The Origins and Prevalence of Texture Bias in Convolutional Neural Networks

Katherine L. Hermann, Ting Chen, Simon Kornblith

Stanford University, Google Research

NeurIPS 2020

Dec 15, 2020

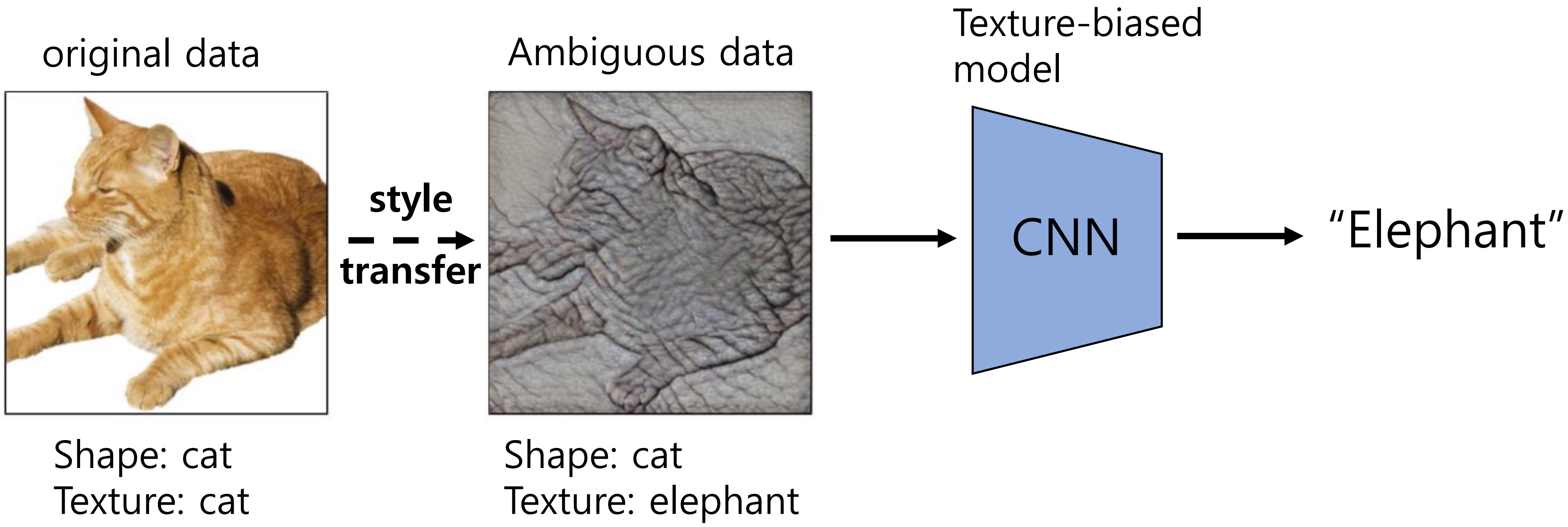
Presented by EungyeupKim

Vision Study

Backgrounds Textual Bias

Textural Bias

- Convolutional neural networks appear to make classifications based on superficial **textural** features, rather than shape information preferentially used by humans.



Motivations

Textual Bias

1. It may be related to the vulnerability of CNNs to adversarial examples.
2. Preference of texture by CNN indicates an inductive bias, different from that of humans.
3. It is difficult to generalize to different distributions than those on which the model is trained.

$\Rightarrow$ What is the most important factor for the texture bias in ImageNet-trained CNNs?

