Метод одновременной локализации и распознавания объектов на изображении

Т. С. Лугуев, Х. С. Муртузаалиев
(Дагестанский государственный университет)

В работе предложен алгоритм локализации и распознавания объектов на изображениях, основанный на нейронных сетях с глубокими архитектурами. Предложенная в настоящей работе архитектура нейронной сети решает обе задачи локализации и распознавания за один проход вычислений. Это позволяет вывести все вычисления на графических процессорах и существенно ускорить локализацию и распознавание.

Ключевые слова: распознавание образов, локализация объектов, нейронные сети, глубокое обучение.

Необходимость понимания объектов, которые изображены на изображениях возникает во многих областях приложений. Недавние прорывы в области глубоких нейронных сетей открывают большие возможности приложения их к изображениям. Значительное накопление размеченных визуальных данных и возможность использования GPU вычислений для обучения открывает возможности использования глубоких архитектур нейронных сетей. Тем не менее, быстрый поиск и распознавание объектов на изображении остается сложной задачей. Из-за значительной вычислительной сложности методов, основанных на глубоких нейронных сетях важным является вычислительная оптимизация методов и распознавания объектов на изображениях.

Существующие алгоритмы локализации объектов зачастую являются операцией с большим временем выполнения. Самые быстрые существующие реализации этого алгоритма реализованы на языке C++ и не использует вычислительные возможности графических процессоров для ускорения вычислений, требуя до 2 секунд на обработку одного изображения.

Так как сами нейронные сети запускаются на графических процессорах, и на обработку одного изображения потребляют на порядок мень-

ше времени, это становится «бутылочным горлышком» при реализации системы распознавания изображений. Для решения этой проблемы представляется актуальной разработка новых методов, которые могут на вход принимать изображение любого размера и возвращать множество прямоугольных областей с вероятностями принадлежности данного региона к каждому классу объектов, решая обе задачи локализации и распознавания за один проход вычислений.

Предлагаемые методы и подходы

В настоящей работе для решения задачи локализации и распознавания объектов на изображениях предложен подход, основанный на нейронных сетях с глубокими архитектурами.

В основе предлагаемого метода, является сверточная нейронная сеть [1], которая является естественным расширением нейронных сетей для обработки изображений. Основные преимуществами глубоких нейронных сетей являются хорошая масштабируемость, быстрая скорость работы, и гибкость использования. Глубокие нейронные сети могут масштабироваться до миллионов параметров, давая им возможность обучаться очень сложным концепциям и тысячам категорий объектов. Вместе с возможностями современного оборудования и в условиях переизбытка исходных данных удается обучать большие и эффективные сети. Для глубоких нейронных сетей отсутствует необходимость хранить дополнительные данные для выполнения предсказаний на новых входных данных. Это означает, что мы можем легко использовать их для встраивания в мобильные устройства и выполнения распознавания за доли секунды.

Сверточные нейронные сети являются иерархическими моделями машинного обучения, которые выявляют сложные представления изображений, используя огромные объемы данных. Стандартная сверточная нейронная сеть состоит из множества слоев двух чередующихся видов: сверточных и субдискретизирующих. Архитектура сверточной нейронной сети обладает двумя ключевыми свойствами, которые делают ее очень полезной для работы с изображениями: общие веса в рамках одного слоя и пространственная субдискретизация. В отличие от полносвязной нейронной сети, где для каждого пикселя входного изображения настраиваются свои весовые коэффициенты в сверточной нейронной сети, для всего изображения используется одно и то же ядро весов. Такой подход позволяет гораздо лучше производить обобщение информации, содержащейся на изображении. На выходе каждого слоя получается набор

двумерных массивов, называемых картами признаков. Отдельные нейроны субдискретизирующего слоя соединены лишь с некоторой частью входной карты признаков, что позволяет получать каждый раз карты признаков меньшей размерности. После каждого слоя субдискретизации применяется нелинейная функция активации (например, гиперболический тангенс). Обучение всей сети может производится с помощью стандартного алгоритма обратного распространения ошибки и его модификациями.

Использование сверточной нейронной сети позволяет добиваться малой чувствительности результата к небольшим сдвигам и искажениям входных данных. Кроме того, структура сверточной нейронной сети позволяет эффективно распараллеливать алгоритмы обучения, поэтому реализацию работы и обучения сети предполагается выполнять для графических процессоров, поддерживающих технологию CUDA.

