

АВТОМАТИЗАЦИЯ РАЗМЕТКИ НАБОРА ДАННЫХ ДЛЯ НЕЙРОННЫХ СЕТЕЙ

Н. О. Бесшапошников¹, М. А. Кузьменко^{1,3}, А. Г. Леонов^{1,2,3}, М. А. Матюшин^{1,3}

¹*Федеральный научный центр Научно-исследовательский институт
системных исследований Российской академии наук*

²*Московский педагогический государственный университет*

³*Московский государственный университет имени М. В. Ломоносова*

nikita.beshaposhnikov@gmail.com, gmk@infomir.ru, dr.l@math.msu.ru, itsaprrank@yandex.ru

Процесс цифровой трансформации затрагивает все сферы жизни современного общества и деятельности человека: экономика, экология, научная деятельность и т. д. Возникают новые технологии, требующие эффективных методов распознавания элементов окружающей среды, отслеживания движений управляемых объектов во взаимодействии, в том числе и с дополненной реальностью. При этом все большую популярность для решения задач распознавания объектов на статических изображениях, а также в видеопотоках приобретают нейронные сети. Разметка набора данных для последующего глубокого обучения – неотъемлемая часть любого проекта, который предполагает использование нейросетевых технологий компьютерного зрения. Для обучения нейронных сетей задаче классификации изображений или нахождения объектов требуется огромное количество размеченных данных. Чем больше классов объектов, тем больше данных требуется для обучения. Естественно, что трудоемкая ручная разметка имеет ряд недостатков: длительное время и возможные ошибки, возникающие в силу рутинности выполняемой задачи. В статье рассмотрены различные подходы к автоматизации задачи разметки изображений для глубокого обучения, которые могут быть использованы для интеграции технологий распознавания образов и создания дополненной реальности с возможной реализацией на персональных компьютерах и мобильных устройствах.

Ключевые слова: нейросети, глубокое обучение, наборы данных, автоматическая разметка, распознавание образов.

AUTOMATION OF DATA MARKUP FOR NEURAL NETWORKS

N. O. Besshaposhnikov¹, M. A. Kuzmenko^{1,3}, A. G. Leonov^{1,2,3}, M. A. Matyushin^{1,3}

¹*System Research Institute, Russian Academy of Sciences*

²*Moscow Pedagogical State University*

³*Lomonosov Moscow State University*

nikita.beshaposhnikov@gmail.com, gmk@infomir.ru, dr.l@math.msu.ru, itsaprrank@yandex.ru

The process of digital transformation affects all spheres of life in modern society. All areas of human activity are now subject to digitalization in one form or another: economics, ecology, scientific activities, etc. Although some technologies, together with their accompanying competencies, become outdated, new technologies are emerging that require effective methods for recognizing elements of the environment, tracking the movements of controlled objects in interaction, including augmented reality. At the same time, neural networks are becoming increasingly popular for solving problems of recognizing objects in static images, as well as in video streams. The data markup for further deep learning is an integral part of any project that involves the use of a neural network of computer vision technologies. For training neural networks the task of classifying images or finding objects requires a huge amount of marked data. The more classes of objects, the more data is re-

quired for training. Naturally, time-consuming manual marking has a number of drawbacks: the long marking time itself and possible errors due to the routine of the task. The article discusses various approaches for automating the task of marking images for deep learning, which can be used to integrate image recognition technologies and create an augmented reality, with possible implementation on personal computers and mobile devices.

Keywords: neural networks, deep learning, data sets, automatic markup, pattern recognition.

За последние несколько лет технологии по распознаванию образов с помощью нейронных сетей все чаще используются в различных отраслях, таких как робототехника, автомобилестроение, торговля, медицина, образование. Вычислительные мощности современных персональных компьютеров, смартфонов и планшетов позволяют не только распознавать элементы статических картинок, но и обрабатывать потоковое видео, например, с камеры, определяя и запоминая расположение объекта (человека или робота), а также его движения и взаимодействие с окружающей средой [1].

При этом многие задачи по-прежнему сводятся к разработке и реализации законченного эвристического алгоритма. В некоторых случаях, когда в силу сложности задачи нахождение приемлемого эвристического решения невозможно, могут быть применены методы машинного обучения. Стандартные модели машинного обучения (SVM, Random Forest, модель логистической регрессии и т. д.) [2] достаточно хорошо показывают себя в задачах, связанных с обработкой скалярных или векторных признаков небольшой размерности [3]. Для обработки изображений в последнее время все чаще используются нейронные сети большой глубины, что порождает новую область исследований, так называемое глубинное обучение [4].

