

Детекция и классификация марок и моделей автомобиля на изображении

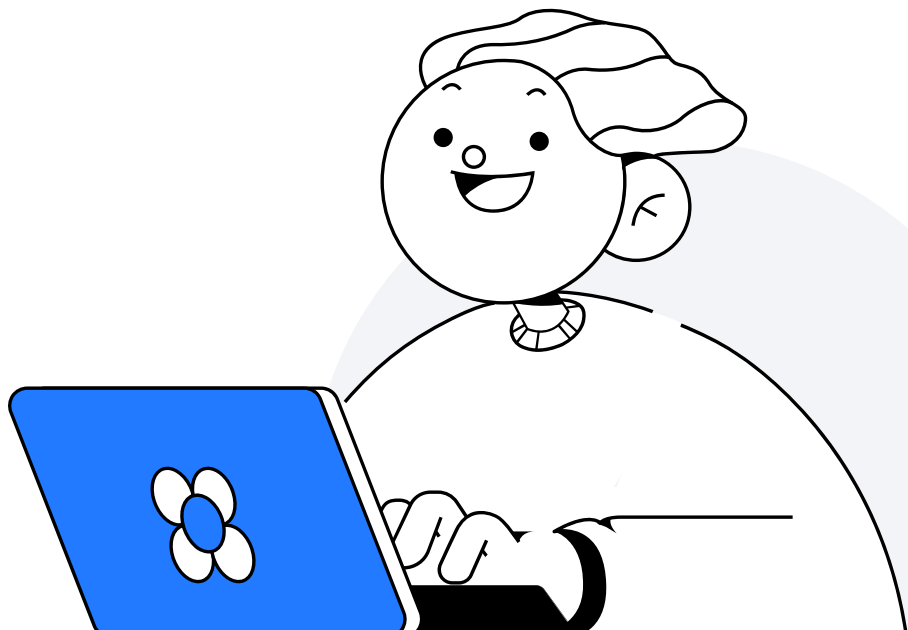
Итоговая работа по курсу «Deep Learning» (профессия «Data Scientist: с нуля до middle»)

Александр Воробьев
Группа: DSU-31



Содержание

- 1 Постановка задачи для глубокого обучения
- 2 Анализ данных
- 3 Методика реализации
- 4 Итоги обучения модели
- 5 Выводы



Постановка задачи для глубокого обучения



1

**Исходная
задача:**

С целью повышения эффективности процесса контроля в данном проекте реализована автоматизация анализа изображения автомобиля на картинке с помощью нейронных сетей (далее - НС) по следующим направлениям:

- детекция (выявление наличия автомобиля и его местоположения на источнике данных – фотоснимке);
- классификация (отнесение выявленного автомобиля к одной из категорий).

**Актуальность
задачи, ее место
в предметной
области:**

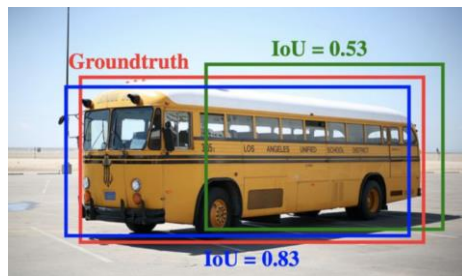
- В настоящее время контроль приходящих фото через Яндекс feeds со стороны дилеров автобрендов публикуемые отделами оценки и подготовки автомобилей не осуществляется. Они автоматически загружаются на платформы агрегаторов и иных коммерческих организаций, которые видит свои каталоги по продаже на своих сайтах.
- В рамках 2023 года было около 5-7% не корректно загруженных фото автомобилей от общего объема загрузок Яндекс feeds. Данные ошибки были связаны с человеческим фактором. Типы ошибок:
 1. Не корректно загружены марка и модель;
 2. Не корректный цвет автомобиля.



