Generalizable Implicit Neural Representations via Instance Pattern Composers

Chiheon Kim*, **Doyup Lee***, Saehoon Kim (Kakao Brain)

Minsu Cho, Wook-Shin Han (POSTECH)

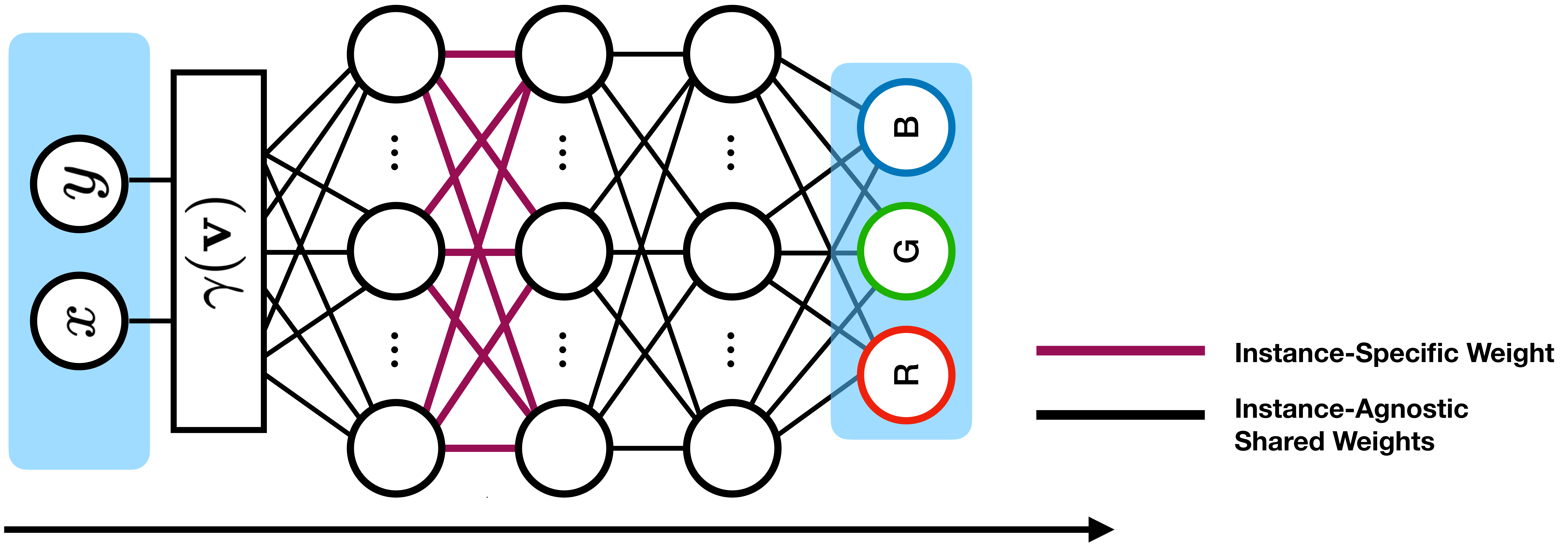
* Equal Contribution

Poster ID: 341
Tag: WED-AM-341

Generalizable INRs via Instance Pattern Composers

- Our simple modulation method can leverage **the powerful modulation capacity** of weight modulation and **the low computational cost** of feature modulations.

- **The shared structures of representations across instances** are the **pattern composition rule** of **the shared weights**, while a data instance is characterized by one modulation weight.



How the modulation weight, Instance Pattern Composers,

can be predicted during training & inferences?

It's compatible with both

1) hypernetworks & 2) meta-learning

Kakao Brain

Paper

Code

Generalizable Implicit Neural Representations via Instance Pattern Composers

Chiheon Kim*, **Doyup Lee***, Saehoon Kim (Kakao Brain)

Minsu Cho, Wook-Shin Han (POSTECH)

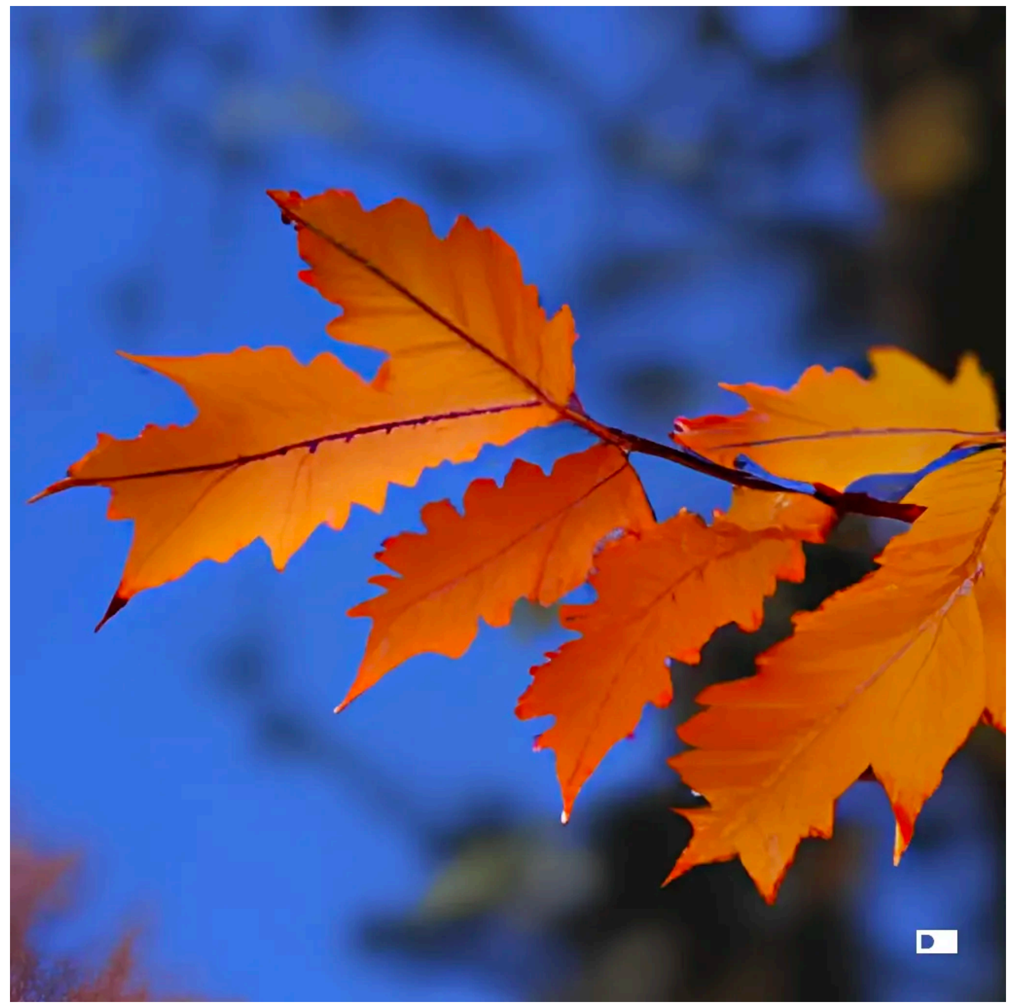
* Equal Contribution

Implicit Neural Representations (INRs)

- For an implicit representation of data instance, **a parameterized neural network** (*e.g.* coordinate-based MLP) is trained to **map a coordinate into its corresponding features**.
- That is, **a data instance** is represented as **a continuous function**.