Для последующего эффективного использования нейронная сеть должна быть обучена, что означает (в машинном обучении) фактически подстройку коэффициентов или весов модели под определенную задачу. Методы обучения делятся на 3 типа: обучение с учителем, обучение без учителя и обучение с подкреплением. Остановимся подробнее на первом из них.

Метод обучения с учителем предполагает наличие обучающей выборки – множества пар, где первый элемент является информацией, подаваемой на вход модели, а второй является точным ответом, который мы ожидаем получить. К такой обучающей выборке предъявляется целый ряд требований, таких как сбалансированность различных случаев, корректность, полнота покрытия задачи, возможность последующей аугментации и т. д. В задачах компьютерного зрения, а именно в задаче локализации [5], остро стоит вопрос о так называемой разметке выборки, поскольку зачастую ее размеры исчисляются тысячами и десятками тысяч изображений. Чтобы получить размеченный датасет изображений (набор данных), могут применяться совершенно разнообразные подходы: от использования ручного человеческого труда вплоть до предобученных нейронных сетей с последующей ручной проверкой. В данной работе рассмотрим некоторые методы создания и разметки крупного датасета для обучения нейронной сети на детектирование объектов в рамках ограниченного семейства задач компьютерного зрения, когда целевые объекты детектирования находятся в пространстве в одной плоскости. Авторы предполагают, что предложенный подход имеет 3-мерное обобщение, что является предметом дальнейших исследований.

Автоматический OpenCV подход. Первоначально были проведены исследования возможности использования решения OpenCV [6] для детектирования границ целевых объектов по их цвету и их дальнейшей классификации по положению на плоскости, которое было заранее установлено перед началом съемки [7].

У такого подхода есть очевидные преимущества. Во-первых, он полностью автоматический, что означает крайне высокую скорость разметки и низкую ее стоимость. Во-вторых, в условиях хорошей освещенности границы целевых объектов определяются максимально точно, что в дальнейшем гарантирует хорошее качество регрессионной части датасета.

Однако минусы данного подхода оказались слишком существенными: в условиях низкой или неравномерной освещенности достигнуть устойчивой работоспособности решения слишком сложно, как и в случае присутствия в кадре посторонних объектов, имеющих в своей цветовой палитре цвета границ целевых объектов. Также проблемой является классификация целевых объектов, то есть определение класса задетектированного объекта по его положению, так как даже при незначительных перспективных искажениях эвристические алгоритмы определения относительного местоположения целевых объектов (так называемой табуляции) перестают работать корректно. Разметка этим способом позволяет автоматически разметить датасет, снятый в условиях освещенности, близких к идеальным, и при минимальных перспективных искажениях. При обучении на таком датасете нейронная сеть не научится детектировать искаженные целевые объекты, и качество обучения будет низким. Кадры, которые не прошли автоматическую разметку, предлагалось размечать полностью вручную, что крайне затратно по времени. В результате использования такого подхода в проведенном эксперименте было размечено 2 500 изображений с 12–16 представителями на каждом, полностью ручной разметке при этом подверглись около тысячи изображений.

Полуавтоматический подход. Если при сборе изображений для датасета есть возможность обеспечить в некоторых пределах неподвижность плоского объекта, то можно впоследствии воспользоваться полуавтоматическим подходом разметки. Он состоит в следующем: плоские целевые объекты обучения помещаются на плоскость (поверхность) и затем производится съемка нескольких кадров (изображений).

Полученные изображения формируют один набор, названный здесь пачкой. В пачке может быть от 100 до 500 изображений, в зависимости от целевого размера датасета. Создание большего количества кадров не имеет смысла в виду того, что целевые объекты обучения не меняют своего положения, следовательно, кадры получаются очень похожими друг на друга.

Затем производится инициальная разметка. Для этого из пачки выбирается изображение, являющееся наиболее фронтальным к поверхности съемки (ось камеры максимально приближена к перпендикуляру плоскости, на которой расположены целевые объекты, в этом случае перспективные искажения минимальны). На выбранном изображении размечаются опорные точки (в количестве от 3 и до произвольного, заранее выбранного числа), а также точные границы целевых объектов обучения. Позиции опорных точек могут быть произвольными, однако никакие три из них не должны лежать на одной прямой. Рекомендуется располагать их так, чтобы их выпуклая оболочка охватывала область, в которой расположены целевые объекты обучения.

