

Отчет практическое ДЗ-3

Дешеулин Олег, 161

27 ноября 2019 г.

1 Описание модели

В качестве статьи для имплементации я выбрал статью о модели cycle-gan, Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. В качестве архитектуры генератора использовался оригинальный U-Net со всеми оригинальными гиперпараметрами, схема архитектуры приведена ниже, схема 1:

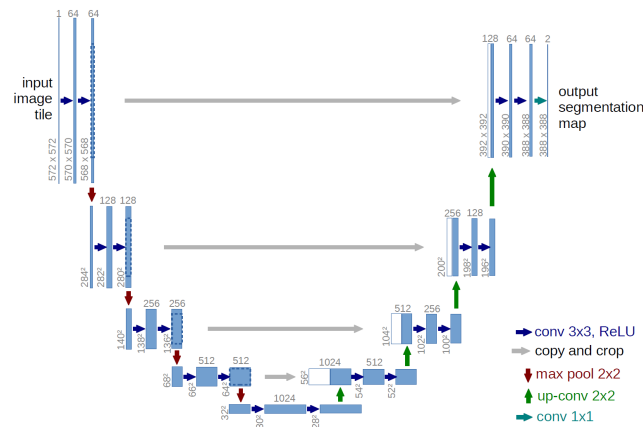


Рис. 1: U-net based архитектура генератора

Для upsamplinga использовалась билинейная интерполяция, либо Transposed конволюции. Был проведен ряд экспериментов и выяснено, что билинейная интерполяция позволяет достичь лучших результатов.

Для дискриминатора использовалась архитектура Patch Discriminator, схема 2. В конце дискриминатора встроена сигмоида и для получения фидбека дискриминатора делается усреднение по всем признакам.

Сам дискриминатор имеет ощутимо меньший размер, чем генератор, это сделано для того чтобы ускорить обучение и позволяет достичь максимального качества за минимальное время.

Далее проведены результаты экспериментов на трех датасетах в unpaired, paired режимах с адверсариал функцией потерь и без. Во всех экспериментах, где была адверсариал функция потерь замерялся усредненный фидбек генератора на каждом из датасетов для того чтобы отслеживать течение

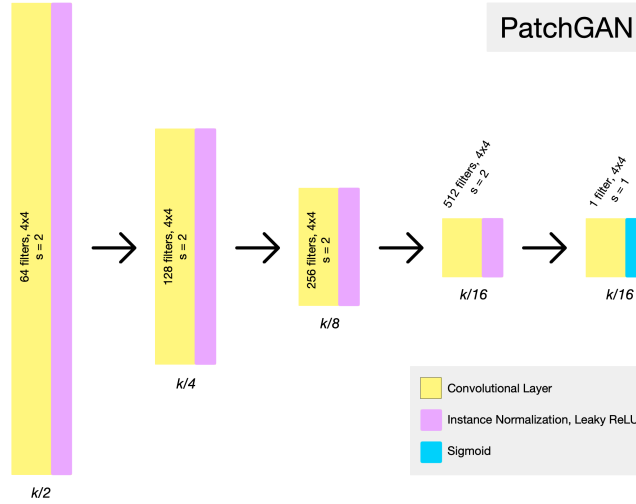


Рис. 2: Patch-based архитектура дискриминатора

обучения. Если датасет был так же paired замерялось l2 расстояние между таргет картинкой и сгенерированной для левого и правого датасетов.

Важно отметить так же, что в оригинальной статье в режиме без дискриминатора применяется только cycle loss. Очевидным локмином тогда становится генерация и картинки самой себя, так как ничто не заставляет модель генерировать вторую картинку из другой выборки. На моих экспериментах ровно это и происходит. Поэтому рассматривать такую постановку задачи абсолютно не интересно и в модель был добавлен переосмысленный idt loss: $\|G_{AB}(A) - B\|_1 + \|G_{BA}(B) - A\|_1$. С таким лоссом проводились все эксперименты. Веса при лосс функциях (если они были необходимы в эксперименте) приведены ниже в таблице 1:

loss name	weight
identity loss (l1)	5
cycle loss (l1)	10
adversarial loss	1

Далее приведена таблица, отражающая значения использованных базовых гиперпараметров:

гиперпараметр	значение
batch size	16
lr	0.0002
beta1	0.5

2 Cityscapes

Этот датасет представляет из себя снимки немецких городов, снятые на проезжающую по улице машину. Они сегментированны и обычно используются для проверки результатов моделей решающих задачу panoptic segmentation.

