

Конкурс Avito-2016:
«Распознавание марки и модели
автомашин на изображениях»

Евгений Нижибицкий

ВМК МГУ, Rambler&Co

28 мая 2016 г.

Вместо введения

Описание задачи

- Цель конкурса — создание эффективного алгоритма классификации изображений автомобилей
- «Эффективность» измеряется в Accuracy



Марка: Land Rover

Модель: Range Rover Evoque



Марка: Mercedes-Benz

Модель: С-класс

Вместо введения

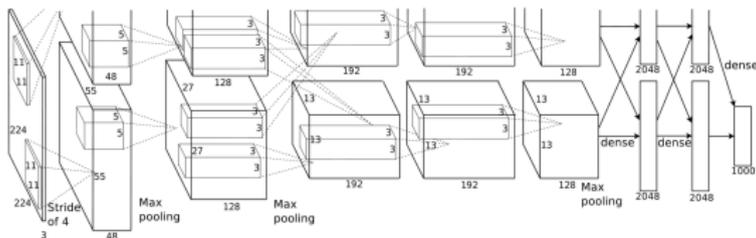
Особенности данных

- Три выборки по 310000, 92000 и 217000 картинок (A/B/C)
- В использовалась для «public leaderboard»'а до 1-го дедлайна
- После 1-го дедлайна ответы для В тоже были выданы
- С использовалась для «private leaderboard»'а
- Всего 236 классов, но все они безымянные

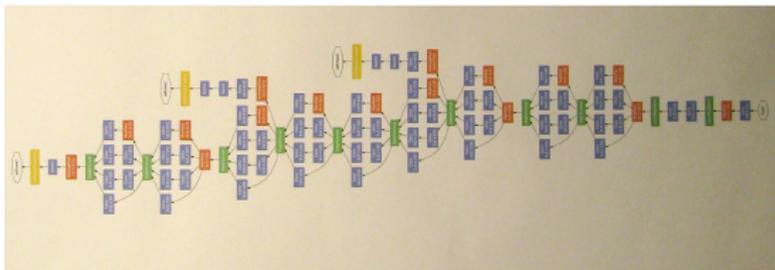


Вместо введения

Нейросети наше всё



- Обученный на ImageNet человек — 5%
- 4x144 Inception-v3¹ (Декабрь 2015) — 3.58%



¹Rethinking the Inception Architecture for Computer Vision



- Intel i7-4790K
- GeForce GTX 980 Ti
- SSD 500 MB/sec

RAMBLER&Co

- Intel i7-6700
- 2× GeForce GTX TITAN X
- SSD 500 MB/sec

Подходы к решению

Обзор моделей I

Model	Source	Scheme	Top5	Top1
AlexNet	caffe	1x1	19.80	42.90
GoogLeNet	caffe	1x1	11.10	31.30
ResNet-18	torch	1x1	10.76	30.43
GoogLeNet	pdf	1x1	10.07	-
ResNet-18	torch	1x10	9.42	28.22
GoogLeNet	pdf	1x10	9.15	-
ResNet-34	torch	1x1	8.74	26.73
Inception-v3	tf	1x1	-	26.50
GoogLeNet	pdf	7x1	8.09	-
GoogLeNet	pdf	1x144	7.89	-
GoogLeNet	pdf	7x10	7.62	-

