ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness

Каратаева Екатерина
Исаев Сергей
Сусла Диана

ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness

- Исследуется особенности восприятия сверточных сетей, обученных на ImageNet
- Существует две гипотезы: shape hypothesis и texture hypothesis
- CNN, обученные на ImageNet, склонны давать ответ на основе текстуры объекта, а не его формы или очертания

Texture vs shape bias in humans and ImageNet-trained CNNs

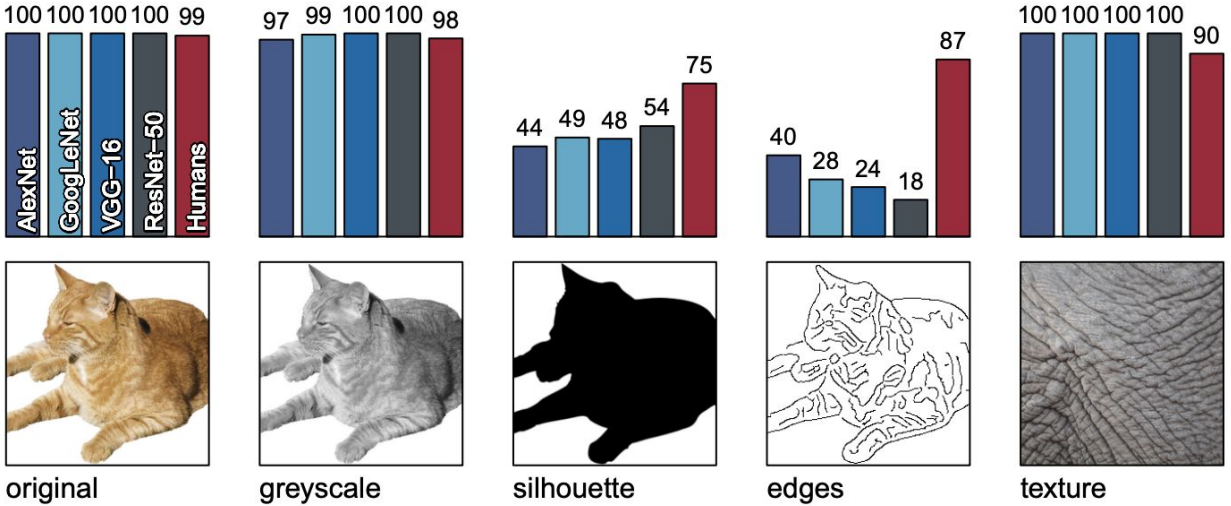
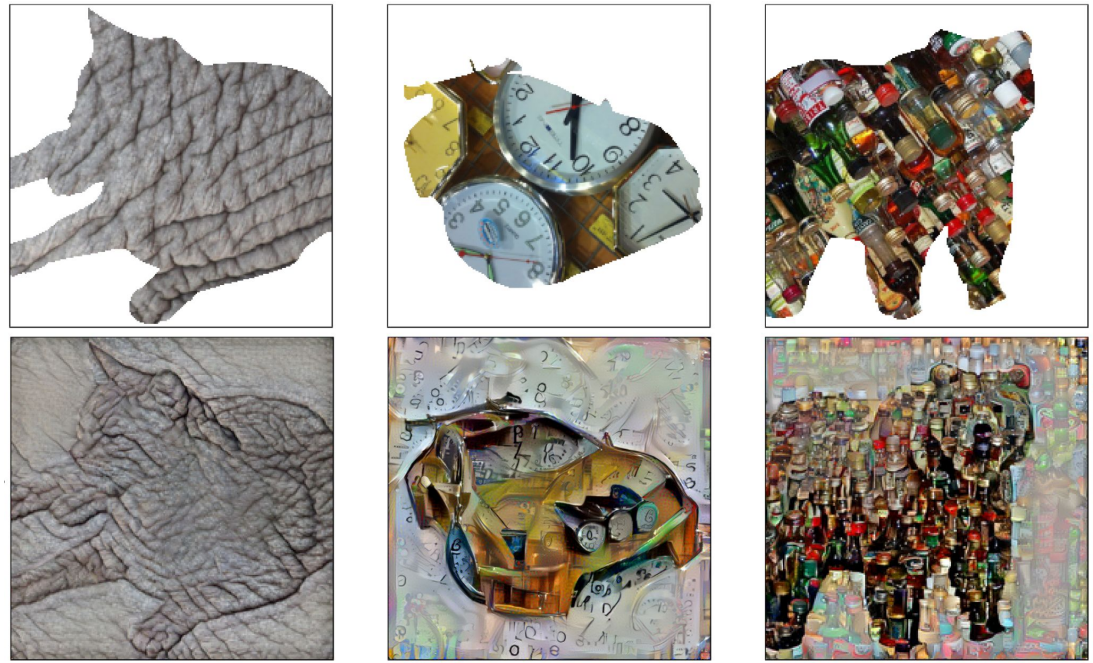
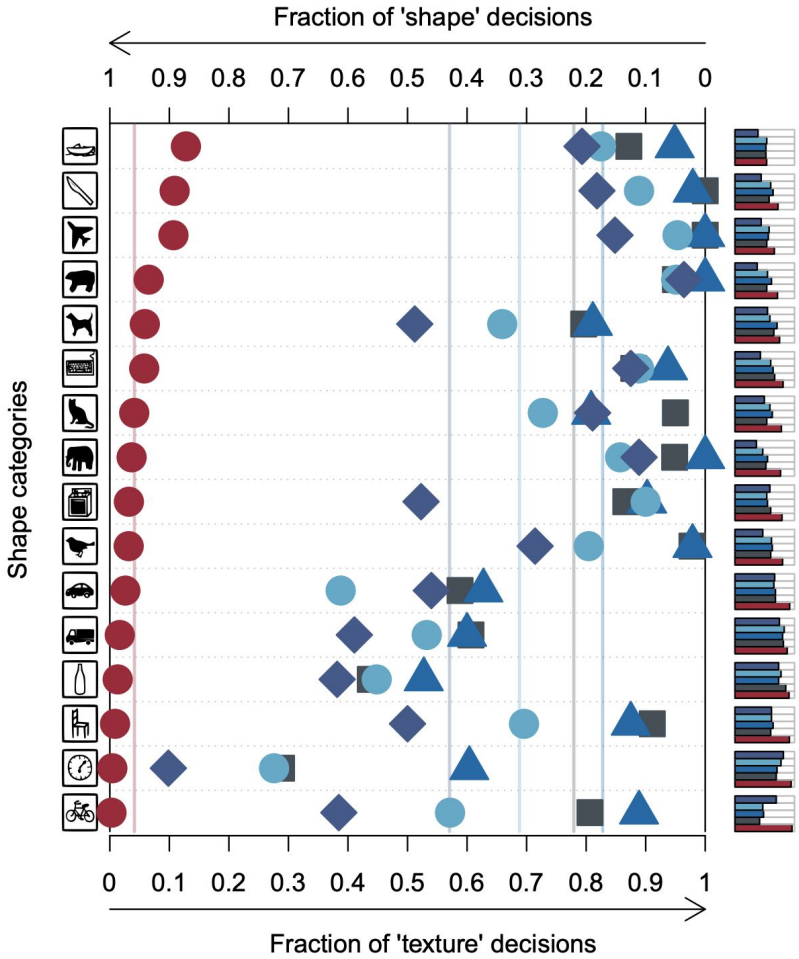


