Автоматическое детектирование и распознавание объектов на изображениях и видео

(2018 г., версия 1.3)

Введение.

Автоматическое детектирование и распознавание объектов на изображениях и видео является одной из основных задач компьютерного зрения. Как правило, эти задачи разбиваются на несколько подзадач: предобработка, выделение характерных свойств изображения объекта и классификация.

Этап предобработки обычно включает некоторые операции с изображением, такие как фильтрация, выравнивание яркости, геометрические корректирующие преобразования для облегчения устойчивого выделения признаков.

Под характерными свойствами изображения объекта понимается некоторый набор признаков, приближённо описывающий интересующий объект. Признаки можно разбить на два класса: локальные и интегральные. Преимуществом локальных признаков является их универсальность, инвариантность по отношению к неравномерным изменениям яркости и освещённости, но они не уникальны. Интегральные признаки, характеризующие изображение объекта в целом, не устойчивы к изменению структуры объекта и сложным условиям освещения. Существует комбинированный подход - использование локальных признаков в качестве элементов интегрального описания, когда искомый объект моделируется набором областей, каждая из которых характеризуется своим набором признаков – локальным текстурным дескриптором. Совокупность таких дескрипторов характеризует объект в целом.

Под классификацией понимают определение принадлежности объекта к тому или иному классу путём анализа вектора признаков, полученного на предыдущем этапе, разделения признакового пространства на подобласти, указывающие на соответствующий класс. Существует множество подходов к классификации: нейросетевые, статистические (Байеса, регрессия, Фишера и др.), решающие деревья и леса, метрические (ближайшие К-соседей, парзеновские окна и т.д.) и ядерные (SVM, RBF, метод потенциальных функций), композиционные (AdaBoost). Для задачи обнаружения объекта на изображении оценивается принадлежность двум классам – классу изображений, содержащих объект, и классу изображений, не содержащих объект (изображениям фона).

Важнейшим фактором, влияющим на качество классификации, ее устойчивость, является выбранное множество признаков, их способность отделять изображения объектов разных классов друг от друга. Чем большей разделительной способностью обладают признаки (возможно, благодаря более сложной структуре), тем проще устроено признаковое пространство, и классификатор может иметь простой вид. Наоборот, чем менее уникальны признаки (но проще по структуре),

тем сложнее устроено признаковое пространство и требуется более сложный классификатор для его успешного разделения. Композиция большого количества “простых” признаков может аппроксимировать небольшое количество “сложных”. Используя эти идеи, современным и эффективным является подход, объединяющий все этапы анализа изображения: предобработку, одновременное выделение множества “простых” признаков и их классификация на основе оптимизации по обучающей базе изображений многослойных сверточных нейронных сетей глубокого обучения (CNN), в которых процедура выделения признаков осуществляется в начальных слоях, являясь частью классификатора, структура признаков формируется автоматически в процессе его обучения и определяется моделью и архитектурой сети. Чем больше “простых” признаков необходимо использовать для характеристики (аппроксимации) целевых объектов, тем больше параметров требуется для задания модели сети, тем она вычислительно “сложнее”. CNN можно рассматривать как обобщенный метод моделирования признакового пространства, однако он требует значительных вычислительных ресурсов и объема обучающей выборки изображений, репрезентативно представляющих все необходимые классы объектов. Это связано с тем, что в CNN признаковая модель объекта формируется по правилам архитектуры сети на основе только той информации, которая содержится в обучающей базе. Напротив, в классическом подходе признаковая модель строится исследователем на базе априорной информации о характере изображений объекта. Таким образом, классическая подзадача разработки эффективного множества признаков, смещается на подзадачу разработки оптимальной архитектуры CNN, которая сама выделяет необходимые признаки из изображений в задаче классификации.

Основным критерием качества детектирующей или распознающей системы являются показатели ошибок классификации (и их производные) – объем ложно положительных (FAR, False Accept Rate) и ложно отрицательных решений (FRR, False Reject Rate) на тестовой выборке положительных и отрицательных примеров (изображений, содержащий или не соответствующий класс объектов). Для количественной характеристики качества строятся графики функций ROC (TPR/FAR), DET (FRR/FAR), Precision-Recall, по которым можно сопоставить результаты различных подходов и работы систем, решающих поставленную задачу.