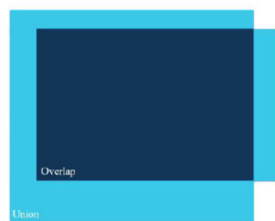
Целевые метрики:

Основная: mAP(mean Average Precision) с IoU=0.5 (**mAP@0.5**), т.е. среднее значение Precision при различных Recall, с условием перекрытия (Intersection-over-Union) не менее 0.5, где:

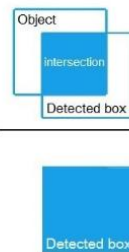
- Precision = True positive / (True positive + False positive)
- Recall = True positive / (True positive + False negative)



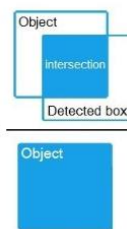
$IoU = \frac{\text{area of overlap}}{\text{area of union}}$



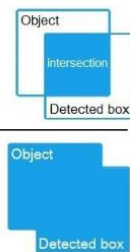
Precision =



Recall =

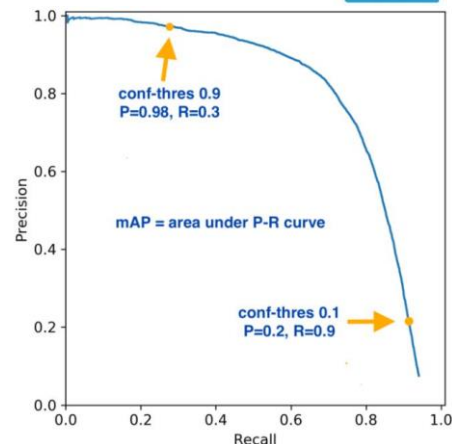


$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}}$



Дополнительная: **F1-score**
(гармоническое среднее между Precision и Recall)

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$



Анализ данных



2

Источник данных: полный датасет с фотографиями, общедоступный <https://universe.roboflow.com/mxk/car-model-detection/dataset/1>

Набор данных: 8 132 (315mb) цветных фотографий, сделанных в хороших погодных условиях, при отсутствии осадков (дождя, снега и иные):

- **Train-датасет: 6 506** (80%) размеченных в формате PASCAL VOC. К каждому jpg-файлу есть xml-файл разметки bounding boxes с указанием автомобиля (марка, модель и год);
- **Val-датасет: 1 626** (20%) - без разметки. В рамках дипломного проекта использовался для визуальной оценки точности предсказания (инференсов) моделей.



Классификация автомобилей (все классы):

Hyundai Accent Sedan 2012
FIAT 500 Abarth 2012
Chevrolet Express Cargo Van 2007
Maybach Landaulet Convertible 2012
Acura RL Sedan 2012
Rolls-Royce Phantom Drophead Coupe Convertible 2012
Aston Martin Virage Convertible 2012
Infiniti QX56 SUV 2011
Bugatti Veyron 16-4 Convertible 2009
FIAT 500 Convertible 2012
Infiniti G Coupe IPL 2012
Hyundai Sonata Hybrid Sedan 2012
BMW ActiveHybrid 5 Sedan 2012
Chevrolet Express Van 2007
Bentley Continental GT Coupe 2012
Buick Regal GS 2012
BMW 1 Series Convertible 2012
Lamborghini Gallardo LP 570-4 Superleggera 2012
Mazda Tribute SUV 2011
Lamborghini Reventon Coupe 2008
Bentley Mulsanne Sedan 2011
Mercedes-Benz SL-Class Coupe 2009
Audi RS 4 Convertible 2008
Chevrolet HHR SS 2010
Bentley Continental Supersports Conv- Convertible 2012
MINI Cooper Roadster Convertible 2012
Chevrolet Tahoe Hybrid SUV 2012
Chevrolet Silverado 2500HD Regular Cab 2012
Suzuki Aerio Sedan 2007
Aston Martin Virage Coupe 2012
Rolls-Royce Ghost Sedan 2012
Buick Verano Sedan 2012
Chrysler Town and Country Minivan 2012
Ford E-Series Wagon Van 2012
Chevrolet Corvette Ron Fellows Edition Z06 2007
Toyota Sequoia SUV 2012
Dodge Challenger SRT8 2011
Honda Accord Sedan 2012
Tesla Model S Sedan 2012
Ferrari California Convertible 2012
Acura ZDX Hatchback 2012
BMW X3 SUV 2012
Nissan NV Passenger Van 2012
Bentley Arnage Sedan 2009
Lincoln Town Car Sedan 2011
HUMMER H3T Crew Cab 2010
Dodge Dakota Club Cab 2007
Honda Accord Coupe 2012
Chevrolet Malibu Hybrid Sedan 2010