Figure 2: Accuracies and example stimuli for five different experiments without cue conflict.




Fraction of 'shape' decisions
1 0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2 0.1 0
Shape categories
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Fraction of 'texture' decisions

Overcoming the texture bias of CNNs

Figure 3: Visualisation of Stylized-ImageNet (SIN), created by applying AdaIN style transfer to ImageNet images. Left: randomly selected ImageNet image of class `ring-tailed lemur`. Right: ten examples of images with content/shape of left image and style/texture from different paintings. After applying AdaIN style transfer, local texture cues are no longer highly predictive of the target class, while the global shape tends to be retained. Note that within SIN, every source image is stylized only once.

architecture	IN→IN	IN→SIN	SIN→SIN	SIN→IN
ResNet-50	92.9	16.4	79.0	82.6
BagNet-33 (mod. ResNet-50)	86.4	4.2	48.9	53.0
BagNet-17 (mod. ResNet-50)	80.3	2.5	29.3	32.6
BagNet-9 (mod. ResNet-50)	70.0	1.4	10.0	10.9

Table 1: Stylized-ImageNet cannot be solved with texture features alone. Accuracy comparison (in percent; top-5 on validation data set) of a standard ResNet-50 with Bag of Feature networks (BagNets) with restricted receptive field sizes of 33×33, 17×17 and 9×9 pixels. Arrows indicate: train data→test data, e.g. IN→SIN means training on ImageNet and testing on Stylized-ImageNet.

Robustness and accuracy of shape-based representations

name	training	fine-tuning	top-1 IN accuracy (%)	top-5 IN accuracy (%)	Pascal VOC mAP50 (%)
vanilla ResNet	IN	-	76.13	92.86	70.7
	SIN	-	60.18	82.62	70.6
	SIN+IN	-	74.59	92.14	74.0
Shape-ResNet	SIN+IN	IN	**76.72**	**93.28**	**75.1**

Table 2: Accuracy comparison on the ImageNet (IN) validation data set as well as object detection performance (mAP50) on PASCAL VOC 2007. All models have an identical ResNet-50 architecture. Method details reported in the Appendix.

