

Отчет к тестовому заданию.

Задание заключалось в реализации алгоритма обнаружения 68 особых точек на лице человека (face alignment) и тестировании данного алгоритма на общедоступных датасетах и сравнении с аналогами.

В первую очередь, были удалены экземпляры с количеством размеченных ключевых точек меньшим 68. Затем данные были обработаны с помощью детектора лиц из библиотеки dlib. Все изображения были обработаны данным алгоритмом и результирующие прямоугольники лиц (боксы) были сохранены. Заметим, что в случае нескольких лиц на фотографии выбирался тот у которого наибольшее пересечение с боксом, построенному по максимальным и минимальным координатам набора ключевых точек. В случае отсутствия найденных лиц на изображении оно отбрасывалось.

Также дополнительно были подсчитаны количества точек непопадающих в бокс.

Методы решения.

В ходе решения задачи было опробовано несколько подходов:

1. Onet.

Данный подход заключается в реализации архитектуры нейронной сети *Onet* из статьи «*Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks*». Это решение было реализовано с помощью фреймворка Pytorch. Ввиду того, что было много боксов в которые не попадали все ключевые точки было принято решение обучать данную сеть на боксах дополнительно расширенных в 1.5 раза. Дополнительно были отброшены фотографии, где не попадала хотя бы одна точка в увеличенный бокс.

Для обучения был использован оптимизатор Adam с переменным обучающим шагом, начальный шаг был принят равным 0,001. Во время обучения было замечено, что loss на валидации уменьшается не так быстро как на тренировочной выборке, а в последствии и вовсе начал расти, что свидетельствует о возможном переобучении модели. Также возможен вариант с недостаточной сложностью модели, что не позволяет ей уловить обобщающие фичи.

2. HRNetv2

Данное решение является реализацией алгоритма из статьи «*High-Resolution Representations for Labeling Pixels and Regions*» (<https://arxiv.org/pdf/1904.04514.pdf>). Сам алгоритм является адаптацией архитектуры High-ResolutionNetwork для задачи обнаружения ключевых точек.

В данном методе также используются увеличенные размеры детекций лиц, но в 1.25 раза. Для обучения был использован оптимизатор Adam с переменным обучающим шагом, начальный шаг был принят равным 0,0001.

В этом методе валидационный лосс во время обучения перестал убывать и обучение было остановлено.

3. HRNet, с предобученными весами.

Аналогичное решение описанному выше, но дополнительно обученное на предоставленных тренировочных датасетах. В данном методе для получения

решения были взяты предобученные веса HR18-300W.pth (<https://1drv.ms/u/s!AiWjZ1LamlxzeYLnza1XU-4WhnQ>)

Результаты.

Все полученные алгоритмы были применены к тестовым частям датасетов 300W и Менро. На основе полученных результатов были построены графики метрики *cummulative error distribution* (CED), они представлены на рисунках 1 и 2 ниже. Также дополнительно была подсчитана площадь под каждой из кривой.

