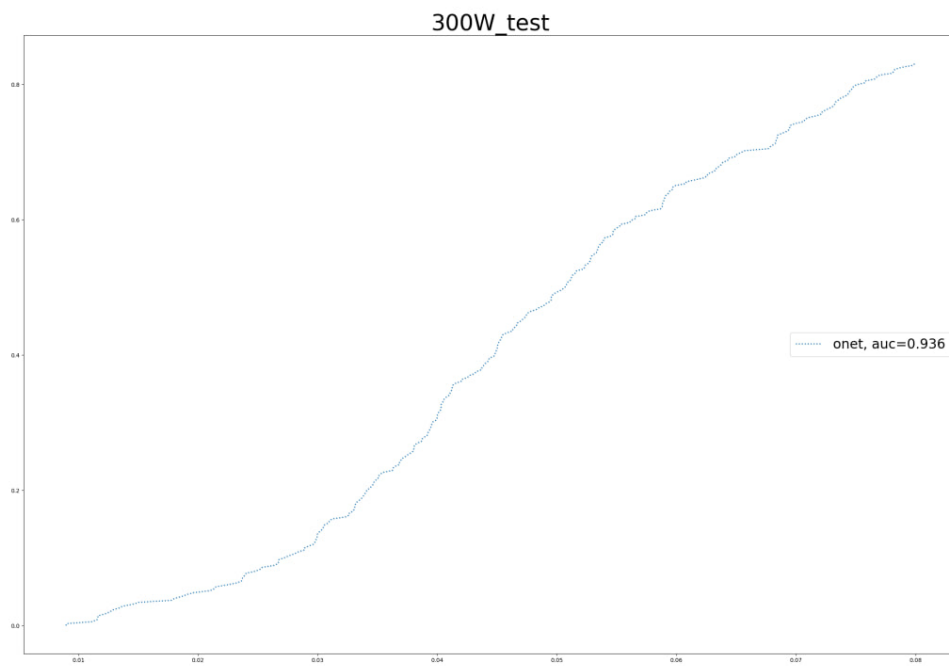


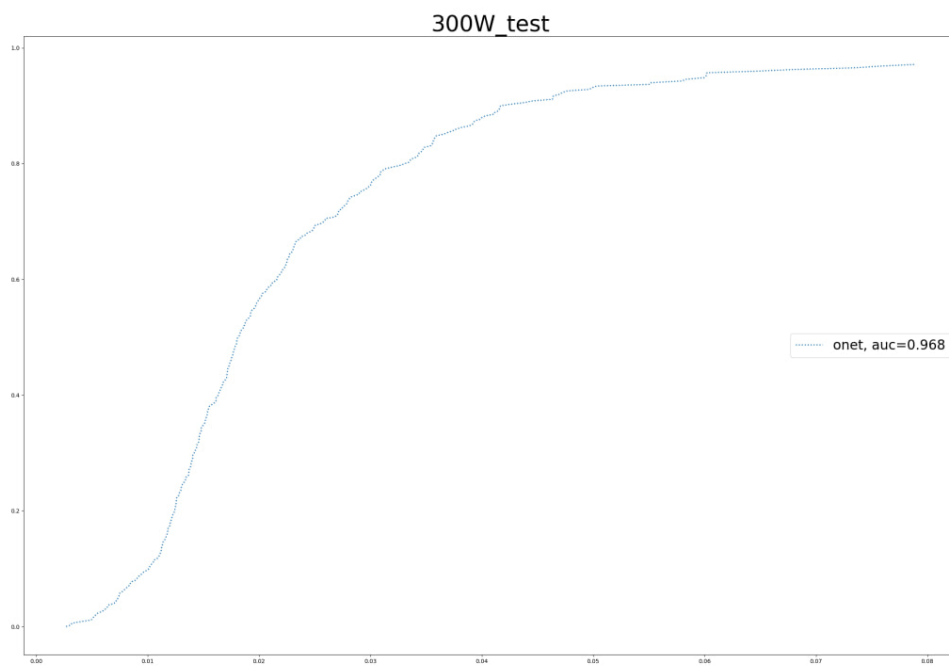
Отчет о выполнении тестового задания VisionLabs

Задачи: обнаружение 68 особых точек, сравнение метрик на датасетах 300W и Menpo, построение графиков по метрике CED.