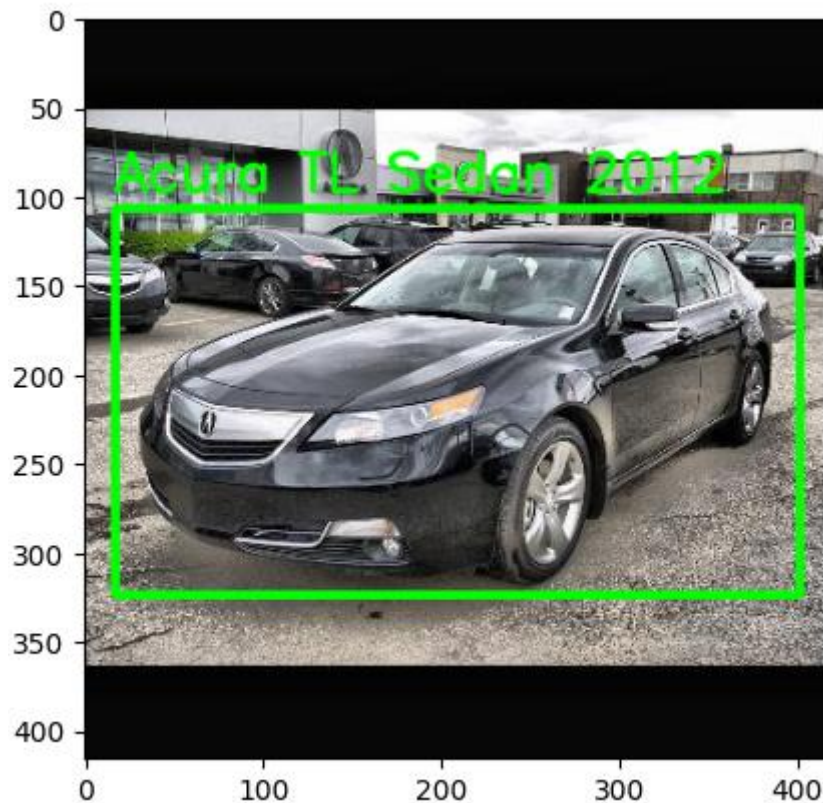
smart fortwo Convertible 2012
Chevrolet Silverado 1500 Hybrid Crew Cab 2012
Chevrolet TrailBlazer SS 2009
Dodge Magnum Wagon 2008
Dodge Sprinter Cargo Van 2009
Audi TT RS Coupe 2012
Hyundai Sonata Sedan 2012
Ferrari 458 Italia Convertible 2012
Isuzu Ascender SUV 2008
Chevrolet Corvette Convertible 2012
GMC Canyon Extended Cab 2012
Audi S4 Sedan 2012
Audi TTS Coupe 2012
Toyota 4Runner SUV 2012
Mercedes-Benz Sprinter Van 2012
BMW M5 Sedan 2010
Audi A5 Coupe 2012
Chrysler Sebring Convertible 2010
Honda Odyssey Minivan 2007
Dodge Charger Sedan 2012
Audi 100 Sedan 1994
BMW 1 Series Coupe 2012
Hyundai Veloster Hatchback 2012
BMW M6 Convertible 2010
Cadillac SRX SUV 2012
Dodge Dakota Crew Cab 2010
Audi TT Hatchback 2011
Dodge Caliber Wagon 2012
Acura TSX Sedan 2012
BMW Z4 Convertible 2012
Aston Martin V8 Vantage Coupe 2012
Dodge Caliber Wagon 2007
Suzuki SX4 Sedan 2012
Ram C-V Cargo Van Minivan 2012
Buick Enclave SUV 2012
Hyundai Veracruz SUV 2012
Ferrari FF Coupe 2012
Scion xD Hatchback 2012
Chevrolet Cobalt SS 2010
Volvo C30 Hatchback 2012
Audi S5 Convertible 2012
Ford F-450 Super Duty Crew Cab 2012
Hyundai Elantra Sedan 2007
Hyundai Santa Fe SUV 2012
Hyundai Azera Sedan 2012
Acura TL Type-S 2008
Dodge Charger SRT-8 2009
Nissan Leaf Hatchback 2012
Ford Ranger SuperCab 2011

