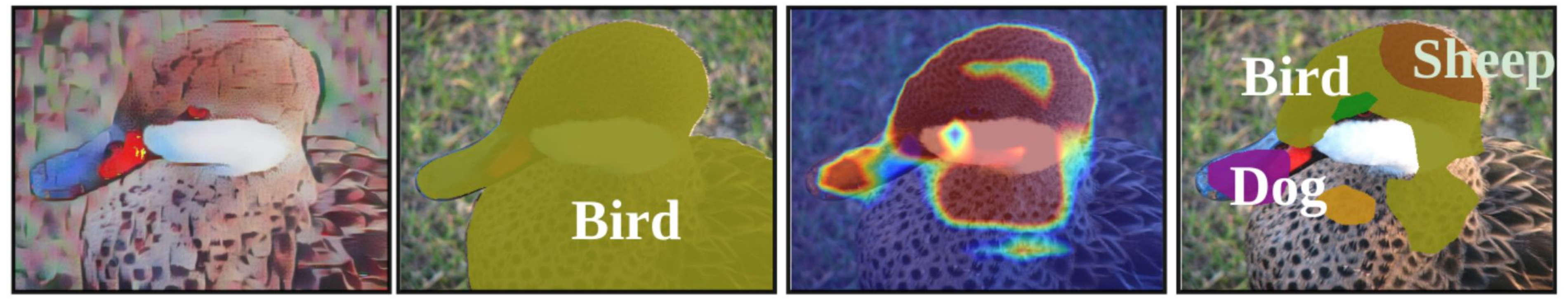
Shape or Texture: Understanding Discriminative Features in CNNs

Md Amirul Islam[1,6], Matthew Kowal[1], Patrick Esser[3], Sen Jia[2], Björn Ommer[3], Konstantinos G. Derpanis[1,5,6], Neil D. B. Bruce[4,6]

[1]Ryerson University, [2]University of Waterloo, [3]Heidelberg University, [4]University of Guelph, [5]Samsung AI Centre Toronto, [6]Vector Institute