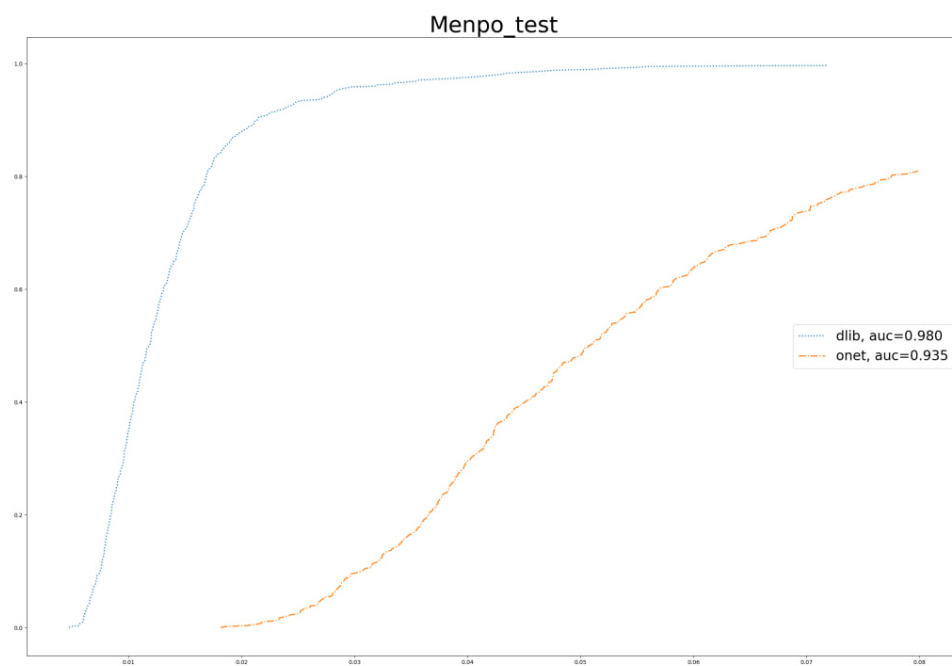
1. Удаляются изображения с размеченным числом точек не равным 68.
2. К исходным картинкам применялся детектор dlib для получения более точных прямоугольников, ограничивающих лица.
3. Если на изображении лиц, выбирается только то, у которого совпадение по IoU с прямоугольником, построенным по размеченным ground-truth точкам, больше 0.3.
4. Если таких нет, изображение исключается.
5. Далее прямоугольник увеличивается в 1.75 раз по длине и ширине, а после выполняется проверка, все ли точки попали в увеличенный прямоугольник. Если нет, то изображение также исключается из рассмотрения.
6. После записываются файлы с координатами задетектированных прямоугольников с лицами, координаты увеличенных прямоугольников, а также файлы с пересчитанными координатами ground-truth точек для увеличенных прямоугольников.
7. Из полученных файлов формируется датасет для обучения Facial Landmark Detection модели (предложена архитектура O-Net). Координаты нормируются по длине и ширине увеличенных прямоугольников, также сохраняется мета-информация для восстановления координат на исходном изображении.
8. Обучается O-Net с функцией потерь, отвечающей только за особые точки. Было проведено 2 эксперимента:
 - a. Модель из статьи. 150 эпох, размер батча 16, lr 0.001
 - b. Модель из статьи с измененными функциями активации с PReLU на SiLU, с измененным оптимизатором с Adam на AdamW, с измененным методом инициализации весов с метода Ксавье на метод Хе. 150 эпох, размер батча 16, lr 0.001
9. Тестирование проводится на тестовых частях датасетов 300W и Menpo. Также на тестовой части датасета Menpo тестируется детектор точек dlib, строятся соответствующие графики, вычисляются площади под графиками.