Подходы к решению

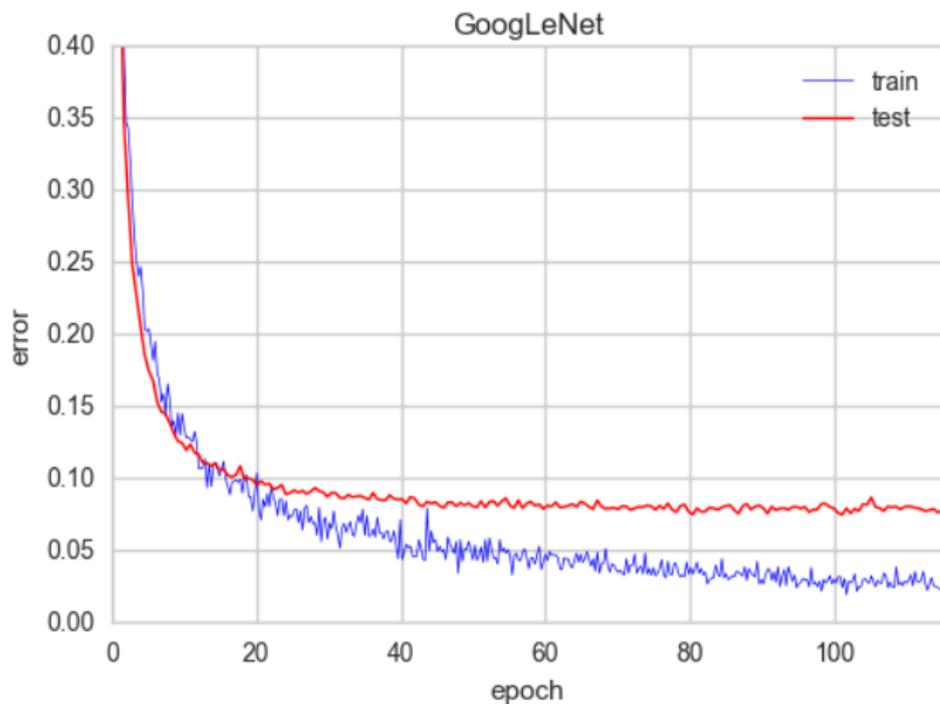
Обзор моделей II

Model	Source	Scheme	Top5	Top1
GoogLeNet	pdf	7x10	7.62	-
ResNet-34	torch	1x10	7.35	24.76
ResNet-50	torch	1x1	7.02	24.01
GoogLeNet	pdf	7x144	6.67	-
ResNet-50	torch	1x10	6.21	22.44
ResNet-101	torch	1x1	6.08	22.24
Inception-v3	pdf	1x1	5.60	21.20
ResNet-101	torch	1x10	5.35	21.08
Inception-v3	pdf	1x10	4.48	19.47
Inception-v3	pdf	1x144	4.20	18.77
Inception-v3	pdf	4x144	3.58	17.20

Подходы к решению

GoogLeNet

Обучение на 90% выборки A с валидацией на остатке:



Подходы к решению

GoogleNet

Качество на 10%-валидации:

Итерации	Кропы	Точность
400k	1x1	91.06
400k+200k	1x1	92.13
400k	1x10	92.63
400k+200k	1x10	93.54
400k+200k+500k	1x10	93.71

Сделанные посылки:

Итерации	Кропы	Точность
200k	1x1	89.78
400k+200k	1x10	92.89

Подходы к решению

Небольшое отступление

После GoogleNet в последующих моделях на большинстве графиков будет рассматриваться качество для такого разбиения подвыборок:

- 384000 изображений — обучение
- 12000 изображений — валидация
- ~7000 изображений — проверка скриптов для теста

Подходы к решению

ResNet-34 / Torch

Microsoft Research Asia — Deep Residual Learning for Image Recognition

- <https://github.com/facebook/fb.resnet.torch>

Что делаем:

- Используем 256x256 без сохранения пропорций
- Используем стандартную стратегию LR из репозитория
- Модифицируем несколько lua-файлов для получения возможности сохранять айдишники картинок вместе с содержимым и записывать вероятности в HDF5-файлы
- В остальном все в основном как в инструкции по ссылке

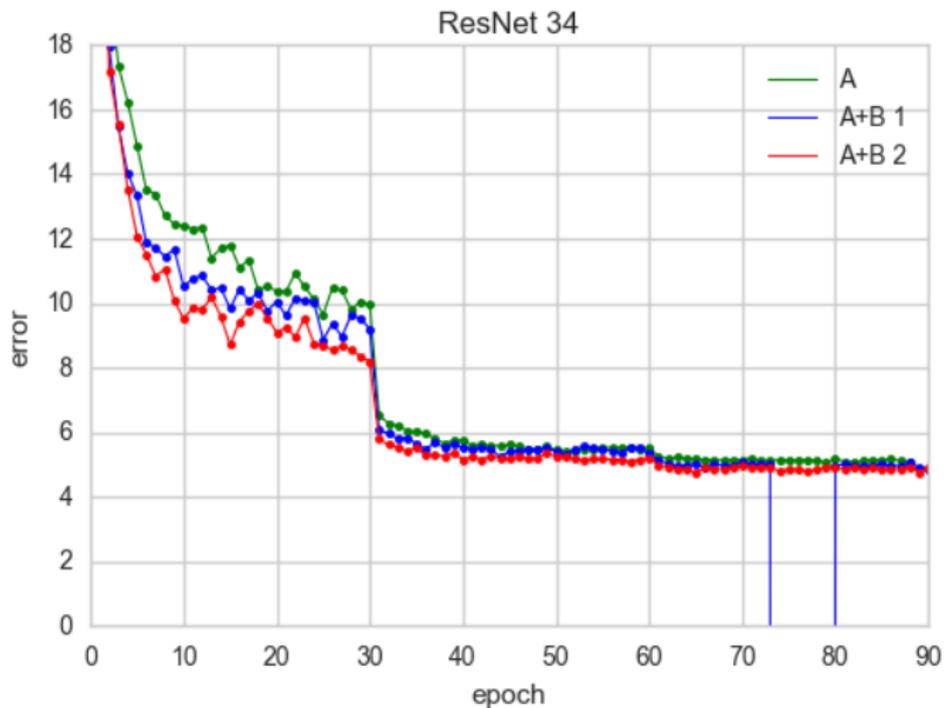
Что не понравилось:

