

Введение в нейронные сети.

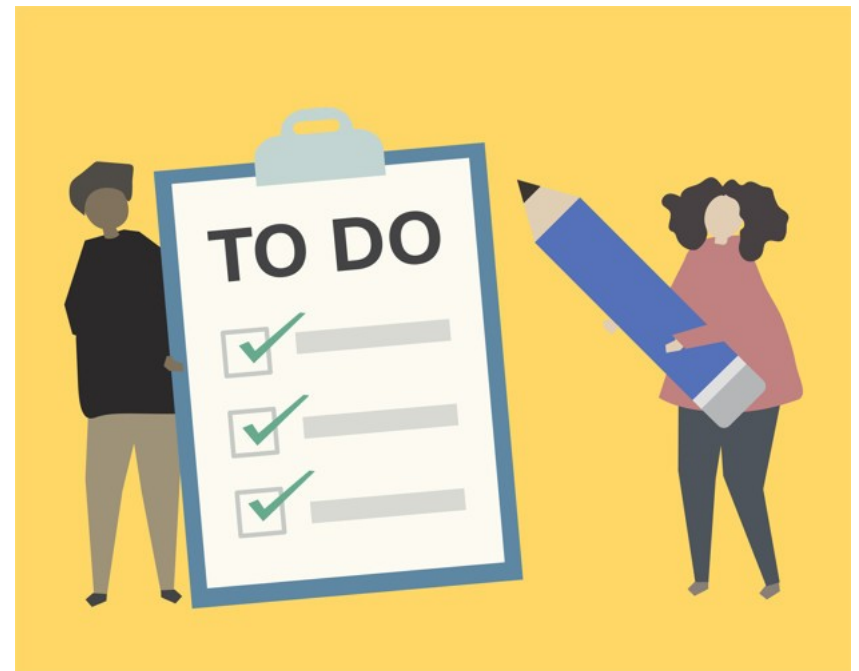
Урок 7. Детектирование объектов



GeekBrains

План вебинара

1. Общие сведения о задаче детектирования объектов
2. MaP
3. Примеры архитектур
4. Практика



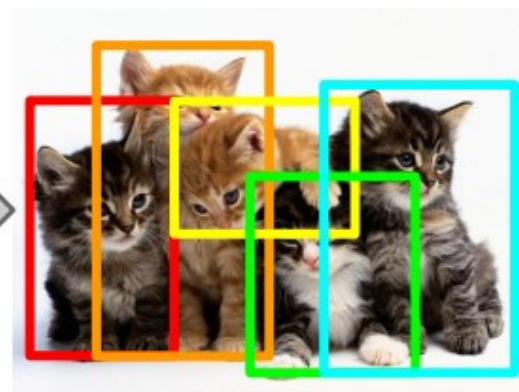
Общие сведения о задаче детектирования объектов.



Объекты



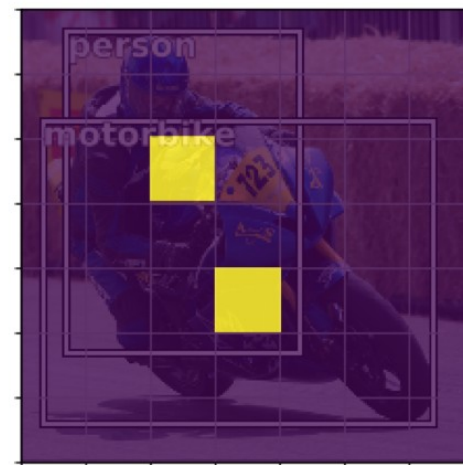
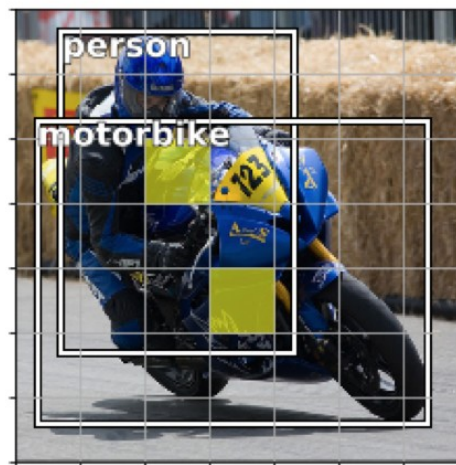
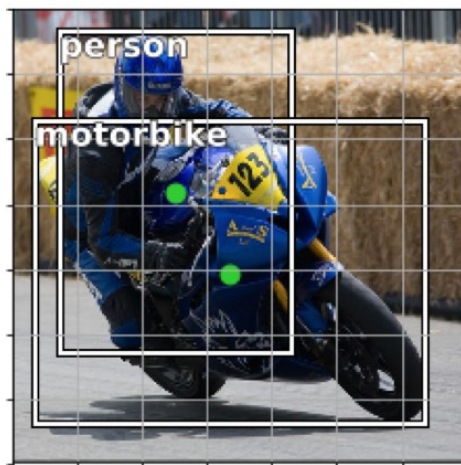
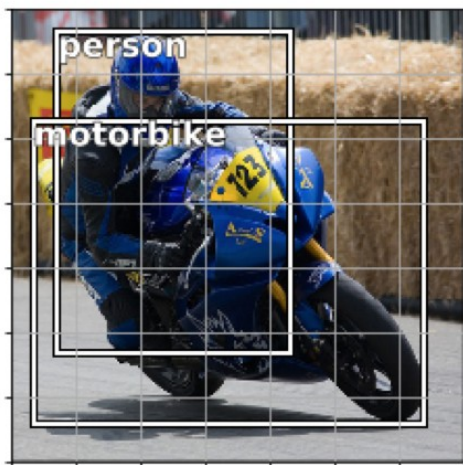
Сегменты



Рамки



Общие сведения о задаче детектирования объектов.