2.1 Метрики

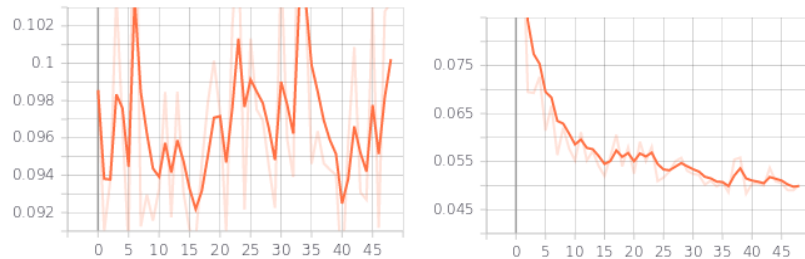


Рис. 3: Pixel Error для задач восстановления города и сегментации соответственно в модели без adversarial функции потерь на валидационной выборке против номера эпохи

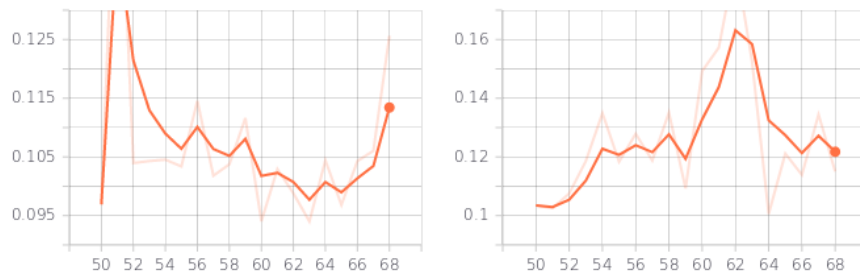


Рис. 4: Pixel Error для задач восстановления города и сегментации соответственно в модели с adversarial функцией потерь на валидационной выборке против номера эпохи

При сравнении графиков по픽сельных ошибок при использовании дискриминатора (фигура 4) и без использования одного (фигура 3) видно, что качество сегментации и значительно лучше, если не использовать adversarial потери, а качество восстановления города выглядит примерно одинаково плохим.

Проверим с помощью фидбека от дискриминатора действительно ли наш генератор обучился при использовании adversarial потерь, фигура 5. Хорошо видно, что на протяжении всего обучения дискриминатор высоко оценивает качество работы. И за 60 эпох он вполне себе сошелся. Теперь когда вполне ясны количественные оценки и проведено количественное сравнение можно провести уже качественное.

2.2 Примеры

Здесь и далее все примеры работы модели имеют одну и ту же структуру: слева результат работы модели, справа исходная картинка.

Сравним для начала результаты работы сегментационной части. Без дискриминатора (фигура 6) и с дискриминатором (фигура 7). Видно, что как и ожидалось по метрикам модель без дискриминатора значительно лучше обучается сегментации как трейне, так и на валидации. Это связано с

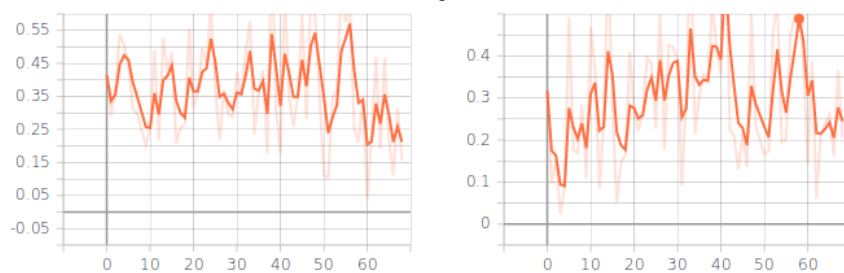


Рис. 5: Усредненная оценка от дискриминатора вероятности быть реальным изображением для сгенерированных для задач восстановления города и сегментации соответственно в модели с адверсариал функцией потерь на валидационной выборке против номера эпохи