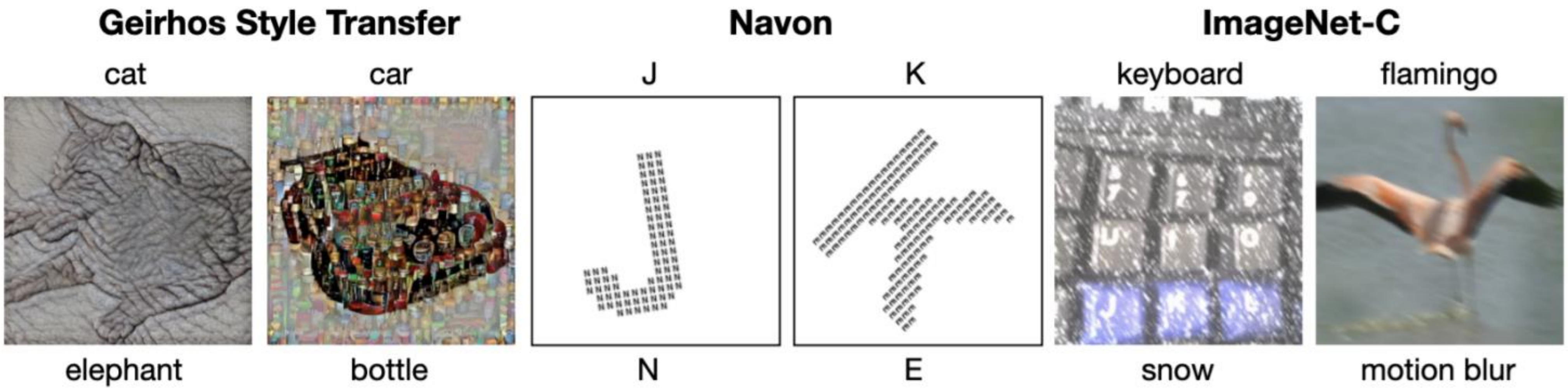
This paper

- Demonstrates that dataset is the most important factor for the texture bias.
- Investigates the relationship between bias and several factors including augmentation techniques, architectures and objective functions.

Datasets

GST, Navon, ImageNet-C

Ambiguous Datasets



- Each image contain independent set of shape and texture labels.
- GST contains both contents (shape) and textures transferred from the real photos (texture).
- Navon consists of a large letter (shape) rendered in small copies of another letter (texture).
- ImageNet-C contains original images (shape) corrupted with noise of texture.

Evaluations

Shape bias, texture bias, shape match, texture match

Evaluation metrics

- Test sets are sampled from the GST.

- Shape Bias
  - The percentage of classification according to shape, provided the model classified either shape or texture correctly.
  - We call a model *shape-biased* if its shape bias is >50%, and *texture-biased* if <50%.

