

ДЛ БДЗ-1 Отчет

Аржанцев Андрей

БПМИ202

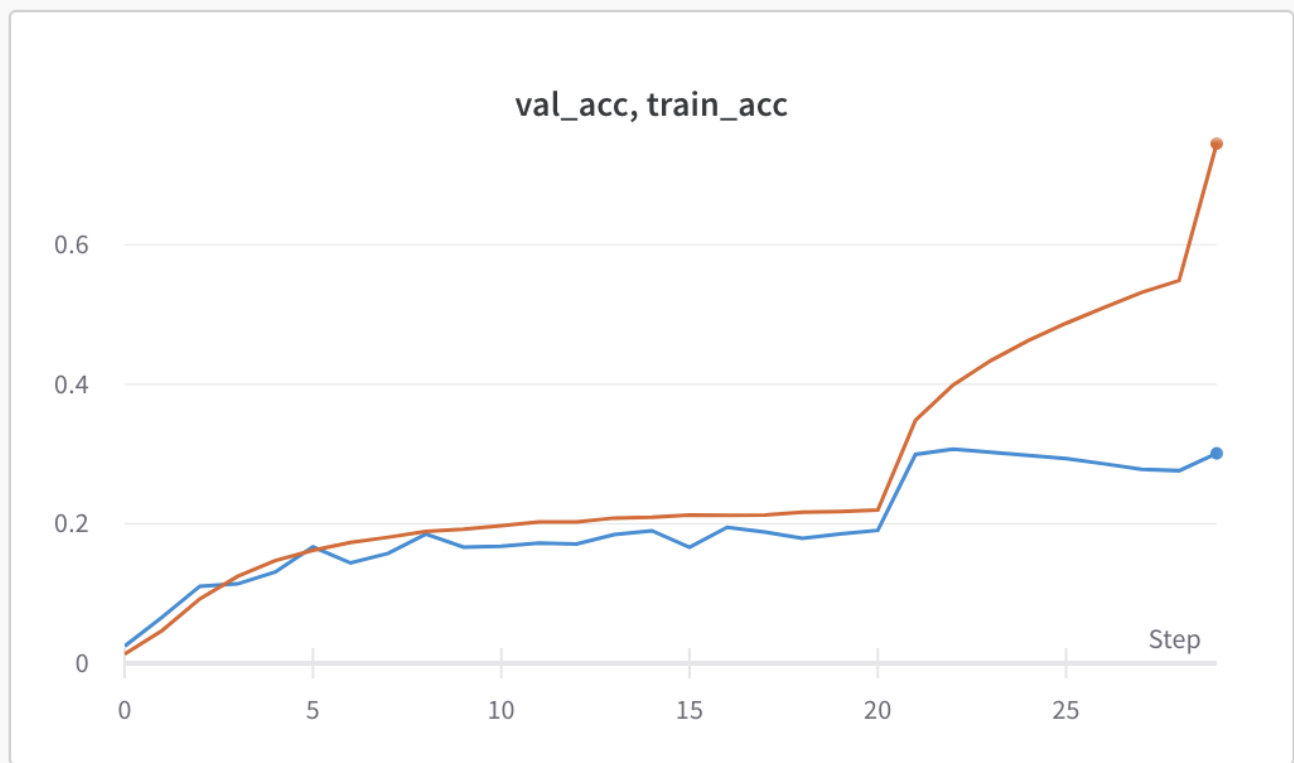
Начнем с чекпоинта - я просто взял код из своей МДЗ-2, где была несложная модель с двумя conv layer-ами и батч нормой после них. Ну и CyclicLR, а так все параметры дефолтные. Собственно, мне хватило 5 эпох, чтобы в итоге выбить точность 0.17 на тесте.

Далее для уже нормального сора я выбрал как базу resnet34 из торча. Ну и последний лэйер я сначала выбрал просто линейным
Версия без дополнительных аугментаций и с дефолтными параметрами получилась не очень, я ее стопнул на 20 эпохах, когда уже стало заметно, что модель подгоняется под train и уже не дает никакого прироста на val:



