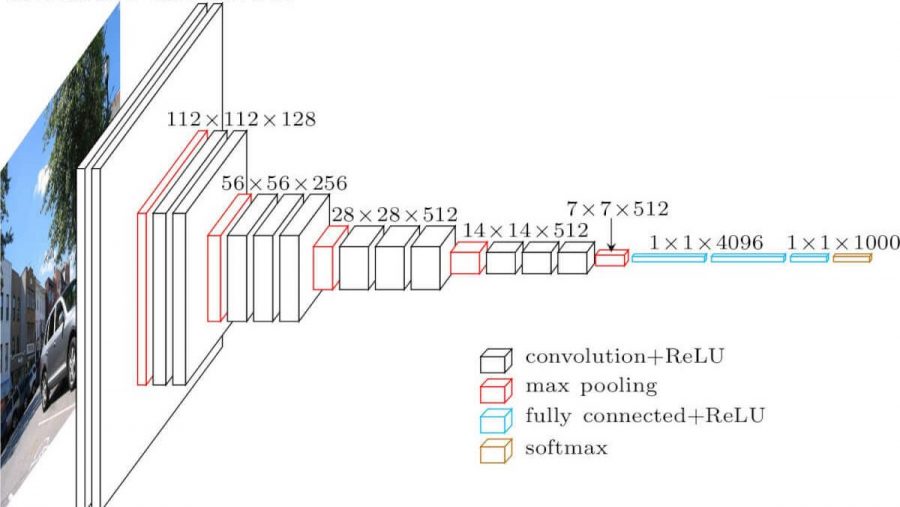
VGG16

VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION

Авторы сети: K. Simonyan и A. Zisserman

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”. Модель достигает точности 92.7% — топ-5, при тестировании на ImageNet в задаче распознавания объектов на изображении.



VGG16 — одна из самых знаменитых моделей, отправленных на соревнование ILSVRC-2014. Она является улучшенной версией AlexNet, в которой заменены большие фильтры (размера 11 и 5 в первом и втором сверточном слое, соответственно) на несколько фильтров размера 3х3, следующих один за другим. Сеть VGG16 обучалась на протяжении нескольких недель при использовании видеокарт NVIDIA TITAN BLACK.

На вход слоя conv1 подаются RGB изображения размера 224х224. Далее изображения проходят через стек сверточных слоев, в которых используются фильтры с очень маленьким рецептивным полем размера 3х3 (который является наименьшим размером для получения представления о том,где находится право/лево, верх/низ, центр).

В одной из конфигураций используется сверточный фильтр размера 1х1, который может быть представлен как линейная трансформация входных каналов (с последующей нелинейностью). Сверточный шаг фиксируется на значении 1 пиксель. Пространственное дополнение (padding) входа сверточного слоя выбирается таким образом, чтобы пространственное разрешение сохранялось после свертки, то есть дополнение равно 1 для 3х3 сверточных слоев. Пространственный пулинг осуществляется при помощи пяти max-pooling слоев, которые следуют за одним из сверточных слоев (не все сверточные слои имеют последующие max-pooling). Операция max-pooling выполняется на окне размера 2х2 пикселей с шагом 2.

После стека сверточных слоев (который имеет разную глубину в разных архитектурах) идут три полносвязных слоя: первые два имеют по 4096 каналов, третий — 1000 каналов (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 снабжены ReLU. Отметим также, что сети (за исключением одной) не содержат слоя нормализации (Local Response Normalisation), так как нормализация не улучшает результата на датасете ILSVRC, а ведет к увеличению потребления памяти и времени исполнения кода.

VGG16 существенно превосходит в производительности прошлые поколения моделей в соревнованиях ILSVRC-2012 and ILSVRC-2013. Достигнутый VGG16 результат сопоставим с победителем соревнования по классификации (GoogLeNet с ошибкой 6.7%) в 2014 году и значительно опережает результат Clarifai победителя ILSVRC-2013, который показал ошибку 11.2% с внешними тренировочными данными и 11.7% без них. Что касается одной сети, архитектура VGG16 достигает наилучшего результата (7.0% ошибки на тесте), опережаю сеть GoogLeNet на 0.9%.

Недостатками сети принято выделять большой объем сети и как следствие недостаточную скорость обучения.

Структурным элементом сети считается сочетание свертка свертка пулинг.

DeepWeeds

**DeepWeeds** - набор данных мультиклассовых видов сорняков для глубокого обучения. Создан для увеличения производительности в сельском хозяйстве через автоматизацию борьбы с вредителями и избирательном выборе гербецидов. Всего 17509 изображений.

Датасет имеет 8 классов изображений + негативы. По факту получаем 9 классов.  ( **а** ) китайское яблоко, ( **б** ) лантана, ( **в** ) паркинсония, ( **г** ) парфений, ( **д** ) колючая акация, ( **е** ) каучуковая лоза, ( **ж** ) сиамский сорняк , ( **h** ) Змеиный сорняк. Каждая картинка имеет 256\*256\*3 точек.

По распределению изображений можно заметить что 1 класс сильно больше остальных.

Данные собирались на специальных приборах. Железом выступили Raspberry Pi 3, Arduino Uno и GPS приемник. Снимали на цветную камеру высокого разрешения FLIR Blackfly 23S6C Gigabit Ethernet.



Пример случайных изображений.