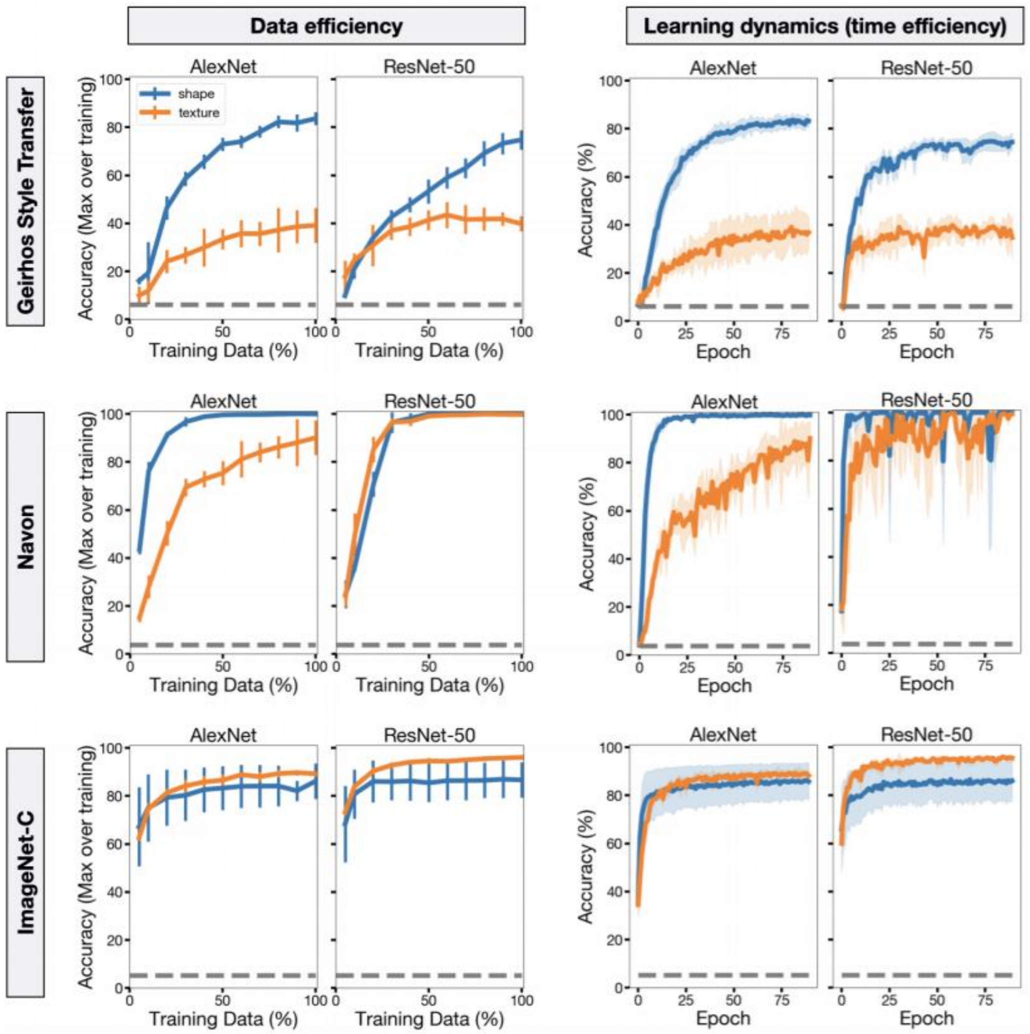
- Shape Match
  - The percentage of correct classification according to shape.

- Texture Match
  - The percentage of correct classification according to texture.

Experiments

CNNs can learn shape as easily as texture

Is the texture-driven classification attributed to the model, or the data?



- AlexNet and ResNet-50 are trained to classify the shape or texture label in the test set, using 5%, 10%, 20%,…, and 100% of the training set.

- As shown in the figure, the models learn to classify ambiguous images by their shape as much as their texture.

- Therefore, the source of texture bias seems to lie in training on ImageNet, rather than on CNN's inductive biases.

Experiments
The role of data augmentation in texture bias

Random-crop augmentation.

Model	Shape Bias		Shape Match		Texture Match		ImageNet Top-1 Acc.	
	Random	Center	Random	Center	Random	Center	Random	Center
AlexNet	28.2%	**37.5%**	16.4%	**19.3%**	**41.8%**	32.1%	**56.4%**	50.7%
VGG16	11.2%	**15.8%**	7.6%	**10.7%**	**60.1%**	57.1%	**71.8%**	62.5%
ResNet-50	19.5%	**28.4%**	11.7%	**16.3%**	**48.4%**	41.1%	**76.6%**	70.7%
Inception-ResNet v2	23.1%	**27.9%**	15.1%	**19.8%**	50.2%	**51.2%**	**80.3%**	77.3%

- Cropping the data points renders the shape of them less reliable feature compared to the texture, resulting in higher texture-biased representations in the training.
- Naturalistic appearance augmentations decrease texture bias relative to baseline.
- However, there is a trade-off between ImageNet-top1 accuracy and shape-bias.

Experiments
The role of data augmentation in texture bias

Appearance augmentation.

Augmentation	Shape Bias	Shape Match	Texture Match	ImageNet Top-1 Acc.
Baseline	19.5%	11.7%	48.4%	76.6%
Rotate 90°, 180°, 270°	19.4%	10.8%	45.1%	75.7%
Cutout	**21.4%**	12.3%	45.2%	76.9%
Sobel filtering	**24.8%**	12.8%	38.9%	71.2%
Gaussian blur	**25.2%**	14.1%	41.7%	75.8%
Color distort.	**25.8%**	15.3%	44.2%	76.9%
Gaussian noise	**30.7%**	17.2%	38.8%	75.6%

- Naturalistic appearance augmentations decrease texture bias relative to baseline.
- Accumulations of appearance augmentations further enhance the shape bias.

