

Отчёт

Датасет для обучения

1. Прямоугольники выдаваемые детектором dlib, расширенные в каждую сторону на 40%. Процент выбирался исходя из количества лиц, все точки которых попадали в расширенную область (Таблица 1).
2. Для уменьшения искажений при преобразовании прямоугольника к квадрату, прямоугольник дополнялся следующим образом: длинная сторона по-возможности увеличивалась на 40% в обе стороны, а меньшая сторона по-возможности дополнялась в обе стороны до квадрата.

Таблица 1 - Объем датасета в зависимости от выбранного прямоугольника лица. n - число изображений в исходном датасете, k - коэффициент расширения.

	Менпо				300W			
	train		test		train		test	
n	6018		661		2521		281	
k	прям.	квадрат	прям.	квадрат	прям.	квадрат	прям.	квадрат
0,15				445				
0,2			531	516	2231			
0,25			561	553	2345			
0,3	5192		572	567	2380		263	
0,35	5224		575	572	2395		265	
0,4	5236	5204	578	575	2401	2365	266	260
0,45	5247			575	2404			
0,5	5250		578	575	2406			
0,55	5253				2406			
0,6								
0,65								
0,7							267	

Сравниваемые модели

1. ONet - модель, основанная на предложенной статье.
2. ResNet50
3. MyNet - собственная модель, основанная на идее углубления ONet для передачи на вход более крупного прямоугольника.

Метрики

1. Среднеквадратическая ошибка.

$$MSE = \frac{\sum^n (targets - preds)^2}{n \cdot norm}, \text{ где } norm = \sqrt{height \cdot width}$$

targets, preds - целевая и предсказанная координата,
height, width - высота и ширина прямоугольника, предсказанного детектором лиц dlib.

2. Среднее расстояние между предсказанными точками и целевыми

$$AvDist = \frac{\sum^n dist}{n \cdot norm}, \text{ где } dist = \sqrt{dx^2 + dy^2},$$

$$norm = \sqrt{(max(x) - min(x)) \cdot (max(y) - min(y))},$$

x, y - целевые координаты точек,

dx, dy - разность между целевыми и предсказанными координатами одной точки

Эксперименты

Onet со входом 48x48 пикселей показала худший результат, чем Onet 99x99, но дальнейшее увеличение входного изображения до 300x300 не улучшило результат. Тогда были проведены эксперименты для ResNet50, результаты существенно лучше, чем Onet, но увеличение входного изображения до 300x300 также не улучшило качество модели. Было решено создать свою модель, основываясь на идее углубить Onet для лучшего выделения признаков.

Помимо разного размера входного изображения изменялся формат координат, были проведены эксперименты как с нормированными [0..1] координатами, так и без нормировки. Наблюдения показали, что добавление сигмоиды (для нормированных координат) на выходе нейронной сети сглаживает валидационный график ошибки, но приводит к большей ошибке.

Для улучшения полученных значений на ResNet (99x99) был добавлен планировщик, управляющий скоростью обучения и аугментация данных (RandomCrop).

Результаты

Результаты сведены в таблицу 2. Расшифровка описания:

NxN - размер входа

n - нормировка координат, сигмоида на выходе

s - добавление планировщика, управляющего скоростью обучения

a - аугментация данных

nl - нормировка координат без сигмоиды (линейный выход)

Лучший результат был достигнут с использованием ResNet50, входом 99x99, планировщиком и аугментацией данных на эпохе обучения.