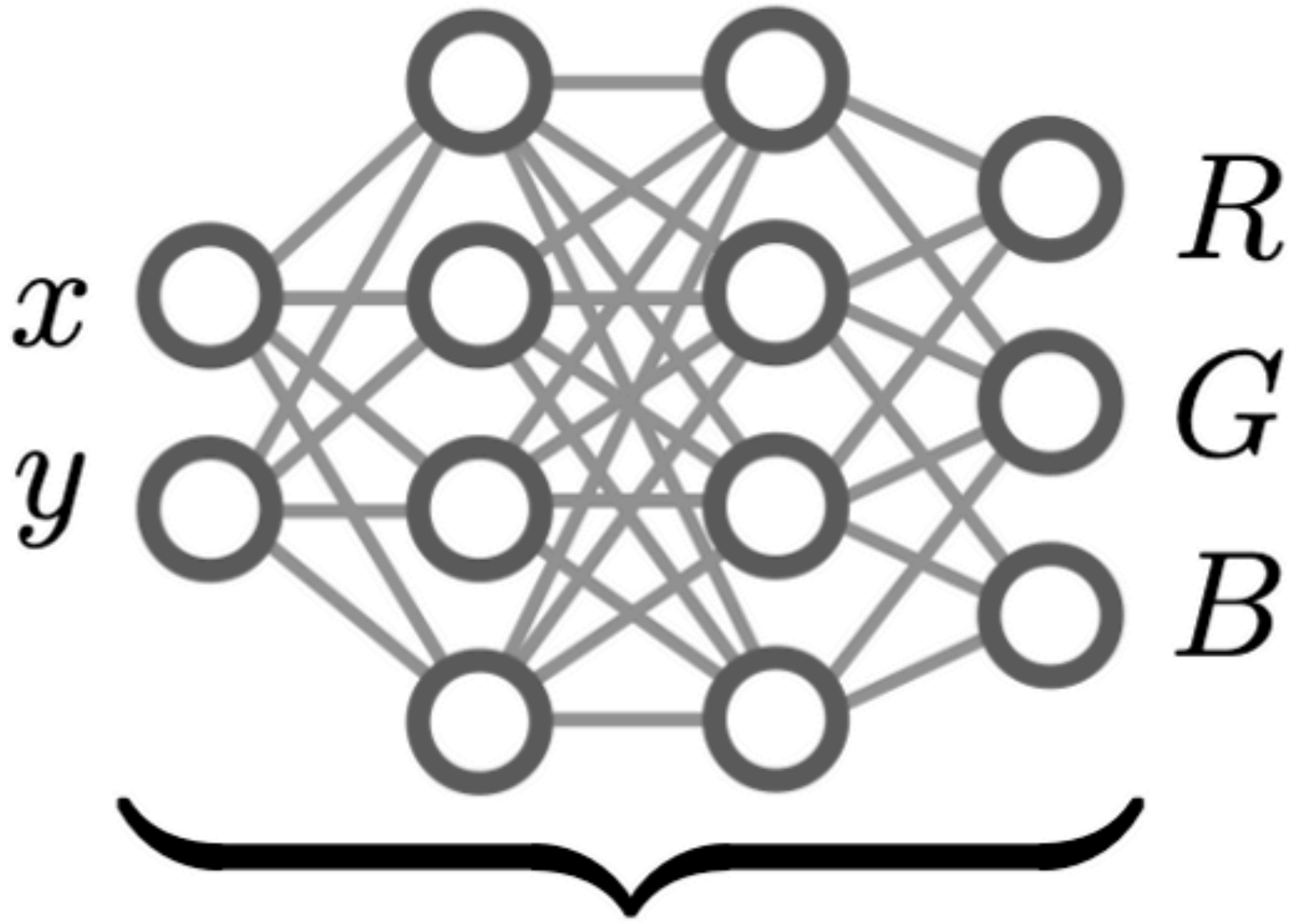
one data instance

$$\mathbf{x} = \{(\mathbf{v}_i, \mathbf{y}_i)\}_{i=1}^{M_n}$$



$\rightarrow$

coordinate-based MLP



$$\mathbf{y}_i = f_\phi(\mathbf{v}_i)$$

one continuous function
(and a set of parameters)

$$\phi = \min_\phi \frac{1}{M_n} \sum_{i=1}^{M_n} ||\mathbf{y}_i - f_\phi(\mathbf{v}_i)||_2^2$$

Each MLP is **separately trained** to represent each data point.

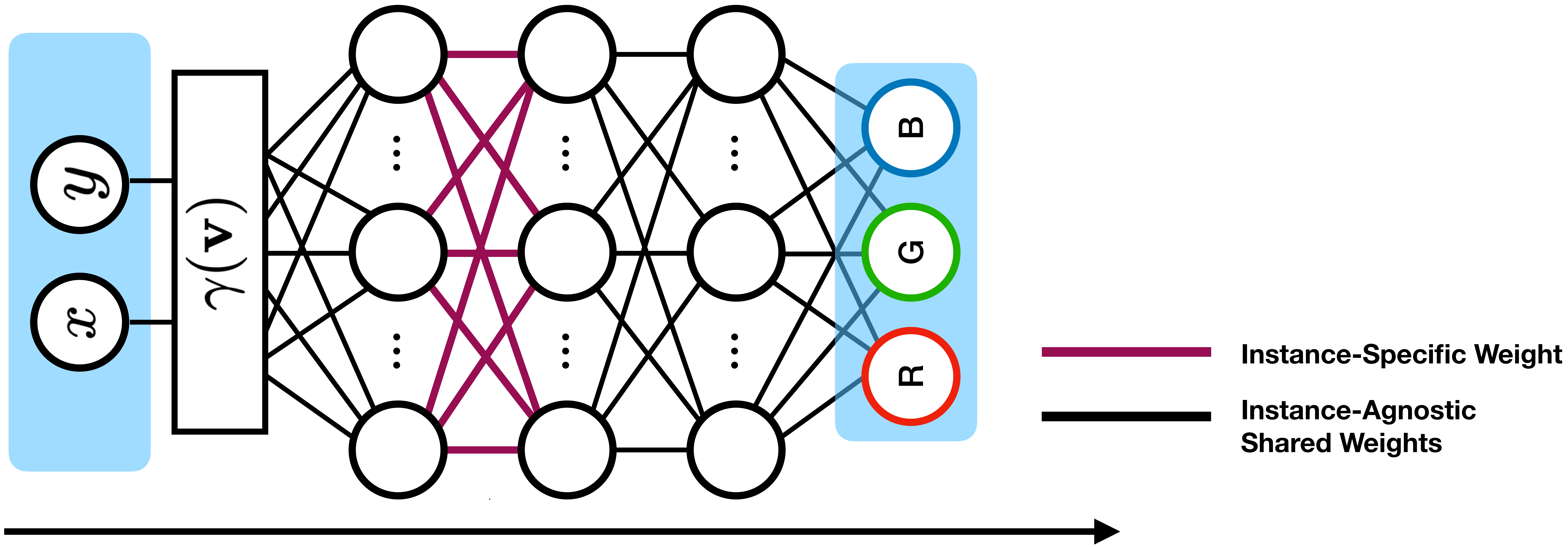
It **cannot learn shared structures** , representations, and knowledge **across** data instances.