BMW X6 SUV 2012
BMW 3 Series Wagon 2012
Suzuki SX4 Hatchback 2012
BMW X5 SUV 2007
GMC Terrain SUV 2012
Honda Odyssey Minivan 2012
Ford F-150 Regular Cab 2012
Audi S5 Coupe 2012
Dodge Ram Pickup 3500 Crew Cab 2010
Ford Fiesta Sedan 2012
Audi 100 Wagon 1994
Chrysler Crossfire Convertible 2008
Volvo XC90 SUV 2007
Cadillac CTS-V Sedan 2012
Land Rover Range Rover SUV 2012
Chevrolet Silverado 1500 Classic Extended Cab 2007
Volkswagen Golf Hatchback 2012
Land Rover LR2 SUV 2012
Ferrari 458 Italia Coupe 2012
Chevrolet Silverado 1500 Regular Cab 2012
Jeep Compass SUV 2012
Jeep Patriot SUV 2012
Buick Rainier SUV 2007
HUMMER H2 SUT Crew Cab 2009
Jeep Wrangler SUV 2012
Acura TL Sedan 2012
Audi R8 Coupe 2012
Hyundai Elantra Touring Hatchback 2012
Bugatti Veyron 16-4 Coupe 2009
Spyker C8 Coupe 2009
Volkswagen Beetle Hatchback 2012
Chevrolet Impala Sedan 2007
GMC Yukon Hybrid SUV 2012
Ford Edge SUV 2012
BMW 3 Series Sedan 2012
Plymouth Neon Coupe 1999
BMW 6 Series Convertible 2007
Chevrolet Silverado 1500 Extended Cab 2012
Chevrolet Traverse SUV 2012
Audi V8 Sedan 1994
Ford Freestar Minivan 2007
Rolls-Royce Phantom Sedan 2012
Nissan Juke Hatchback 2012
Dodge Durango SUV 2012
Toyota Corolla Sedan 2012
Dodge Journey SUV 2012
Hyundai Genesis Sedan 2012
McLaren MP4-12C Coupe 2012
Lamborghini Aventador Coupe 2012