Основными преимуществами глубоких нейронных сетей являются хорошая масштабируемость, быстрая скорость работы, и гибкость использования. Глубокие нейронные сети могут масштабироваться до миллионов параметров, давая им возможность обучаться очень сложным концепциям и тысячам категорий объектов. Вместе с возможностями современного оборудования и в условиях переизбытка исходных данных удается обучать большие и эффективные сети. Для глубоких нейронных сетей отсутствует необходимость хранить дополнительные данные для выполнения предсказаний на новых входных данных.

Для реализации алгоритма одновременной локализации и распознавания объектов было выполнено внедрение подсети RPN [2] в нейронную сеть VGG [3]. Для этого убраны последние три слоя слой softmax *prob*, слой отброса *drop7* и *fc8*. Вместо этого добавлены после слоя *relu7* два сестринских слоя ***bbox_pred*** и ***cls_score***. **bbox_pred** представляет собой полносвязный слой, с числом выходов, равным $4*n$, где n — это число классов в наборе данных, над которым производилось обучение. Например, для набора данных MS COCO [4], содержащего 80 классов число выходов равно 324, так как был добавлен дополнительный класс, для обозначения фона.

Еще одна особенность архитектуры сети, предложенной нами — это внедрение подсети rpn, между слоем *relu5_3* и слоем *pool5* модели VGG.

Результаты экспериментов

В данной работе для обучения модели используется набор данных MS COCO [4]. Это набор данных, от компании Microsoft, для обучения и те-

стирования моделей для решения задач понимания изображений. Содержит более трехсот тысяч изображений, распределенных по восьмидесяти категориям. На одном изображении расположено несколько объектов, так же каждое изображение снабжено пятью разными описаниями для решения задачи описания изображений. Всего же отмечено более двух миллионов объектов во всем наборе данных.

Выполнение обучения нейронной сети проводилось на компьютере со следующими характеристиками:

- 48 ядерный процессор Intel(R) Xeon(R) CPU E5-2690 v3 @ 2.60GHz;
- 64 гигабайта оперативной памяти;
- графический процессор NVidia Tesla K80 с 11 гигабайтами оперативной памяти.

Нами было проведено обучения одной и той же нейронной сети двумя способами: обучение сети с нуля,и обучение сети используя подход transfer learning [5].

В первом случае, нейронная сеть обучается не используя никаких априорных данных о весах, во втором же случае, обучение продолжается, используя веса модели, натренированной авторами VGG.

Обучение нейронной сети, используя веса предобученной модели, заняло 3,5 суток, обучение же с нуля заняло около 6 суток.

IoU	**area**	**maxDets**	**Точность**
0.50:0.95	all	100	0.731
0.50	all	100	0.806
0.75	all	100	0.793
0.50:0.95	small	100	0.835
0.50:0.95	medium	100	0.642
0.50:0.95	large	100	0.898

Таблица 1. Средняя точность распознавания.

В таблице **??** приведены результаты тестирования. Значения точности и полноты в таблице приведены для разного набора параметров, используемых при оценке:

IoU (*Intersection-Over-Union* — отношение пересечения прямоугольников к их объединению) — порог меры, определяющая какие прямоугольники считать совпадающими, начиная с которого считается, что объект обнаружен.

area — показывает, объекты какой площади рассматривать. Может принимать четыре значения: small — маленький, medium — средний, large — большой, all — все.

maxdets — максимальное число объектов, определяемых на одном изображении.

Заключение

Предложенная в настоящей работе архитектура нейронной сети решает обе задачи локализации и распознавания за один проход вычислений. Это позволяет вывести все вычисления на графических процессорах и существенно ускорить локализацию и распознавание.

Также при использовании предлагаемых нейронных сетей с глубокими архитектурами не требуется специального конструирования признаков для различных категорий изображений, что позволяет применять одну и ту же систему распознавания к изображениям, которые могут содержать объекты из тысяч классов.

Работа выполнена в рамках госзадания Минобрнауки РФ и поддержана грантом компании NVidia.

Список литературы

[1] Zeiler M. D., Fergus R. Visualizing and understanding convolutional networks // ECCV. — 2014. — P. 818–833.

[2] Shaoqing R., Kaiming H., Ross G., Jian S. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks // Neural Information Processing Systems (NIPS). — 2015.

[3] Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition // International Conference on Learning Representations (ICLR). — 2015.

[4] Tsung-Yi L., et al. Microsoft COCO: Common Objects in Context // Lecture Notes in Computer Science. — Vol. 8693. — P. 740–755.

[5] Yosinski J., et al. How Transferable are Features in Deep Neural Networks? // Advances in Neural Information Processing Systems. — 27. — P. 3320–3328.