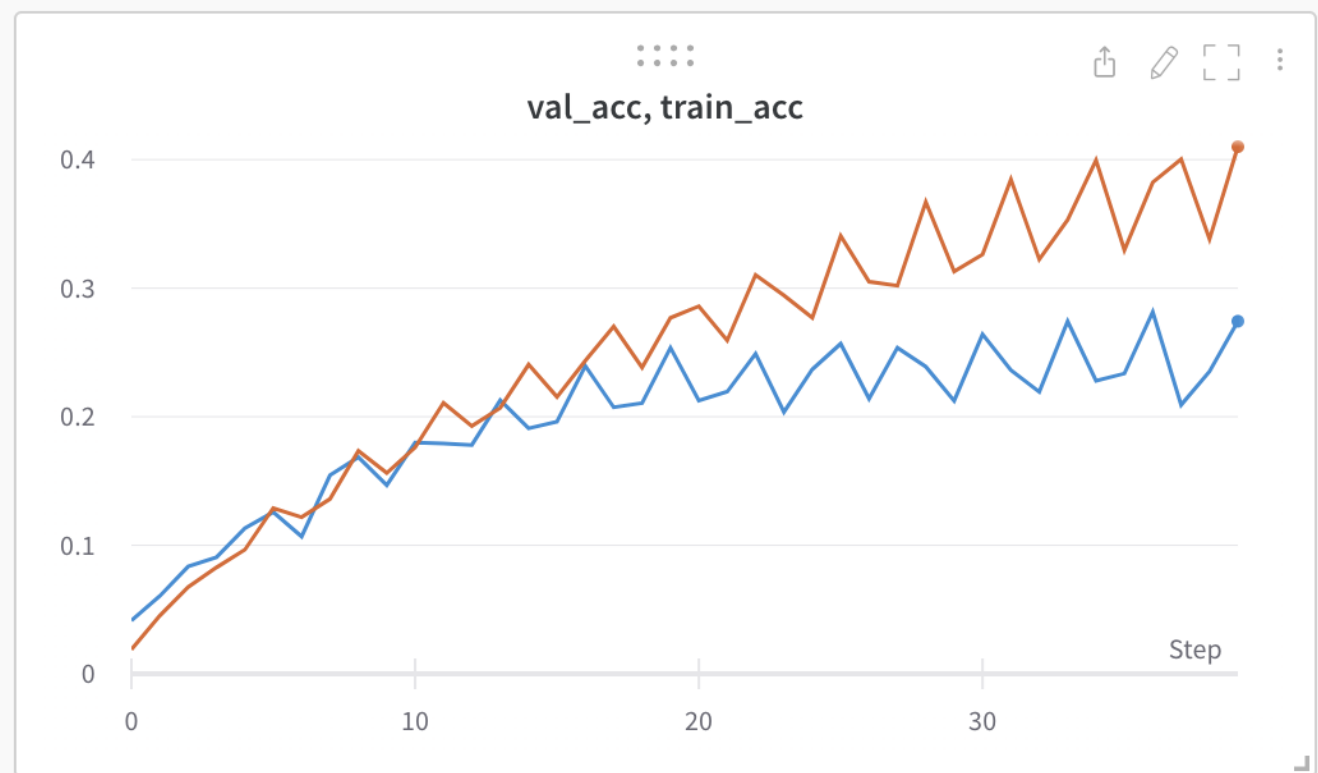
С проблемой переобучения было решено бороться путём довешивания различных аугментаций и перебора гиперпараметров. Также хотелось посмотреть и на другие скедулеры - еще во время выполнения МДЗ-2 мне приглянулся ReduceByPlato, потому что он позволяет на длительной обучении получать большие скачки в точности, когда "надежд уже почти нет", приятная штука (собственно из названия понятно, как это должно работать).

Времени у меня было немного, зато были какие-никакие ресурсы - Датасфера, кагль, колаб (хоть он и падал 24/7, и в итоге на нем я только тестил модели, обученные на первых двух ресурсах). В общем то поставил обучаться свой резнет для ReduceByPlato, WeightDecay=5e-4 и с AutoAugment, получалось так:



Ну видно, что ReduceByPlato получился норм, но после скачка уже стал жестко переобучаться, поэтому я его стопнул. AutoAugment тут видимо не особо лучше перформил, нежели RandomHorizontalFlip, который я юзал изначально. В целом тут удалось добиться точности на тесте over 0.25, но такое.

Параллельно у меня бежало обучение для CyclicLR, RandAugment и WeightDecay=1e-4:



Тут хоть и выбивался скор также больше 0.25, но выглядит все крайне подозрительно. Скорее