Generalizable INRs via Instance Pattern Composers

- We **categorize the weights of MLP** into the following two types:

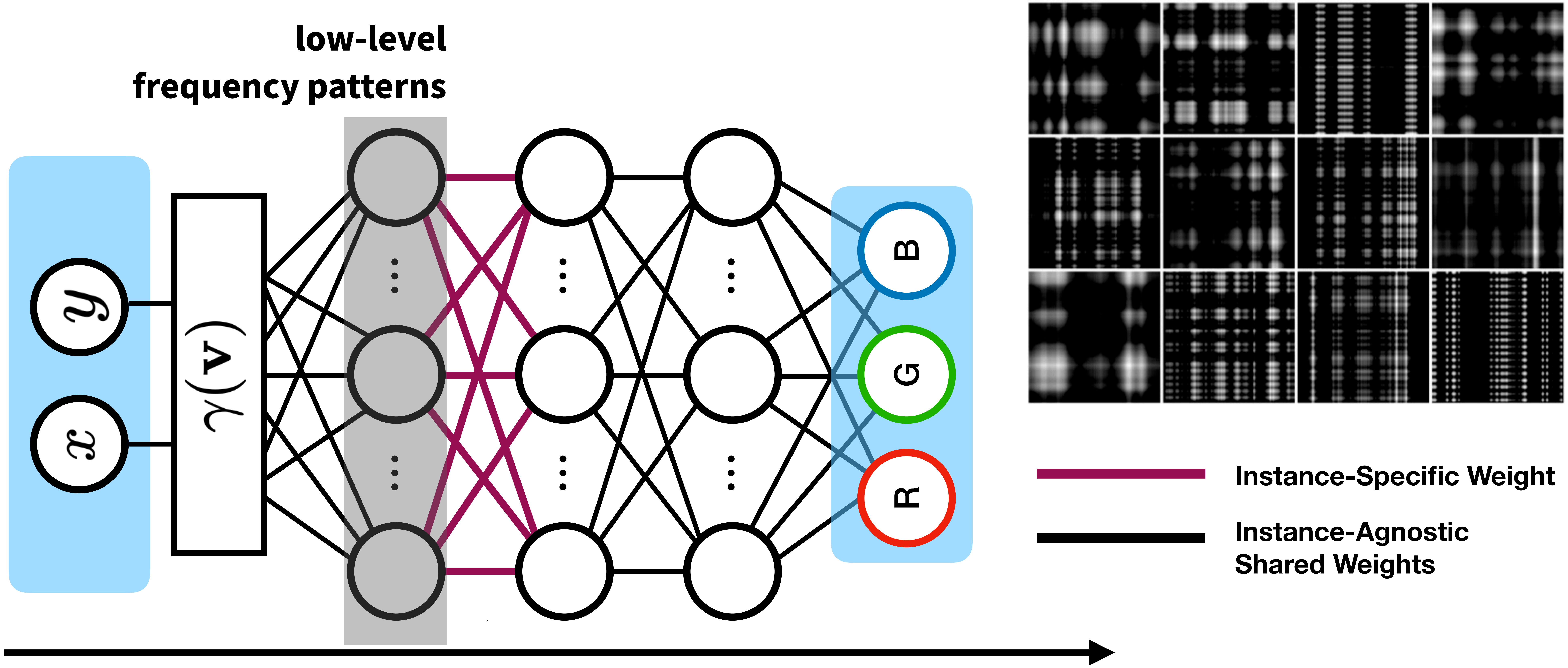
  i) **Instance Pattern Composer** as instance-**specific** parameter

  ii) **Pattern Composition Rule** as instance-**agnostic** parameter.



Generalizable INRs via Instance Pattern Composers

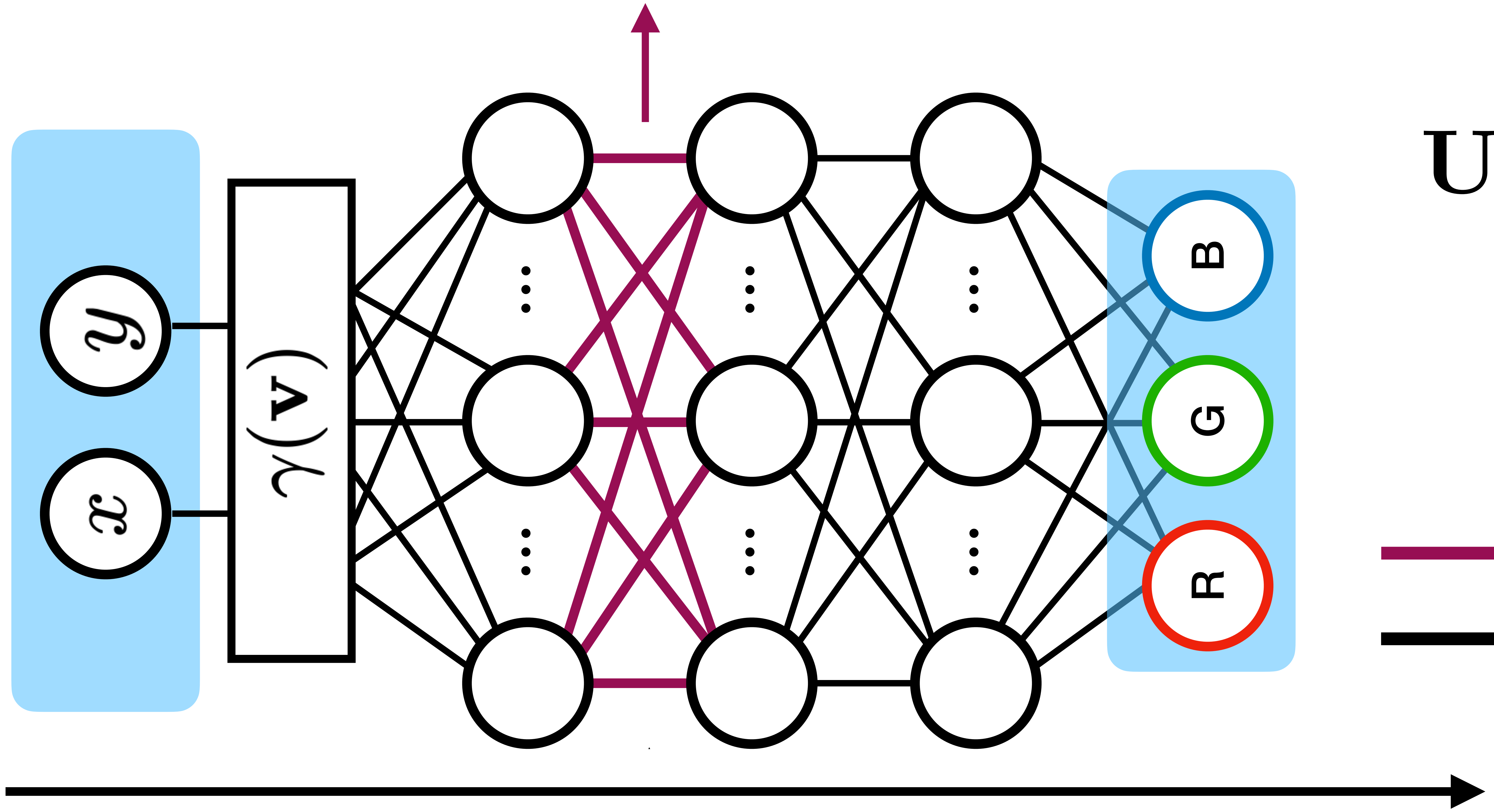
- After the Fourier feature mapping, **instance-agnostic** **low-level frequency patterns** are extracted for irregular and non-periodic frequency patterns.