Augmentation(s)	Shape Bias	Shape Match	Texture Match	ImageNet		IN-Sketch		SIN	
				top-1	top-5	top-1	top-5	top-1	top-5
Baseline	19.5%	11.7%	**48.4%**	76.6%	**93.3%**	22.4%	39.3%	7.7%	17.0%
+ Color distortion	25.8%	15.3%	44.2%	**76.9%**	**93.3%**	28.1%	46.6%	9.9%	20.5%
+ Gaussian blur	30.7%	17.2%	38.8%	76.8%	**93.3%**	29.0%	47.9%	11.1%	21.9%
+ Gaussian noise	36.1%	20.1%	35.5%	75.9%	92.8%	29.8%	48.9%	12.6%	24.3%
+ Min. crop of 64%	48.7%	29.1%	30.7%	73.5%	91.5%	**30.9%**	**51.4%**	14.5%	28.2%
+ Stronger aug.	55.2%	33.3%	27.1%	72.0%	90.7%	30.4%	50.5%	**15.1%**	**28.8%**
+ Longer training	**62.2%**	**38.3%**	23.3%	71.1%	90.0%	30.5%	50.4%	14.9%	28.4%

Experiments

Effect of training objectives

Therefore, texture bias is driven by the joint image-label statistics of the ImageNet dataset.

- To correctly label the many dog breeds in the dataset, for instance, a model would have to make texture-based distinctions between similarly shaped objects.

- To test this, this paper compares the shape bias of supervised models with that of self-supervised models.

- For the self-supervised methods, *Rotation classification*, *Exemplar*, *BigBiGAN*, and *SimCLR* are utilized.

- A trainable classifier is located on top of the learned representations from those methods using ImageNet dataset.

Experiments
Effect of training objectives

Self-supervised learning methods

Objective	Shape Bias		Shape Match		Texture Match		ImageNet Top-1 Acc.	
	AlexNet	ResNet-50	AlexNet	ResNet-50	AlexNet	ResNet-50	AlexNet	ResNet-50
Supervised	29.8%	21.9%	17.5%	13.5%	41.2%	48.2%	57.0%	75.8%
Rotation	47.0%	32.3%	21.6%	14.2%	24.3%	29.8%	44.8%	44.4%
Exemplar	29.9%	14.4%	12.6%	7.5%	29.5%	44.7%	37.2%	41.8%
BigBiGAN	-	31.9%	-	17.7%	-	37.7%	-	55.4%
SimCLR	-	37.0%	-	17.3%	-	29.4%	-	69.2%
Supervised w/ SimCLR aug.	-	40.4%	-	23.1%	-	34.0%	-	76.3%

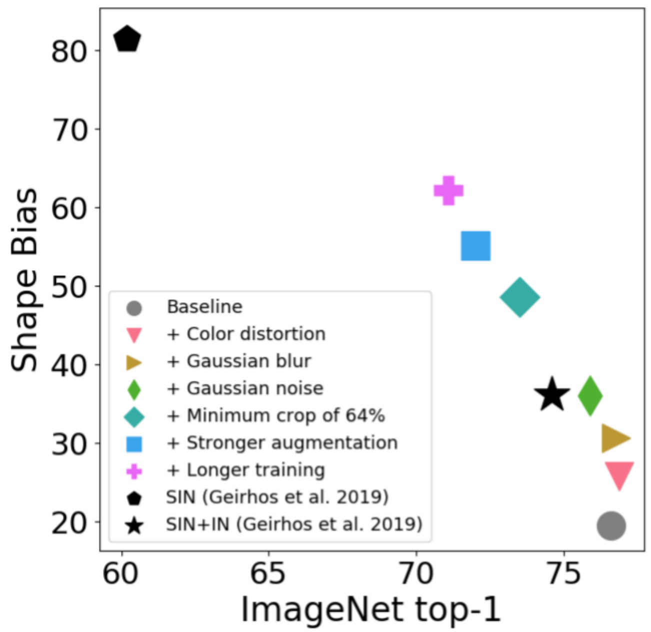
- All the methods except *Exemplar* encourage the model to be less texture-biased compared to the supervised manner.