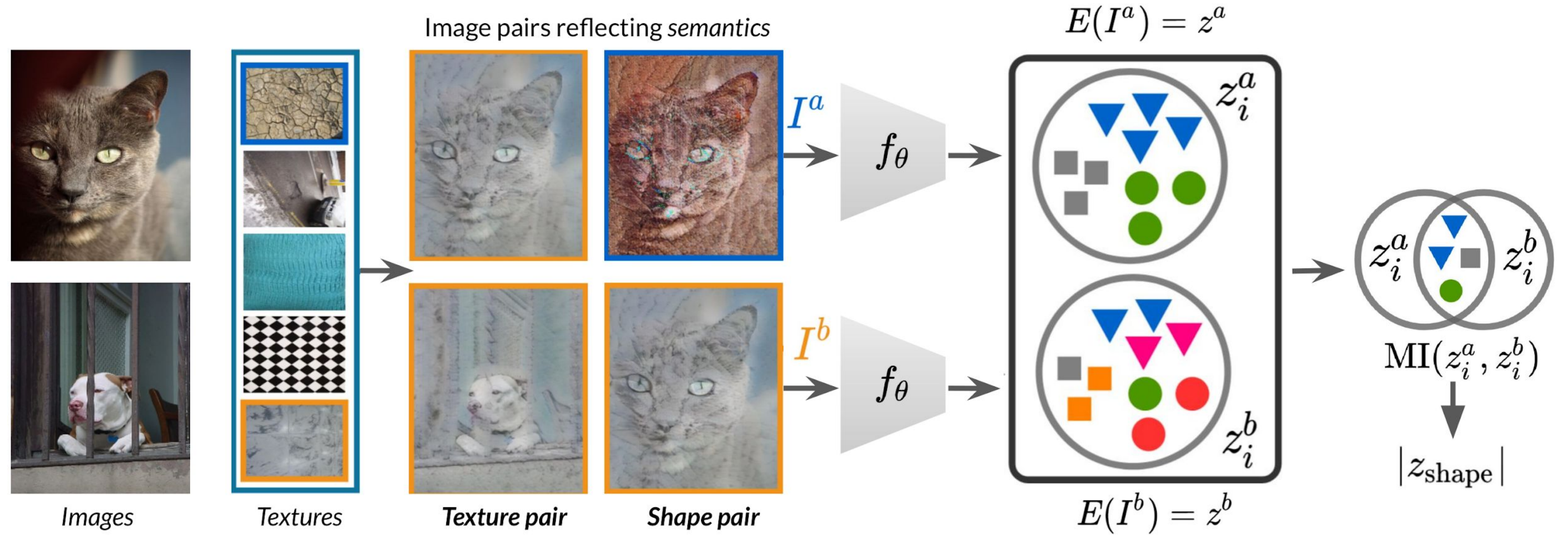
Shape and Texture Bias IN CNNs



Stylized Image | Segmentation GT | Shape | Semantic

- "Shape Bias models make predictions based on object's shape" -> Do they?
- We lack metrics for measuring the amount of *shape* encoded in CNNs.
- We propose the following **two new metrics**:

Estimating Shape and Texture Encoding Neurons (1)



Compute Mutual Information → $\mathrm{MI}(z_i^a, z_i^b) \geq -\frac{1}{2}\log(1-\rho_i^2), \quad \text{where } \rho_i = \frac{\mathrm{Cov}(z_i^a, z_i^b)}{\sqrt{\mathrm{Var}(z_i^a)\,\mathrm{Var}(z_i^b)}}.$

- We calculate the mutual information between the latent representations to produce an estimate of which neurons encode *shape*.

Decoding Per-Pixel Shape Information (2)

- Quantifying per-pixel shape information in the latent representation of CNNs

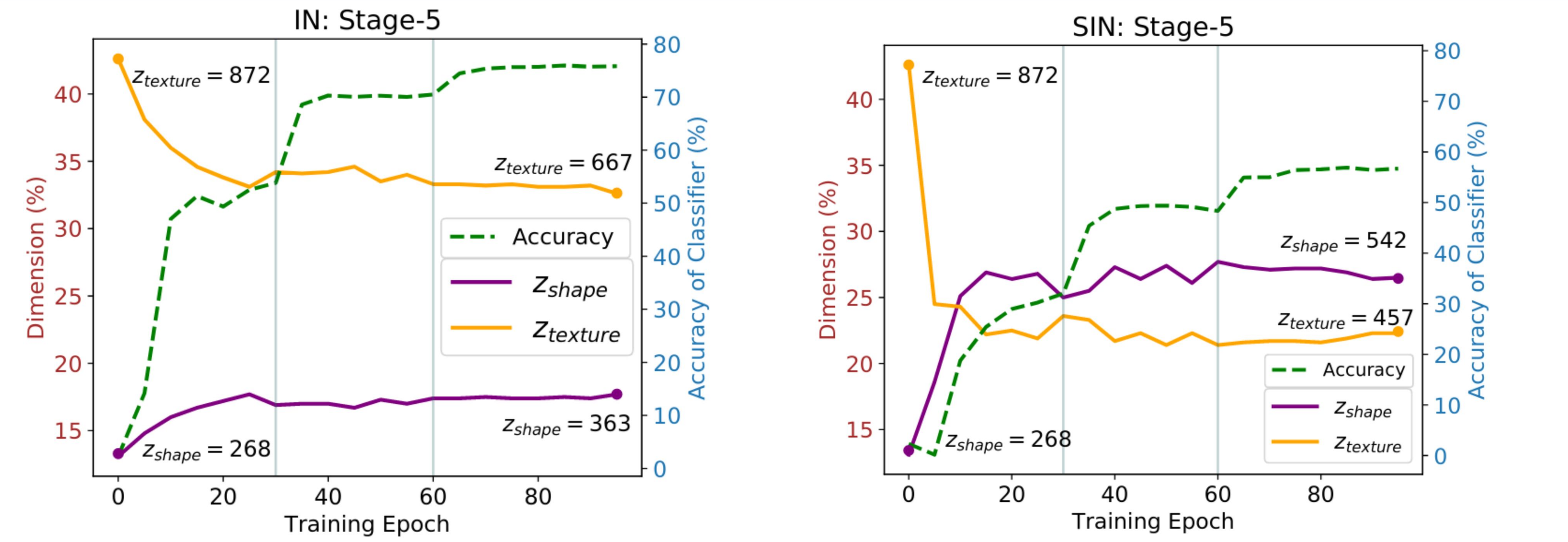
Dimensionality Estimation of Shape and Texture