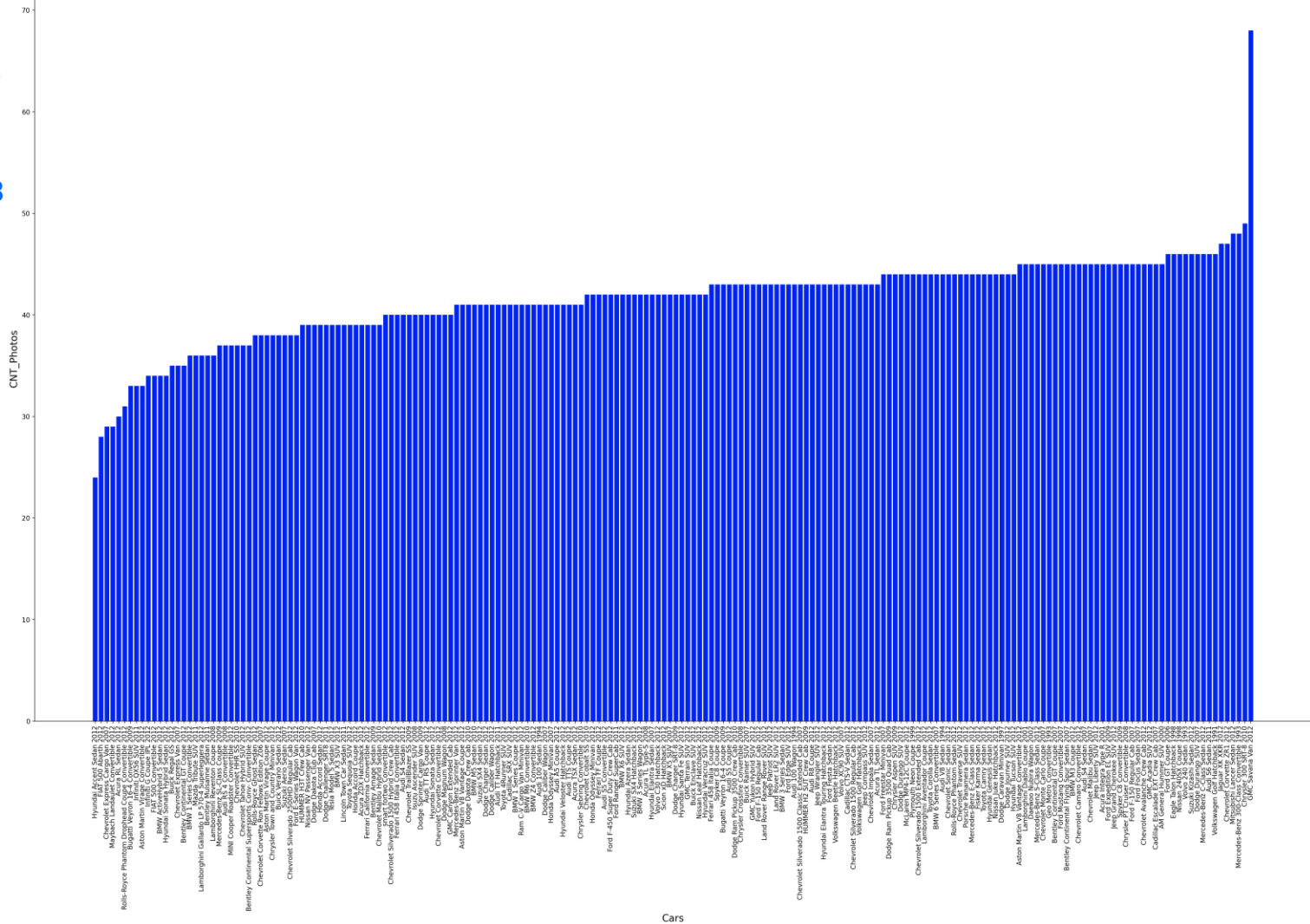
Mercedes-Benz E-Class Sedan 2012
Chrysler Aspen SUV 2009
Dodge Caravan Minivan 1997
Fisker Karma Sedan 2012
Chevrolet Sonic Sedan 2012
Hyundai Tucson SUV 2012
Toyota Camry Sedan 2012
Porsche Panamera Sedan 2012
Dodge Ram Pickup 3500 Quad Cab 2009
Chevrolet Camaro Convertible 2012
Chevrolet Monte Carlo Coupe 2007
GMC Acadia SUV 2012
Audi S4 Sedan 2007
AM General Hummer SUV 2000
Lamborghini Diablo Coupe 2001
Chevrolet Malibu Sedan 2007
Mercedes-Benz S-Class Sedan 2012
Daewoo Nubira Wagon 2002
Bentley Continental GT Coupe 2007
Chrysler PT Cruiser Convertible 2008
Chevrolet Avalanche Crew Cab 2012
Spyker C8 Convertible 2009
Jeep Grand Cherokee SUV 2012
Geo Metro Convertible 1993
Cadillac Escalade EXT Crew Cab 2007
Ford Focus Sedan 2007
Aston Martin V8 Vantage Convertible 2012
Ford F-150 Regular Cab 2007
Bentley Continental Flying Spur Sedan 2007
Jeep Liberty SUV 2012
Acura Integra Type R 2001
Ford Expedition EL SUV 2009
BMW M3 Coupe 2012
Ford Mustang Convertible 2007
Nissan 240SX Coupe 1998
Audi S6 Sedan 2011
Volkswagen Golf Hatchback 1991
Eagle Talon Hatchback 1998
Mercedes-Benz C-Class Sedan 2012
Suzuki Kizashi Sedan 2012
Volvo 240 Sedan 1993
Dodge Durango SUV 2007
Ford GT Coupe 2006
Jaguar XK XKR 2012
Chevrolet Corvette ZR1 2012
Mercedes-Benz 300-Class Convertible 1993
Mitsubishi Lancer Sedan 2012
Chrysler 300 SRT-8 2010
GMC Savana Van 2012



Примеры на фотографиях из размеченного (Train) датасета:



Распределение
автомобилей
по классам в
размеченном
(train) датасете:



Основной вывод по результатам разведочного анализа данных:

- Классы немного не сбалансированы между собой.
- Весь датасет имеет полную разметку для всех изображений.



Методика реализации



3

- YOLOv8 Backbone**

The backbone consists of 9 stages (P1 to P9) that process the input image. The output of the backbone is a feature map of size $640 \times 640 \times 3$.

YOLOv8 Head

The head consists of 3 stages (P3 to P5) that process the feature map. The output of the head is a feature map of size $640 \times 640 \times 3$.

