Отчёт о проделанной работе

1 Введение

Выбранная мной задача - LFW. Так как metric learning для меня была относительно новая область, работа началась с изучения статей на эту тему. После прочтения пары десятков метров текста про различные архитектуры и подходы к задаче face recognition, я решил реализовать два метода: проверка непосредственно пар с использованием contrastive loss и обучение глубокой нейронной сети с использованием COSFACE-loss. Общая идея была сравнить, насколько второй метод лучше справляется с задачей верификации. Думаю, ответ очевиден.

2 Contrastive loss

Идея cl в том, чтобы максимально отдалить друг от друга разные пары и приблизить одинаковые. Для этого используется следующая формула:

$$L = y * p(x_1, x_2)^2 + (1 - y) * min(m - p(x_1, x_2), 0)^2$$

Пройдёмся по обозначениям: y - таргет список (0 или 1), $p(x_1, x_2)$ - расстояние между двума векторами, m - параметр margin, отвечающий за расстояние, после которого лосс не одинаковых фотографий не будет увеличиваться.

Немного подробнее про расстояние. В большинстве статей говорится либо о cosine sim, либо о euclidian distance. Я в своей работе использовал именно последнее, так как оно лучше отражало разницу итоговых значений на тестах и может разнести два вектора по пространству дальше друг от друга, что должно положительно сказаться на негативных сэмплах.

Дальше стоит сказать о выбранной архитектуре CNN. Во время обучения в тинькофф у нас была задача колоризации. Именно оттуда я и взял эту сеть - оно хорошо показала себя на той задаче и, я думал, хорошо проявит себя в этой. К сожалению, чуда не случилось.

Какой итог train-a: лосс упал очень быстро и потом особо никуда не двигался. На тесте (причём не важно, были ли люди из той же выборки или новые фотографии) сеть почти не различала негативные и позитивные сэмплы. Я связываю это с тем, что сеть выучилась на лёгких примерах и не смогла "обобщить" результаты на новой выборке.

3 DCNN + COSFACE

Здесь выбор был побольше - глубоких нейросетей в torchvision достаточно, а вариантов лоссов как минимум 4. Я же выбрал resnet50 и COSFACE соответственно. Можно спросить, почему?
Посмотрев много примеров решения задач верификации лиц и прочитав оригинальную статью [2], я понял, что люди предпочитают resnet чаще других сетей (как минимум для "учебных"примеров). Основная идея "проброса"градиентов в достаточно глубоких сетях стала одной из главных причин становлений resnet. А 50 я выбрал, потому что это первая сеть с использование проброса через 3 слоя, а не 2, как было в 18 и 34.
Что же касается COSFACE, то здесь лукавить не буду - это просто первая рабочая(!!) модификация cross-entropy loss, которая при этом не сложная в реализации. Конечно, большую точность показал бы скорее всего arcface или sphere, но преумалять заслуги COSFACE я всё таки не буду. На нескольких тестовых примерах он показал себя очень хорошо. И это при том, что в силу ограничений по gpu (так как всё обучение проходило в Google Collab), при мало количестве эпох (30) сеть смогла верифицировать примеры из датасета. Код для реализации я взял из репозитория [4]

Вывод: до продакшена ещё далеко, но перспективы есть :)

Список литературы

[1] https://towardsdatascience.com/the-why-and-the-how-of-deep-metric-learning-e70e16e199c0

[2] https://arxiv.org/pdf/1512.03385.pdf

[3] https://arxiv.org/pdf/1801.09414.pdf

[4] https://github.com/ronghuaiyang/arcface-pytorch/blob/master/models/metrics.py