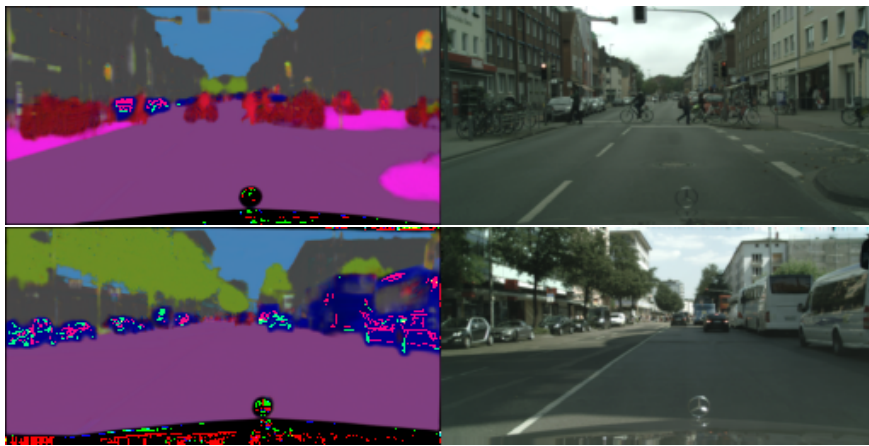


Рис. 6: Примеры работы модели без адверсариал потерь на задаче сегментации, сверху трейн, внизу валидация

тем, что задача сегментации не требует никакого "переноса стиля" и лосс дискриминатора только мешает обучению, замедляя его.

Теперь сравним результаты работы модели на задаче генерации города по сегментационной маске без адверсариал потерь (фигура 8) и с оной (фигура 9). Хорошо видно, что модель без адверсариал потерь неспособна улавливать характерные черты домов, например выдумывать окно, или структуру деревьев, чему вполне с успехом обучается наша модель с адверсариал потерями. Видно, что такие различия неспособна улавливать просто попиксельная L2 метрика, и для их отслеживания необходимо считать например Inception Score или FID score (Frechet Inception distance), эти метрики способны улавливать семантические различия между картинками, а не только цветовые. И в целом общеприняты для тестирования генеративных моделей.



Рис. 7: Примеры работы модели с адверсариал потерями на задаче сегментации, сверху трейн, внизу валидация

3 Edges2shoes

Датасет представляет из себя контуры ботинок, называемые edges и собственно сами ботинки. Задача ставится соответственно в две стороны - по ботинку генерировать его edges и наоборот по edges собственно сам ботинок. Будем действовать аналогично предыдущему пункту, сначала рассмотрим количественные результаты, а затем качественные.

3.1 Метрики

Если обобщать, результаты очень похожи на результаты на датасете cityscapes, однако поговорим подробнее. Давайте снова посчитаем попиксельную ошибку для обоих способов обучения: фигура 10 отражает картину без дискриминатора, фигура 11 отражает картину с дискриминатором. Хорошо видно, что как и с сегментацией задача построения очертаний не требует адверсариал потерь и ее получается хорошо решить и без них. В это же время видно, что модель без адверсариал потерь не справляется с задачей в которой необходимо, что либо придумать и имеет даже меньшую метрику на ботинках.

На фигуре 12 видно, что модель можно продолжать обучать, дискриминатор еще не заполнил всю тренировочную выборку и генератор достаточно сильно дискриминируется, продолжая получать полезные фидбеки, но обучение к моменту остановки продолжалось уже 12 часов, так что неудивительно, что я его прервал.