Generalizable INRs via Instance Pattern Composers

- We define **instance pattern composers** **to characterize the INR of a data instance.**
- **Instance pattern composers** are **the only instance-specific part** of our coordinate-based MLP.

modulated weight by instance pattern composer $V^{(n)}$



$$\mathbf{U}\mathbf{V}^{(n)} = \mathbf{W}_{\mathrm{m}}^{(n)}$$

$$\mathbf{U} \in \mathbb{R}^{d \times r}$$

$$\mathbf{V}^{(n)} \in \mathbb{R}^{r \times d}$$

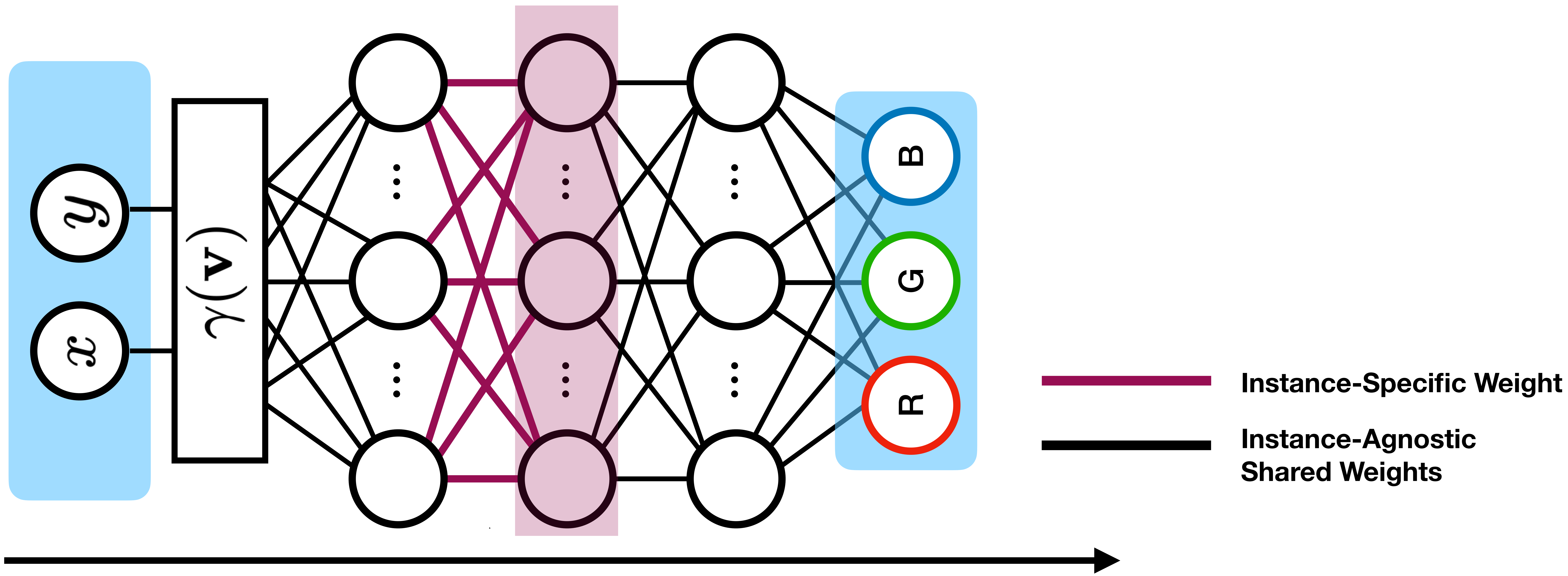
Instance-Specific Weight

Instance-Agnostic Shared Weights

Generalizable INRs via Instance Pattern Composers

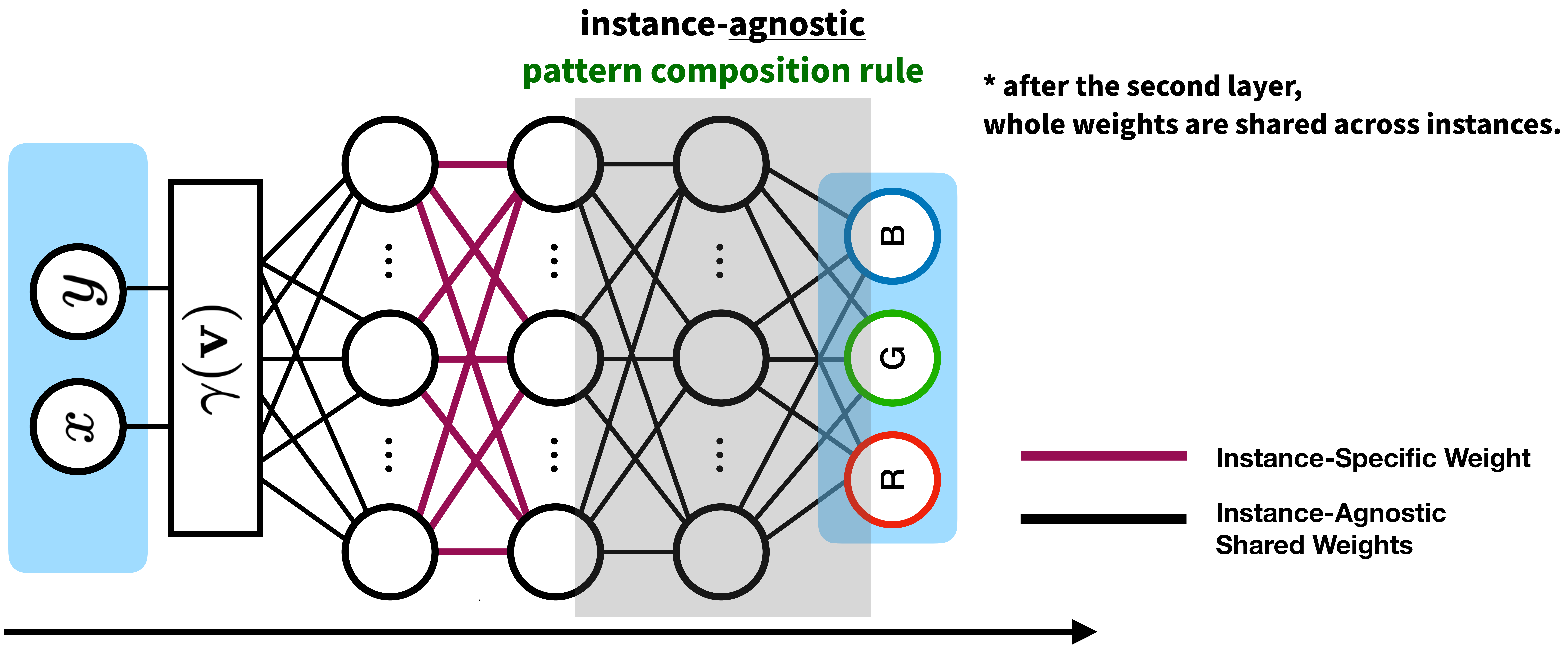
- We **categorize the weights of MLP** into the following two types:
  i) **Instance Pattern Composer** as instance-specific parameter
  ii) **Pattern Composition Rule** as instance-agnostic parameter.

instance content patterns