Таблица 2 - результаты обучения

				порог ошибки = 0.08				порог ошибки = 0.8			
				MSE		AvDist		MSE		AvDist	
модель	описание		ер	Menpo	300W	Menpo	300W	Menpo	300W	Menpo	300W
dlib		п		0,031	0,018	0,055	0,064				
		к		0,03	0,018	0,054	0,063	0,686	0,67	0,777	0,786
ONet	48x48	п	26	0,001	0	0,014	0,016				
	99x99n	к	14	0,003	0	0,02	0,022				
	99x99ns	к	19	0,003	0	0,02	0,021				
MyNet	300x300ns	к	31	0,004	0	0,022	0,023				
MyNet7	300x300ns	к	23	0,005	0	0,025	0,026	0,464	0,203	0,742	0,745
	300x300ns2	к	29	0,006	0	0,028	0,028	0,493	0,215	0,744	0,747
	300x300nsl	к	49	0,003	0	0,021	0,022	0,432	0,149	0,736	0,739
	300x300s	к	42	0,003	0	0,019	0,02	0,415	0,152	0,734	0,737
ResNet50	48x48	п	86	0,01	0	0,036	0,036				
	99x99	п	82	0,02	0,001	0,046	0,047				
	99x99s	к	42	0,019	0,002	0,044	0,044	0,616	0,391	0,766	0,766
	99x99sa	к		0,022	0,002	0,047	0,046	0,634	0,413	0,769	0,768
	99x99sn	к	54	0,006	0	0,029	0,027	0,5	0,198	0,748	0,744
	300x300	к	138	0,02	0,002	0,043	0,043	0,604	0,387	0,763	0,764

Ниже на рисунках 1, 2 представлены CED графики при пороге 0.008, дополнительные графики, соответствующие таблице, можно найти в папке CED.