Details

The details section shows the internal structure of the backbone and head layers, including Conv, C2f, SPPF, and Bottleneck layers. The diagram illustrates the flow of data through these layers, with dimensions and stride values provided for each stage.

Model Configuration Table

model	d (depth, multiple)	w (width, multiple)	r (ratio)
n	0.33	0.25	2.0
s	0.33	0.50	2.0
m	0.67	0.75	1.5
l	1.00	1.00	1.0
x	1.00	1.25	1.0

Backbone Layers

 - P1: Conv, $k=3, s=2, p=1$
 - P2: Conv, $k=3, s=2, p=1$
 - P3: C2f, shortcut=True, $n=3 \times d$
 - P4: Conv, $k=3, s=2, p=1$
 - P5: C2f, shortcut=True, $n=6 \times d$
 - P6: Conv, $k=3, s=2, p=1$
 - P7: C2f, shortcut=True, $n=6 \times d$
 - P8: Conv, $k=3, s=2, p=1$
 - P9: C2f, shortcut=True, $n=3 \times d$
 - SPPF

Head Layers

 - P3: Conv, $k=3, s=1, p=0$
 - P4: Conv, $k=3, s=1, p=0$
 - P5: Conv, $k=3, s=1, p=0$

Loss Calculation

The loss is calculated using the Bbox, Cls, and BCE components. The Bbox loss is calculated as $\text{Bbox} = \text{d} \times \text{reg} \times \text{max}$. The Cls loss is calculated as $\text{Cls} = \text{nc}$. The BCE loss is calculated as BCE .

Шаги в ходе подготовки данных и обучения НС на YOLOv8:

1. Размеченный (train) датасет вручную разбит на трейн-(80%) и тест-(20%) выборки с сохранением %-ного соотношения между автомобилями;
2. Разметка файлов сконvertирована из формата PASCAL VOC (XML) в формат YOLO (TXT, координаты bounding boxes отнормированы по центру изображения, все классы объектов сохранены);
3. Склонирован с github репозиторий Ultralytics, содержащий pretrained-модели YOLOv8 разной степени сложности (от Small до Large), с целью дальнейшего обучения на нашем ограниченном датасете;
4. Сформирован настроечный YAML-файл (пути к данным + перечисление классов);
5. Проведено предварительное обучение на малом датасете (данные только по Чехии) с целью выбора подходящей модели и гиперпараметров для всего датасета. Для контроля процесса обучения использовался Tensorboard, а также файлы с метриками, формируемые непосредственно моделью YOLOv8 в ходе обучения;
6. После подбора подходящей модели и гиперпараметров проведено обучение на всем датасете (по всем странам);
7. Проведен референс модели на тестовом датасете (фото без разметки).



Вариант реализации N°2 (дополнительный) – SSD300 @VGG16

(репозиторий: <https://github.com/sgrvinod/a-PyTorch-Tutorial-to-Object-Detection>)

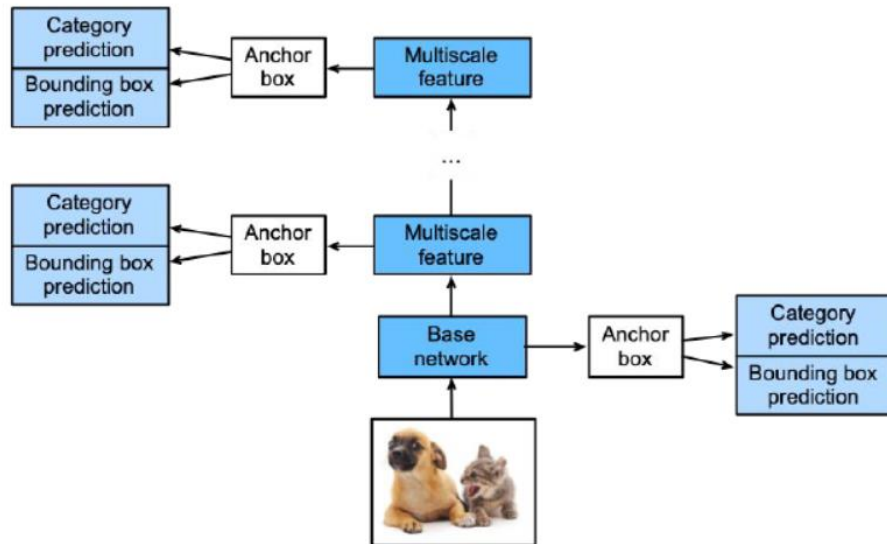
Причины выбора модели SSD300 @VGG16 в качестве дополнительного варианта:

- Сравнение с основным вариантом на YOLOv8;
- Возможность реализовать НС «с нуля»;
- Является «Single-Shot-Detector»-ом, показывающим хорошие результаты в детекции объектов.

