

В статье рассматривается получение новой self-supervised модели с трансформером для получения предобученных признаков. Модель создается в основе предположения, что предобученные признаки могут получить лучшую модель для обработки изображения наподобие BERT. Особенность данной статьи состоит в том, что удастся получить признаки с высокой обобщающей способностью относительно других self-supervised моделей. Это показывает, что подход self-supervised обучения может получить модель по аналогии BERT, но для изображений. Данный факт делает статью, как нельзя актуальной.

Плюсы:

1. В сравнении с ранними подходами признаки, полученные при self-supervised обучении, имеют лучшую обобщающую способность. Показано, что такие признаки эффективны в распознавании изображения с помощью k-NN (78.3% точность на ImageNet).
2. Показана применимость признаков в различных задачах: классификации, нахождения почти одинаковых изображений, сегментации изображений.
3. Сравнивается архитектура модели с другими self-supervised моделями с использованием трансформеров, объяснена мотивация создания такой архитектуры, в статье поясняется как можно воспроизвести результаты и имеется открытый код модели.
4. Представлены подробные анализ и описание, проведенных экспериментов, которые явно показывают преимущества новой модели на примере задачи сегментации изображения.

Минусы:

1. Не показаны примеры неудачной сегментации изображений и анализ таких неудач с целью выявления узких мест в архитектуре.

Оценка статьи:

В данной статье подробно приводится процесс обучения, архитектура модели и различные эксперименты с исследованием эффективности модели, поэтому я считаю, что статья заслуживает оценки 9. В своей оценке уверен на 4 из 5.