- However, supervised method with the same augmentations as *SimCLR* demonstrates the similar texture-bias as others, indicating the importance of data augmentation in reducing texture bias.

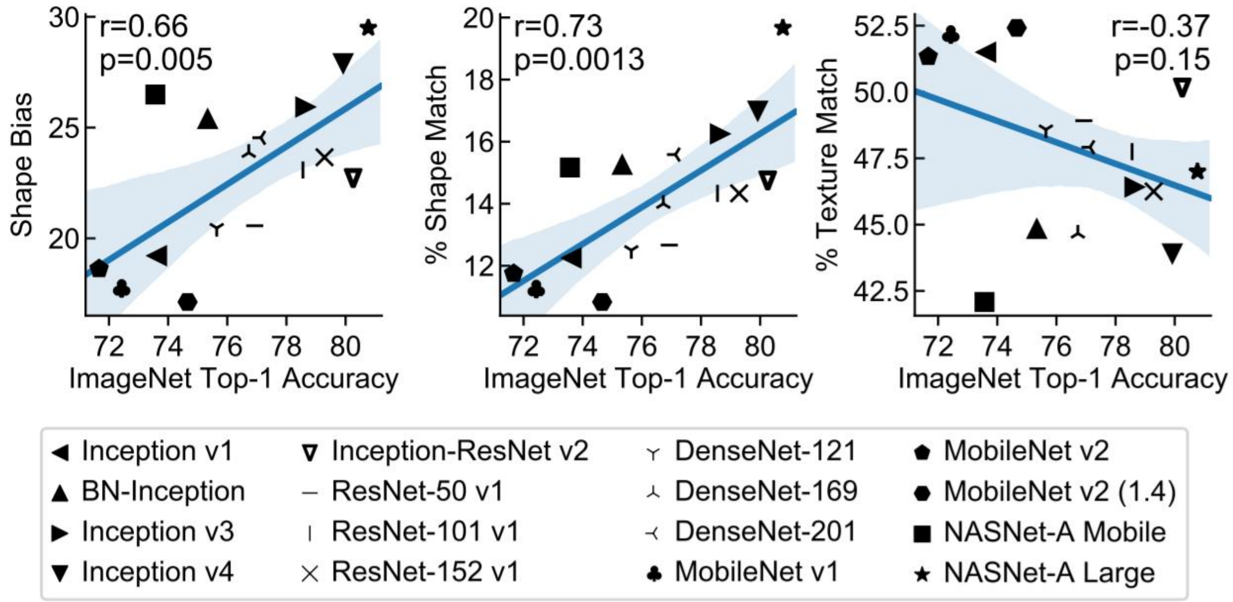
Experiments

Effect of shape bias in ImageNet top1 accuracy

Trade-off between Top1 accuracy and shape bias



Correlation between Top1 accuracy and shape bias

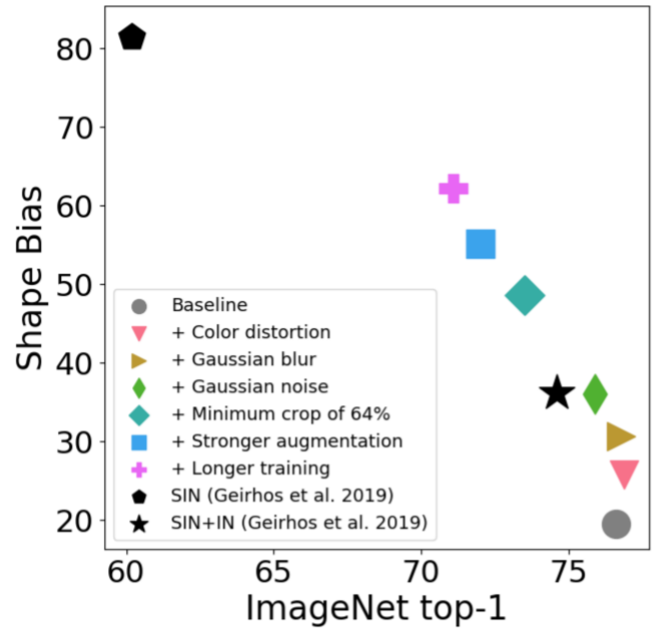


- As the model is trained with ImageNet, increased shape bias tends to harm the top1 accuracy for the same model.
- However, selected high-performing ImageNet models show the positive correlation with the shape-bias, while no significant relationship with the texture-bias.