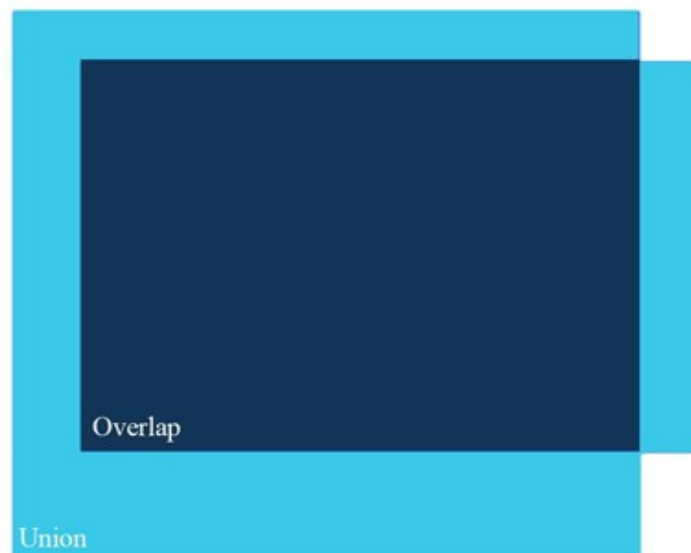
MaP



GeekBrains



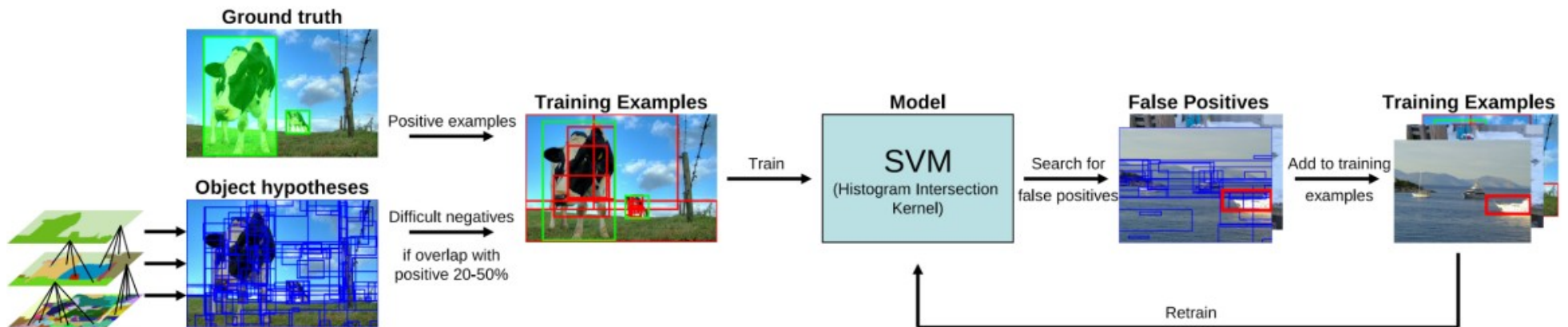
$$IoU = \frac{\text{area of overlap}}{\text{area of union}}$$



Selectiv Search.



GeekBrains

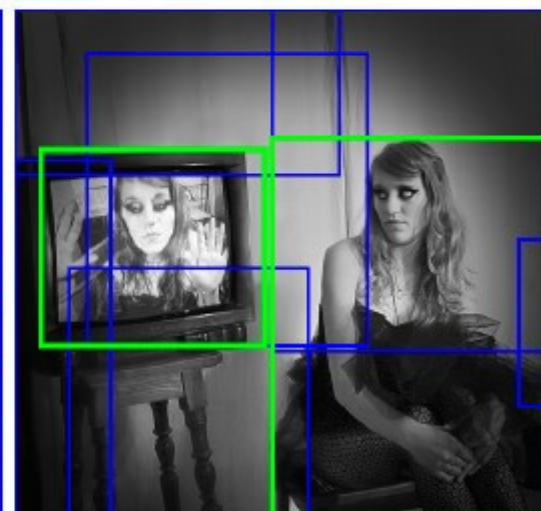
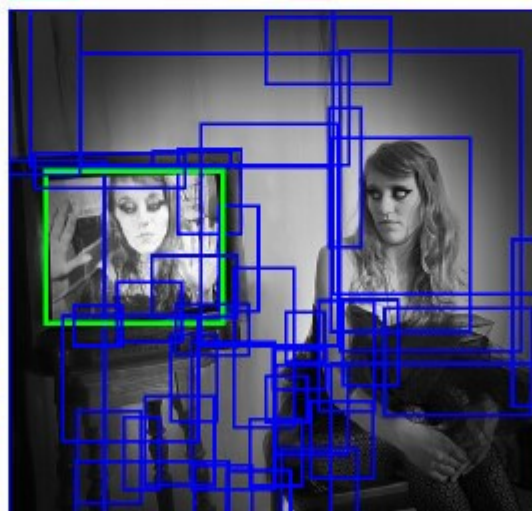
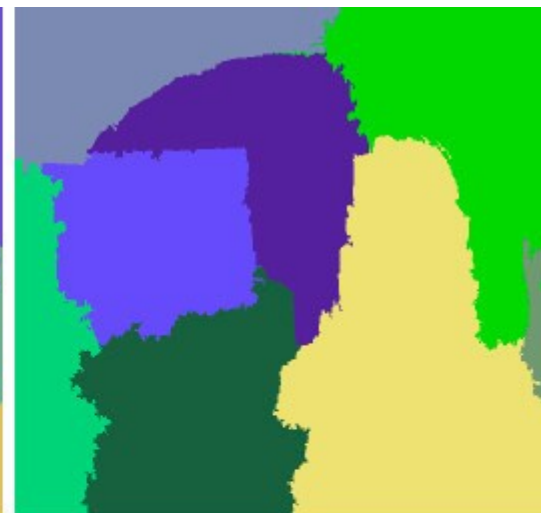


http://vision.stanford.edu/teaching/cs231b_spring1415/slides/ssearch_schuyler.pdf

Selectiv Search.



GeekBrains



Selectiv Search.

