

Курс Введение в нейронные сети

Урок 7. Детектирование объектов

Практическое задание

1. Сделайте краткий обзор какой-нибудь научной работы посвященной тому или иному 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?

Ответ

В научной работе «Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks» (Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun) описана нейронная сеть для детектирования и классификации объектов Faster R-CNN.

Архитектура Faster R-CNN объединяет Region Proposal Network (RPN) и Fast R-CNN в единую сеть путем совместного использования их свертывающих характеристик, используя недавно популярную терминологию нейронных сетей "механизмы внимания". RPN используется для генерации гипотез, получая на вход карту признаков изначального изображения. Сеть Faster R-CNN строится на одних и тех же весах с помощью следующего итеративного обучения:

1. **Тренировка RPN сети.** Свёрточные слои инициализируются весами, натренированными на ImageNet. Дообучаем на задаче определения регионов с каким-либо классом (уточнением класса занимается часть Fast R-CNN).
2. **Тренировка Fast R-CNN сети.** Свёрточные слои инициализируются весами, натренированными на ImageNet. Дообучаем, используя гипотезы об объектах с помощью RPN сети, обученной в п.1. В этот раз задачей обучения является уточнение координат и определение конкретного класса объекта.

После выполнения п.2 мы имеем две нейронных сети с разными весами для свёрточных слоёв.

3. **Обучаем только RPN часть**, используя веса из п.2 (слои, идущие до RPN сети, принадлежащие feature extractor, замораживаются и никак не изменяются).
4. **Обучаем слои для Fast R-CNN**, используя веса из п.3 (то есть, уже более точно настроенный RPN), (остальные веса – идущие ранее или относящиеся к RPN — заморожены).

Таким образом, архитектура Faster R-CNN отличается от других архитектур тем, что генерация гипотез выполняется с помощью отдельно дифференцируемого модуля и в архитектуре реализовано итеративное обучение.

Преимуществом архитектуры Faster R-CNN являются быстрота и точность. Архитектура Faster R-CNN справляется немного хуже с локализацией. RPN может быть трудно работать с объектами разных масштабов. RPN имеет фиксированное приемное поле. Таким образом, малые объекты могут занимать очень небольшую часть рецептивного поля или, если объекты очень большие, то рецептивное поле будет содержать только часть объекта. Проблему можно решать, тренируя множество RPN для различных масштабов: каждая часть RPN будет принимать различные свёрточные слои или наборы слоёв в качестве входа, так что приемное поле будет разного размера. Это значительно улучшит одновременное обнаружение малых и крупных объектов моделью Faster R-CNN.