

# Self-supervised learning

Автор: Виданов Андрей  
Руководитель: Барабанщикова Полина

8 марта 2023 г.

## Аннотация

С каждым днем в мире появляется все больше и больше данных, которых мы можем использовать для целей машинного обучения. Однако для классических методов обучения с учителем нужна разметка на данных, которая требует затрат человеко-часов и, соответственно, денег. Вот почему в последние годы все активнее развиваются техники обучения без учителя, которые не требуют предварительной разметки на данных. Для изображений является полезной задача создания эмбедингов, поскольку они могут использоваться в различных задачах машинного обучения для повышения качества финальных моделей. Переход к подобным методам бесспорно является перспективным, но все же таит в себе много подводных камней с которыми придется столкнуться современной науке. в данной статье планируется подробнее рассмотреть влияние лосс-функции на итоговые представления.

## Введение

Развитие техник self-supervised learning'a идет семимильными шагами. В данной области ставится множество задач и применяются различные, в том числе генеративные, методы которые подробно освещены в статье [4] Как показывают эксперименты, применение классических для традиционного машинного обучения техник, таких как batch norm, может улучшать результаты до передовых в данной области знания [5].

Однако мотивы применения тех или иных подходов зачастую являются сугубо практическими(повышение метрики), но не теоретически обоснованными, что ставит под вопрос универсальность их применения для задач своего типа. Теоретические изыскания в данном вопросе активно проводятся [2], тем не менее большинство работ сосредоточено на получении эффективных эвристик. Что не удивительно, поскольку конкретизировав задачу можно использовать ее особенности для повышения метрик. Специфика задачи генерации эмбедингов для картинок требует особых методов В частности, для задач связанными с изображениями нет методов для создания отрицательных пар, обучаясь на которых модель могла бы

улавливать чем отличаются кардинально разные картинки. Избежать данной проблемы можно спроектировав модели, не требующие негативных пар. Модель минимизирует расстояние между эмбедингами похожих картинок. Но данные методы склонны к вырождению, иначе говоря сталкиваются с коллапсом (явление, когда модель сопоставляет всем картинкам одно и то же представление). В результате чего стоит крайне аккуратно подбирать лосс-функции [2] или проектировать модель. [1], [3].

Приведенные выше техники отлично показали себя на практике, но теоретическая мотивация их использования до сих пор остается не прозрачной. Задачей исследования является улавливания зависимостей между применениями различных техник и качеством получаемых эмбедингов, а также риска столкнуться с коллапсом. Возможно, модификации лосс-функции приведут к улучшению финальных показателей в прикладных задачах.

Предлагаемые изменения в функции потерь являются точечными из-за чего получится лучше определить их влияние на результат модели. Но называть данный подход комплексным сложно, что безусловно является недостатком предлагаемого решения.

Измеримыми целями исследования являются численные показатели полученные в ходе применения модификаций к функции потерь.

## Список литературы

- [1] Xinlei Chen Kaiming He - "Exploring Simple Siamese Representation Learning"
- [2] Adrien Bardes, Jean Ponce, Yann LeCun - "VICREG: VARIANCE-INVARIANCE-COVARIANCE REGULARIZATION FOR SELF-SUPERVISED LEARNING"
- [3] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, Stephane Deny
- [4] Xiao Liu, Fanjin Zhang, Zhenyu Hou, Li Mian, Zhaoyu Wang, Jing Zhang, Jie Tang - "Self-supervised Learning: Generative or Contrastive"
- [5] Abe Fetterman Josh Albrecht - Understanding Self-Supervised and Contrastive Learning with "Bootstrap Your Own Latent"(BYOL)