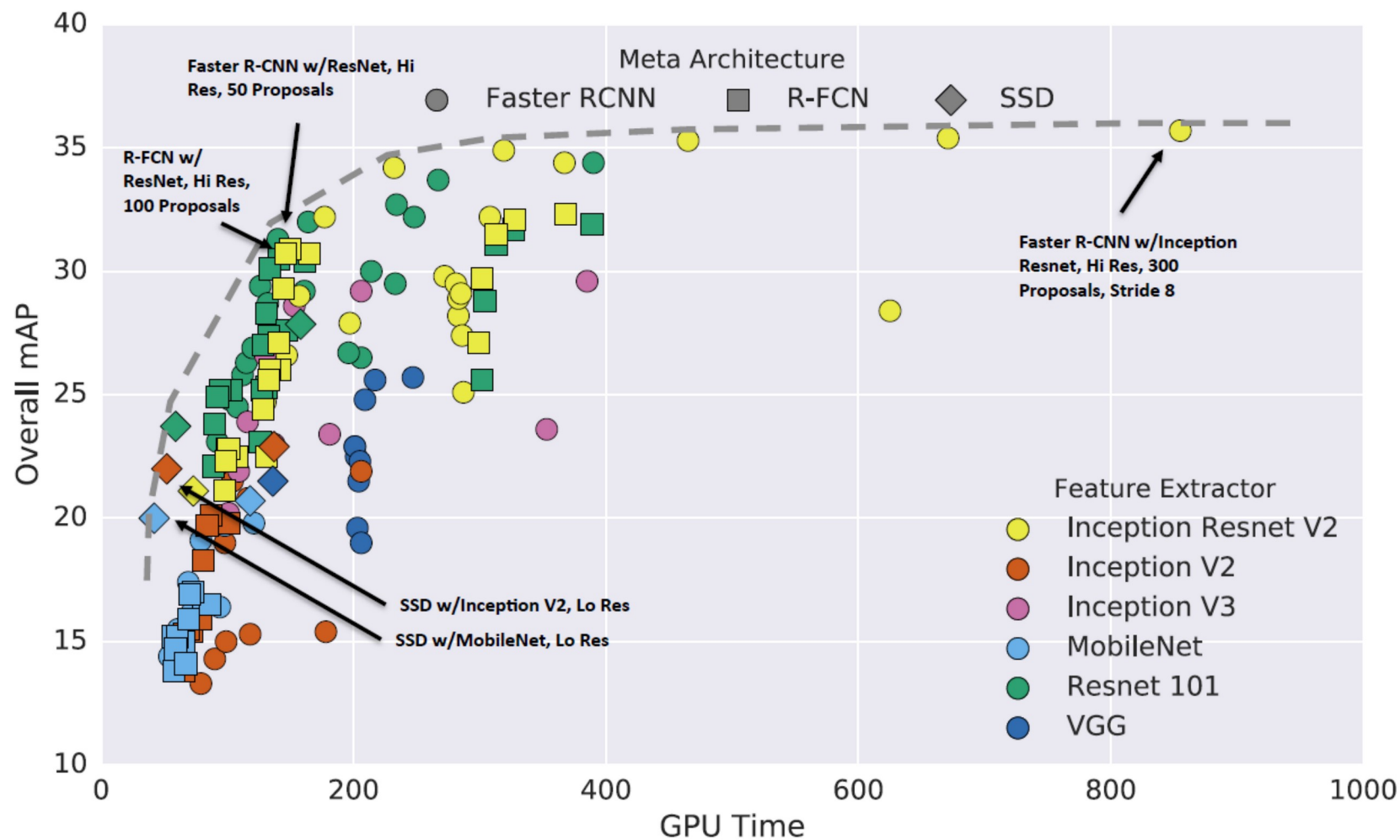


GeekBrains



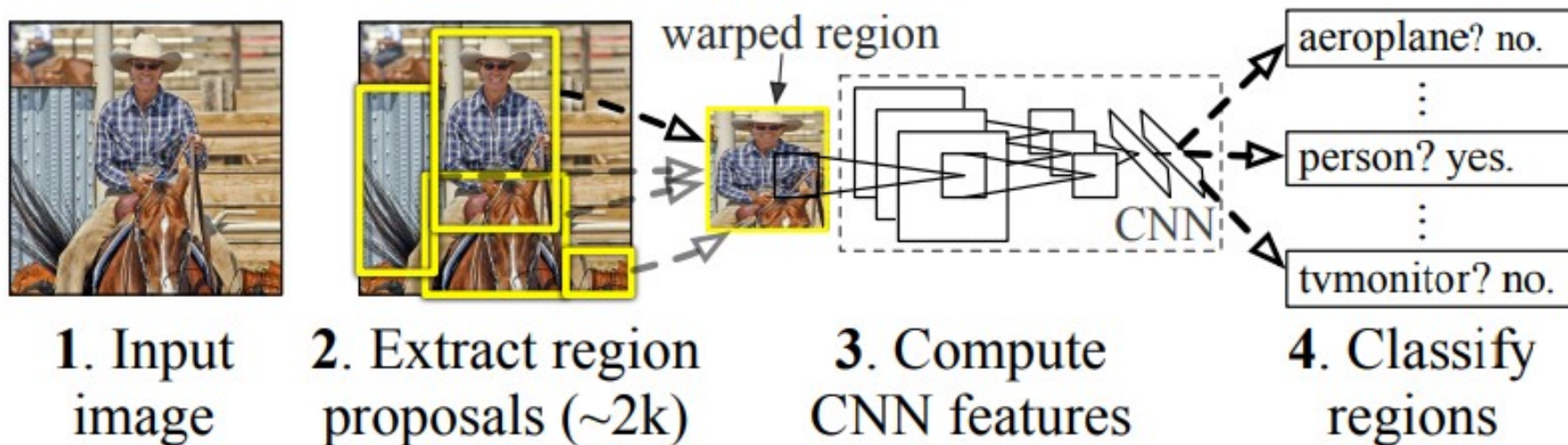
- 1) Построим признаки: цвет, текстура и т. д.
- 2) Построим метрику близости объектов
- 3) Построим иерархию объединений

Примеры архитектур



Примеры архитектур

R-CNN: *Regions with CNN features*



Примеры архитектур

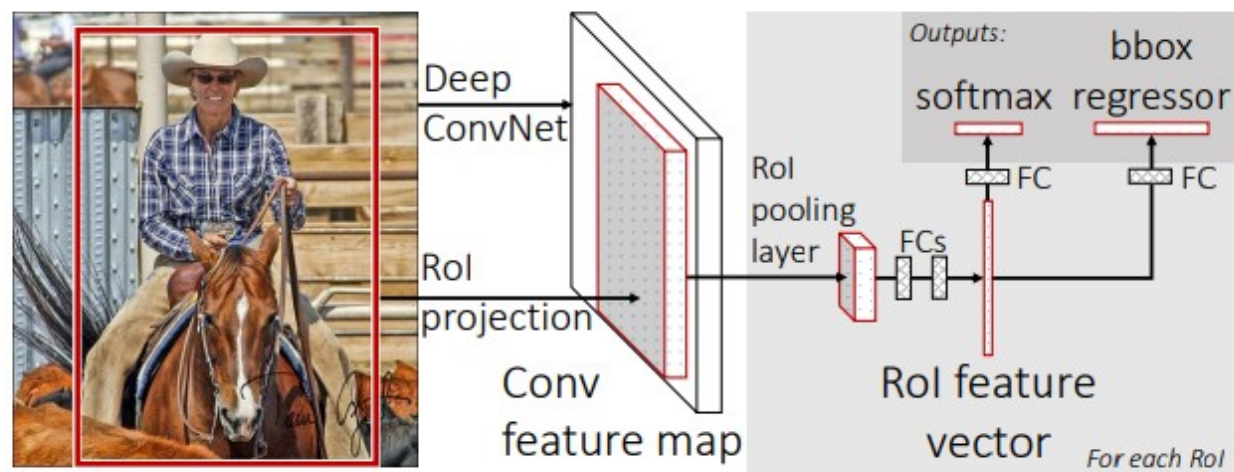


Figure 1. Fast R-CNN architecture. An input image and multiple regions of interest (RoIs) are input into a fully convolutional network. Each RoI is pooled into a fixed-size feature map and then mapped to a feature vector by fully connected layers (FCs). The network has two output vectors per RoI: softmax probabilities and per-class bounding-box regression offsets. The architecture is trained end-to-end with a multi-task loss.