- This indicates that model architectures with high top1 accuracy are effective at extracting shape information.

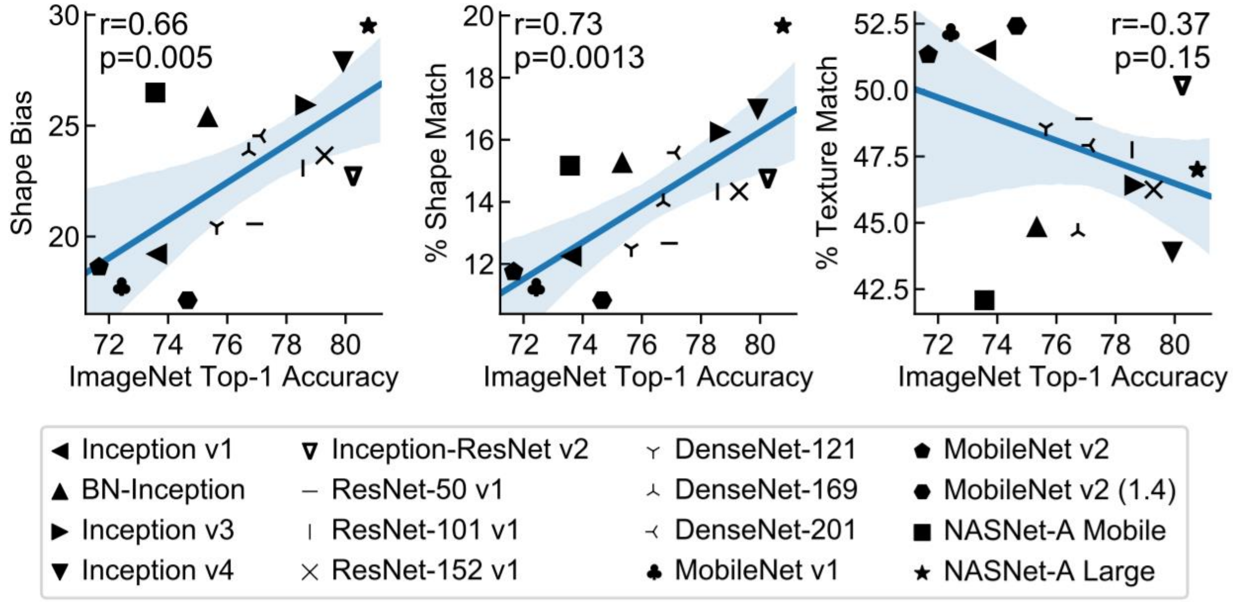
Experiments

Effect of shape bias in ImageNet top1 accuracy

Trade-off between Top1 accuracy and shape bias



Correlation between Top1 accuracy and shape bias



- As the model is trained with ImageNet, increased shape bias tends to harm the top1 accuracy for the same model.
- However, selected high-performing ImageNet models show the positive correlation with the shape-bias, while no significant relationship with the texture-bias.
- This indicates that model architectures with high top1 accuracy are effective at extracting shape information.