Model	Shape	Texture
ResNet50	349	692
BagNet33	284	825
BagNet9	276	841

Dataset*	Shape	Texture
ImageNet	349	692
Stylized ImageNet	536	477
IN + SIN	376	640

- BagNets have more neurons encoding *texture* than the *shape*
- *Shape* biased models have more *shape* encoding neurons than traditional ImageNet pretrained model

When Does Shape Become Relevant During Training?



- Shape encoding neurons increase only marginally for IN but grows much larger and faster in case of SIN over the course of training.
- Texture factor decreases as the training progresses.

Decoding Per-pixel Shape from a Pretrained Network

Training Initialization	Shape	Semantic
Random	48.0	6.1
IN-Freeze	80.2	62.7
IN	70.6	50.9

Training Initialization	Shape	Semantic
IN	79.8	61.6
SIN	76.4	53.7
IN+SIN	77.8	58.0

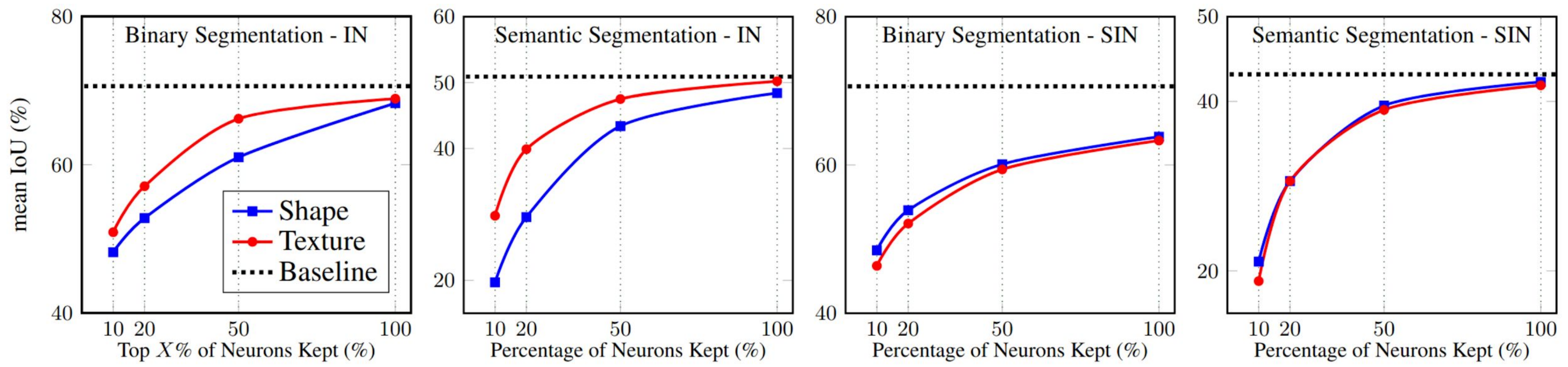
- Measure the amount of decodable *shape* from frozen CNN by training a one layer convolutional readout module.

Where is Shape Information Stored?

		Shape	Texture
Stage2 (f2)	IN	14.1	40.2
	SIN	14.1	42.6
Stage5 (f5)	IN	17.0	33.8
	SIN	26.2	23.3

- Examine if *shape* information is ***equally distributed*** across different stages
- CNNs encode a surprising amount of *shape* information at all stages

Targeting Shape and Texture Neurons



- Validate if the most *texture* or ***shape-specific neurons*** can influence the shape decoding performance when keeping these specific neurons during training.
- Shape biased model is *more reliant on shape neurons* than a texture biased model

Conclusions

- Introduced two new methods for quantifying *shape* information in the latent representation of CNNs in terms of **Neurons** and **Pixels**
- *Shape* is mostly learned during the first part of training
- *Shape* bias models do not encode global object *shape*
- Biasing a CNN towards shape predominantly changes the number of *shape* encoding neurons in the last feature encoding stage.