Примеры архитектур

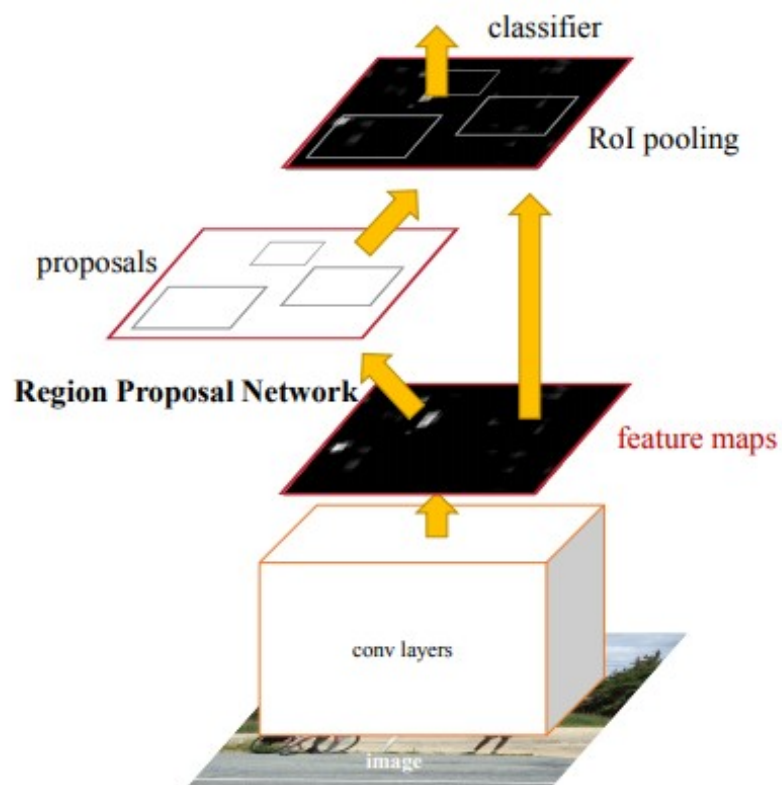
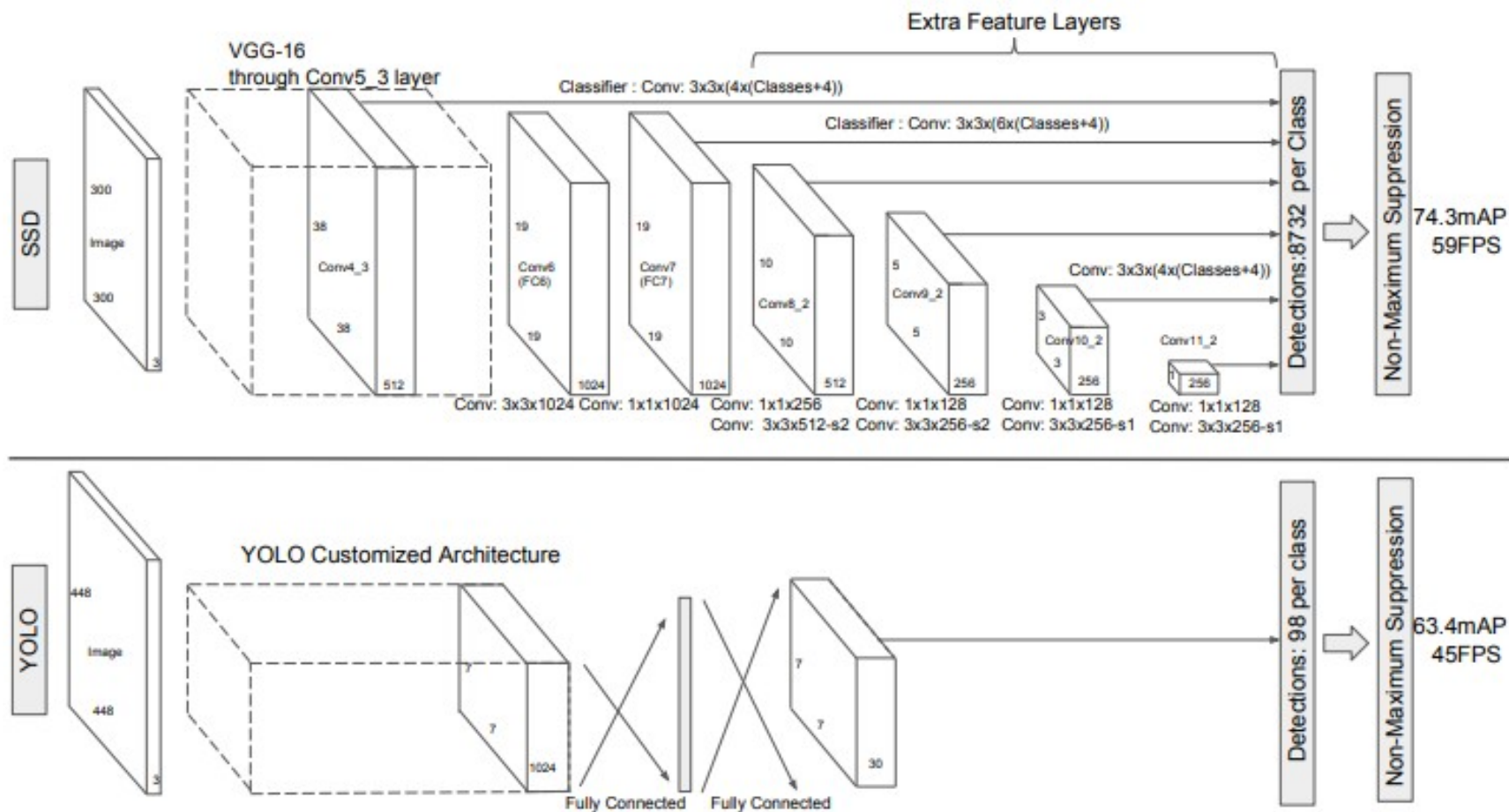


Figure 2: Faster R-CNN is a single, unified network for object detection. The RPN module serves as the 'attention' of this unified network.

Примеры архитектур



GeekBrains



The diagram illustrates the YOLO v3 network architecture. It starts with an input image of a dog, which is processed by a series of layers. The architecture is divided into three main scales for detection:

- Scale 1 (Stride: 32):** This scale produces a detection layer with 82 channels. The corresponding bounding box on the dog image is the smallest and least accurate.
- Scale 2 (Stride: 16):** This scale produces a detection layer with 94 channels. The corresponding bounding box is larger and more accurate than Scale 1.
- Scale 3 (Stride: 8):** This scale produces a detection layer with 106 channels. The corresponding bounding box is the largest and most accurate, closely matching the ground truth.

The legend identifies the components:

- Concatenation:** Represented by a circle with an asterisk (*).
- Addition:** Represented by a circle with a plus sign (+).
- Residual Block:** Represented by a pink rectangle.
- Detection Layer:** Represented by an orange rectangle.
- Upsampling Layer:** Represented by a green rectangle.
- Further Layers:** Represented by black dots.

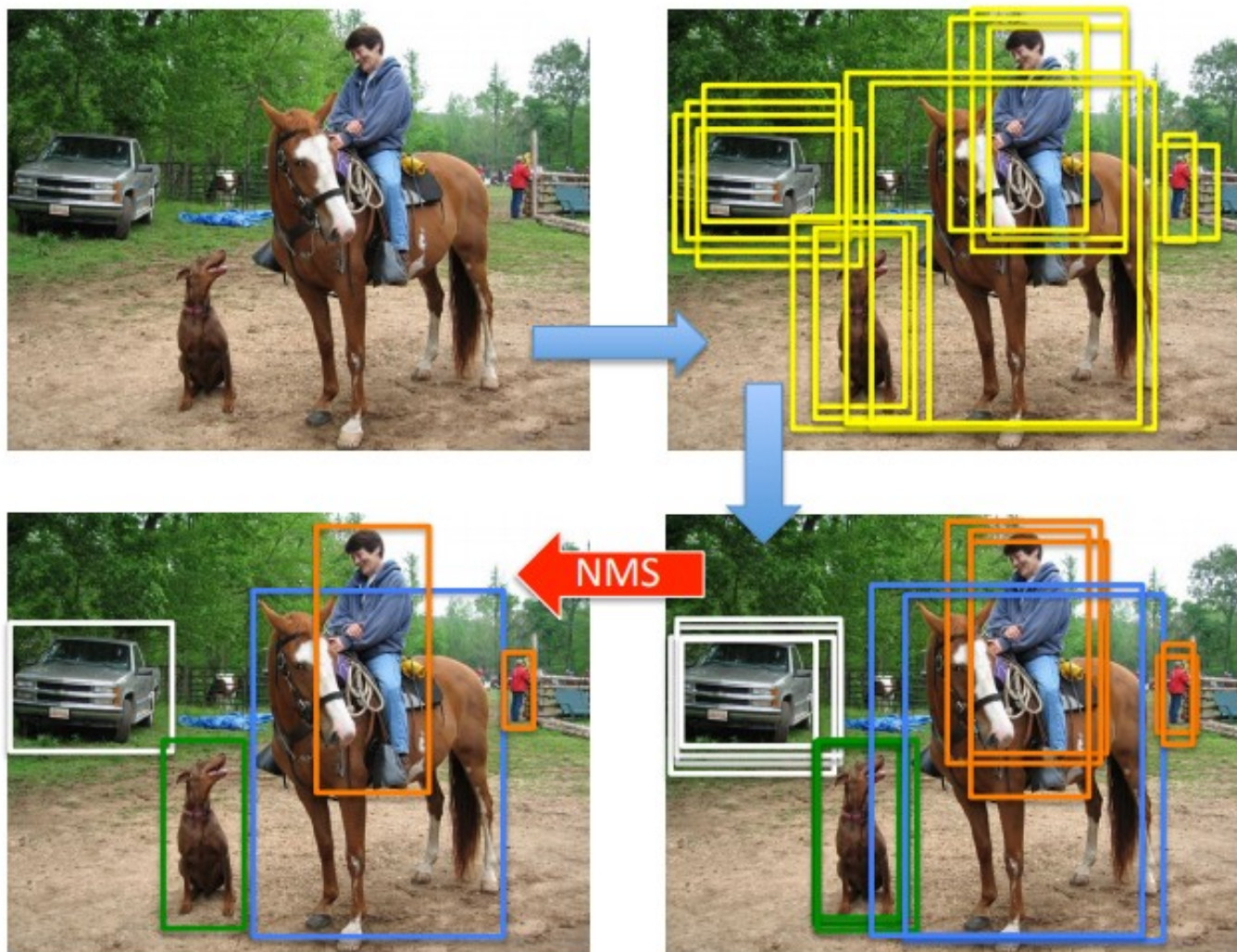
The diagram shows the flow of data from the input image through various layers, including residual blocks, upsampling layers, and detection layers, leading to the final bounding box predictions at three different scales.

