

State farm distracted driver detection(Kaggle)

Моржаков Василий

Мальцев Антон

Андранинов Николай

18 место / 1400 (2%)

C0: нормальное состояние



C1: смс, правая рука



C2: телефон, правая рука



С3: смс, левая рука



C4: телефон, левая рука



С5: тянется к радио



С6: что-то пьет



C7: тянется назад



C8: прическа, очки, нос



С9: разговаривает с пассажиром



Параметры базы и метрика

- Тренировочная база 10 классов, около 25000 изображений
- Тестовая база – 75000 изображений

Метрика – multiclass logloss

$$\text{logloss} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(p_{ij}),$$

Какая точность достижима человеком на тренировочной базе?

Всего 94-96%

Почему? C0/C9

Разговаривает с соседом? Нормальное состояние?



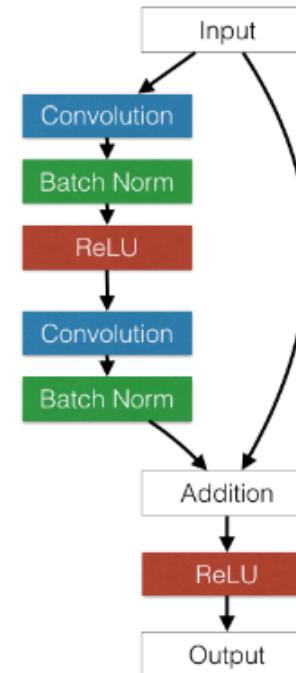
Решили участвовать

1. Повторить опубликованный результат Fine-tuning VGG16 в caffe

Не удалось

2. Использовать pretrained 224x224 на ImageNet Resnet 50

Удалось получить 0.25



Confusion matrix

x	0	1	2	3	4	5	6	7	8	9
0	0.723	0	0	0	0	0	0	0	0.014	0.284
1	0.008	0.972	0	0	0	0	0	0	0	0
2	0	0.009	0.99	0	0	0	0	0.02	0.041	0
3	0.042	0.009	0	0.981	0	0	0	0	0	0.025
4	0.008	0	0	0.019	0.991	0	0	0	0.027	0.012
5	0	0	0	0	0	1	0	0.01	0	0.025
6	0	0	0	0	0	0	0.972	0	0.027	0
7	0	0	0	0	0	0	0	0.96	0.014	0.012
8	0.05	0.009	0.01	0	0	0	0.028	0	0.878	0.099
9	0.168	0	0	0	0.009	0	0	0.01	0	0.543

Можно признаки отличать по лицу



Pretrained
Resnet100
C0,C8,C9

Помогло на
0.01

Поиск лица:

- Haar/Hog
- CNN
- SegNet (или другие сегментационные сети)