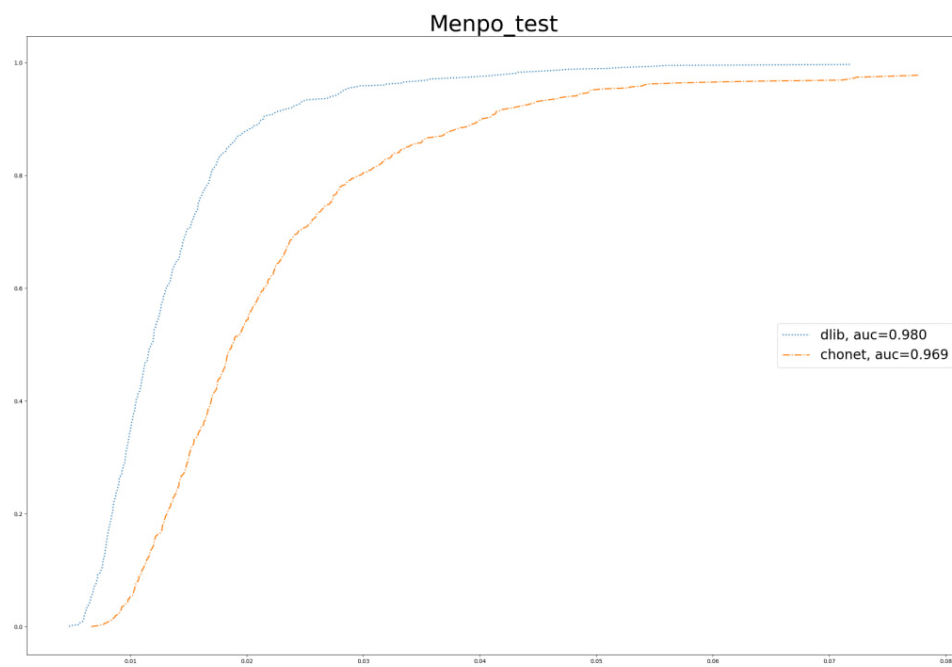
O-Net (a) на тесте 300W



O-Net (b) на тесте 300W



O-Net (a) и dlib на тесте Менпо



O-Net (b) и dlib на тесте Менпо

Выводы

По результатам экспериментов можно сказать, что dlib лучше справился с задачей Facial Landmark Detection на тестовой части датасета Menpo.

Однако стоит заметить, что использование другой функции активации в архитектуре O-Net, другого метода инициализации весов и другого оптимизатора во время обучения позволило получить более высокую метрику, приближающуюся к метрике dlib.

Следует предположить, что дальнейшие эксперименты с архитектурой позволят еще сильнее приблизиться к dlib. Либо же стоит взять архитектуру из топа PaperWithCode по задаче Facial Landmark Detection.

UPD:

Дополнительно были проведены следующие эксперименты:

- Wing Loss/Adaptive Wing Loss <https://arxiv.org/pdf/1904.07399.pdf> / <https://arxiv.org/abs/1711.06753>
- Аугментации (RGBShift, RandomBrightnessContrast, CLAHE, RandomGamma, AdvancedBlur, ISONoise, MultiplicativeNoise, ChannelDropout, GaussNoise)
- WarmUp (https://github.com/Tony-Y/pytorch_warmup/tree/master)
- Косинусный scheduler / ReduceLROnPlateau scheduler
- Dropout/DropBlock (<https://github.com/miguelvr/dropblock/tree/7fb8fbfcb197a4bb57dc9193bcd6f375ff683f85>)
- SiLU/ReLU
- Входные размеры 64, 72, 128, 256, соответственно углублял сеть
- BatchNorm2d
- Residual блок (onetResidal.py)

По итогу экспериментов удалось добиться следующих результатов:

