Shape or Texture: Understanding Discriminative Features in CNNs

Ryerson University

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

UNIVERSITY OF WATERLOO

UNIVERSITY of GUELPH

VECTOR INSTITUTE

SAMSUNG

Motivation

Geirhos et al. (ICLR 2019)

	Texture Image	Shape	Texture-Shape Cue Conflict
ResNet50	**81.4%** *Indian elephant*	**71.1%** *Tabby cat*	**63.9%** *Indian elephant*
pe-Biased ResNet50	-	**53.2%** *Tabby cat*	**50.6%** *Tabby cat*

Motivation

Class Label: Bird

Stylized Image

“Shape biased models make predictions based on the object’s shape”

Motivation

Class Label: Bird

Stylized Image

Segmentation GT

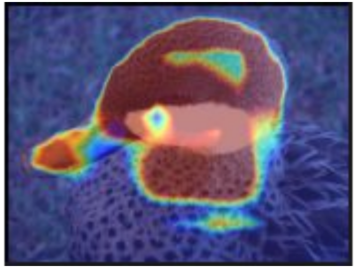
“Shape biased models make predictions based on the object’s shape”

Motivation

Class Label: Bird

Stylized Image

Segmentation GT



‘Shape’

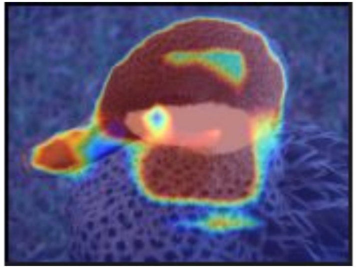
“Shape biased models make predictions based on the object’s shape”

Motivation

Class Label: Bird

Stylized Image

Segmentation GT



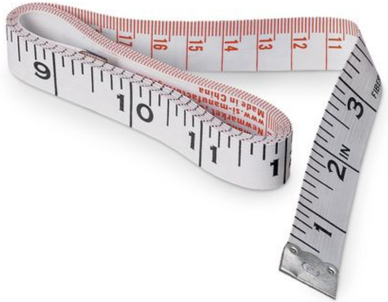
‘Shape’

‘Semantic’

“Shape biased models make predictions based on the object’s shape”

How much shape information do CNNs encode?

Problem:



Lack of metrics for shape and texture information!

Network Dissection

Bau et al. (CVPR 2017)

Based on semantic segmentation ground truth

↓

'Shape' segmentation maps do not exist
Unable to measure shape neurons

Shape Bias

Geirhos et al. (ICLR 2019)

Measures accuracy of Shape-texture cue conflict images

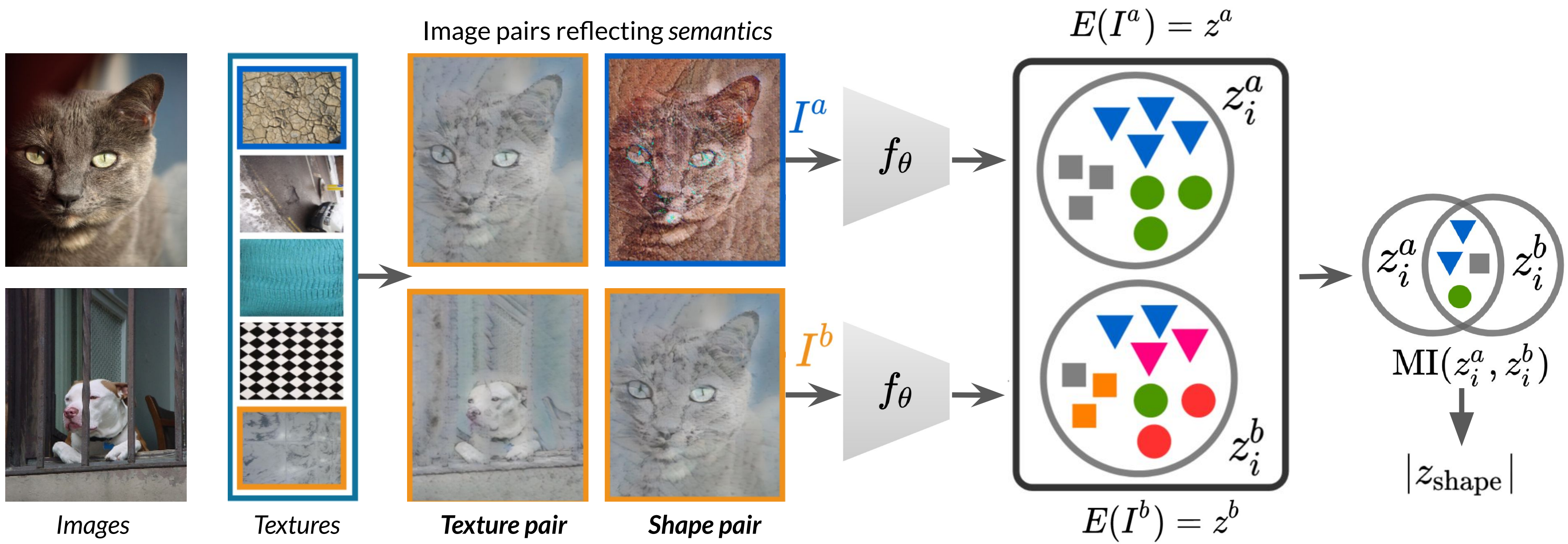
↓

Lacks ability to measure shape on a *per-pixel level*

Our Solution

Two new metrics for measuring shape information

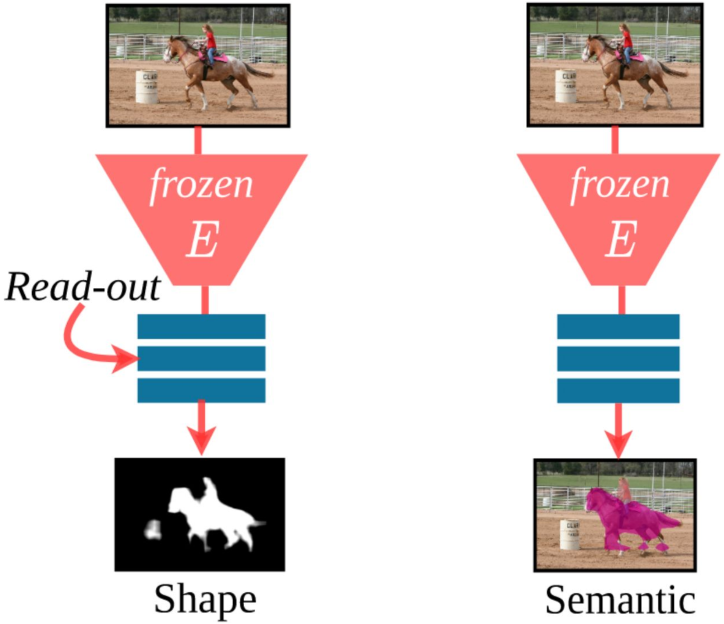
Method 1 - Estimating the number of shape encoding neurons



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

(Kraskov et al., 2004; Foster & Grassberger, 2011)

Method 2 - Decoding Per-Pixel Shape from A Pre-Trained Network



Estimating Shape & Texture Dimensionality

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

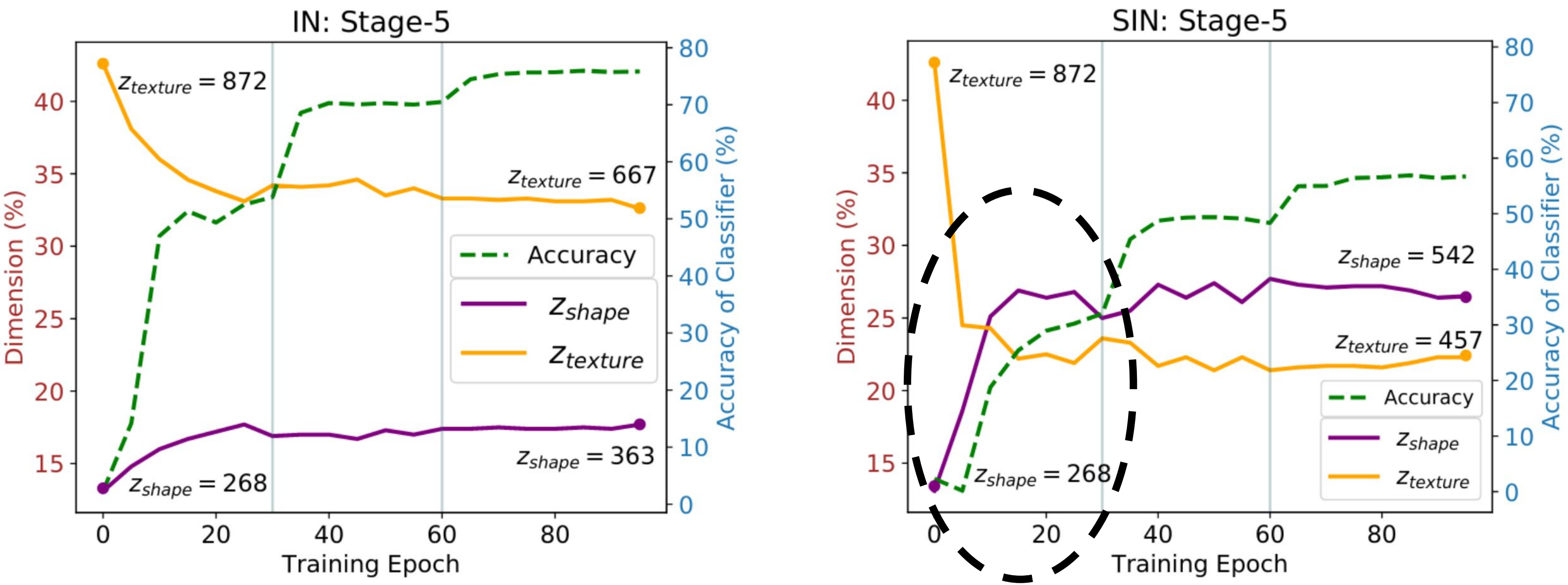
All models trained on ImageNet

Estimating Shape & Texture Dimensionality

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

*Models trained on datasets are ResNet50s**

When Does Shape Become Relevant During Training?



Quantifying Shape Information in CNN Latent Representation

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

Quantifying Shape Information in CNN Latent Representation

Training Initialization	*Bin.*	*Semantic*
IN	79.8	61.6
SIN	76.4	53.7
IN+SIN	77.8	58.0

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

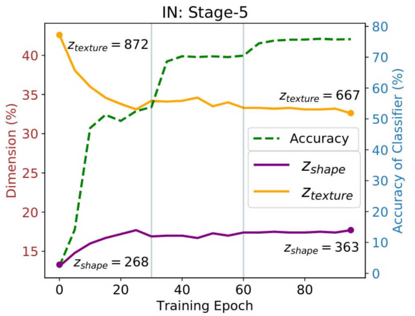
Where is Shape Information Stored?

		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

Conclusion

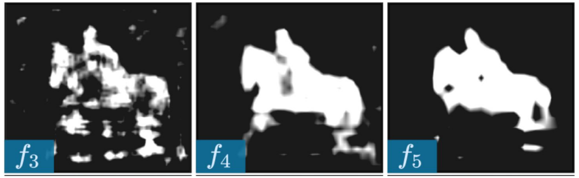
Introduced **two** new methods for quantifying shape information in the latent representations of neural networks in terms of **Neurons and Pixels**

1) Shape is learned early during training



2) Shape-Biased models do not encode global object shape

3) Shape-Bias -> effects last feature encoding stage



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]Department of Computer Science, Ryerson University, Canada
[2]University of Waterloo, Canada
[3]IWR, HCI, Heidelberg University, Germany
[4]School of Computer Science, University of Guelph, Canada
[5]Samsung AI Centre Toronto, Canada
[6]Vector Institute for AI, Canada

Poster Session 5, May 04