YOLO v3 network Architecture



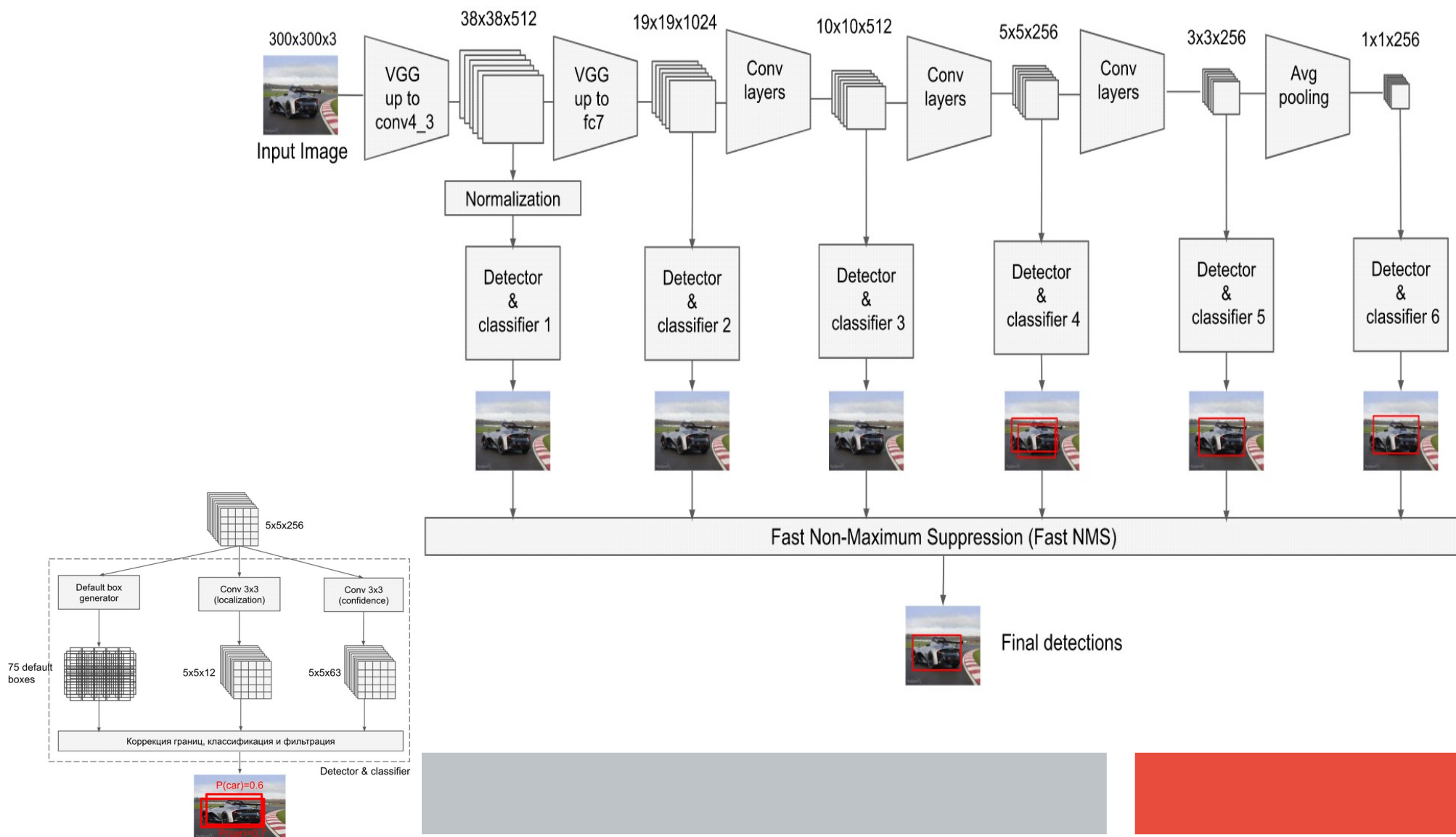
GeekBrains

Примеры архитектур NMS

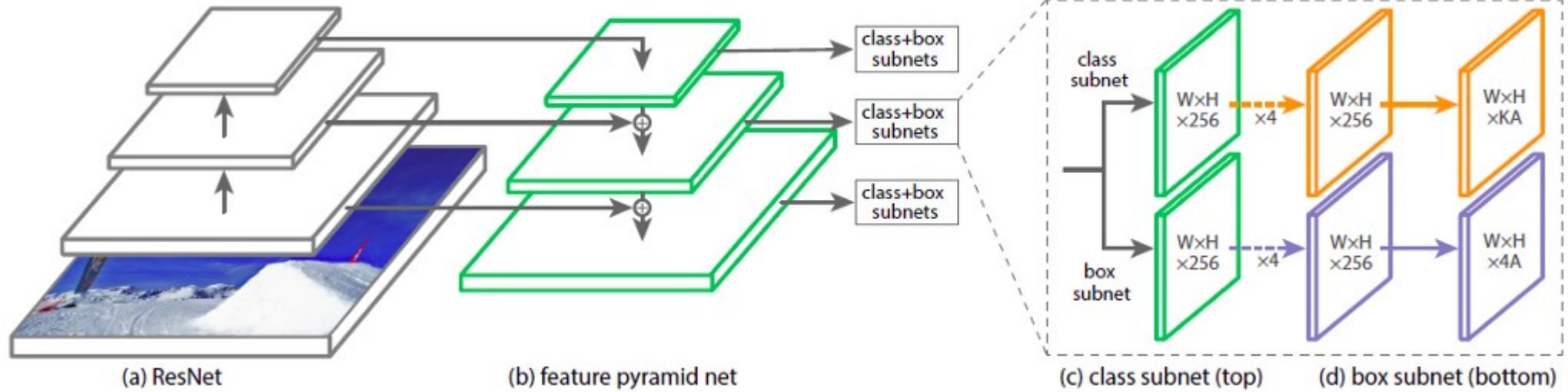


<https://arxiv.org/pdf/1704.04503.pdf>

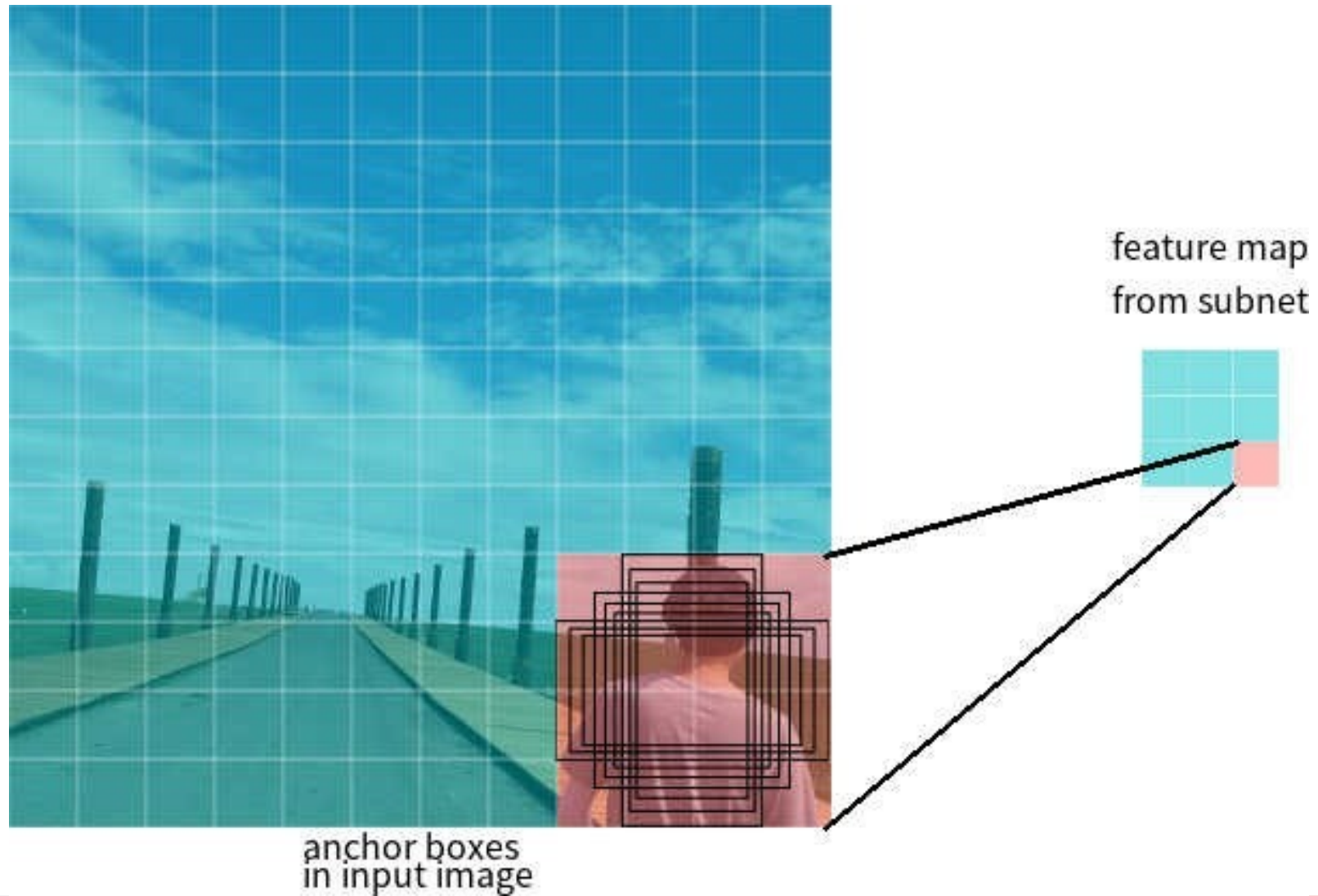
