

The image features a dark blue background with several geometric shapes. In the top-left corner, there is a blue parallelogram and a light green parallelogram, both tilted at an angle. Below these, there are several dark gray parallelograms of varying shades and orientations, creating a layered, abstract effect.

VGG16



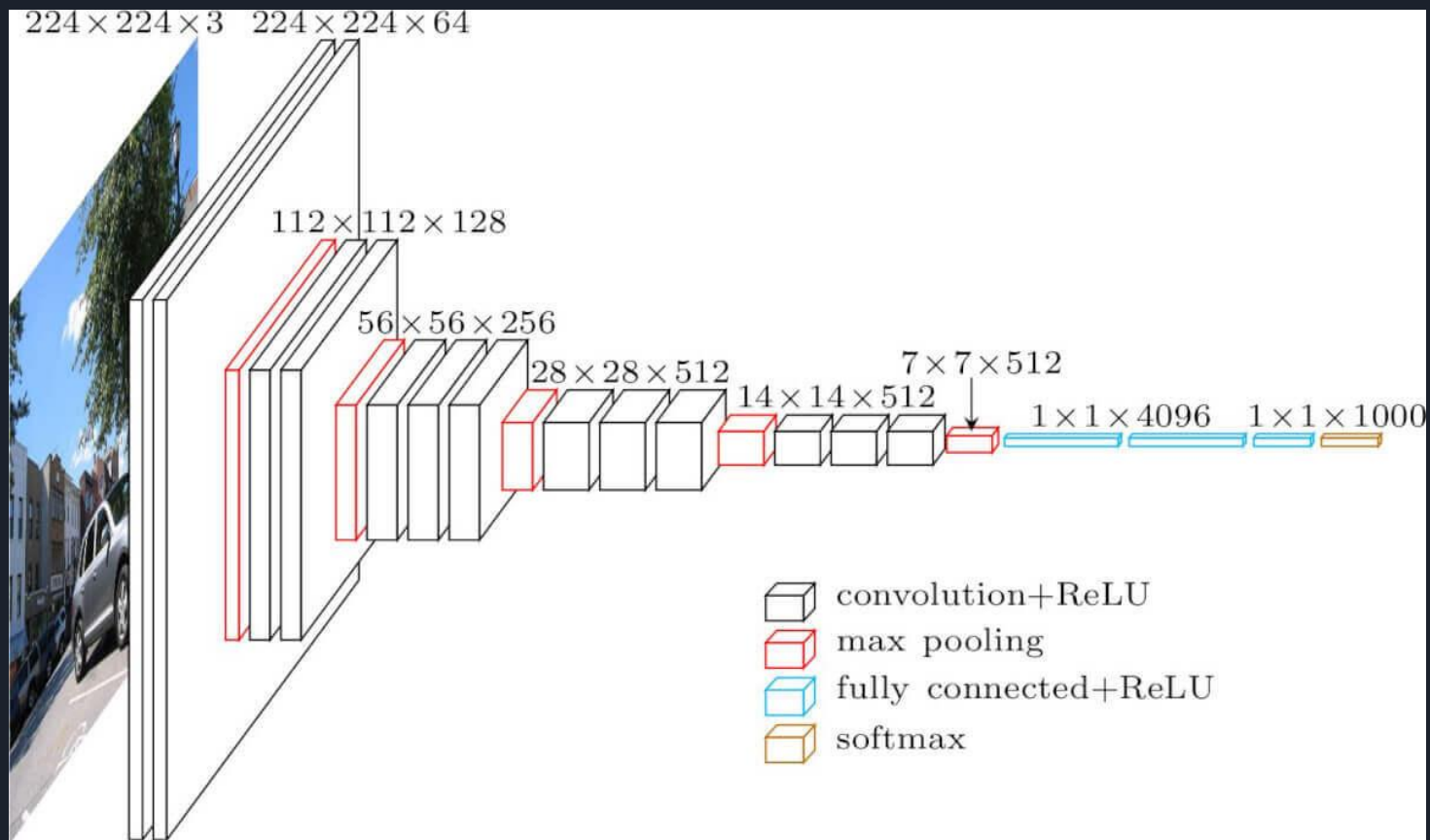
# Создание

VGG16 — модель сверточной нейронной сети, предложенная K. Simonyan и A. Zisserman из Оксфордского университета в статье “Very Deep Convolutional Networks for Large-Scale Image Recognition”.