Generalizable INRs via Instance Pattern Composers

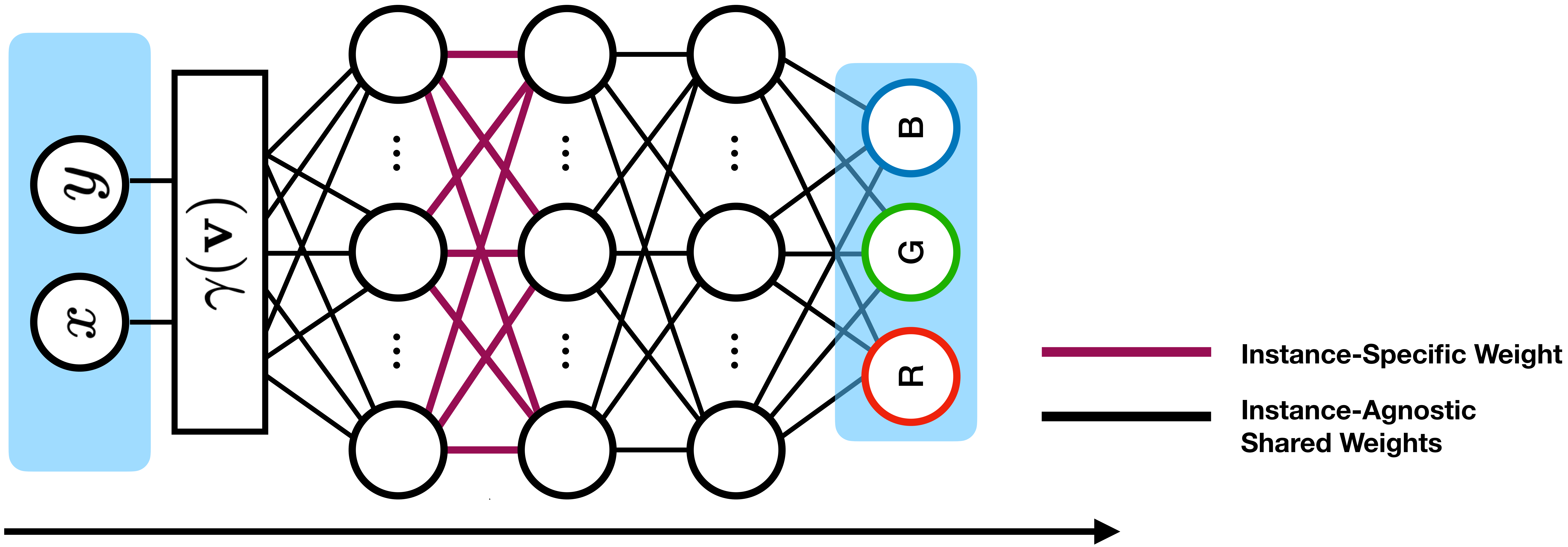
- We **categorize the weights of MLP** into the following two types:
  - i) **Instance Pattern Composer** as instance-specific parameter
  - ii) **Pattern Composition Rule** as instance-agnostic parameter.



Generalizable INRs via Instance Pattern Composers

- Our simple modulation method can leverage **the powerful modulation capacity** of weight modulation and **the low computational cost** of feature modulations.

- **The shared structures of representations across instances** are the **pattern composition rule** of **the shared weights**, while a data instance is characterized by one modulation weight.



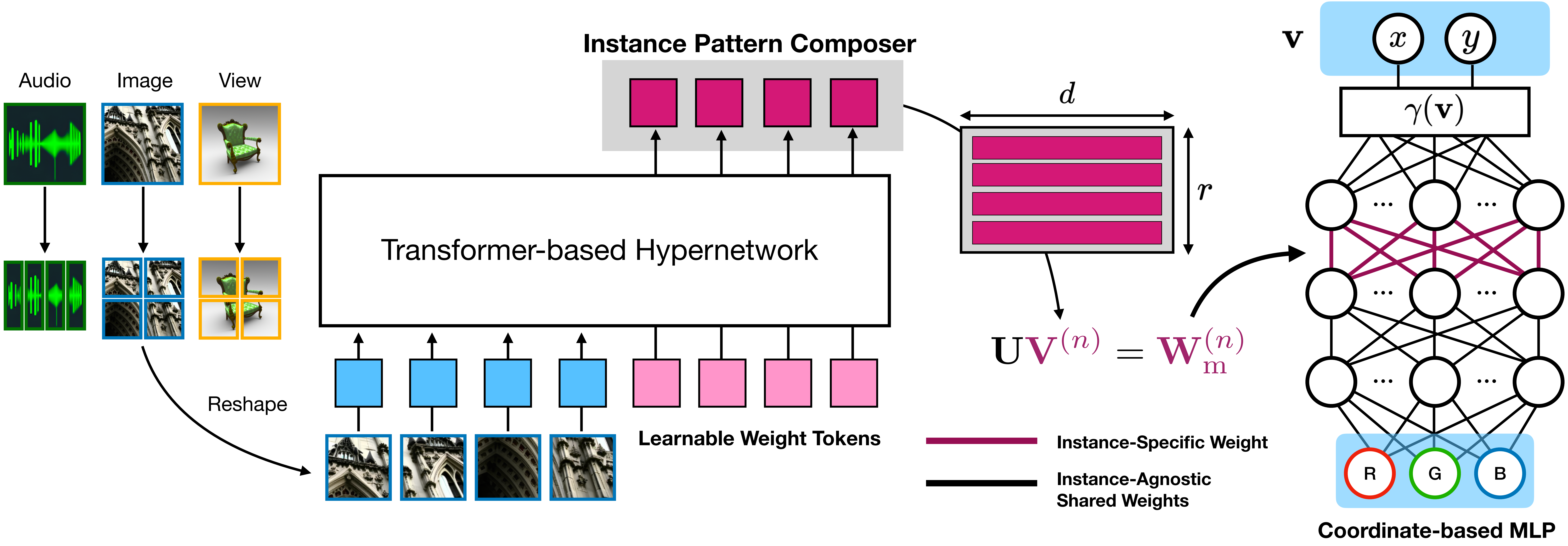
How the modulation weight, Instance Pattern Composers, can be <u>predicted during training & inferences</u>?

