

РЕШЕНИЕ ЗАДАЧИ ДЕТЕКЦИИ С ПОМОЩЬЮ ОБУЧЕНИЯ С ЧАСТИЧНЫМ ПРИВЛЕЧЕНИЕМ УЧИТЕЛЯ

Лесцов Борис Андреевич

Студент

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: boris.lestsov@graphics.cs.msu.ru

Научный руководитель — Конушин Антон Сергеевич

В последние годы сверточные нейронные сети (СНС) стали повсеместно применяться для решения множества задач компьютерного зрения, таких как, например, классификация изображений, детекция и сегментация объектов на изображениях. Для обучения СНС с учителем требуются большие объемы данных, размеченные человеком, которые, к сожалению, доступны не всегда: разметка данных часто является трудоемким и продолжительным процессом, особенно в таких задачах как детекция и сегментация. В связи с этим за последние годы был разработан ряд методов, позволяющих использовать неразмеченные данные в процессе обучения СНС для повышения их итогового качества работы. Неразмеченные данные используются вместе с доступной размеченной выборкой, поэтому это семейство подходов называется обучением с частичным привлечением учителя. Целью данной работы является проверка возможности использования данных методов для решения задачи детекции.

Существует широкий класс современных методов обучения с частичным привлечением с учителя, использующих специальную функцию потерь, называемую функцией согласованности (consistency loss function) [1]. Использование функции согласованности позволяет добиться того, что предсказания СНС становятся инвариантными к некоторому классу преобразований исходного изображения, что повышает качество работы СНС. Использование функции согласованности не требует разметки изображений, что позволяет использовать ее в применении к неразмеченным данным в процессе обучения.

В методе предполагается наличие некоторого обратимого стохастического преобразования F (аугментации), применимого к изображению и предсказанию детектора. В рассматриваемом подходе в качестве F используются геометрические преобразования.

Обозначим за $nn(x^{sup})$ — результат работы детектора nn на размеченном изображении x^{sup} , а за $prob^{sup}$ и reg^{sup} — предсказанные

векторы вероятностей и регрессии соответственно:

$$prob^{sup}, reg^{sup} = nn(x^{sup}).$$

В режиме обучения с учителем, оптимизируется функционал:

$$L^{sup} = BCE(y, prob^{sup}) + \|reg^{sup} - reg_{true}\|_2^2,$$

где $BCE(y_{true}, prob^{sup})$ — функция потерь для классификации (бинарная кросс-энтропия), y_{true} — метка класса, а reg_{true} — истинный вектор регрессии.

При обучении с частичным привлечением учителя дополнительно вычисляются предсказания вероятностей $prob^{unsup}$ и регрессии reg^{unsup} для неразмеченного изображения x^{unsup} :

$$prob^{unsup}, reg^{unsup} = nn(x^{unsup}),$$

а также предсказания $prob_{aug}^{unsup}$ и reg_{aug}^{unsup} для $F(x^{unsup})$ — аугментированной версии исходного неразмеченного изображения:

$$prob_{aug}^{unsup}, reg_{aug}^{unsup} = F^{-1}(nn(F(x^{unsup}))),$$

где F^{-1} применяется к предсказаниям детектора.

Суть метода состоит в добавлении в итоговую функцию потерь компоненты L^{unsup} , отвечающей за обучение без учителя:

$$L^{unsup} = JS(prob^{unsup}, prob_{aug}^{unsup}) + \|reg^{unsup} - reg_{aug}^{unsup}\|_2^2,$$

где JS — расстояние Дженсона-Шеннона.

Тогда, итоговый оптимизируемый функционал L :

$$L = L^{sup} + L^{unsup}.$$

При тестировании на открытом наборе изображений Pascal VOC [2] метод позволил получить на валидационной выборке метрику 72.5 mAP, против 69.4 mAP при обыкновенном обучении с учителем.

Литература

1. Qizhe Xie, Zihang Dai, Eduard H. Hovy, Minh-Thang Luong, and Quoc V. Le. Unsupervised Data Augmentation, // CoRR, abs/1904.12848, 2019
2. Страница конкурса Pascal VOC Challenge: <http://host.robots.ox.ac.uk/pascal/VOC/>