Примеры архитектур



RetinaNet

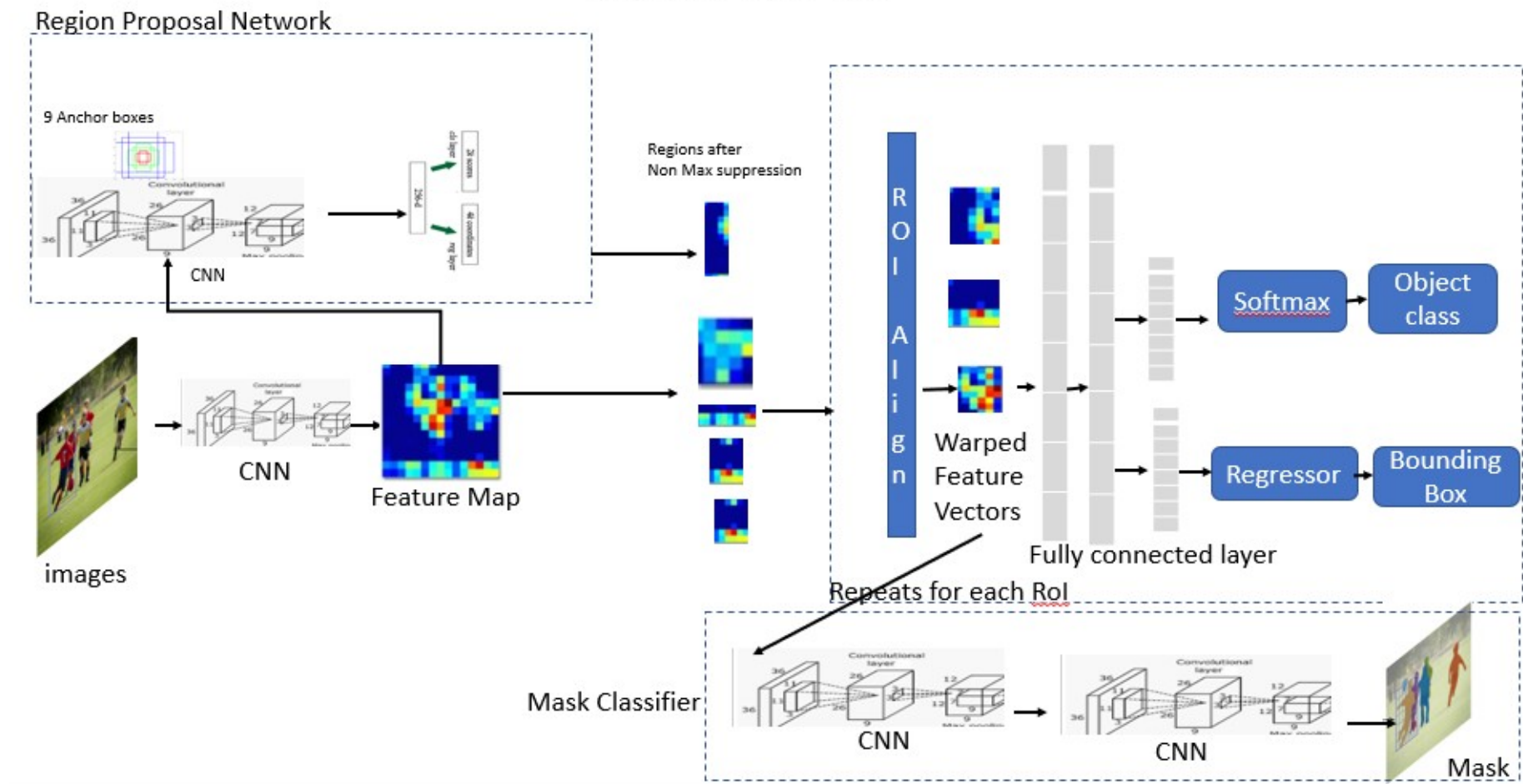


RetinaNet



Примеры архитектур

Mask RCNN



Вопросы



GeekBrains



Сложные модели



ResNet

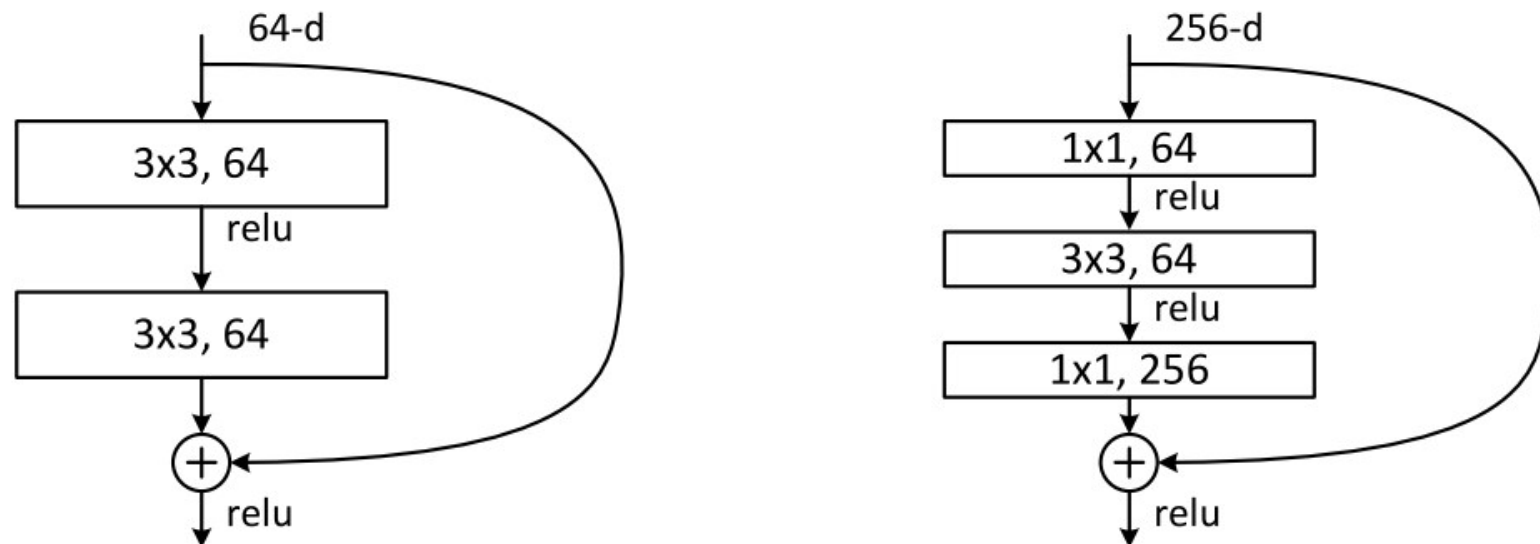
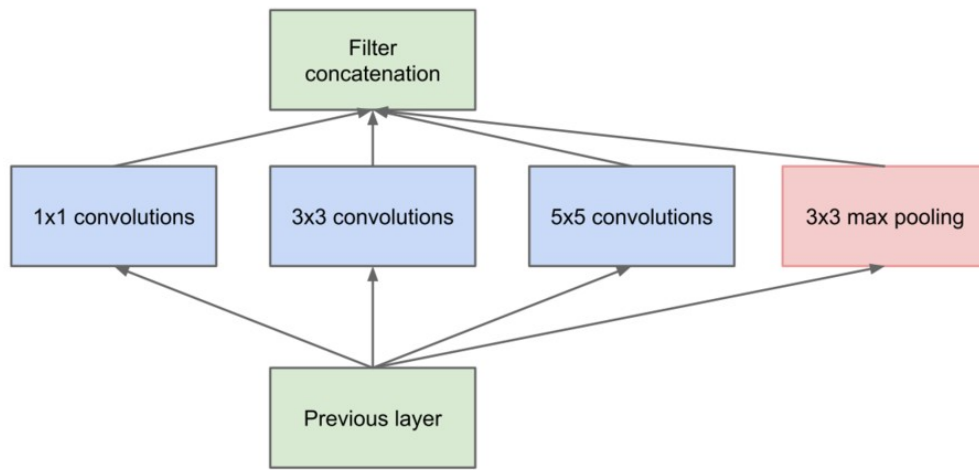