Для разработки алгоритмов, обучения методов и тестирования существуют открытые базы изображений, которые пополняются исследователями. Некоторые базы доступны по адресам:

VOC2012
http://host.robots.ox.ac.uk/pascal/VOC/pubs/everingham10.pdf
http://host.robots.ox.ac.uk/pascal/VOC/
https://github.com/Microsoft/CNTK/tree/master/Examples/Image/DataSets/Pascal

INRIA Person Dataset
http://pascal.inrialpes.fr/data/human/

MNIST

https://en.wikipedia.org/wiki/MNIST_database

EMNIST
https://arxiv.org/abs/1702.05373

Fashion-MNIST
https://arxiv.org/abs/1708.07747

YouTube-8M
https://research.google.com/youtube8m/workshop2018/

KITTI
http://www.cvlibs.net/publications/Geiger2012CVPR.pdf
http://www.cvlibs.net/datasets/kitti/eval_object.php

1. Разработка методов построения локальных дескрипторов изображения для использования в качестве признаков в задачах детектирования и распознавания объектов на изображении.

Существуют разные подходы к выбору признаков для распознавания объекта. Как правило, тип признаков определяется характером объекта, который нужно распознать. Изображения объектов можно разделить на два вида: содержащие кусочно-линейные геометрические элементы и имеющие сложную внутреннюю структуру. Первый класс может быть описан с помощью контуров и границ, второй - в терминах текстуры. Выделение признаков объектов первого класса, как правило, опирается на дескрипторы особых точек изображения - углов, границ, пятен. Для объектов второго класса используется разложение области изображения или окрестностей точек в регулярной сетке по системе базисных функций, либо с помощью каскадной обработки фиксированным банком фильтров. Возможна комбинация этих подходов. Одна из таких комбинаций в настоящее время широко используется на основе Гистограммы Ориентаций Градиентов (HOG), показывающий в настоящее время один из лучших результатов для рассматриваемых задач и до сих пор конкурирующий со сверточными нейронными сетями. Существуют и другие более-менее удачные дескрипторы: Haar, LBP, SIFT, и др., которые в разной степени адекватности описывают локальные особенности изображения.

Ставится задача – улучшить один из этих дескрипторов или разработать свой при зафиксированном (заданном) методе классификации (например, SVM, AdaBoost) с целью получить графики ROC или DET лучше, чем у других методов или классификаторов.

Статьи про HOG:

http://pascal.inrialpes.fr/soft/olt

http://vc.cs.nthu.edu.tw/home/paper/codfiles/hkchiu/201205170946/

Histograms%20of%20Oriented%20Gradients%20for%20Human%20Detection.pdf

https://chrisjmccormick.wordpress.com/2013/05/09/hog-person-detector-tutorial/

Одним из вариантов обобщения HOG: обобщение на основе преобразования Радона.

Статьи про LBP:

http://www.scholarpedia.org/article/Local_Binary_Patterns

http://www.ee.oulu.fi/mvg/files/pdf/pdf_740.pdf

http://cs229.stanford.edu/proj2008/Jo-FaceDetectionUsingLBPfeatures.pdf

http://www.cse.oulu.fi/CMV/Downloads/LBPMatlab

Одним из вариантов обобщения LBP: вместо гистограмм распределения пикселей по LBP-коду, использовать гистограмму распределения модулей градиентов точек в блоке по LBP-коду (HG-LBP).

Статьи про другие дескрипторы:

http://www.scholarpedia.org/article/Scale_Invariant_Feature_Transform

http://www.miksik.co.uk/papers/miksik2012icpr.pdf

https://wwwpub.zih.tu-dresden.de/~cvweb/teaching/Courses/WS_2014_15/HS/UpdateOnFeatures_StefanHaller.pdf

Возможные задачи:

1.1. Разработка метода построения текстурных признаков на основе HOG/LBP/Haar/… дескриптора для задачи обнаружения объекта на изображении.

В настоящий момент хорошим выбором может быть детектор на основе алгоритма AdaBoost с Haar-дескрипторами, который является одним из самых быстрых методов, однако обладает большой специфичностью, результат сильно зависит от ракурса съемки, он не может быть обучен для различных ракурсов одновременно. Предлагается разработать новый дескриптор, обладающий меньшей специфичностью и, возможно, в дальнейшем предложить более гибкий алгоритм классификации, позволяющий разделять разные ракурсы объекта. В качестве целевых объектов можно взять: лица людей**,** пешеходов, автомобили. Задача остро востребована для нахождения частично загороженных пешеходов, автомобилей и сцен с плохими условиями освещения (пасмурно, сумерки).

1 **1.2. Разработка метода построения текстурных признаков на основе HOG/LBP/Haar/… дескриптора для задачи распознавания модели автомобиля на изображении.**

В задаче требуется не только распознать тип кузова, а также модель и производителя транспортного средства. Необходимо оценить влияние ракурса на качество классификации. Возможно сужение области интереса вплоть до радиаторной решетки, найденной на этапе локализации.

1.3. Разработка метода построения текстурных признаков на основе HOG/LBP/ Haar/… дескриптора для задачи распознавания мимики/пола по изображению лица.

Задача востребована как первый этап распознавания личности по лицу, так как обычно данные методы чувствительны к мимическим искажениям лица, целесообразно сначала оценить мимику и далее использовать эту информацию для увеличения надежности распознавания (на видеоизображении можно использовать фильтрацию по мимическому состоянию и пропускать только кадры с нейтральным выражением, лица либо использовать модель распознавания с учетом мимической модели); так же задача востребована для исследовательских целей в области рекламы и психологии, автоматической сортировки фотографий и т.д. Как можно классифицировать эмоции показано здесь: https://www.projectoxford.ai/demo/emotion#detection

2. Многослойные решающие деревья или каскады решающих деревьев.

На практике, труднопреодолимой проблемой является построение классификатора, позволяющего распознавать объекты инвариантно к их ракурсу на изображении, вследствие того, что в разных ракурсах (при поворотах по глубине) изображения объектов очень разнообразны. Один из возможных подходов - предварительная подготовка обучающей выборки, кластеризация ее на множества, содержащие изображения объектов в определенном ракурсе, и определения своего класса для каждого такого ракурса. Однако, такой подход для большего количества возможных ракурсов довольно трудоемкий. Ставится задача построения методов классификации на основе каскадов решающих деревьев, как альтернативы нейросетям глубокого обучения для обнаружения или распознавания интересующих объектов на изображении в различных ракурсах, и сравнение их с существующими решениями. Целью является построение такого обучения алгоритма классификации, который бы помог разделить объекты одного класса в различных ракурсах, от объектов другого ракурса без предварительной кластеризации по ракурсам обучающей выборки. В качестве признаков могут быть использованы низкоуровневые локальные дескрипторы, такие как HOG, LBP, Haar и др или адаптивное получение признаков из процедуры обучения. Эта тема требует серьезного исследования.

Предлагаемые задачи:

3.1 Разработка метода классификации на основе многослойных решающих деревьев/ каскадов решающих деревьев для задачи обнаружения объекта на изображении.

В качестве целевых объектов можно взять: лица людей**,** пешеходов (база INRIA), автомобили. Задача остро востребована для нахождения частично загороженных пешеходов, автомобилей и сцен с плохими условиями освещения (пасмурно, сумерки).

3.2 Разработка метода классификации на основе многослойных решающих деревьев/ каскады решающих деревьев для задачи распознавания модели

автомобиля на изображении.