Что помогло

Приращения/изменения выборки



Кроп до 15% с каждой
стороны



Поворот до 15 градусов

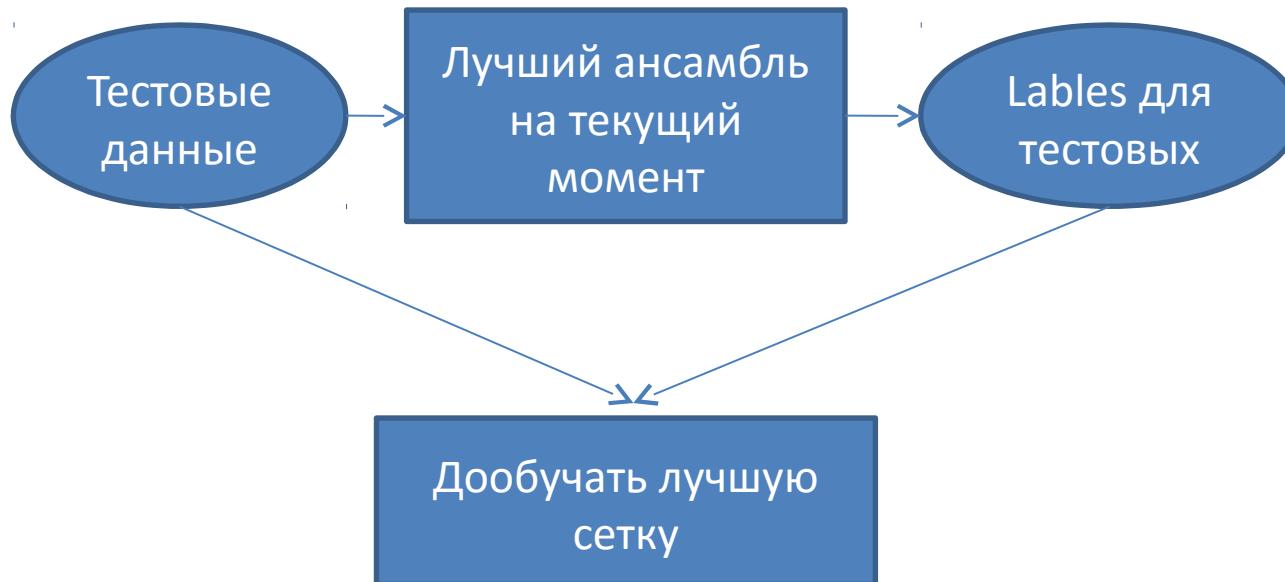


Зашумление
данных



Что помогло

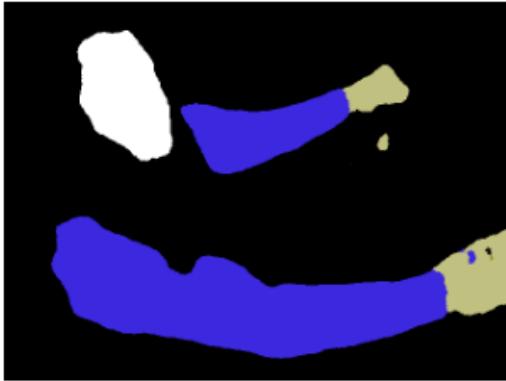
Pseudo-lables



Ансамбль (арифметическое среднее)

- Худшие: resnet 50 0.25, resnet 100 (50/50)
- Лучше: resnet 100 + augmentation, худшие (50/50) (около 0.18)
- Добавили лицо (см.дальше) по классам 0,8,9 (50/50) (около 0.17)
- Pseudo-lables + все остальное (50/50) (около 0.16)

Что не помогло



SegNET (+ закрыт)

Метод перспективный для
редукции информации на
входном изображении.



Работало несколько хуже,

т.к.:

- 1) Выражение лица
- 2) Козырек
- 3) Руль и положение руки
относительно него для
определения класса
«Радио»

Головастики



Не помогло

А на что обучилась сеть?



Простой способ определить на что обучилась сеть



Двигаем
прямоугольник
и смотрим, как
изменяется
выходной слой

Не на то

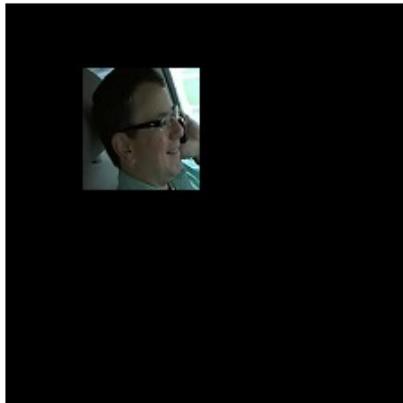


Здесь фокус внимания должен быть на лице – обернулся говорить с пассажиром, но нет

Примешаем в обучающую выборку

90% только лица, +20% полных картинок за эпоху до 50/50

Примешали:



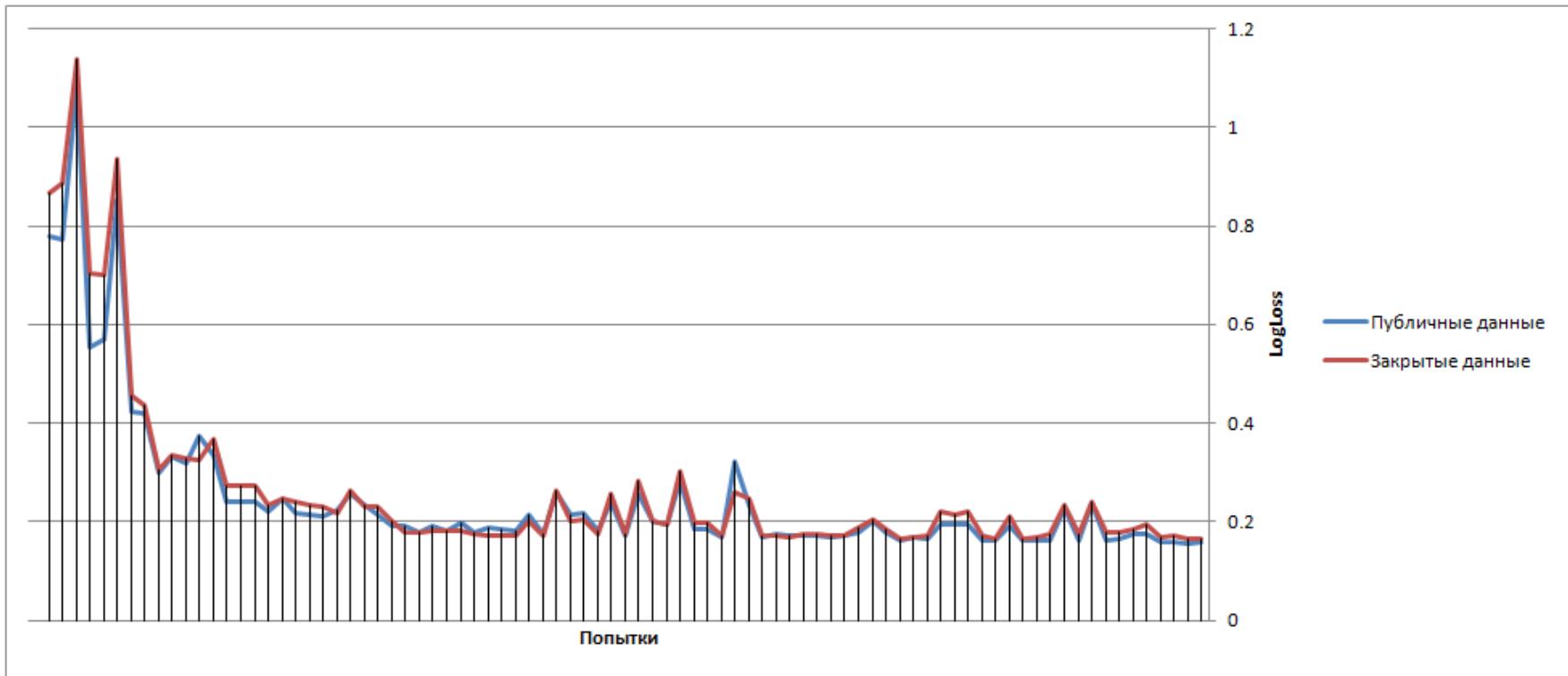
Стало:



И НЕ ПОМОГЛО

Но подход очень полезный!