- По умолчанию (частые) дампы моделей занимают гигабайты
- Пришлось написать adhoc-скрипт в crontab

Подходы к решению

ResNet-34 / Torch

Качество одной модели ResNet-34 на валидации



Подходы к решению

Inception-v3 / Tensorflow

Google Inc. — Rethinking the Inception Architecture for Computer Vision

- <https://github.com/tensorflow/models>

Что делаем:

- Используем исходные картинки — делаем бинарные шарды
- Делаем много своих файлов на основе таковых для имаджнета с более мягкими условиями искажений для создания картинок на лету, добавляем возможность сохранять айдишники картинок вместе с содержимым и записывать вероятности в HDF5-файлы
- Разные интервалы для генерации искажений уменьшаем раза в 2
- Оставляем как минимум 40% исходных изображений (а не 8!)

Что не понравилось:

- Не удалось получить валидную выдачу всех предсказаний с id
- Пришлось гонять сессии tensorflow с 1 картинкой

Подходы к решению

Inception-v3 / Tensorflow

Качество одной модели Inception-v3 на валидации

Итерации	Точность
5000	0.7735
10000	0.9031
15000	0.9262
20000	0.9347
30000	0.9446
60000	0.9518
90000	0.9544
120000	0.9571

Подходы к решению

Объединение различных моделей

Качество для различных смесей (рассматривались различные сценарии того, что успеет обучиться и предсказаться):

Модель	Точность
ResNet34 _{1-crop}	0.9497
ResNet34 _{10-crop}	0.9537
2 × ResNet34 _{10-crop}	0.9559
Inception-v3 _{1-crop}	0.9564
Inception-v3 _{10-crop}	0.9588
$\sqrt{\text{ResNet34}_{10\text{-crop}} \cdot \text{Inception-v3}_{1\text{-crop}}}$	0.9594
$\sqrt{\text{ResNet34}_{10\text{-crop}} \cdot \text{Inception-v3}_{10\text{-crop}}}$	0.9608
2 × Inception-v3 _{10-crop}	0.9613
$\sqrt{2 \times \text{ResNet34}_{10\text{-crop}} \cdot 2 \times \text{Inception-v3}_{10\text{-crop}}}$	0.9627

Результат

Итоговая модель

Предсказания — $\arg \max$ вероятностей, полученных по формуле

$$\text{ResNet34}_{10\text{-crop}}^{0.15} \cdot \text{ResNet34}_{10\text{-crop}}^{0.15} \cdot \text{Inception-v3}_{10\text{-crop}}^{0.35} \cdot \text{Inception-v3}_{10\text{-crop}}^{0.35}$$



Результат

Вычислительная сложность

ResNet-34:

- 32 картинки в батче
- 3250MB памяти на батч
- 12000 батчей на эпоху
- 1 эпоха за 45 минут
- Полное обучение (90 эпох) — 60 часов

Inception-v3:

- 32 картинки в батче
- 11750MB памяти на батч
- 12000 батчей на эпоху
- 1 эпоха за 4 часа
- Полное обучение (12 эпох) — 48 часов

Результат

Финальный рейтинг

#	Никнейм участника	Результат
1	Евгений Нижибицкий	0.9615
2	Lesseps	0.9557
3	n01z3	0.9405
4	BucketNet	0.9355
5	rasim	0.9339
6	kwentar	0.9229
7	sovcharenko	0.9092
8	Expasoft	0.9000
9	Osborn	0.8840

Спасибо!