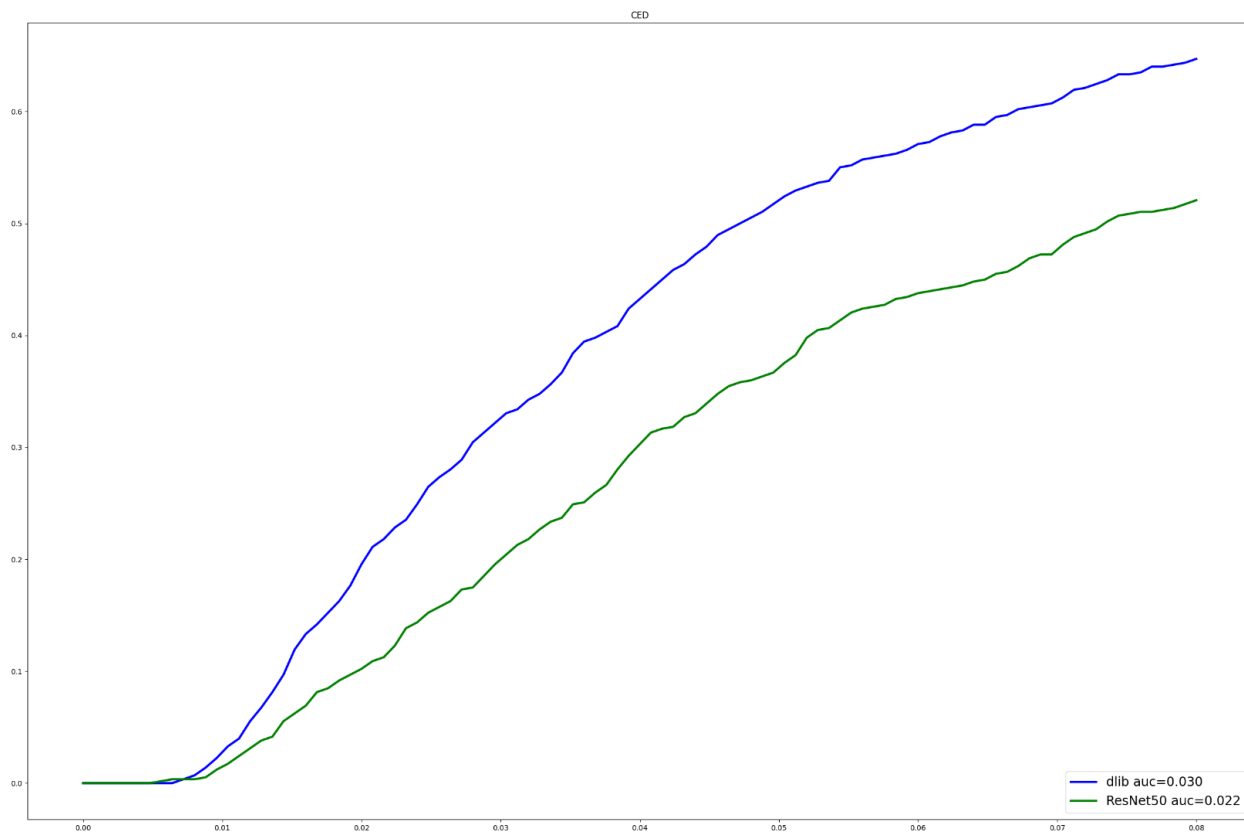


Рисунок 1 - CED график для датасета Метро

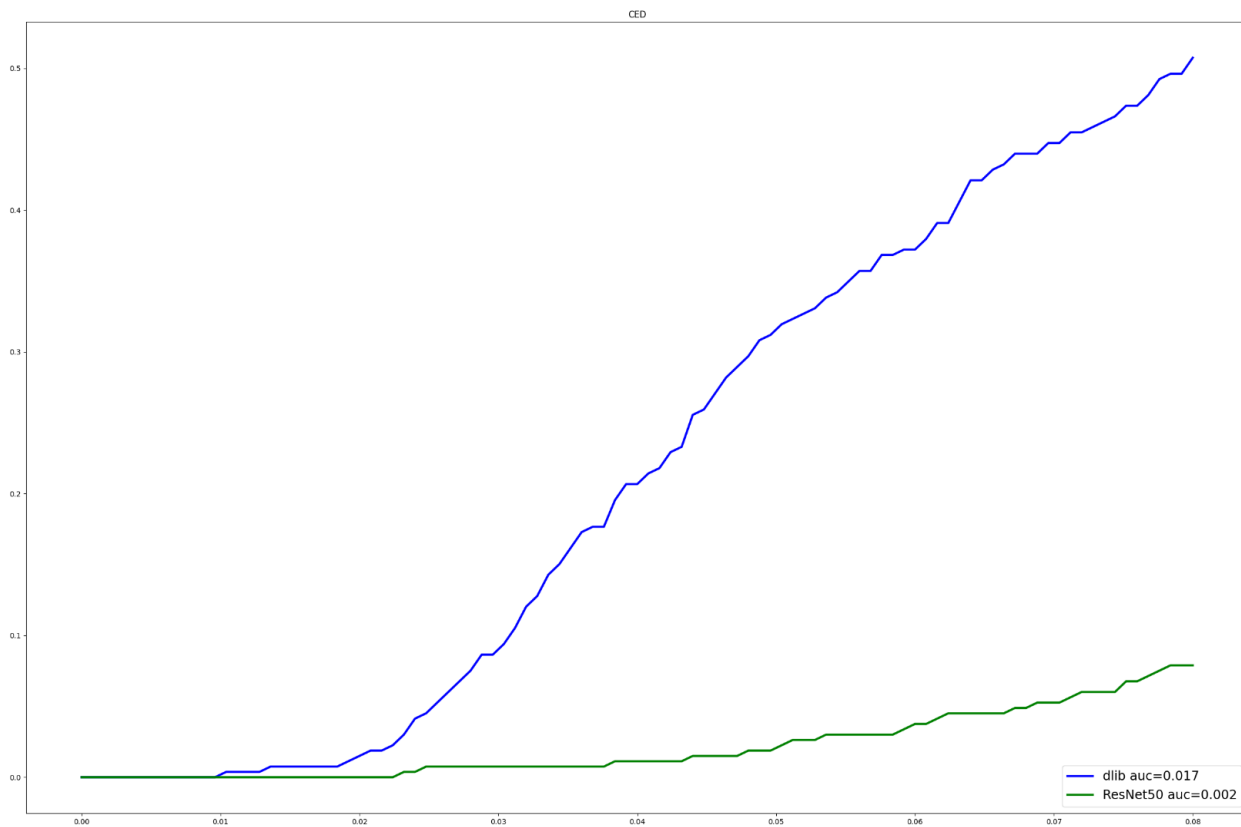


Рисунок 2 - CED график для датасета 300W

Скрипт

Для запуска обучения и тестирования достаточно запустить *train_test.py* с параметрами *menpo_path* и *w_path* (путь к датасетам). Напрмиер:

```
./train_test.py -m landmarks_task/Menpo -w landmarks_task/300W
```

Дополнительные параметры:

--description - описание модели

--epochs - число эпох для обучения

--num_workers - число потоков в загрузчике датасета

--batch_size