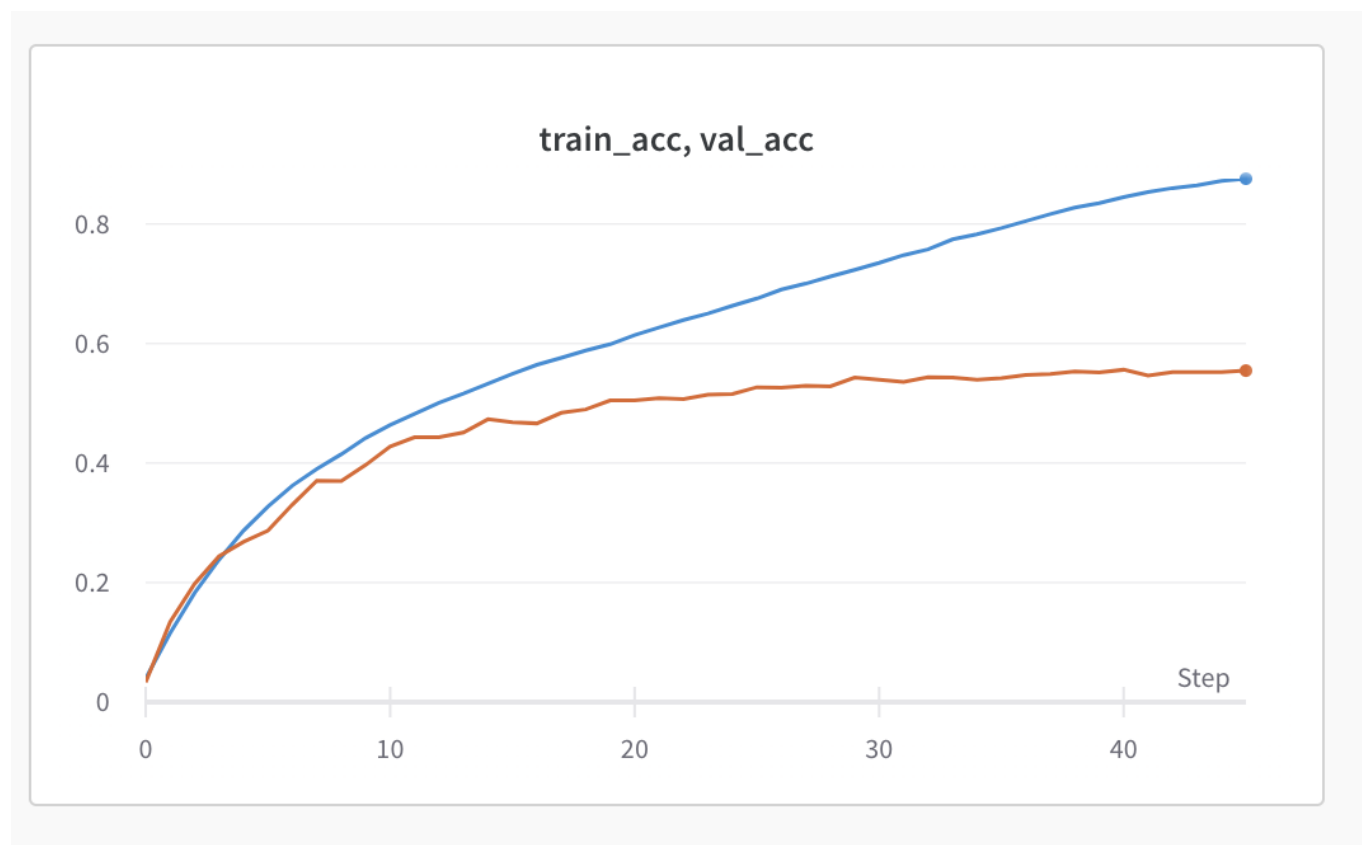
всего такие скачки получались из-за большого изначального $lr=0.1$, но в моменте я об этом не подумал, да.

Дальше я еще немного повозился с резнетом, пытаюсь доводить больше аугментаций и еще добавлял дропаут перед последним линейным слоем, но оно в принципе обучалось плохо, поэтому решил менять модель кардинально.