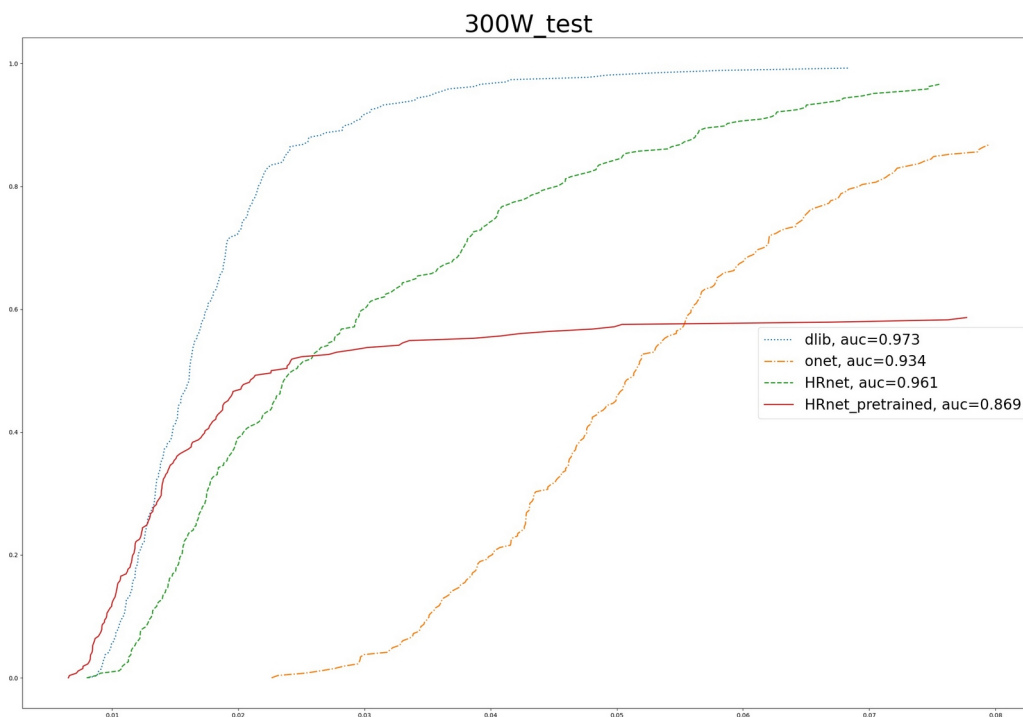


Рисунок 1. Кривые метрики CED для тестовой части датасета 300W

Для сравнения дополнительно приводятся результаты работы решения из библиотеки dlib.

Как видно из них, на тестовой части датасета 300w лучший результат показал алгоритм dlib — площадь под кривой ошибок составляет 0.973, немного отстал от него HRNet обученный на датасетах (0,961), далее идут Onet (0,937) и предобученный HRNet(0,869), несмотря на то что HRNet с предобученными весами имеет большее количество элементов с низкой ошибкой, но при этом больше чем на половине примеров его нормированная ошибка была больше 0.08.

На датасете Менро видим другую ситуацию, лучший результат у обученного HRNet (0.971), далее близкие результаты показали dlib (0,962) и предобученный Hrnet (0.957), худший результат показал Onet (0,937).

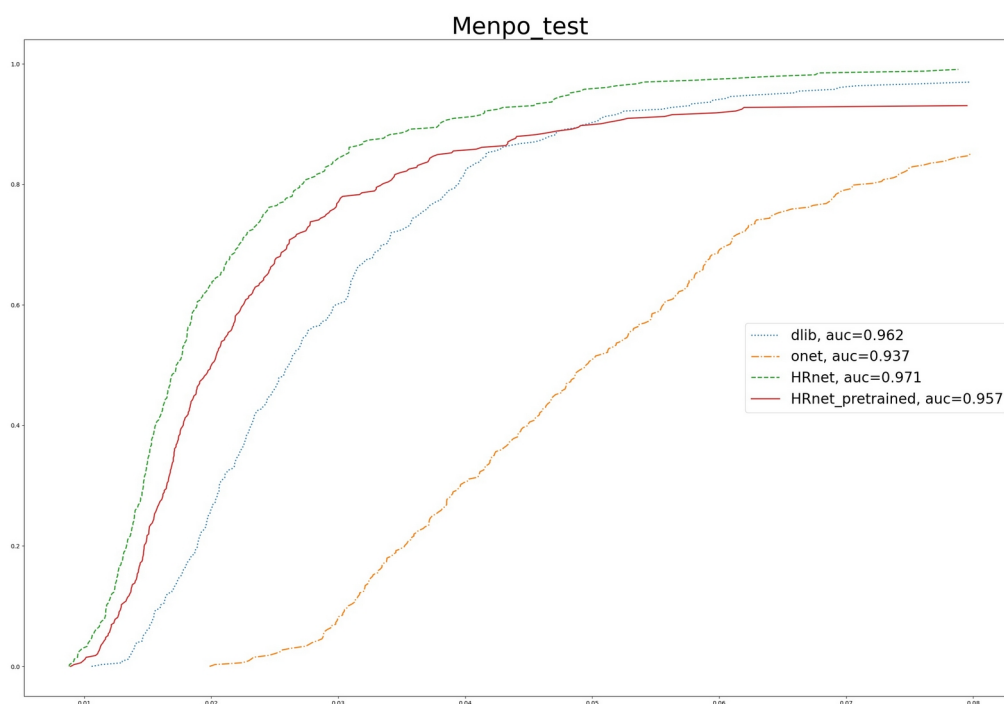


Рисунок 2. Кривые метрики CED для тестовой части датасета Менро

На основе полученных результатов можно сделать вывод что лучше всего себя показывают dlib и обученный HRNet. Предобученному HRNet скорее всего не хватает дообучения на новых данных, а сеть Onet по моему предположению имеет слишком слабую архитектуру, которая не позволяет учитывать сложные зависимости.

В дальнейшем имеет смысл рассмотреть влияние изменений размеров лицевых детекций и использование других детекторов лиц.