



УНИВЕРСИТЕТ ИТМО

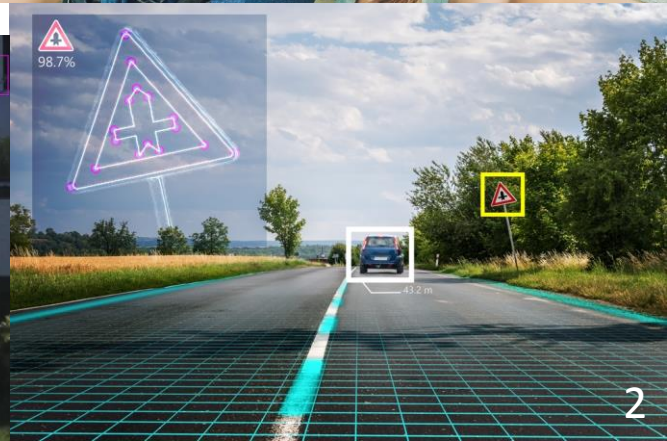
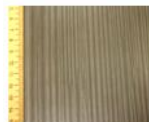
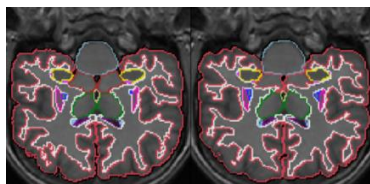
# Исследование архитектур нейронных сетей для классификации и локализации объектов на изображении

Литвинцева Анна Викторовна  
Группа R3435

Научный руководитель:  
Шаветов Сергей Васильевич к.т.н.

# Актуальность использования компьютерного зрения

- ✓ системы видеонаблюдения;
- ✓ системы управления автомобилями;
- ✓ распознавание объектов с воздуха;
- ✓ медицинские системы анализа изображений;
- ✓ сортировка, поиск брака и другие операции в серийном производстве;
- ✓ системы геопозиционирования и картографические системы;
- ✓ анализ эмоционального состояния человека;



**Задача:** анализ и сравнение современных архитектур нейронных сетей для классификации и локализации объектов на изображении

**Цель:** формирование рекомендаций по использованию нейронных сетей для задач робототехники, в частности технического зрения

# Классификация

## ✓ Что было сделано:

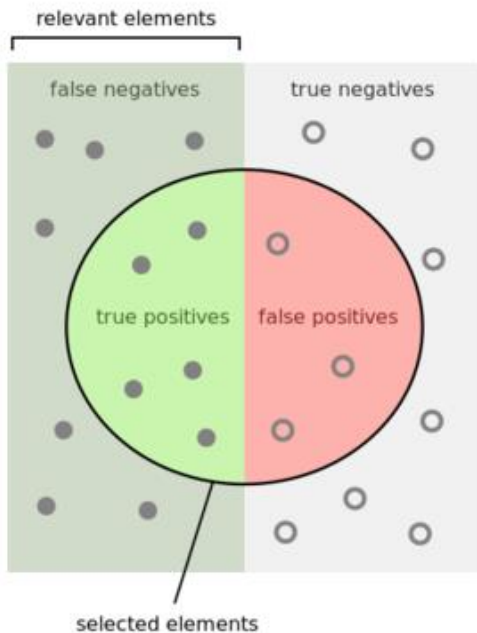
- Разработаны классы для обучения моделей, расчета метрик, загрузки и предобработки данных.
- Проведено сравнение архитектур по статическим характеристикам
- Проведено сравнение для многоклассовой классификации
- Проведено сравнение для бинарной классификации на маленьком наборе данных

## ✓ Инструменты:

- Python 3.6
- PyTorch
- Torchvision
- NumPy
- Matplotlib
- Seaborn

# Метрики

	$y = 1$	$y = 0$
$\hat{y} = 1$	True Positive (TP)	False Positive (FP)
$\hat{y} = 0$	False Negative (FN)	True Negative (TN)



