcal{L}(\Theta, \phi; M) = \frac{1}{N} \sum_{n=1}^{N} \mathcal{L}_n(\theta^{(n)}, \phi; m^{(n)})$$

Results

- We employ three datasets for experiments, namely ***CelebA, Imagenette***, and ***MIT-Indoor67***. Notably, these datasets were selected due to their varied nature, for the evaluation of the robustness of our method across different data distributions.

In-domain performance

Method		CEL	IMG	IND	# Param.
MAE	Large	15.018	14.693	15.181	313.6M
	Base	15.401	14.452	14.370	106.2M
MINR	TransINR	**21.865**	18.737	17.756	44.5M
	GINR	21.680	**19.358**	**18.622**	**43.7M**

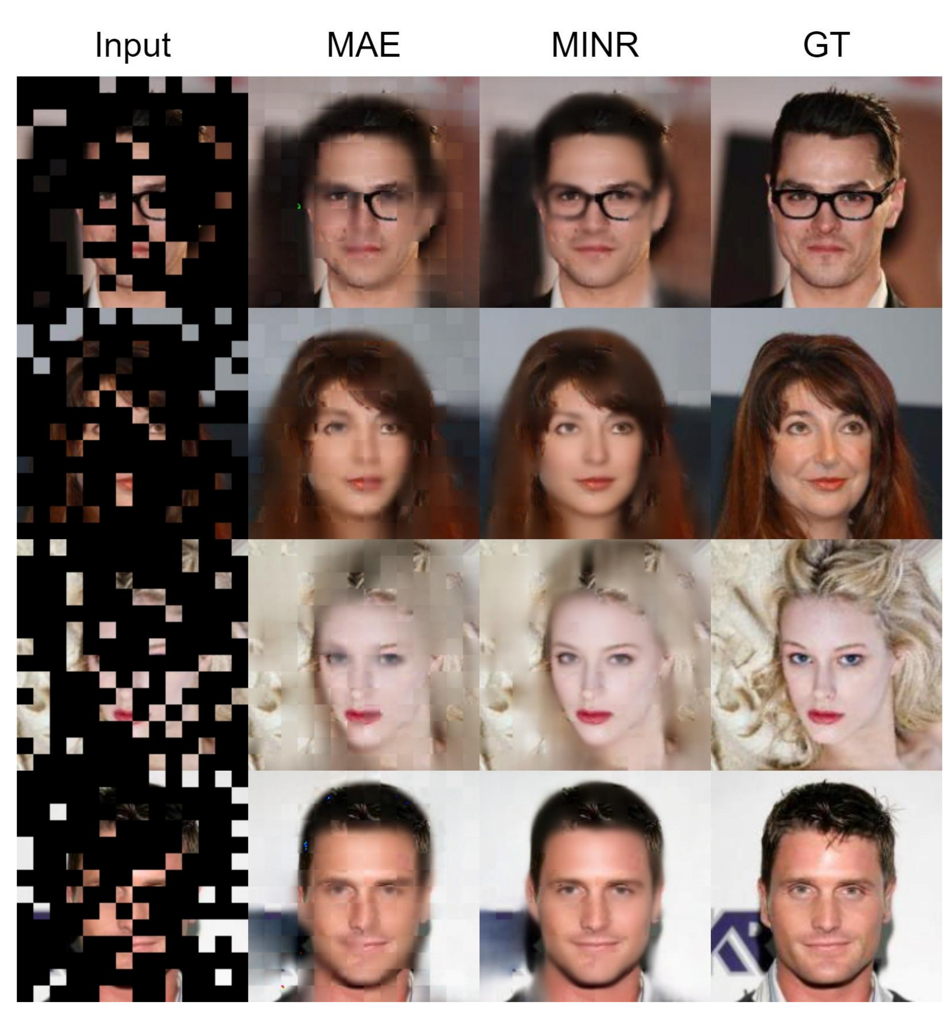
Out-of-domain performance

Method		CEL →		IMG →		IND →	
		IMG	IND	CEL	IND	CEL	IMG
MAE	Large	14.262	14.300	14.853	14.779	14.858	14.949
	Base	14.508	14.464	14.499	14.558	13.831	14.069
MINR	TransINR	**18.058**	**17.361**	19.929	17.920	18.992	18.103
	GINR	18.041	17.336	**19.994**	**18.045**	**19.509**	**18.573**

Qualitative results

For a fair comparison, we visualize the results by pasting unmasked patches onto the reconstruction results.

Our approaches clarity enhanced the reconstruction performance for all experiment settings.



References

[1] He, Kaiming, et al. "Masked autoencoders are scalable vision learners." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022.
[2] Chen, Yinbo, and Xiaolong Wang. "Transformers as meta-learners for implicit neural representations." *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2022.
[3] Kim, Chiheon, et al. "Generalizable Implicit Neural Representations via Instance Pattern Composers." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023.

Acknowledgement

Myungjoo Kang was supported by the NRF grant [2012R1A2C3010887] and the MSIT/IITP ([1711117093], [NO.2021-0-00077], [No. 2021-0-01343, Artificial Intelligence Graduate School Program (SNU)]).