3.2 Примеры

Продолжим количественное исследование качественным. Для начала посмотрим на сгенерированные очертания. Видно, что как на обучающей, так и на валидационной выборке качество очертаний лучше у полностью supervised модели без дискриминатора, фигуры 13 и 14, без дискриминатора и с, соответственно. Для задачи генерации все наоборот, во-первых

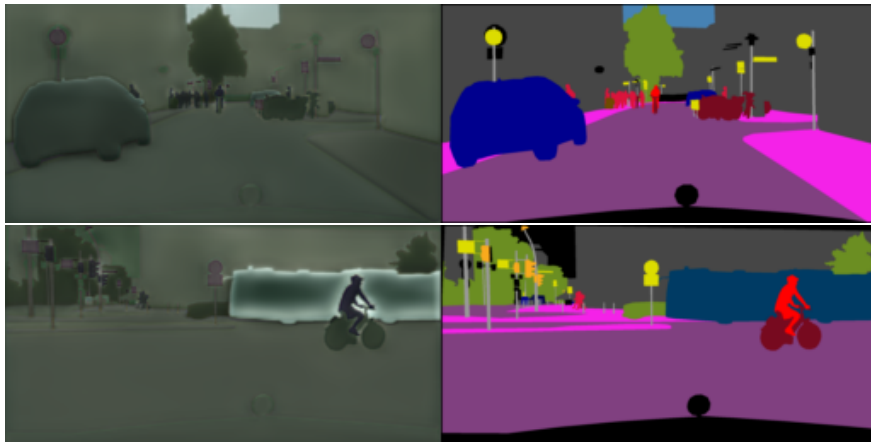


Рис. 8: Примеры работы модели без адверсариал потерь на задаче генерации города по сегментации, сверху трейн, внизу валидация

supervised модель переобучается уже за три эпохи как видно на фигуре 15, однако на трейне хорошо восстанавливает ботиночки. В это же время модель с адверсариал потерями сходится значительно дольше, однако имеет сравнимое визуальное качество на трейне и значительно лучшее на валидации, как видно на фигуре 16.

4 Face2simpson (бонус)

В качестве дополнительной задачи предлагается генерировать из лица и верхней части фигуры человека такую же только симпсона. Для этой задачи были найдены датасеты imbd-wiki и симпсонов. Первый представляет из себя фотографии известных людей, актеров, вырезанные из фильмов. Он представляет из себя весьма разнородный датасет людей разных возрастов, полов и рас, он довольно большой, но я использовал только около 250 тысяч картинок. Аналогично, вырезанные из симпсонов их фотографии собранные по всем фильмам представляют из себя второй датасет. Примерный вид этих датасетов будет понятен по примерам, которые были из них сгенерированны.

На моих экспериментах очень важно, чтобы даже unpaired датасеты были плюс-минус одинаковые в смысле структуры. Например, я пробовал использовать датасет лиц UTKFaces, который представляет из себя значительно более аккуратные кропы лиц людей. При использовании такого датасета модель абсолютно не обучалась, поскольку генератору было выгодно генерировать одинаковые картинки людей, не привязываясь к симпсонам в целом, поскольку, если генерировать их из симпсонов отступы в кропе будут значительно отличаться от тех, которые есть в датасете. В связи с этим при обучении использовались RandomCrops картинок, причем параметры были различными в зависимости от датасета: для симпсонов размер кропа был ограничен всего лишь 0.95 снизу, а для Imdb 0.55. Аналогично предыдущим пунктам здесь далее приводятся количественные и качественные результаты.



Рис. 9: Примеры работы модели с адверсариал потерями на задаче генерации города по сегментации, сверху трейн, внизу валидация

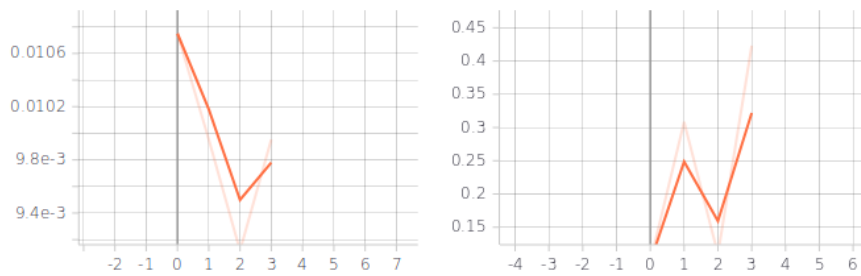


Рис. 10: Pixel Error для задач восстановления структуры ботинка и его самого соответственно в модели без адверсариал функции потерь на валидационной выборке против номера эпохи

4.1 Метрики

Поскольку используется unpaired датасет и производится unsupervised обучение смотреть на попиксельную ошибку нет никакого смысла, да и возможности, поэтому для отслеживания течения обучения я использовал средний фидбек от дискриминатора, а для проверки результатов провел небольшой опрос, как часто принято делать в статьях о генерации, когда нужно оценить визуальное качество и хочется получить какие-то метрики.