How many selected items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

**Достоверность:**

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

**Точность:**

$$\text{precision} = \frac{TP}{TP + FP}$$

**Полнота:**

$$\text{recall} = \frac{TP}{TP + FN}$$

**F-мера:**

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

# Результаты

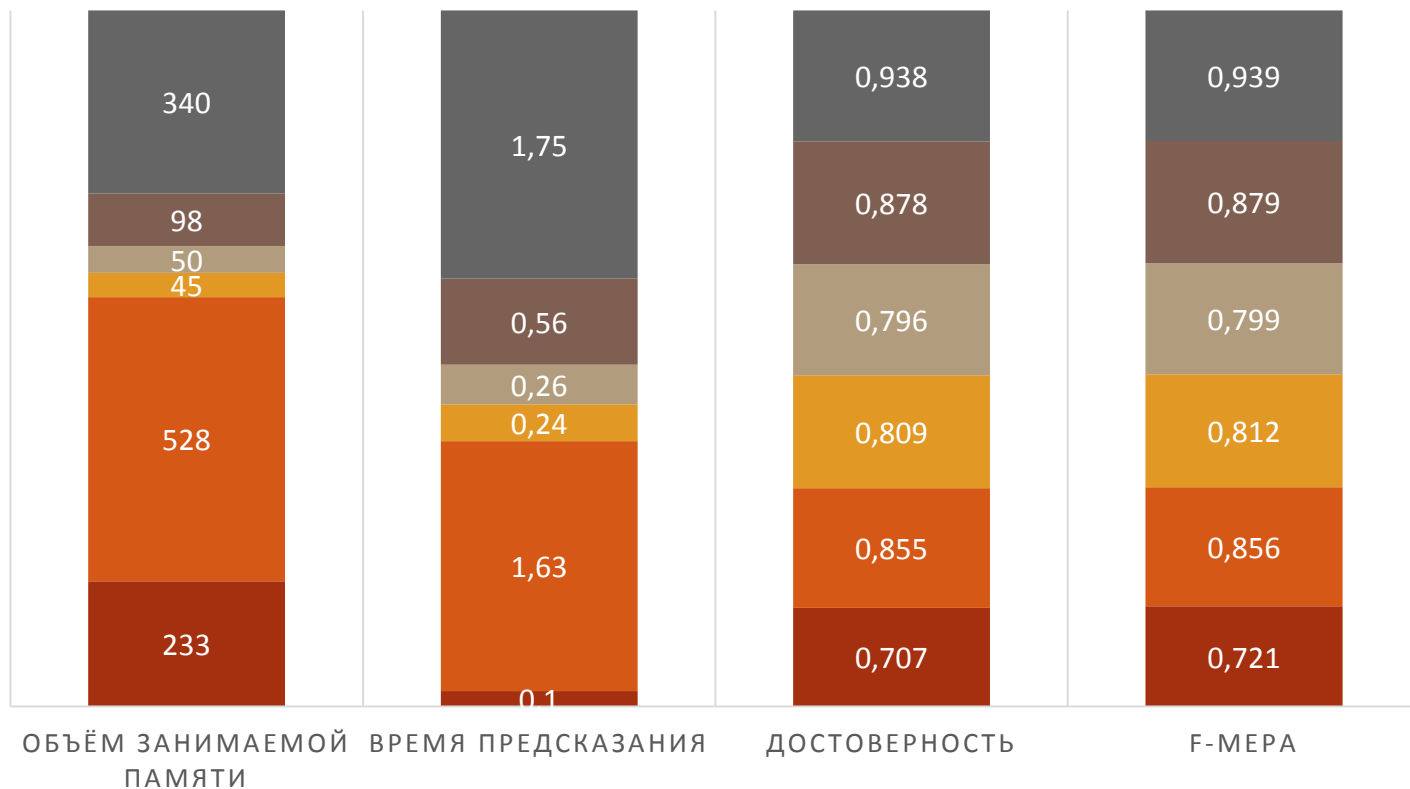
(многоклассовая классификация)

	GoogleNet	ResNet-18	ResNet-50	AlexNet	ResNeXt-101	VGG16
Количество слоёв	22	18	50	8	101	16
Количество параметров	6 624 904	11 689 512	22 557 032	61 100 840	88 791 336	138 357 544
Объём занимаемой памяти	49.7 Мб	44.7 Мб	97.8 Мб	233.1 Мб	340 Мб	527.8 Мб
Время предсказания	0.26 с	0.24 с	0.56 с	0.10 с	1.75 с	1.63 с
Достоверность	0.796	0.809	0.878	0.707	0.938	0.855
F-мера	0.799	0.812	0.879	0.721	0.939	0.856

