

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Расчетно-графическая работа

«Сверточные нейронные сети в решении задачи классификации изображений» ПО ДИСЦИПЛИНЕ «Интеллектуальные системы»

Факультет: АВТ
Группа: АПИМ-17
Студент: Демянчук А.И.

Преподаватель: Барилло Д.В.

НОВОСИБИРСК
2017 г.

Оглавление

1 Нейронные сети	3
2. Описание структуры исходных данных и задачи в терминах предметной области и машинного обучения	6
2.1 Первый этап (VGG16)	6
2.2 Второй этап - ансамбли нейросетей	8
Заключение	10
Список литературы	11

1 Нейронные сети

Нейронные сети (НС), и в особенности глубокие нейронные сети, используются сейчас для решения очень многих задач машинного обучения. Нейросеть является универсальным аппроксиматором практически любой гладкой функции. Стандартная структура сети подразумевает наличие нескольких слоев, как минимум это один входной, один скрытый и один выходной слой. Такая простейшая сеть иногда называется многослойный персептрон. Происхождение названий входной и выходной не требует объяснений. Обозначение "скрытый" подразумевает то, что конечный пользователь не взаимодействует с результатами, полученными от этого слоя. Термин глубокая в характеристики нейросети используется в том случае, когда архитектура ее подразумевает наличие как минимум двух скрытых слоев. НС используются как в задачах обучения с учителем, так и в задачах обучения без учителя. Простым примером обучения с учителем является задача классификации изображений, т.е. обнаружение объектов определенного класса. Нейросеть должна, имея на входе массив изображений и зная класс каждого из них, обучиться на этих данных и уметь отобразить набор ранее не виденных изображений на набор соответствующих им классов с минимальной ошибкой. Примером обучения без учителя является задача кластеризации. Решая такую задачу, НС должна в массиве данных, каждый экземпляр которого имеет определенный набор признаков, обнаружить характерные паттерны и разбить данные на кластеры. Стандартная внутренняя реализация нейронной сети подразумевает последовательное применение к данным линейных (с использованием настраиваемых параметров) и нелинейных функций на каждом слое с предсказанием на выходном слое; оценку ошибки предсказания с использованием какой-либо функции потерь и последующую минимизацию функции потерь с использованием метода обратного распространения. Обратное распространение необходимо для того, чтобы найти градиент с учетом весов для каждого слоя и, последовательно изменяя веса на величину этого градиента, прийти к минимальной ошибке (градиентный спуск).

Сверточные нейронные сети являются одной из разновидностей глубоких сетей и лидером в задачах из области компьютерного зрения в настоящий момент. Сверточная нейросеть обладает такой архитектурной особенностью, которая позволяет подавать на вход двумерные или трехмерные массивы, что в свою очередь оптимально для работы с изображениями. Сверточные нейросети наследуют

основные черты от обычных нейросетей прямого распространения и также состоят из входного, выходного и ряда скрытых слоев. Существенным отличием является наличие в архитектуре специальных сверточных слоев, за счет которых осуществляется процесс выделения характерных признаков и снижается вычислительная нагрузка.

Сверточный слой представляет собой набор случайным образом сгенерированных весов, которые также часто называются фильтрами или ядрами свертки. Эти веса представляют собой матрицу определенного размера, например 3 x 3. Такая матрица поэлементно перемножается с блоком соответствующего размера в изображении. Выходом является сумма полученных значений. Далее фильтр перемещается на следующий блок и процедура повторяется. Размер шага фильтра может варьировать и обычно составляет 1-3 пикселя. Все полученные значения формируют выходную матрицу, которую называют картой признаков. Пример описанной процедуры представлен на рис. 8. Применение таких фильтров во время тренировки нейросети позволяет настроить веса этих фильтров так, чтобы выделить в изображении какие-то характерные признаки: прямые или наклонные линии, окружности, границы; и в последующем распознавать эти особенности на этапе тестирования.