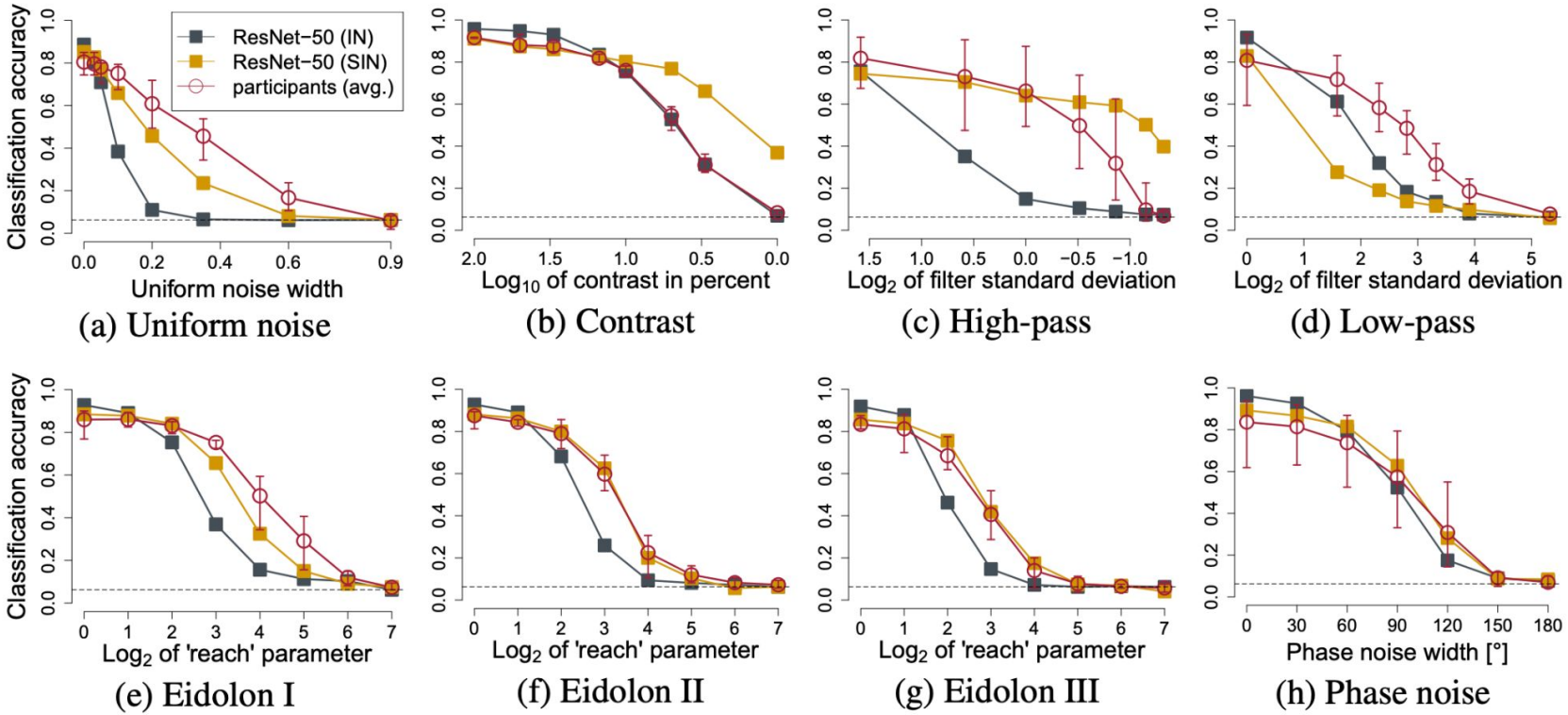


Figure 6: Classification accuracy on parametrically distorted images. ResNet-50 trained on Stylized-ImageNet (SIN) is more robust towards distortions than the same network trained on ImageNet (IN).

			Noise			Blur			
training	ft	**mCE**	Gaussian	Shot	Impulse	Defocus	Glas	Motion	Zoom
IN (vanilla ResNet-50)	-	76.7	79.8	81.6	82.6	74.7	88.6	78.0	79.9
SIN	-	77.3	71.2	73.3	72.1	88.8	85.0	79.7	90.9
SIN+IN	-	**69.3**	**66.2**	**66.8**	**68.1**	**69.6**	**81.9**	**69.4**	80.5
SIN+IN	IN	73.8	75.9	77.0	77.5	71.7	86.0	74.0	**79.7**

		Weather				Digital			
training	ft	Snow	Frost	Fog	Brightness	Contrast	Elastic	Pixelate	JPEG
IN (vanilla ResNet-50)	-	77.8	74.8	66.1	56.6	71.4	84.8	76.9	76.8
SIN	-	71.8	74.4	66.0	79.0	**63.6**	81.1	72.9	89.3
SIN+IN	-	**68.0**	**70.6**	**64.7**	57.8	66.4	**78.2**	**61.9**	**69.7**
SIN+IN	IN	74.5	72.3	66.2	**55.7**	67.6	80.8	75.0	73.2

Table 4: Corruption error (lower=better) on ImageNet-C (Hendrycks & Dietterich, 2019), consisting of different types of noise, blur, weather and digital corruptions. Abbreviations: mCE = mean Corruption Error (average of the 15 individual corruption error values); SIN = Stylized-ImageNet; IN = ImageNet; ft = fine-tuning. Results kindly provided by Dan Hendrycks.

Плюсы статьи

- Статья хорошо написана и легко читается
- Проведено объёмное исследование
- Есть репозиторий для получения стилизованных картинок
- Подробное описание экспериментов позволяет воспроизвести их в точности
- Придумана аугментация и метод обучения, улучшающий качество на ImageNet

Минусы статьи

- Мало новизны. Это скорее структуризация и воспроизведение уже существовавших предположений
- Не нашел подробного описания обучения Shape-ResNet или кода. Возможны проблемы с воспроизводимостью

Оценка

Низкая уверенность из-за того, что не разбираюсь в экспериментах с людьми, поэтому не могу сказать, насколько они были правильно проведены.

Оценка: **7/10**

Уверенность: **3/5**

Контекст

Работа написана в 2018 году. Представлена на ICLR 2019 в виде орал выступления.

