

Efficient Visual Pretraining with Contrastive Detection, Olivier J. Hénaff, Skanda Koppula, Jean-Baptiste Alayrac, Aaron van den Oord, Oriol Vinyals, João Carreira

Автор исследования: Булатова Катя

1. Когда написана работа? Опубликована ли она на какой-то конференции? Если да, то как она была представлена на конференции? Есть ли у нее какая-то история публикации?

Первая версия была опубликована на архиве 19 марта 2021, вторая и последняя версия – 5 августа того же года.

Конференция: ICCV2021, Day 2 (2021/10/13).

В той конференции оба первых места (Marr Prize и Best Student Paper) оказались у Microsoft. Marr Prize получил Swin Transformer.

2. Кто авторы статьи? Где и над чем они работают? Как другие проекты авторов связаны с этой статьей? Основываясь на тексте статьи и истории публикаций авторов, подумайте как авторы пришли к идее статьи – может быть это прямое улучшение их предыдущей работы, а может это больше выглядит как случайная находка.

1. Olivier J Henaff: UK, работает в DeepMind с 2018 года. Получил PhD в NYU, до этого учил математику и физику в Lycée Sainte Geneviève и École Polytechnique. В основном изучает теорию ИИ, биологические процессы. В DeepMind его обязанностью является исследование self-supervised алгоритмов, работа над распознаванием (data-efficient recognition) и пониманием сцен (scene understanding). Первый автор Data-efficient image recognition with contrastive predictive coding, по сути статьи-предшественника; соавтор Divide and Contrast: Self-supervised Learning from Uncurated Data. 4 статьи за 2021, 2 из них про комбинацию

contrastive learning и self-supervised методов, 1 про создание “универсальной модели” (модификация perceiver), и 1 про анализ последовательностей кадров в видео.

2. Skanda Koppula: UK, DeepMind, BS в Computer Science в MIT, выпуск 2016. Занимается в основном теорией машинного обучения, безопасностью, архитектурой компьютера, CV, SR. Работал в Google, NVidia, Yahoo и Square. Его прошлые работы не так сильно связаны конкретно с темой нашей статьи, видно, что именно первый автор задавал ее тему. Его статьи больше про теорию машинного обучения, оптимизацию, например, по памяти или повышению эффективности операций. 3 статьи в 2021, 1 – про perceiver вместе с первым автором, 1 про выявление стадии слияния галактик (просто специальная regression model), и наша. Любит рисовать, даже хранит фотографии картин импрессионистов на гитхабе. У него есть канал на ютубе с 6 подписчиками; его видео все (все 3) про его исследования, также есть плейлист с классической музыкой с концертов.
3. Jean-Baptiste Alayrac: UK, DeepMind. Выпустился из Ecole polytechnique (там же получил PhD) и Telecom ParisTech в 2015, получив магистра в CV. Работал в DeepMind, Oxford University и MIT. Увлекается пешим туризмом и альпинизмом. Основные области интересов – nlp и обработка видео. 9 статей в 2021, много про трансформеры и аудио, тоже работал над продолжением perceiver. Женился на Pauline Métivier 6 мая 2018 года.
4. Aäron van den Oord: UK, DeepMind. Первый автор WaveNet. Получил PhD в Гентском университете. В основном интересуется генеративными моделями наподобие PixelCNN и WaveNet, также занимается Self-Supervised Learning. Является первым автором Representation Learning with Contrastive Predictive Coding, статьи, которая, вероятно, повлияла на нашу. 18 статей за 2021 год, много про self-supervision, генерацию (текста, аудио, видео), обработку последовательностей и CV.
5. Oriol Vinyals: Team Lead, DeepMind. Получил PhD в Беркли. Один из соавторов тензорфлоу, seq2seq и knowledge distillation. Его ИИ победил чемпиона в StarCraft. Всего у него 137,928 цитирований, по google scholar кажется, что нет тем, которыми бы он не занимался. Также внес большой вклад в биоинформатику. 17 статей за 2021 год, много про языковые

модели и про биоинформатику, так же немало про графы. Статей, напрямую связанных с нашей, за этот год нет. Тоже занимался Perceiver.

6. Joao Carreira: DeepMind. Получил степень Doctor Of Philosophy в Боннском университете, но сейчас занимается Computer Science. Занимается сегментацией, распознаванием объектов и локализацией, что уже ближе к теме статьи. 7 публикаций за 2021 год, в основном про видео и аудио, статей, близких к нашей теме, нет. Тоже занимался Perceiver.