Базовая модель для извлечения признаков, за которой следуют свёрточные блоки по уменьшению высоты и ширины в два раза

Генерируем anchor box'ы на каждом размере

Предсказываем класс и bounding box для каждого anchor box'a



Шаги в ходе подготовки данных и обучения НС на SSD300@VGG16:

1. На основании размеченного train/val-датасета, использованного ранее для YOLOv8, были сформированы файлы разметки в формате JSON (список классов, список train-файлов и train-объектов в них, список валидационных файлов и val-объектов в них);
2. Проведена аугментация загружаемых данных - random crop, переворот, изменение размера и фотометрических параметров (яркости, насыщенности, контраста);
3. Описана структура SSD300 на архитектуре VGG16;
4. На датасете были подобраны гиперпараметры обучения;
5. Проведено обучение на всём датасете;
6. Рассчитаны значения метрики mAP@50-95 для сравнения с YOLOv8;
7. Проведен референс модели на выборочных снимках из тестового датасета.



Общий шаг перед обучением моделей:

В рамках дипломного проекта для более скорейшего обучения моделей YOLOv8 и SSD300 VGG16 было принято решение ограничить объем классификации до 10 автомобилей (марок, моделей и года):

- Chevrolet Traverse SUV 2012
- Dodge Dakota Club Cab 2007
- Buick Verano Sedan 2012
- Acura TL Sedan 2012
- Audi TTS Coupe 2012
- Mitsubishi Lancer Sedan 2012
- Hyundai Sonata Hybrid Sedan 2012
- Ford F-450 Super Duty Crew Cab 2012
- Geo Metro Convertible 1993
- Dodge Charger Sedan 2012

После ограничений набор данных содержит 415 цветных фотографий, в т.ч. 336 (80%) размеченных (Train) и 79 (20%) (Valid). Автомобили были выбраны, чтобы постараться соблюсти небольшой дисбаланс классов.



Итоги обучения модели



4

Итоговые (наилучшие) результаты серии экспериментов

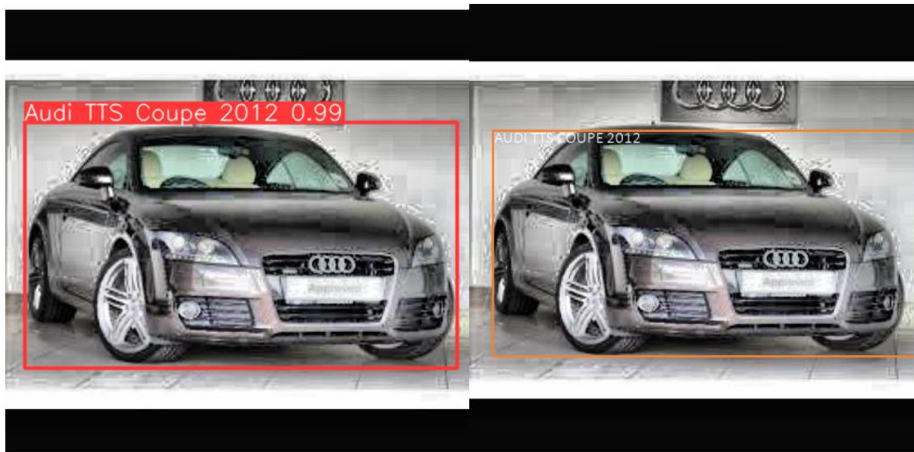
Обучение сетей проводилось с использованием GPU на домашнем ПК в конфигурации Intel i9 и GPU nVidia Geforce 4070 Ti (Python-3.11.7, torch-2.2.2+cu118).