Авторы статьи:

- **Robert Geirhos**
- **Patricia Rubisch**
- **Claudio Michaelis**
- **Matthias Bethge**
- **Felix Wichmann**
- **Wieland Brendel**

Все работают или учатся в Тюбингенском университете (Германия).

Авторы статьи

Robert Geirhos

Тюбингенский университет. Международная исследовательская школа интеллектуальных систем Макса Планка. (IMPRS-IS)

Научные интересы:

Understanding CNNs, Deep Learning, Human Vision, Psychophysics, Robustness

h-индекс: 9

Patricia Rubisch

Тюбингенский университет. Аспирантка кафедры вычислительной нейронауки Эдинбургского университета

Научные интересы:
Synaptic Plasticity, Supervised and unsupervised learning for Spiking Neural Networks, Reservoir Computing

h-индекс: 2

Авторы статьи

Claudio Michaelis

Аспирант в Тюбингенском университете.

Научные интересы:

Machine Learning, Computer Vision

h-индекс: 8

Matthias Bethge

Приглашенный профессор в Тюбингенском университете и EPFL (École polytechnique)

Научные интересы:

Computational Neuroscience, Machine Learning, Vision

Соавтор статьи
Image style transfer using convolutional neural networks

h-индекс: 61

Авторы статьи

Felix Wichmann

Тюбингенский университет.

Научные интересы:

Psychophysics, Vision, Visual Perception, Human Vision

В основном занимается психофизикой и психометрическими функциями.

h-индекс: 38

Wieland Brendel

Emmy Noether Group Leader, University of Tübingen

Научные интересы:

Machine Learning, Computer Vision

В основном занимается исследованиями различных методов, влияющих на надежность модели.

h-индекс: 22

Связанные работы

- **Texture and art with deep neural networks.** Leon A Gatys, Alexander S Ecker, Matthias Bethge, 2017

- **On the performance of GoogLeNet and AlexNet applied to sketches.** Pedro L. Ballester, R. M. Araújo 2016

- **Synthesising Dynamic Textures using Convolutional Neural Networks**. Christina M. Funke, Leon A. Gatys, Alexander S. Ecker, Matthias Bethge, 2017

- **Texture Synthesis Using Convolutional Neural Networks.** Leon Gatys, Alexander S. Ecker, Matthias Bethge, 2015

- **Comparing deep neural networks against humans: object recognition when the signal gets weaker.** Robert Geirhos, David H. J. Janssen, Heiko H. Schütt, Jonas Rauber, Matthias Bethge, Felix A. Wichmann, 2017

Цитирования

Всего цитирований 1228. Прямых продолжений работы нет.

- **YOLOv4: Optimal Speed and Accuracy of Object Detection.** Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao, 2020
- **Analyzing and Improving the Image Quality of StyleGAN.** Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, Timo Aila, 2020
- **Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey.** Naveed Akhtar, Ajmal Mian, 2018
- **Adversarial Examples Are Not Bugs, They Are Features.** Andrew Ilyas, Shibani Santurkar, Dimitris Tsipras, Logan Engstrom, Brandon Tran, Aleksander Madry, 2019.

Вклад и практическое применение

- Статья помогает лучше понимать работу нейронных сетей на основе CNN. Сделан вывод о том, что смещение на текстуре можно сдвигать в сторону формы, если обучать модель на подходящем наборе данных.

- Авторы показывают, что ориентированность на структуру в CNN не является свойством архитектуры по умолчанию, а скорее вызвано особенностями обучающих данных.

- В статье был предложен новый набор данных ImageNet, который называется Stylized-ImageNet (SIN), где текстура заменяется случайно выбранным стилем рисования.

- ShapeResNet - это первая сеть, которая приблизилась к надежности на уровне человеческой классификации при искажениях, которые не были частью обучающих данных.

Идеи для исследования

- Провести эксперименты на других наборах данных и с другими, более широкими и глубокими нейронными сетями, чтобы понять, сохранятся ли результаты (например, ResNet-152)

- Использовать знания для создания крутых adversarial моделей