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

(Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A. Wichmann, Wieland Brendel)

1. Общие сведения.
   Статья имеет 2 версии, опубликованные, соответственно, 28 сентября 2018 года и 18 февраля 2019 года, и была представлена на конференции ICLR 2019 7 мая в формате oral. После первой версии статьи вышел ряд рецензий, авторы которых задавались вопросами реализации и выборов методов, которые не были раскрыты в статье. Авторы статьи подробно ответили на все вопросы и опубликовали новую версию, имеющую впоследствии положительные рецензии и рекомендованную к устному выступлению на конференции.

2. Авторы статьи.
   Все авторы из Тюбингенского университета (University of Tubingen), трое из которых (Matthias Bethge, Felix A. Wichmann, Wieland Brendel) являются senior авторами. Matthias Bethge является руководителем Bethge Lab, в которой также состоят Robert Geirhos, Claudio Michaelis, Wieland Brendel. Felix A. Wichmann является руководителем Wichmann Lab, в которой состоит Robert Geirhos. Все авторы в той или иной степени уже работали вместе и имеют общие публикации. Для троих из них (Robert Geirhos, Matthias Bethge, Felix A. Wichmann) данная работа является логическим продолжением их предыдущей деятельности.

   Robert Geirhos в своих работах стремится ответить на вопрос “Почему глубокие нейронные сети видят мир так, как видят?”, и данная работа гармонично вписывается в историю его публикаций, которая раскрывает данный вопрос с разных сторон:
   - Wichmann, F. A., Janssen, D. H., Geirhos, R., Aguilar, G., Schütt, H. H., Maertens, M., & Bethge, M. (2017). Methods and measurements to compare men against machines. Electronic Imaging, Human Vision and Electronic Imaging, 2017(14), 36–45.
   - Geirhos, R., Janssen, D. H., Schütt, H. H., Rauber, J., Bethge, M., & Wichmann, F. A. (2017). Comparing deep neural networks against humans: object recognition when the signal gets weaker. arXiv preprint arXiv:1706.06969.
   - Geirhos, R., Medina Temme, C. R., Rauber, J., Schütt, H. H., Bethge, M., & Wichmann, F. A. (2018). Generalisation in humans and deep neural networks. Advances in Neural Information Processing Systems 31 (pp. 7548–7560).

Matthias Bethge занимался физикой в бакалавриате, и, начав работать с моделями глубинного обучения, тоже стремится понять их поведение в сравнении с человеком, что видно в ряде его публикаций (является соавтором во всех статьях выше). Также Matthias Bethge исследует текстуры и style transfer, что напрямую повлияло на методы, использованные в данной работе:

- L. A. Gatys, A. S. Ecker, and M. Bethge. Texture Synthesis Using Convolutional Neural Networks. Advances in Neural Information Processing Systems 28, 2015
- L. A. Gatys, A. S. Ecker, and M. Bethge. Image Style Transfer Using Convolutional Neural Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016
- I. Ustyuzhaninov, W. Brendel, L. Gatys, and M. Bethge. What does it take to generate natural textures? International Conference on Learning Representations, 2017
- L. A. Gatys, A. S. Ecker, and M. Bethge. Texture and art with deep neural networks. Current Opinion in Neurobiology, 46, 178-186, 2017

Felix A. Wichmann занимается исследованиями на стыке глубинного обучения и психофизики. Среди его публикаций также прослеживается цель объяснить глубокие нейронные сети, и есть работы на тему текстур (является соавтором в ряде статей выше).

3. Источники вдохновения.
   Наибольшее влияние на работу оказали предыдущие статьи одного из авторов (Matthias Bethge) на тему текстур:
   - Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Texture and art with deep neural networks. Current Opinion in Neurobiology, 46:178–186, 2017.
   - Wieland Brendel and Matthias Bethge. Approximating CNNs with bag-of-local-features models works surprisingly well on ImageNet. In International Conference on Learning Representations, 2019.

   Обе данные работы несут в себе следующую мысль: “CNN все еще могут идеально классифицировать изображение с текстурой, даже если глобальная форма была полностью уничтожена”. Именно она легла в основу данной работы в противовес противоположному мнению о том, что CNN имеют shape bias:
   - Jonas Kubilius, Stefania Bracci, and Hans P Op de Beeck. Deep neural networks as a computational model for human shape sensitivity. PLoS Computational Biology, 12(4):e1004896, 2016.

     “CNN неявно учат представления формы так же, как это делают люди”

- Samuel Ritter, David GT Barrett, Adam Santoro, and Matt M Botvinick. Cognitive psychology for deep neural networks: A shape bias case study. arXiv preprint arXiv:1706.08606, 2017.

  “CNN имеют shape bias как дети: для них форма важнее цвета”

4. Цитирования.
   На данный момент работа имеет 1127 цитирований и является базовой на тему texture bias (цитируется на данную тему). Также данную работу цитируют из-за созданного авторами набора данных StylizedImageNet: к нему либо обращаются напрямую, либо говорят о нем как о способе аугментации.

   Прямым продолжением темы данной работы можно назвать следующую статью:

   Hermann, K.L., Kornblith, S.: Exploring the origins and prevalence of texture bias in convolutional neural networks. arXiv preprint arXiv:1911.09071 (2019)

   В ней говорится о том, что для избавления от texture bias необязательно использовать нереалистичный style transfer как в данной работе, а достаточно выбрать подходящие стандартные аугментации: использование изменения цвета, яркости, шума, размытости будет приводить к shape bias, а поворотов, сдвигов, вырезания рандомного куска изображения - к texture bias.

   Прямые конкуренты данной работы не были найдены.

5. Дополнительные исследования.
   Авторы статьи используют рандомные текстуры для формирования набора данных. Можно ли улучшить качество, подбирая для каждого изображения текстуры, наиболее ухудшающие качество (использовать своего сорта adversarial атаку)?

6. Применение в индустриальных приложениях.
   Распознавание человека с реалистичным гримом/в реалистичном костюме (например, животного).