Для того, чтобы нейросеть могла оптимизировать достаточно сложные функции к результатам свертки далее применяется нелинейная функция. Чаще всего в настоящее время используется ReLU (Rectified Linear Unit). Она поэлементно превращает любое отрицательное значение карты признаков в ноль - $\max(0, x)$, где x - это значения карты признаков.

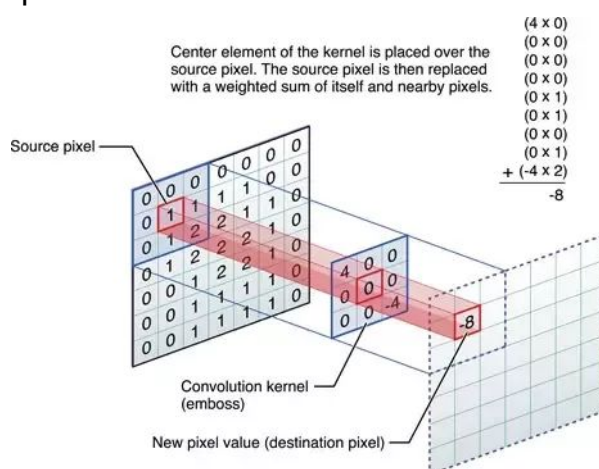


Рис. 8 Процедура свертки в применении к одному блоку исходного изображения [1]

Полученные карты далее подвергаются снижению размерности. Из каждого блока размером 2 x 2 выбирается максимальное значения и из этих значений формируется новая матрица. Таким образом снижается размерность матрицы, что приводит к меньшему количеству весов в следующих слоях и уменьшению вычислительной сложности модели. При этом основная информация полученная из предшествующих слоев сохраняется. Наглядно данная операция представлена на рис. 9.

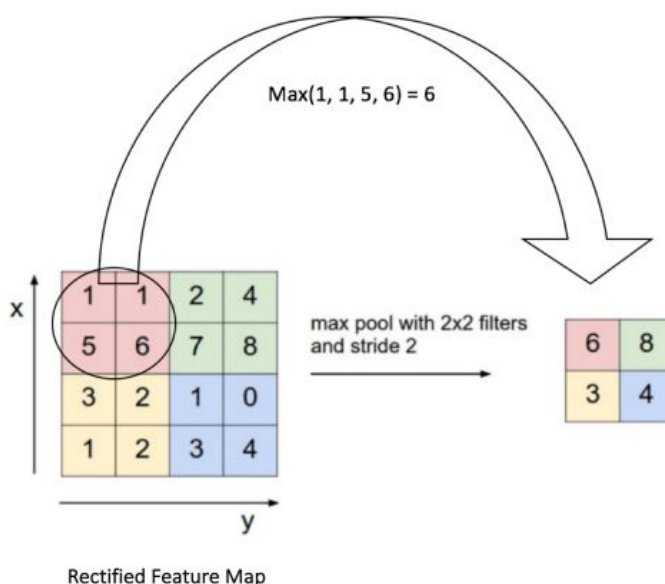


Рис. 9 Операция снижения размерности карты признаков [2]

Построение сверточной нейронной сети завершается одним или несколькими полносвязными слоями. Выходной слой в результате выдает вектор с вероятностями того или иного ответа, в зависимости от задачи, поставленной перед алгоритмом.

Следует отметить, что имеются результаты исследовательских работ, которые показывают очень хорошие результаты по точности распознавания с использованием различных архитектур искусственных нейронных сетей. В настоящее время точность предсказания варьирует от 85% до 98% в зависимости от выбранной модели [3]. Эти показатели превосходят результаты всех описанных ранее методов.

2. Описание структуры исходных данных и задачи в терминах предметной области и машинного обучения

Работа относится к области задач компьютерного зрения. Набор данных Stanford Dogs Dataset представлен на платформе Kaggle [4] состоит из тренировочного и тестового наборов изображений собак. Задача работы заключается в реализации классификатора, который способен распознать 120 различных пород собак на основании фотографий. Тренировочный набор данных состоит из 10220 изображений, тестовый набор из 4010 изображений.