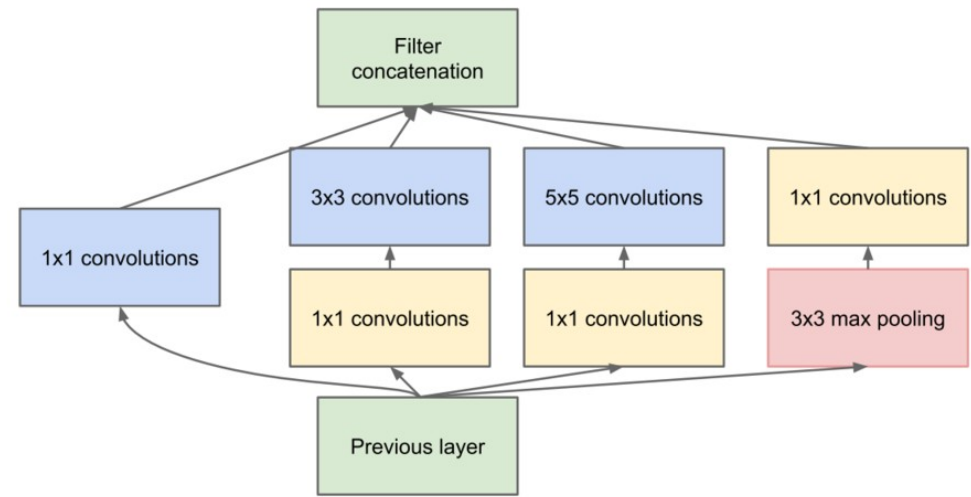


Figure 5. A deeper residual function \mathcal{F} for ImageNet. Left: a building block (on 56×56 feature maps) as in Fig. 3 for ResNet-34. Right: a “bottleneck” building block for ResNet-50/101/152.