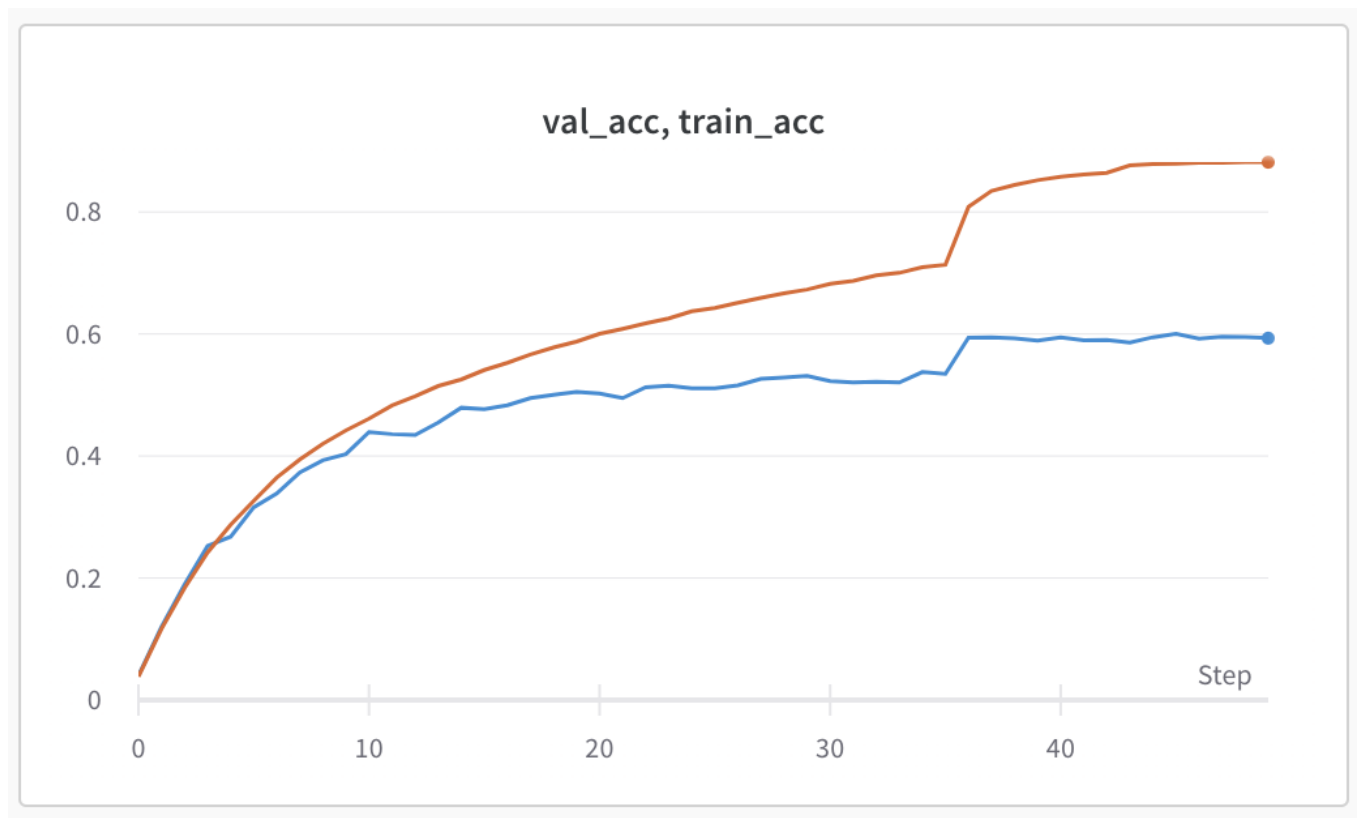
Ну и собственно на помощь пришел 4ый семинар 202 группы, где оказывается уже была написана крутая модель под MobileNetV2. Мне осталось только прогуглить немного mobilenet-ы, взять нужный, ну и в первом приближении просто копияста семинара. И модель наконец начала нормально перформить! Обучив всего на 12 эпохах, я уже выбил в соревновании 0.45. Тут наверное и стало заметно, что я делал не так в резнете, а именно ресайзил картинки в (32,32) вместо (224,224), а также использовал слишком большой lr , впрочем пойти фиксить это уже было менее осмысленно, чем доделать нормально этот мобайл.

Собственно я поставил обучение на 50 эпох для всех тех же параметрах, что и в семинаре, только добавил $WeightDecay=5e-4$ и для двух разных скедулеров - Cosine и ReduceByPlato. Первый скедулер сначала казался лучше - по крайней мере уже через 20 эпох у меня там получилось заслать на точность over 0.5. На 30-ой эпохе этого смог и добиться и ReduceByPlato (сохранял веса на 20,30,35,40,45 эпохе).

После 30-ой эпохи точность на валидации уже почти не поднималась для Cosine, в итоге 0.55 там так и не выбились.



Зато ReduceByPlato не подвел и совершил нереальный скачок на 36-ой эпохе, добив точность на валидации ровно до 0.6, хоть и после скачка сразу стал учиться плохо с уже очень маленьким lr . Собственно версия с Plato для 40 и 45 эпох в итоге и дала мне over 0.55, а именно 0.585, а на 50-ой он довольно неожиданно вырос еще на процент.



Чтобы добить 0.6 я дообучил эту же модель еще на 10 эпохах, но уже с мизерной валидацией (ну можно было и вообще с нулевой), просто поменяв сплит с 0.9 до 0.999. И в итоге вот оно - решение на 60% точности на тесте. Потом я еще на 20 эпох ставил дообучаться, ни к чему это собственно ни привело и всегда было 0.6+eps.

Чувствуется, что можно было еще поулучшать модель - как минимум попробовать посмотреть другие трансформеры.

В целом по выводам могу отметить

- MobileNet оказался у меня заметно лучше, нежели ResNet34
- ReduceByPlato был немного лучше других рассматриваемых скедулеров, все посылки с ним давали прирост в 3-5 % в соревновании
- Выбрать правильные численные параметры важно - без нормального начального lr у меня вообще все было слабо, ну и без WeightDecay также на MobileNet получились результаты хуже.
- Крутая идея дообучать без валидации, я даже как-то сразу не подумал об этом, но сначала смотришь по графикам и дебажишь переобучение, потом уже увеличиваешь размер трейна на хорошо получившейся модели