Данное задание решено с использованием ансамбля различных архитектур сверточных нейронных сетей по методике Transfer Learning. А именно, были использованы предобученные сети из высокоуровневого фреймворка глубокого обучения Keras. Работа была проведена в два этапа. Первым этапом была использована предобученная нейросеть архитектуры VGG16 из библиотеки Keras, вторым этапом был использован ансамбль нейросетей для извлечения признаков и однослойная сеть для классификации.

2.1 Первый этап (VGG16)

VGG16 - это одна из широко известных на сегодняшний день архитектур сверточных нейронных сетей. В нейросети блоками реализован ряд сверточных слоев, в которых последовательно происходит процедура свертки, применяется ReLU и уменьшение размерности (pooling). Последний блок сверточных слоев передает свои результаты на последовательность полносвязных слоев, которая завершается классификатором. Архитектура представлена 19 слоями, 16 из них являются сверточными. В 2013 году этот вариант нейросети оказался лучшим на соревновании по компьютерному зрению, которое проводится на наборе данных ImageNet.

В работе была использована реализация нейросети VGG16 из библиотеки Keras. После необходимой предварительной обработки изображений были загружены веса предобученной сети.

```
model = VGG16(weights="imagenet", include_top=False,
input_shape=(INPUT_SIZE, INPUT_SIZE, 3))
```

Все верхние слои нейросети “заморожены”, чтобы не обучать новые параметры на наших данных.

```
for layer in model.layers[:19]:
    layer.trainable = False
```

Добавлен полносвязный слой и слой классификации. Получена нейросеть с немногим более, чем 2000000 обучаемых параметров.

Layer (type)	Output Shape	Param #
=====		
=		
input_2 (InputLayer)	(None, 64, 64, 3)	0
block1_conv1 (Conv2D)	(None, 64, 64, 64)	1792
block1_conv2 (Conv2D)	(None, 64, 64, 64)	36928
block1_pool (MaxPooling2D)	(None, 32, 32, 64)	0
block2_conv1 (Conv2D)	(None, 32, 32, 128)	73856
block2_conv2 (Conv2D)	(None, 32, 32, 128)	147584
block2_pool (MaxPooling2D)	(None, 16, 16, 128)	0
block3_conv1 (Conv2D)	(None, 16, 16, 256)	295168
block3_conv2 (Conv2D)	(None, 16, 16, 256)	590080
block3_conv3 (Conv2D)	(None, 16, 16, 256)	590080
block3_pool (MaxPooling2D)	(None, 8, 8, 256)	0
block4_conv1 (Conv2D)	(None, 8, 8, 512)	1180160
block4_conv2 (Conv2D)	(None, 8, 8, 512)	2359808
block4_conv3 (Conv2D)	(None, 8, 8, 512)	2359808
block4_pool (MaxPooling2D)	(None, 4, 4, 512)	0

block5_conv1 (Conv2D)	(None, 4, 4, 512)	2359808
block5_conv2 (Conv2D)	(None, 4, 4, 512)	2359808
block5_conv3 (Conv2D)	(None, 4, 4, 512)	2359808
block5_pool (MaxPooling2D)	(None, 2, 2, 512)	0
flatten_1 (Flatten)	(None, 2048)	0
dense_1 (Dense)	(None, 1024)	2098176
dense_2 (Dense)	(None, 120)	123000
=====		
=		
Total params: 16,935,864		
Trainable params: 2,221,176		
Non-trainable params: 14,714,688		

Выше представлены характеристики полученной сети. Проведен ряд испытаний с различным количеством эпох и параметром learning rate. Тем не менее результаты полученные на валидационном наборе данных были неудовлетворительными: точность предсказания составила около 40%, что лучше чем случайный выбор, но все же значительно хуже современных показателей в задачах классификации изображений.