# Результаты

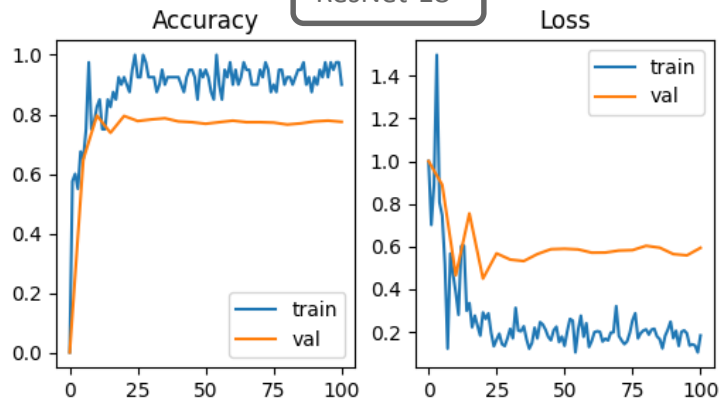
(многоклассовая классификация)

- AlexNet
- VGG16
- ResNet-18
- GoogleNet
- ResNet-50
- ResNeXt-101

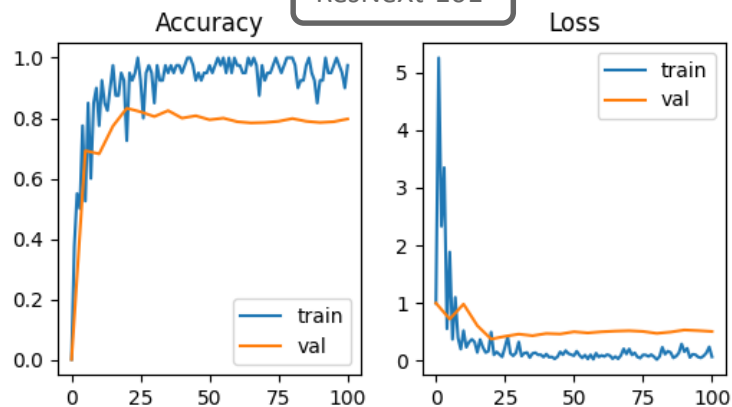


# Значения достоверности и потери

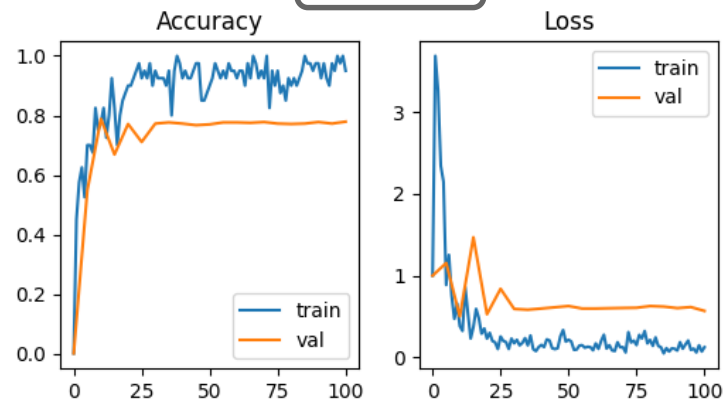
ResNet-18



ResNeXt-101



ResNet-50





# Результаты

		ResNet-18	ResNet-50	ResNeXt-101
Количество слоёв		18	50	101
Количество параметров		11 689 512	22 557 032	88 791 336
Объём занимаемой памяти		44.7 Мб	97.8 Мб	340 Мб
Время предсказания		0.24 с	0.56 с	1.75 с
Бинарная классификация	Достоверность	0.797	0.808	0.868
	Точность	0.813	0.838	0.904
	Полнота	0.889	0.868	0.889
	F-мера	0.849	0.853	0.897
Многоклассовая классификация	Достоверность	0.809	0.878	0.938
	F-мера	0.812	0.879	0.939

# Локализация

## ✓ Что было сделано:

- Изучена структура и особенности архитектуры
- Проведено сравнение на наборе данных COCO 2017
- Проведено сравнение на произвольных изображениях