В задаче требуется не только распознать тип кузова, а также модель и производителя транспортного средства. Классификатор должен быть инвариантен к ракурсу съемки.

https://en.wikipedia.org/wiki/Decision_tree

http://www.machinelearning.ru/wiki/images/a/a5/MOTP14_8.pdf

4. Сверточные нейронные сети глубокого обучения. *(тема под вопросом по техническим причинам – нет мощностей для обучения сетей)*

DeepLearning в последние годы бурно развивается и в мировых тестах с большим преимуществом перегоняет классические методы распознавания, такие как SVM, Байес, Решающий деревья (леса) и т.д. Однако вопрос построения оптимальной архитектуры остается открытым. На примере задачи обнаружения целевого объекта на изображении или на примере задач классификации (например, пола человека) предлагается исследование архитектуры сверточных нейросетей глубокого обучения.

Конкретный выбор прикладной задачи ограничивается только доступностью/ наличием достаточного объема обучающей и тестовой базы изображений. Общим недостатком задач, связанных с работой сверточной нейронной сети, кроме отсутствия достаточного объема обучающей выборки (можно использовать только открытые источники), является большие вычислительные ресурсы для ее глубокого обучения. Обычно для этой цели используется специальные сервера с массивом видео карт с CUDA или вычислительные кластера. Доступность соответствующих вычислительных ресурсов в обучающих и тестовых целях является отдельной организационной проблемой и требует отдельного внимания.

Предлагаемые задачи:

4.1 Разработка архитектуры сверточной нейронной сети глубокого обучения для детектирования объектов на изображении.

В частности, в качестве целевого объекта можно взять очки на лице человека, автомобиль на дороге, пешеход.

4.2 Разработка архитектуры сверточной нейронной сети глубокого обучения для распознавания модели автомобилей на изображении.

В задаче требуется не только распознать тип кузова, а также модель и производителя транспортного средства. Необходимо оценить влияние ракурса на качество классификации. Возможен подход использования известного положения и ракурса, оцененного на этапе локализации объекта, как одного из входов сети, а также возможно сужение области интереса вплоть до радиаторной решетки.

4.3 Разработка сверточной нейронной сети для 3D реконструкции изображения лица по стереопаре.

Использование 3D информации на изображении лица (в дополнении к вертикальной и горизонтальной координате каждой точки добавляется глубина) позволяет построить 3D поверхность лица, что дает значительное преимущество в задаче распознавания личности по сравнению с обычным использованием 2D фотографии. Для нахождения карты глубин используется процедура стереореконструкции по двум стереопарам, изображениям полученных с двух разнесенных на некоторое фиксированное расстояние камер. Зная соответствие точек сцены на стереопарах, их относительное расположение на паре изображениях (диспаратность) можно оценить расстояние каждой точки изображения от камеры в мировой системе координат. Для нахождения этих соответствий, как правило, используют алгоритмы поиска максимальной локальной корреляции окрестностей каждой точки на паре изображений. Этот подход обладает рядом недостатков: требует значительных вычислительных затрат, не гарантирует нахождения гладкого решения на всей области, достижения глобального оптимума. В данной задаче предлагается разработать и реализовать нейронную сеть, позволяющую на основе предварительного обучения построить карту диспаратности для предъявляемой стереопары. Это может быть регрессионная сеть, позволяющая находить гладкие решения, использующая неявную модель лица (полученную на этапе обучения) для заполнения (интерполяции) недостающей информации о положении точек. Нейронная сеть может быть обучена на базе изображений стереопар с известными картами диспаратности, либо с помощью добавления специальных выходных слоев сети, реализующих проверку критерия корректности найденной диспаратности. Задача довольно объемна, и в случае разработки нестандартных слоев, отсутствующих в современных фреймворках (таких как caffe), потребует некоторых программистских усилий для их реализации.

Статьи:

http://vision.stanford.edu/teaching/cs231b_spring1415/slides/alexnet_tugce_kyunghee.pdf

http://mi.eng.cam.ac.uk/~cipolla/publications/inproceedings/2016-PAMI-SegNet.pdf

http://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Szegedy_Going_Deeper_With_2015_CVPR_paper.pdf

https://arxiv.org/pdf/1512.03385v1.pdf

https://arxiv.org/pdf/1409.1556v6.pdf

http://caffe.berkeleyvision.org/

https://github.com/BVLC/caffe/tree/windows