2.2 Второй этап - ансамбли нейросетей

Вторым этапом было решено использовать несколько нейросетей для решения задачи и несколько изменить подход к ее решению. Были взяты готовые архитектуры (Resnet, Inception, Exception) из библиотеки Keras, также предобученные на наборе данных ImageNet. Эти архитектуры являются разновидностью подхода Residual learning в сверточных нейросетях, который предполагает использование входных данных от вышележащих слоев и объединение их с выходом нижележащих. Это объединение служит входными данными для последующих слоев. Так получается тренировать очень глубокие нейросети. Далее каждая из этих сетей была окончена слоем GlobalPooling, который сворачивал карты признаков в один массив. В итоге предсказания нейросети были массивом, который выходил из

последнего полносвязного слоя. Таким образом было выполнено извлечение признаков.

```
features = np.concatenate([inception_features,
xception_features, resnet_features], axis=-1)
```

Следующим шагом простая двухслойная нейросеть была обучена классифицировать эти вектора извлеченных ранее признаков.

```
inputs = Input(features.shape[1:])
x = inputs
x = Dropout(0.5)(x)
x = Dense(num_class, activation='softmax')(x)
model = Model(inputs, x)
```

По итогу данного эксперимента на валидационном сете была достигнута высокая точность (около 92%) и низкая логистическая функция потерь (около 0.27). Ниже представлен фрагмент лога тренировки на завершающих эпохах.

```
Epoch 145/150
9199/9199 [=====] - 0s - loss: 0.2717 - acc: 0.9351 -
val_loss: 0.3003 - val_acc: 0.9218
Epoch 146/150
9199/9199 [=====] - 0s - loss: 0.2711 - acc: 0.9356 -
val_loss: 0.3000 - val_acc: 0.9218
Epoch 147/150
9199/9199 [=====] - 0s - loss: 0.2734 - acc: 0.9321 -
val_loss: 0.3000 - val_acc: 0.9208
Epoch 148/150
9199/9199 [=====] - 0s - loss: 0.2688 - acc: 0.9341 -
val_loss: 0.2998 - val_acc: 0.9218
Epoch 149/150
9199/9199 [=====] - 0s - loss: 0.2721 - acc: 0.9327 -
val_loss: 0.2995 - val_acc: 0.9218
Epoch 150/150
9199/9199 [=====] - 0s - loss: 0.2719 - acc: 0.9313 -
val_loss: 0.2995 - val_acc: 0.9218
```

Заключение

В ходе работы решалась задача классификации изображений. Было необходимо определить вероятность 120 классов (пород собак) для каждого изображения в тестовом наборе данных. Работа была проведена в два этапа, в ходе которых были использованы различные архитектуры сверточных нейросетей (VGG16, ResNet, Inception, Xception). На первом этапе в случае изолированного использования предобученной на наборе изображений ImageNet нейросети VGG16 были получены низкие показатели точности, незначительно превышающие случайный выбор. Второй эксперимент оказался более удачным. В ходе него была использована комбинация нейросетей и извлечение признаков типа (bottleneck features). Такая методика позволила получить отличный результат на валидационной и тестовой выборке. В приложение к отчету приложены pdf-файлы с исходным кодом.

Список литературы

1. Quora из электронного источника <https://qph.ec.quoracdn.net/main-qimg-578748437404fe6733bc7823755e813c.webp>
2. <https://ujwlkarn.files.wordpress.com/2016/08/screen-shot-2016-08-10-at-3-38-39-am.png?w=494>
3. Kasar, Manisha M., Debnath Bhattacharyya, и Tai-hoon Kim. «Face Recognition Using Neural Network: A Review». International Journal of Security and Its Applications 10, вып. 3 (31 март 2016 г.): 81–100. <https://doi.org/10.14257/ijisia.2016.10.3.08>.
4. Электронный ресурс <https://www.kaggle.com/c/dog-breed-identification>
- 5.