Судя по всему, эта работа была предложена и в первую очередь проделана первым автором, так как связана напрямую с его прошлыми работами. Так же у него есть 2 молодых соавтора; один советчик (6) и два важных человека из той же компании с богатым опытом в машинном обучении и большим количеством цитирований. Все авторы писали с какими-то другими авторами хотя бы одну статью, но вообще пересечений очень много. Например, статья про модификацию модели Perceiver пересекается по 5 авторам. Это в первую очередь связано с тем, что они все из DeepMind.

3. Какие из статей в списке ссылок оказали наибольшее влияние на данную работу? Можно ли выделить какие-то 1-3 статьи, которые можно назвать базовыми для этой работы? Опишите в чем связь с этими работами.

Среди источников есть:

1. Базовые статьи (по типу transfer learning – это важно);
2. Много ссылок именно на transfer learning (например, какие есть self-supervised задачи), в том числе большое количество ссылок именно про contrastive objective (максимизирование похожести положительных примеров & минимизирование негативных);
3. Технические статьи (гиперпараметры для обучения);
4. Остальное.

Выделим 2 статьи, с которыми тесно связана наша:

1. SimCLR и

2. BYOL. Эти две статьи, которые мы уже разбирали в курсе, используются как бейзлайны. Из них берут архитектуру, немного меняя contrastive loss.

4. Кто цитирует данную статью? Есть ли у этой работы прямые продолжения?

На статью есть 21 ссылка. Темы ссылающихся статей: contrastive methods, unsupervised learning, pretraining и representation, что неудивительно. Есть некоторое количество работ, которые совершенствуют ту же технику, например, “Object-Aware Cropping for Self-Supervised Learning” (Shlok Kumar Mishra et al) (contrastive representations, улучшение cropping для обучения на менее систематизированных наборах данных) и “DETReg: Unsupervised Pretraining with Region Priors for Object Detection” (Amir Bar et al) (предобучают специальную сеть на обнаружение объектов).

Ссылаются обычно как на один из pixel-level методов получения представлений с помощью contrastive loss через запятую с несколькими другими похожими методами.

Прямого продолжения нет, скорее наша рассматриваемая статья является продолжением “Data-Efficient Image Recognition with Contrastive Predictive Coding” (Olivier J. Henaff et al) (пересечение в два автора), в которой улучшается Contrastive Predictive Coding (unsupervised задача), и в итоге получается обучать на гораздо меньшем количестве данных.

5. Есть ли у работы прямые конкуренты? Опишите как соотносится данная работа с этими конкурентами.

1. “Unsupervised Semantic Segmentation by Contrasting Object Mask Proposals” (Wouter Van Gansbeke et al): появилась на arXiv на месяц раньше рассматриваемой нами статьи. В статье исследуется та же тема, и так же ставится self-supervised задача с масками, полученными unsupervised методами. Хотя в статье тоже обсуждается transfer learning, акцент делается не на повышении эффективности pretraining, как и в

следующей статье-конкуренте. К тому же, этой статье используются специальные архитектуры для задач COCO.

2. “Self-Supervised Visual Representation Learning from Hierarchical Grouping” (Xiao Zhang, Michael Maire): вышла в конце 2020. Отношу ее к конкурентам, так как, вероятно, авторы нашей статьи уже начали писать ее к этому моменту. Там рассматривается похожая идея: на малом количестве данных учатся накладывать маски, которые потом используются в self-supervised задаче.

И еще есть много статей, где предлагают похожие идеи получения представлений с помощью contrastive loss.

6. Попробуйте на основе результатов статьи предложить исследование, не проведенное к текущему моменту.

У статьи нет четкого теоретического обоснования почему она так хороша, кроме интуитивного предположения в секции How can it be. Утверждение о том, что станет хуже, если использовать в качестве негативных примеров области с разных картинок можно проверить. Вероятно, можно попробовать провести исследование в данном направлении.

В целом, было проведено много экспериментов и по разным задачам, и по разным эвристикам сегментации. Возможно, можно еще поэкспериментировать с заменой ResNet как CNN на что-нибудь другое, как делают некоторые конкуренты.

7. Попробуйте найти применение статье в промышленных приложениях.

Новизна статьи заключается в том, что в contrastive learning позитивными примерами объявляются области одинаковой маски с одной картинки. Это довольно конкретная теоретическая вещь, придуманная для contrastive learning в машинном обучении, и я не вижу ей применения в других областях.