Для начала отметим, что не смотря на то что прошло уже 20 часов с момента начала обучения, модель не переобучилась, и качество улучшается, как видно по фидбекам как на тренировочной (фигура 17) так и на валидационной выборках (фигура 18).

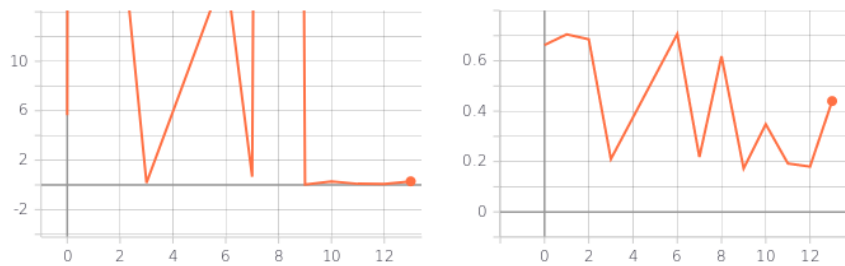


Рис. 11: Pixel Error для задач восстановления структуры ботинка и его самого соответственно в модели с адверсариал функцией потерь на валидационной выборке против номера эпохи

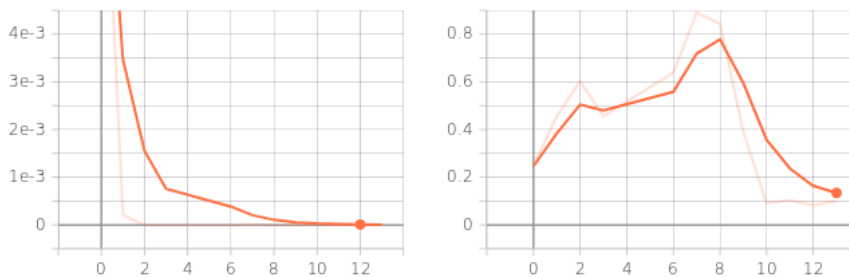


Рис. 12: Усредненная оценка от дискриминатора вероятности быть реальным изображением для сгенерированных для задач восстановления очертаний ботинка и его самого соответственно в модели с адверсариал функцией потерь на валидационной выборке против номера эпохи

user	mark	comment
den	6	качественная глазуня норм, я не понял, где фейк
alyona	7	
akhmad	8	
irina	6.75	
denza	5	
mean	6.55	

user	mark	comment
den	4	такое себе
alyona	4	
akhmad	2	
irina	2.5	
denza	4	
mean	3.3	

Чтобы выяснить реальное качество, как я отметил выше был проведен опрос, людей просили оценить качество картинок по шкале от 1 до 10. Люди видели те же картинки, что и читатели этого отчета. Они так же могли оставлять комменты, но многие воздержались. Было опрошено 5 человек и результаты приведены выше в таблицах 4.1 и 4.1. После такой полноценной



Рис. 13: Примеры работы модели без адверсариал потерь на задаче генерации очертаний ботинка по фотографии, справа трейн, слева валидация



Рис. 14: Примеры работы модели с адверсариал потерями на задаче генерации очертаний ботинка по фотографии, справа трейн, слева валидация

проверки можно уверенно сказать, что задача стилизации симпсонов значительно проще, действительно, не требуется восстанавливать более сложную человеческую кожу, а достаточно только лишь окрасить все в желтый.

4.2 Примеры

Далее приведены примеры работы обоих генераторов, ровно на эти же примеры смотрели люди при построении своей оценки.