It's compatible with both

1) <u>hypernetworks</u> & 2) meta-learning

Transformer-based Hypernetwork

- Transformer-based hypernetwork **predicts <u>the row vectors</u> of instance pattern composers**.
- **Learnable weight tokens** are used as the inputs of the transformer to predict modulation weights.



Meta-Learning for Instance Pattern Composers

- CAVIA can be modified to learn the initialization of instance pattern composers for rapid adaptation in few gradient steps, while the remaining weights are fixed.

Algorithm 1 Optimization-based meta-learning [32] for generalizable INRs via instance pattern composer.

Require: Randomly initialized θ, ϕ, a dataset $\mathcal{X}$, the number of inner steps N_{inner}, and learning rates ϵ, ϵ'.

while not done **do**
    for $n = 1, \cdots, N$ **do**
        Initialize instance-specific parameter $\phi^{(n)} \leftarrow \phi$
    end for
    `/* inner-loop updates for` $\theta^{(n)}$ `*/`
    for all step $\in \{1, \cdots, N_{\text{inner}}\}$ and $\mathbf{x}^{(n)} \in \mathcal{X}$ **do**
        $\phi^{(n)} \leftarrow \phi^{(n)} - \epsilon \|\phi^{(n)}\|^2 \nabla_{\phi^{(n)}} \mathcal{L}_n(\theta, \phi^{(n)}; \mathbf{x}^{(n)})$
    end for
    `/* outer-loop updates for` θ, ϕ `*/`
    Update $\phi \leftarrow \phi - \epsilon' \nabla_\phi \mathcal{L}(\theta, \{\phi^{(n)}\}_{n=1}^N; \mathcal{X})$
    Update $\theta \leftarrow \theta - \epsilon' \nabla_\theta \mathcal{L}(\theta, \{\phi^{(n)}\}_{n=1}^N; \mathcal{X})$
end while

update the initialization
of instance pattern composers

update instance-agnostic
pattern composition rule

Audio Reconstruction

- Five layer MLP is trained to reconstruct 1D audio signal with 1 second and 3 seconds, respectively.
- Our generalizable INRs via Instance Pattern Composers outperforms previous TransINR, validating the effectiveness of our simple weight modulation methods.

Table 1. PSNRs of the reconstruction of the LibriSpeech test-clean dataset whose sample is trimmed into one and three seconds.

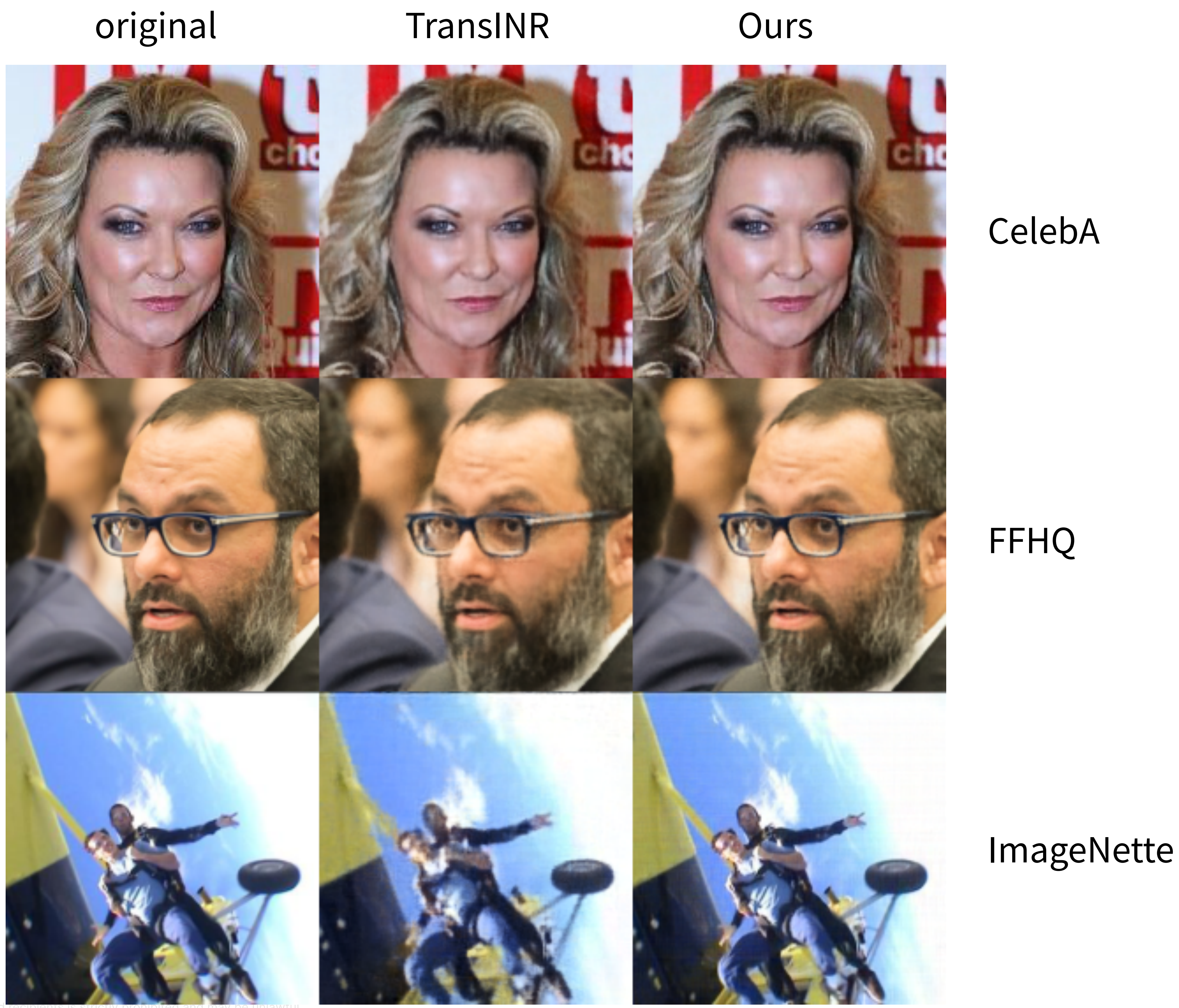
	LibriSpeech (1s)	LibriSpeech (3s)
TransINR	39.22	33.17
Ours	**40.11**	**35.38**

178x178 Image Reconstruction

- Transformer predicts 256 number of weight modulation vectors as instance pattern composers.
- The ImageNette dataset contains 10 classes of images in ImageNet.