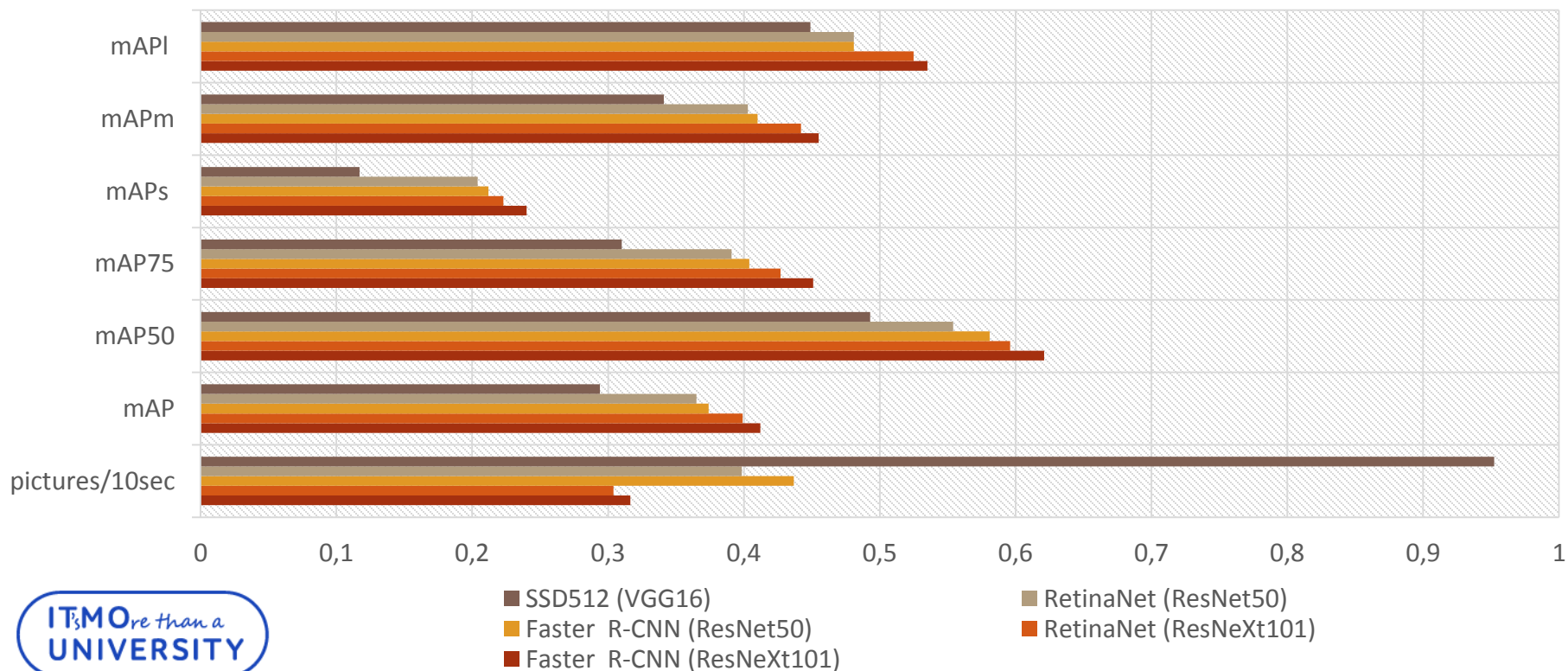
## ✓ Инструменты:

- Python 3.6
- MMDetection

# Результаты сравнения на наборе данных COCO

	Faster R-CNN (ResNeXt101)	RetinaNet (ResNeXt101)	Faster R-CNN (ResNet50)	RetinaNet (ResNet50)	SSD512 (VGG16)
T <sub>CPU</sub>	31.6	32.9	22.9	25.1	10.5
mAP	0.412	0.399	0.374	0.365	0.294
mAP50	0.621	0.596	0.581	0.554	0.493
mAP75	0.451	0.427	0.404	0.391	0.310
mAPs	0.240	0.223	0.212	0.204	0.117
mAPm	0.455	0.442	0.410	0.403	0.341
mAPI	0.535	0.525	0.481	0.481	0.449

## Результаты сравнения на наборе данных COCO





## Рекомендации по использованию рассмотренных решений

- ✓ Системы видеонаблюдения – ResNet50, Faster R-CNN
- ✓ Системы управления автомобилями – ResNet50, Faster R-CNN
- ✓ Системы управления малыми роботами на одноплатных компьютерах – ResNet18, SSD
- ✓ Распознавание объектов на смартфонах – ResNet18
- ✓ Распознавание объектов с воздуха – RetinaNet
- ✓ Медицинские системы анализа изображений – ResNeXt101, Faster R-CNN
- ✓ Поиск брака и другие операции в серийном производстве – ResNet50, Faster R-CNN
- ✓ Системы геопозиционирования и картографические системы – ResNeXt101, Faster R-CNN

# Спасибо за внимание!

[www.ifmo.ru](http://www.ifmo.ru)

IT'sMO<sup>re</sup> than a  
UNIVERSITY