№	Показатель	YOLOv8	SSD300@VGG16
1	mAP@50 по всем классам (%%), в т.ч.:	0.984	1.0
	Acura TL Sedan 2012	0.987	1.0
	Audi TTS Coupe 2012	0.97	1.0
	Buick Verano Sedan 2012	0.987	1.0
	Chevrolet Traverse SUV 2012	0.995	1.0
	Dodge Charger Sedan 2012	0.988	1.0
	Dodge Dakota Club Cab 2007	0.975	1.0
	Ford F-450 Super Duty Crew Cab 2012	0.989	1.0
	Geo Metro Convertible 1993	0.987	1.0
	Hyundai Sonata Hybrid Sedan 2012	0.969	1.0
	Mitsubishi Lancer Sedan 2012	0.995	1.0
2	Оптимальная pretrained- модель	YOLOv8s (291 layers, 20915769 parameters, 20915769 gradients, 48.4 GFLOPs)	N/A
3	Кол-во epoch обучения	100	170
4	Batch Size	16	16
5	Image Size	All (rescale to 416)	All (rescale to 300x300)
6	Learning Rate	0.01	0.0001
7	Метод оптимизации	SGD	AdamW
8	Общее время обучения для достижения mAP (п.1)	~25 минут	~78 минут
9	Итоговый (суммарный) Loss	N/A	1.98

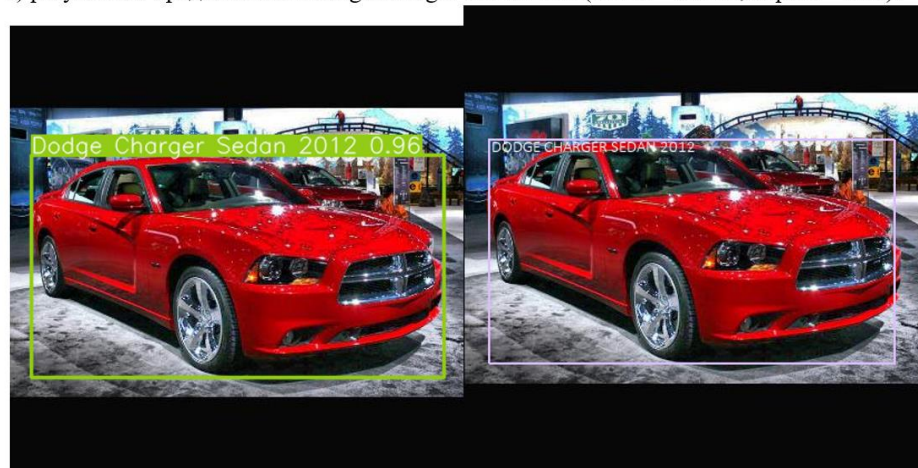


Инференсы(предсказания) на тестовом (неразмеченном) датасете:

а) результаты предсказаний: Audi TT Coupe 2012 (слева – YOLO, справа - SSD):



б) результаты предсказаний: Dodge Charger Sedan 2012 (слева – YOLO, справа - SSD):



Выводы

5



Выводы:

Задача детекции и классификации автомобилей (марка, модель и год) с применением методов глубокого обучения была успешно реализована. Наилучшие результаты показала архитектура YOLOv8. Вместе с высокой скоростью и качеством на обучении, устойчивостью к выбросам, эта архитектура также показывает малое время обучения, что позволяет использовать её для решения задач реал-тайм детекции и классификации.

Тем не менее, реализация на базе SSD300@VGG16 также показала положительные результаты: несмотря на более длительное время обучения при более высоких метриках, визуальное качество детекции находится на высоком уровне.

Примеры неудачных экспериментов:

При обучении SSD300 метод оптимизации SGD показал гораздо худшие результаты по времени сходимости (разница в несколько раз по сравнению с AdamW), а также mAP на уровне 30%, лосс не опускался ниже 4.0. А также LR=0.01, 0.03 приводили к «всплеску» градиентов.



Направления для дальнейшего повышения качества решения задачи:

- Использование вычислений в формате FP16 (вместо FP32) для обучения SSD300@VGG16, что позволит существенно сократить время на обучение модели с сохранением высокой точности предсказаний. При этом YOLOv8 уже по умолчанию обучается в режиме AMP (Automatic Mixed Precision);
- После определения автомобиля (марка, модель и год) определять цвет автомобиля путем добавления дополнительного шага дедукции цвета – потребует дополнительное обучение модели, возможно с разделением на две: детекция автомобиля и детекция цвета;
- До обучить модель на детекцию салона автомобиля исходя из его марки, модели и года.



Спасибо за внимание!