Table 2. PSNRs of reconstructed images for178×178 resolution of images in the CelebA, FFHQ, and ImageNette test dataset.

	CelebA	FFHQ	ImageNette
Learned Init [27]	30.37	-	27.07
TransINR	33.33	33.66	29.77
Ours	**35.93**	**37.18**	**38.46**



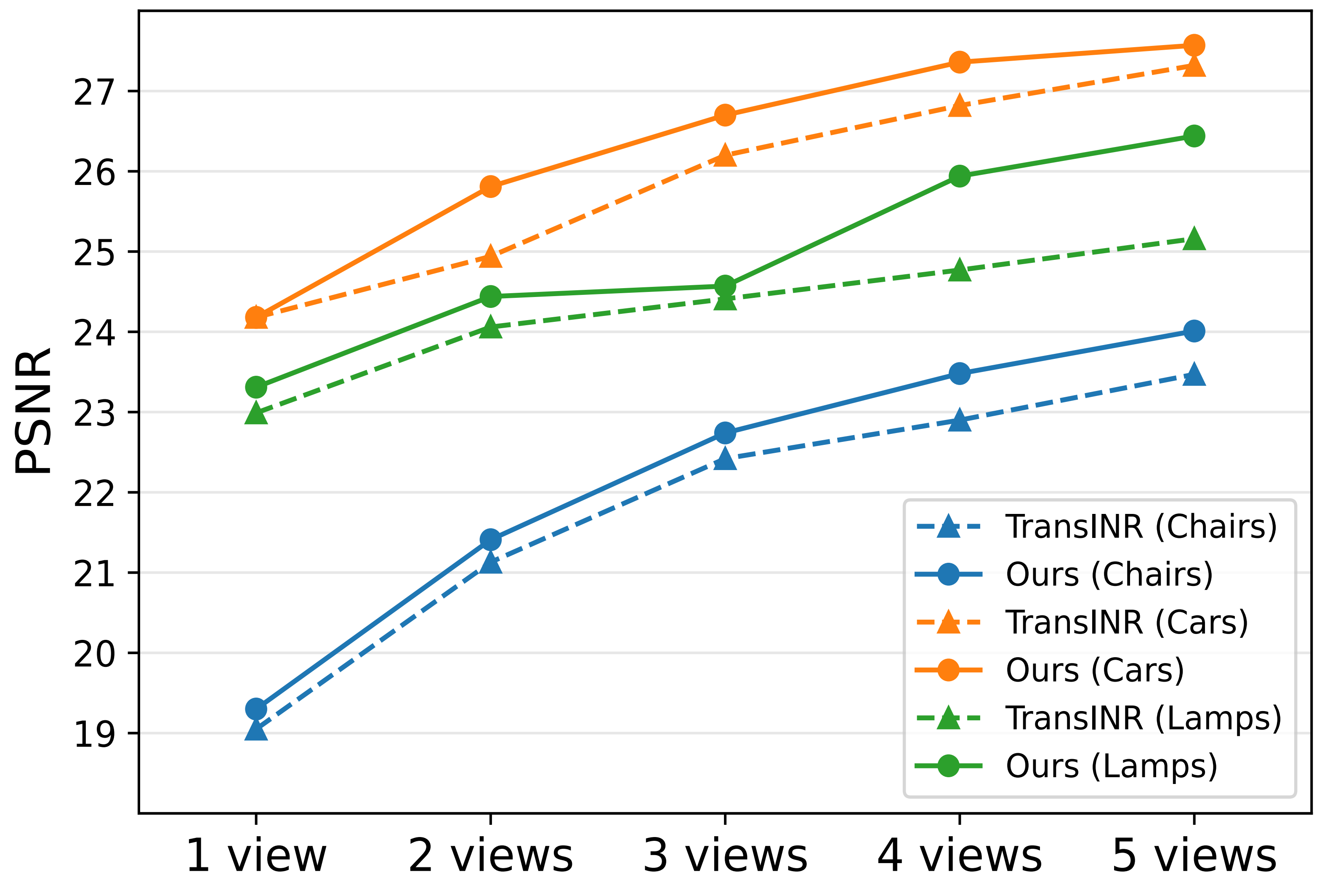
TransINR Ours

Novel View Synthesis

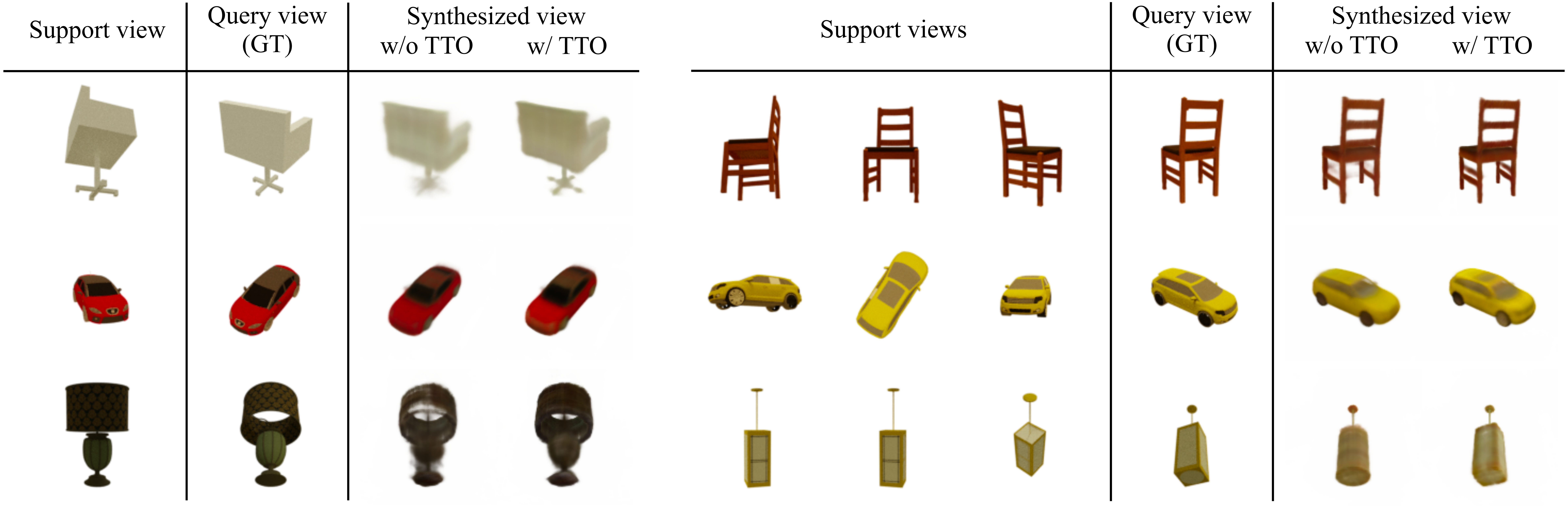
- For novel view synthesis, we use ShapeNet Chairs, Cars, and Lamps.
- Six layer MLPs with 256 hidden dimensions are used to estimate neural field of 3D objects.
- Transformer takes few views of an object to **predict Instance Pattern Composers for neural field.**
- We use simple volumetric rendering, since we focus on validating the efficacy of our modulation.