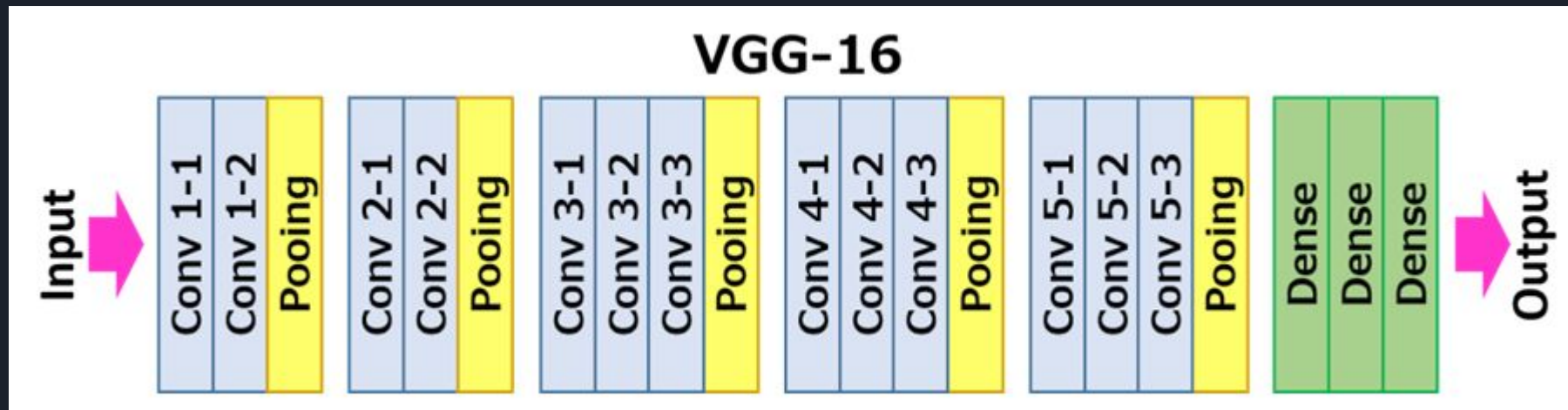
VGG16 является улучшенной версией AlexNet, в которой заменены большие фильтры (размера 11 и 5 в первом и втором сверточном слое, соответственно) на несколько фильтров размера 3x3, следующих один за другим.

Сеть VGG16 обучалась на протяжении нескольких недель при использовании видеокарт NVIDIA TITAN BLACK.

# Архитектура



# Архитектура





# Архитектура

- На вход слоя `block1_conv1` подаются RGB изображения размера  $224 \times 224$  пикселя. Далее изображения проходят через стек сверточных слоев, в которых используются фильтры с очень маленьким окном размера  $3 \times 3$ .
- Сверточный шаг фиксируется на значении 1 пиксель - без пробелов.
- Пространственный пулинг осуществляется при помощи пяти `max-pooling` слоев, которые следуют за одним из сверточных слоев ( не каждым )
- Операция `max-pooling` выполняется на окне размера  $2 \times 2$  пикселей с шагом 2.



# Архитектура

- После стека сверточных слоев (который имеет разную глубину в разных архитектурах) идут три полносвязных слоя:
  - fc1 - 4096 нейронов
  - fc2 - 4096 нейронов
  - predictions - 1000 нейронов(так как в соревновании ILSVRC требуется классифицировать объекты по 1000 категориям; следовательно, классу соответствует один канал).
  - Последним идет soft-max слой.
- Конфигурация полносвязных слоев одна и та же во всех нейросетях.
- Все скрытые слои имеют функцию активации ReLU



# Недостатки

К сожалению, сеть VGG имеет два серьезных недостатка:

- Очень медленная скорость обучения.
- Сама архитектура сети весит слишком много (появляются проблемы с диском и пропускной способностью)



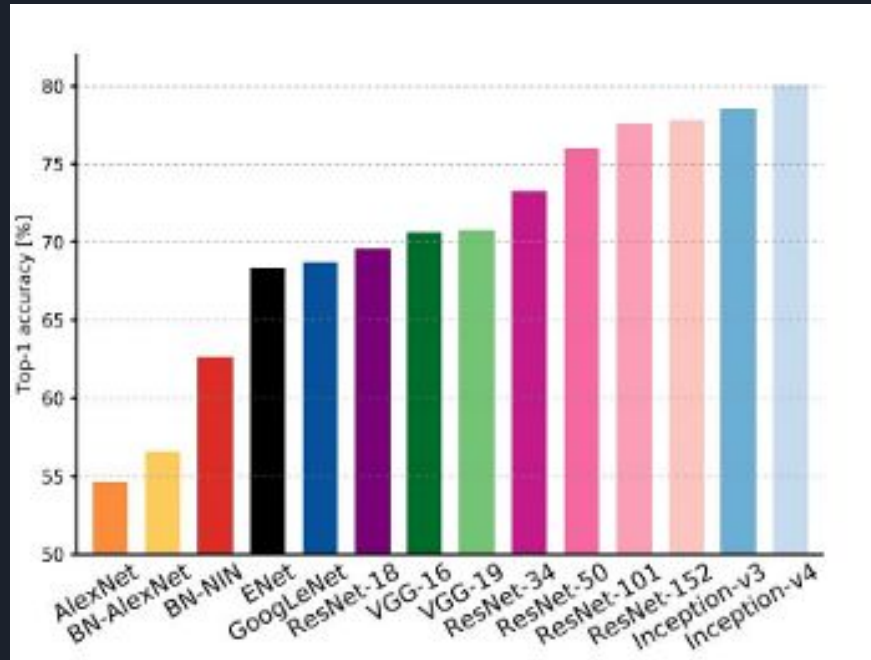
## Преимущества

Несмотря на недостатки, данная архитектура является отличным строительным блоком для обучения, так как её легко реализовать.

Модель достигает точности 92.7% — топ-5, при тестировании на ImageNet в задаче распознавания объектов на изображении. Этот датасет состоит из более чем 14 миллионов изображений, принадлежащих к 1000 классам.



# Сравнение точности



# Сравнение точности и количества операций