График результатов



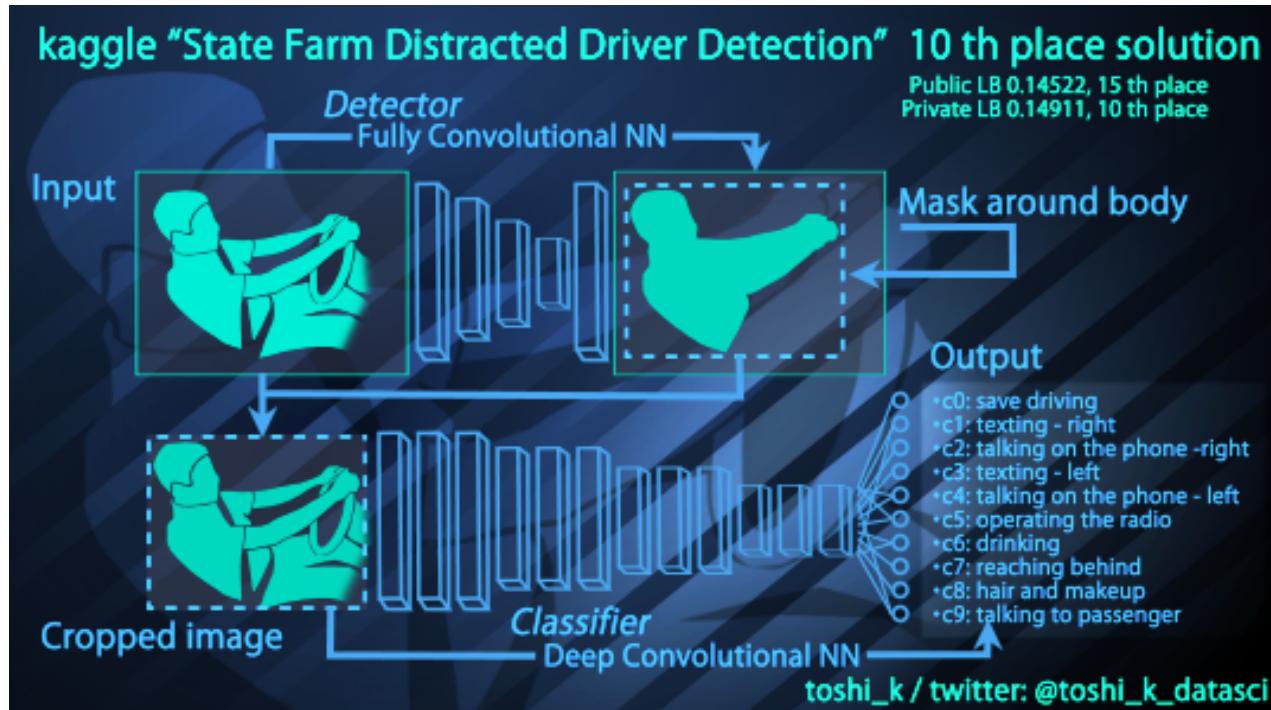
Что мы не сделали

- Не сделали Folds
- Вычитание фона
- В батч не вносили случайных водителей
- не выделяли информацию о водителе, чтобы разделить фичи, касающиеся водителя и его состояния

15 место

- 150 обученных моделей
- Используется область лица
- Используется область руля

10 место



5 место



Интересный
Augmentation,

5млн
изображений

10 титанов

3 и 1 решения



Изображения — это нарезанное видео. Можно собрать его обратно, выбрав наиболее похожие изображения. И выносить решение в среднем по всему классу.

-Третье место: по 5 ближайшим. Поиск за счёт изображения 80×60 , KNN

-Первое место: по 10 ближайшим, поиск не описан

А полезны ли решения?

Как делать законченную систему анализа состояния водителя?



Полезно ли на практике решение, которое ищет ближайшие 5 кадров? Может проще и точнее сделать анализ видео?
Та ли это система где вы хотите использовать 150 сетей?

Возможно ли реальное применение?

Правильно ли собрана база и как далеко полученное решение от реальности?

1) При наборе базы водители машину не водили. Машина ехала на грузовике, водителям давались задания что делать. Водители не крутили головой, не смотрели на неудобно расположенные светофоры, и.т.д. Итоговый результат распознавания таких ситуаций — разговор с соседом:



Возможно ли реальное применение?

2) Вся база набиралась в дневные часы. Судя по всему в рабочие. Нет низкого солнца. Нет вечерних кадров. Сетка начинается путаться. Веса на нейронах гуляют.



Про ночь нет смысла говорить. При ИК освещении всё выглядит иначе

Возможно ли реальное применение?

3) База набиралась в Америке. Вы спросите как я понял? Нет ни одной машины с ручной коробкой...



Вот так вот я тянусь назад. Вероятно за лопатой.