Table 4. Performace comparison of generalizable INRs on novel view synthesis from a single support view.

	Chairs	Cars	Lamps
Matched Init [27]	16.30	22.39	20.79
Shuffled Init [27]	10.76	11.30	13.88
Learned Init [27]	18.85	22.80	22.35
TransINR	19.05	**24.18**	22.89
Ours	**19.30**	**24.18**	**23.41**



Novel View Synthesis (Cont’d)



(a) With one support view.

(b) With three support views.

Effects of the Weight Modulation Locations

- When the location of weigh matrix moves into the output layer, the performance of INRs deteriorates due to the limited power of pattern composition rule.
- **Modulating early layer** is the key of high performance for generalizable INRs.

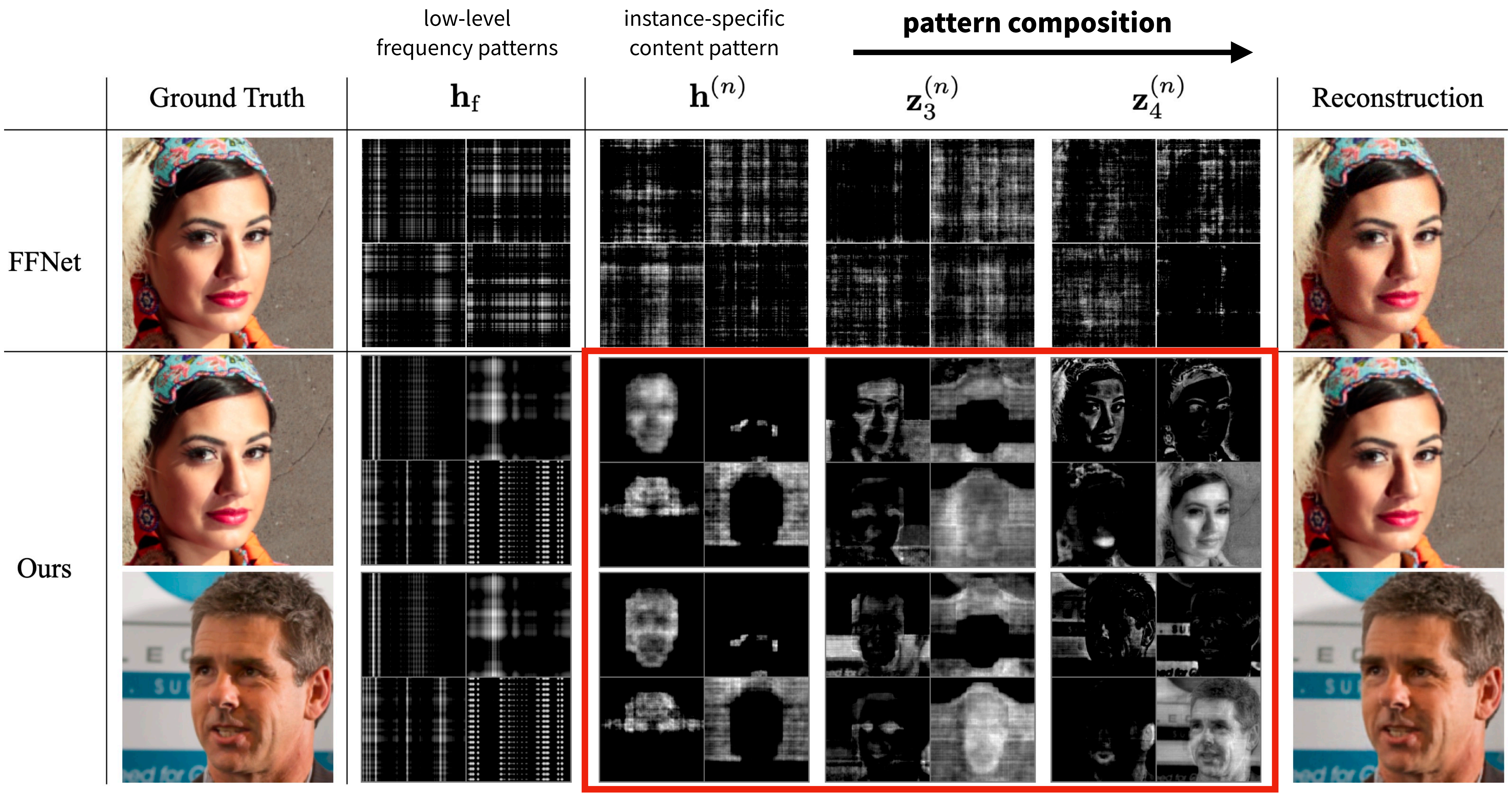
Table 7. PSNRs of our generalizable INRs on image reconstruction according to the location of modulated weights in MLP.

	the modulated layer of MLP				
	1	2	3	4	5
ImageNette	31.00	**35.93**	32.99	31.10	20.26
FFHQ	36.04	**36.20**	34.2	31.09	22.92

too simple frequency features ← → **limited pattern composition rules**

Activation Visualization of INRs

- Our generalizable INRs learns more **interpretable** and **common representations** across instances.



Conclusion

- We have proposed the framework for **generalizable INRs via instance pattern composers**.
- Instance pattern composers **modulate one weight matrix** of the early MLP layer to generalize the learned INRs for unseen data instances.
- **Thanks to the simplicity**, our framework is **compatible with both optimization-based meta-learning and hypernetworks** to significantly improve the performance of generalizable INRs.
- Experimental results demonstrate **the broad impacts of the proposed method on various domains and tasks**, since our generalizable INRs effectively learn **underlying representations across instances.**

Thank You :)