

Название статьи (авторы статьи):

Contrastive Learning for Unpaired Image-to-Image Translation.

Taesung Park, Alexei A. Efros, Richard Zhang, Jun-Yan Zhu.

Автор рецензии: Сусли Диана

1) *Краткое описание статьи.*

- Решается задача непарного перевода изображений.
- Предложенный авторами метод основан на максимизации взаимной информации между патчами входа и выхода, используя структуру, основанную на contrastive learning.
- Используется Patchwise Contrastive Learning.

2) *Актуальность.* Перевод изображений на данный момент крайне актуален. Каждый год выходят новые статьи по этой теме. Результаты таких работ могут быть полезны например для создания различных мобильных приложений, накладывающих на фотографии фильтры в режиме онлайн (например Prisma и т.п.).

3) *Сильные стороны.*

- Сама статья написана очень подробно и довольно понятно. Пояснены все ключевые моменты - от функций потерь и до выбора патчей входного изображения: почему лучше выбирать негативные патчи из того же самого изображения, почему лучше работать с патчами, а не с целыми изображениями. При этом аргументируя это не только своими выводами, но и опытом применения в других направлениях: "We hypothesize that by using internal statistics, our encoder does not need to model large intra-class variation such as white horse vs. brown horse, which is not necessary for generating output zebras. Single image internal statistics has been proven effective in many vision tasks such as segmentation, super-resolution, and denoising".
- Все части объяснения включали в себя конкретный пример (про лошадь и зебру), из-за чего статью было проще воспринимать.
- Модель довольно проста, относительно других работ по данной теме, которые используют большее число функций потерь и гиперпараметры. CUT использует метод Patchwise Contrastive Learning, который в отличие от GCGAN и CycleGAN не требует вспомогательных генераторов и дискриминаторов, учит отображение в одном направлении.
- Как утверждают авторы, до данной работы не было алгоритмов перевода изображений, не использующих какую-либо заранее определенную функцию сходства (L_1 или perceptual loss). Например в CycleGAN использовался L_1 cycle-consistency loss.
- В статье проведено большое количество экспериментов, наглядно доказывающих преимущество метода перед остальными.
- Есть сравнение с основными SOTA моделями (CycleGAN, MUNIT, DRIT, GCGAN, DistanceGAN и т.д.). По всем рассматриваемым метрикам CUT лучше других моделей. FastCUT сокращает потребление памяти в сравнении с другими SOTA моделями.
- Предложенный метод может быть расширен до случая, когда на входе и выходе модели всего одно изображение (SinCUT).
- Представлены как удачные, так и неудачные примеры работы модели.

- Большой ссылочный аппарат (91 работа).

4) Слабые стороны.

- Есть момент с тем, что неясно, зачем утверждается, что для контрастирования берутся L слоёв сети, если в итоге эксперименты проводятся либо для всех, либо для последнего.

deeper layers corresponding to bigger patches. We select L layers of interest and pass the feature maps through a small two-layer MLP network H_l , as used in SimCLR [9], producing a stack of features $\{z_l\}_L = \{H_l(G_{\text{enc}}^l(x))\}_L$, where G_{enc}^l represents the output of the l -th chosen layer. We index into layers $l \in \{1, 2, \dots, L\}$

Method	Training settings					Testing datasets		
	Id	Negs	Layers	Int	Ext	Horse→ Zebra	Cityscapes	
						FID↓	FID↓	mAP↑
CUT (default)	✓	255	All	✓	✗	45.5	56.4	24.7
no id	✗	255	All	✓	✗	39.3	68.5	22.0
no id, 15 neg	✗	15	All	✓	✗	44.1	59.7	23.1
no id, 15 neg, last	✗	15	Last	✓	✗	38.1	114.1	16.0
last	✓	255	Last	✓	✗	441.7	141.1	14.9
int and ext	✓	255	All	✓	✓	56.4	64.4	20.0
ext only	✓	255	All	✗	✓	53.0	110.3	16.5
ext only, last	✓	255	Last	✗	✓	60.1	389.1	5.6

- Не пояснено, как именно патчи выбираются из изображения. Случайно, нет?
- Еще пару вопросов к коду.

- 5) *Воспроизводимость*. Есть open source код с кодом модели. И также в самой статье приводятся пояснения, благодаря чему результаты статьи воспроизводимы. Но пояснения довольно высокоуровневые.
- 6) *Экспертные отзывы*. Отзывов с конференций на данную статью в открытом доступе нет. На openreview ситуация такая: <https://openreview.net/forum?id=OgoJ7mi-1pD>. Были найдены и просмотрены видео-разборы. Но из них добавить к тому, что я уже написала, нечего.
- 7) *Оценка статьи + уверенность в оценке* по НИПС. Моя оценка работе 9. В статье подробно описан процесс обучения, архитектура модели и различные эксперименты, визуализирующие работу модели и ее эффективность. Уверенность в оценке 4, так я знакома не со всеми связанными работами по теме (в том числе не знакома со всеми 91 работами из списка источников).