На картинках симпсонов (фигура 19) видно, что модель хорошо справляется с переносом стиля симпсонов, люди окрашиваются в желтый, изменяется форма лица, глаза, фон размывается и становится более мультяшным. Появляются так же предметы из симпсонов, например ожерелье на картинке справа сверху. При этом сохраняются все ключевые элементы лица, видны результаты даже переноса гримасы. картинки действительно заслуживают высокой оценки, несмотря на некоторые значительные ошибки, например третий глаз.

Результаты значительно хуже в домене людей, как видно на фигуре 20. Средняя оценка меньше почти в два раза и видно что модель сильно размазывает картинки. Однако, лица перестраиваются, но не учитывается пол и возраст, учитывается только лишь форма лица. Прическа иногда переносится правильно, иногда нет, кроме того сами по себе лица сильно более размазанные. Скорее всего это связано с неоднородностями в датасете людей, там есть люди слишком разных возрастов, цветов кожи, полов и причесок.

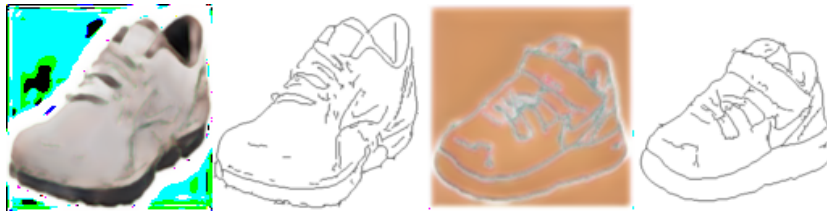


Рис. 15: Примеры работы модели без адверсариал потерь на задаче генерации ботинка по очертаниям, справа трейн, слева валидация



Рис. 16: Примеры работы модели с адверсариал потерями на задаче генерации ботинка по очертаниям, справа трейн, слева валидация

5 Итоги

В результате этого дз я научился некоторым техникам и моделям использующимся в машинном переводе, попробовал для себя многие трюки и современные архитектуры и достиг хорошего сора.

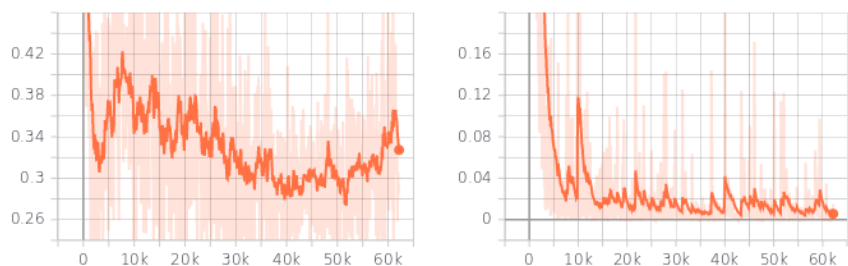


Рис. 17: Усредненная оценка от дискриминатора вероятности быть реальным изображением для сгенерированных для задач восстановления человека из симпсона и наоборот на тренировочной выборке против номера эпохи

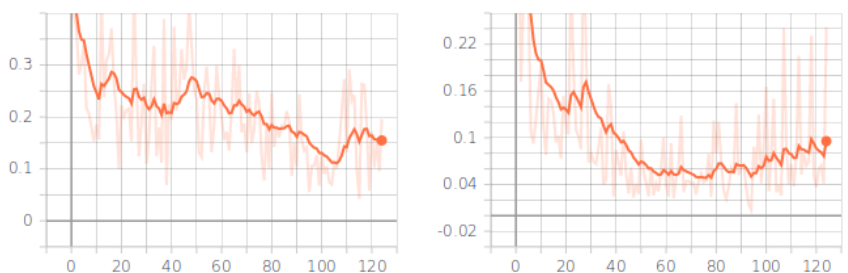


Рис. 18: Усредненная оценка от дискриминатора вероятности быть реальным изображением для сгенерированных для задач восстановления человека из симпсона и наоборот на валидационной выборке против номера эпохи



Рис. 19: Примеры работы модели в задаче генерации симпсонов по человеку, слева сгенерированная, справа исходная, картинки выбраны случайно, средняя оценка 6.55



Рис. 20: Примеры работы модели в задаче генерации человека по симпсону, слева сгенерированная, справа исходная, картинки выбраны случайно, средняя оценка 3.3