Inception



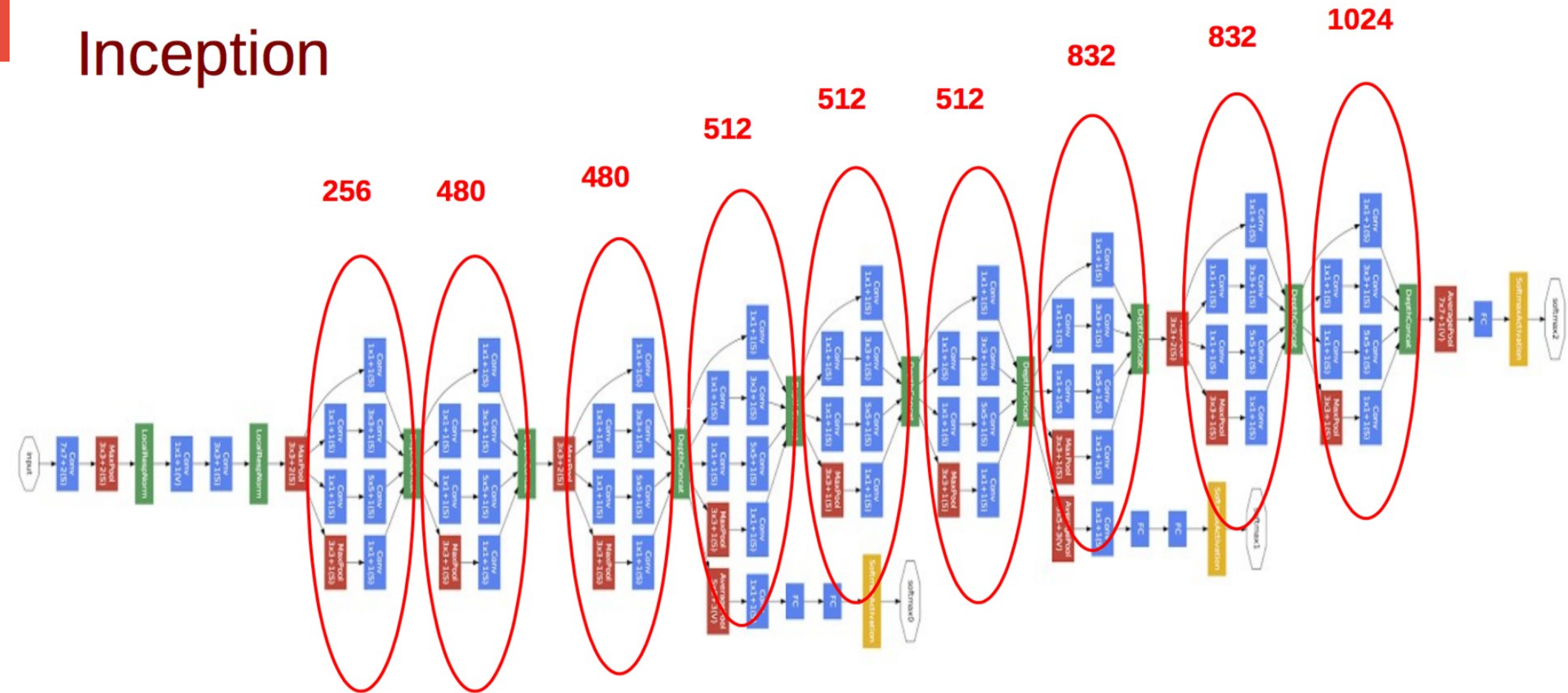
(a) Inception module, naïve version



(b) Inception module with dimension reductions

Figure 2: Inception module

Inception



Width of **inception modules** ranges from 256 filters (in early modules) to 1024 in top inception modules.

ResNet+Inception

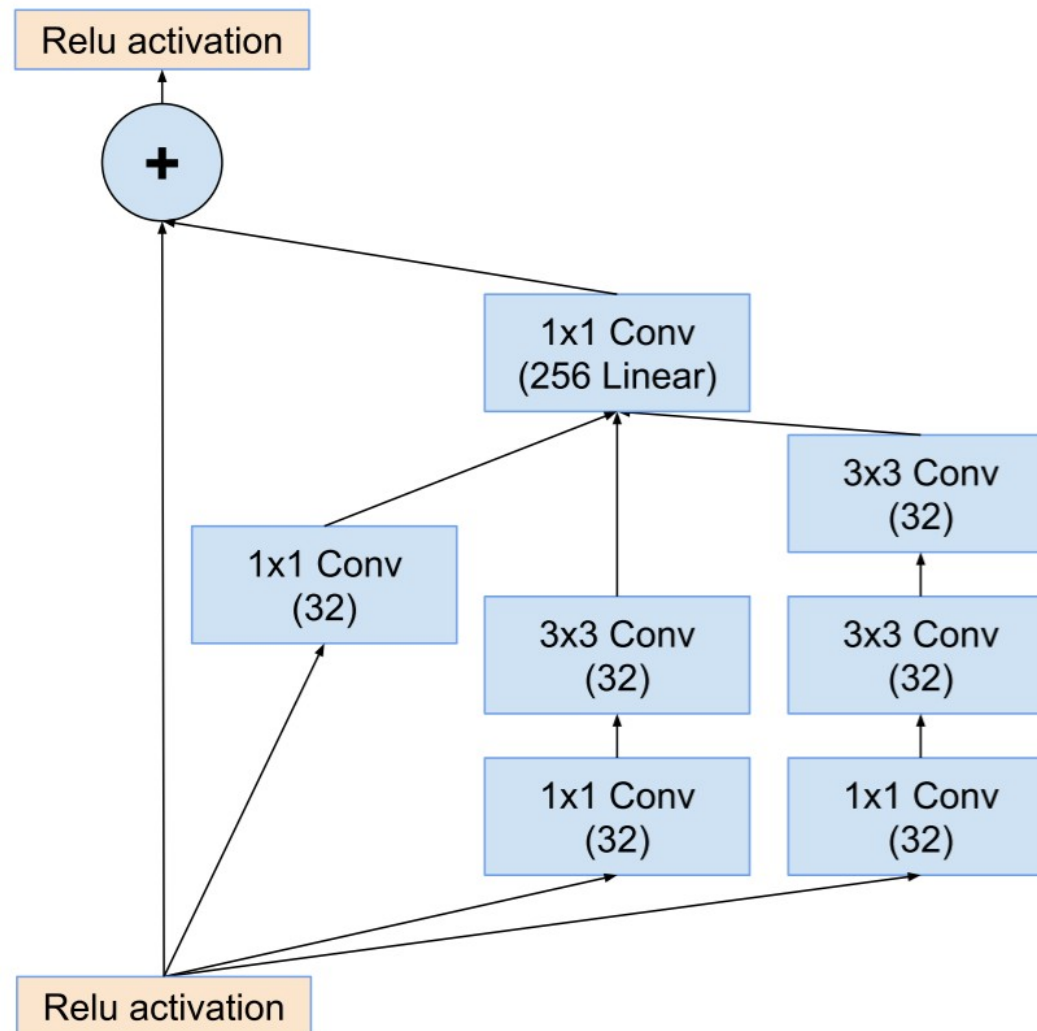


Figure 10. The schema for 35×35 grid (Inception-ResNet-A) module of Inception-ResNet-v1 network.

Практическое задание



1. Сделайте краткий обзор статьи посвященной алгоритму для object detection, который не рассматривался на уроке. Проведите анализ: Чем отличается выбранная вами на рассмотрение архитектура нейронной сети от других архитектур? В чем плюсы и минусы данной архитектуры? Какие могут возникнуть трудности при применении данной архитектуры на практике?

*2. Ссылка на репозиторий с полным кодом для обучения ssd нейросети - <https://github.com/sergeyveneckiy/ssd-tensorflow>. Попробуйте улучшить точность ее работы и напишите отчет, что вы пробовали изменить в ее параметрах и как это отражалось на процессе обучения нейронной сети.

Обратите внимание! Мин. сист. требования для запуска данного проекта - это минимум 8 Gb ОЗУ. Если у вас недостаточно мощности компьютера, то вы можете просто изучить содержимое исходного кода и датасета данного проекта.