

НЕЙРОСЕТЕВОЙ ПОДХОД К РАСПОЗНАВАНИЮ МОДЕЛИ АВТОМАШИН НА ИЗОБРАЖЕНИЯХ

Нижбицкий Евгений Алексеевич

Аспирант

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: nizhibitsky@cs.msu.ru

За последние несколько лет методы глубокого обучения получили широкое распространение и стали эталоном качества сразу в нескольких типах прикладных задач, включая классификацию изображений, текстов и звуков. Так, в задаче категоризации изображений за последние 4 года они не только оставили позади «классические» подходы [1], но и начали справляться лучше человека [2].

В данной работе исследуется адаптация моделей, на 2016-й являвшихся одними из лучших в задаче классификации изображений на наборе данных ImageNet — Inception-v3 от исследователей компании Google и ResNet от исследователей из лаборатории Microsoft Research Asia. Модели дообучаются (используя технику transfer learning) для решения задачи по созданию автоматического рубрикатора объявлений о продаже автомобилей, представленной на конкурсе, организованном компанией Avito в 2016-м году [3].

В качестве данных для экспериментов используется размеченный организаторами набор изображений с автомобилями 236-ти не разглашаемых марок. Всего было представлено 3 массива изображений — 309 710 изображений для обучения, 92 667 файлов для валидации и около 217 тысяч тестовых изображений. Метрикой качества была точность классификации (accuarcy). Базовым её значением для задачи были указаны **83%**, полученные на основе одной из моделей репозитория Caffe (скорее всего это была адаптированная AlexNet).

При первом подходе к решению данной задачи была дообучена на представленных данных предобученная модель GoogleNet также из репозитория Caffe. Всего 600 000 итераций обучения с уменьшением параметра скорости позволили достигнуть точности в **92.89%**.

Затем из двух представленных размеченными массивов изображений были выделены 90% данных для обучения и 10% для контроля. С таким разбиением удалось успеть в рамках конкурса обучить еще две модели с доступными реализациями и весами, предобученными на ImageNet — Inception-v3, реализованная с применением tensorflow/python, и ResNet-34 (с применением torch/luа).

Каждая из двух моделей ResNet-34 после 90 эпох дообучения с

параметрами по умолчанию достигла на валидации точности **95%**. Одна дообученная модель Inception-v3 показала точность **95.64%**.

В то время как еще не все модели успели обучиться, рассматривались различные сценарии того, сколько моделей смогут успеть обучиться и какую часть из них возможно будет успеть применить на тестовой выборке (на это отводилось трое суток). Качество различных комбинаций моделей и подходов к предсказанию приведено в таблице 1. Под обозначением “ $2 \times \text{MODEL}$ ” понимается геометрическое среднее предсказаний двух идентичных по строению моделей, обученных с разной случайной инициализацией. “1-crop” и “10-crop” — подходы, при которых предсказание делается на основе исходного изображения или усреднения полученных вероятностей для различных участков исходного и отраженного изображений соответственно.

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

Таблица 1: Качество композиций из моделей на 10%-контроле.

Лучшее качество на контроле, равное **96.27%**, ожидаемо доставили комбинация и усреднение предсказаний по нескольким участкам изображений для обученных пар моделей. Именно в таком виде и удалось сделать итоговое предсказание, что позволило автору занять первое место в конкурсе с точностью **96.15%** на тестовой выборке.

Литература

1. Krizhevsky A. ImageNet Classification with Deep Convolutional Neural Networks // Advances in Neural Information Processing Systems 25, 2012, P. 1097–1105
2. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition // CVPR, 2016.
3. Страница конкурса «Avito-2016: Распознавание марки и модели автомашин на изображениях»: <http://www.machinelearning.ru/wiki/index.php?title=Avito-2016-Cars>