После инициальной разметки проводится полная разметка всей пачки: на каждом изображении пачки указываются те же опорные точки, что были выбраны на инициальном изображении, в том же порядке, в каком они были отмечены изначально. Порядок обеспечивает взаимно-однозначное соответствие между инициальными и смещенными опорными точками. Разметка опорных точек на всех изображениях из пачки завершает ручную часть полуавтоматического подхода.

Далее для каждого целевого объекта инициального изображения, для каждой отмеченной точки его границы формируется ее опорный базис. Ближайшая к точке границы опорная точка объявляется началом координат опорного базиса, две следующие по удаленности точки формируют его оси. Если было выбрано N опорных точек, то количество опорных базисов будет составлять $N!/(N-3)!$, то есть количество возможных выборок 3 точек из N точек без учета порядка. Опорные точки, таким образом, разбивают плоскость изображения на свои области действия. Границы областей действия опорных базисов проходят по серединным перпендикулярам к отрезкам, соединяющим опорные точки.

После данного вычисления для каждого изображения из пачки для каждого инициального опорного базиса и соответствующего ему сдвинутого опорного базиса производится расчет линейного преобразования, переводящего первый базис во второй. Полученное пре-

образование затем применяется ко всем точкам разметки инициального изображения, попавшим в область действия данного инициального опорного базиса, что дает в результате новые точки границы целевого объекта обучения на новом изображении пачки. Данная процедура продлевается для всех точек границ целевых объектов обучения, отмеченных на инициальном изображении.

Таким образом, автоматическим образом из полной разметки всего лишь одного изображения и разметки опорных точек на остальных изображениях пачки методом склейки $N!/(N-3)!$ линейных преобразований формируется полная разметка всех изображений пачки.

Преимущества полуавтоматического подхода перед полностью ручным подходом очевидны: трудозатраты на разметку сокращаются в разы, полностью вручную необходимо разметить всего несколько десятков изображений. В дальнейшем при разметке опорных точек на каждую точку уходит порядка N секунд времени. При количестве представителей на изображении в большем количестве, чем $N/4$ (предполагается, что для выделения одного целевого объекта обучения достаточно отметить 4 точки), разметка опорных точек будет менее затратной по времени.

По сравнению с полностью автоматическим данный подход смотрится выигрышно, если имеется необходимость детектирования целевых объектов в условиях сильных перспективных искажений или неравномерной, плохой освещенности, так как все подобные изображения в автоматическом подходе не могут быть размечены.

Также стоит отметить низкую цену ошибки классификации, так как она наличествует либо на всех фотографиях пачки, либо ни на одной из них, и ее исправление достаточно сделать в одном изображении (инициальном). Остальные изображения при этом размечать заново не нужно.

Количество опорных точек и их местоположение в данном подходе являются гиперпараметрами. Для формирования хотя бы одного опорного базиса должно быть отмечено больше 2 точек. Местоположение должно обеспечивать равноудаленность всех точек границ от множества опорных точек, то есть опорные точки должны образовывать ерс-сеть на области изображения, содержащей целевые объекты обучения; чем меньше ерс, тем выше будет конечное качество разметки. Эксперименты показывают, что при отсутствии в пачке сильных перспективных искажений и небольших относительных размеров области изображения, содержащей целевые объекты обучения, достаточно выбора трех опорных точек. При умеренных перспективных изображениях рекомендуется использовать четыре опорные точки. Сильные перспективные искажения практически полностью нивелируются шестью опорными точками.

Такой подход имеет перспективу усовершенствования полуавтоматического подхода в следующих направлениях:

- автоматизация нахождения опорных точек на изображении (тренировка сверточной нейросети на нахождение физических объектов опорных точек с дальнейшим расположением их в кадре) с целью исключить стадию ручной разметки опорных точек;
- при оптических искажениях линзы камеры перед применением линейных преобразований можно производить корректировку линзовых эффектов на изображении, что увеличит конечную точность разметки;
- использование двумерных линейных преобразований не способно обеспечить стопроцентную точность при разметке. Восстановление реального движения камеры в трехмерном пространстве становится возможным при разметке 12 опорных точек, однако по затратам времени она сравнима с полной разметкой трех целевых объектов обучения, и в дальнейшем данное усовершенствование целесообразно только при автоматизации процесса разметки опорных точек. На рис. 1–4 приведены результаты разметки изображения с протяженной областью целевых объектов обучения по 4 (рис. 1, 2) и по 6 (рис. 3, 4) опорным точкам, которые отмечены на рисунках большими белыми точками, а маленькими – точки границ.

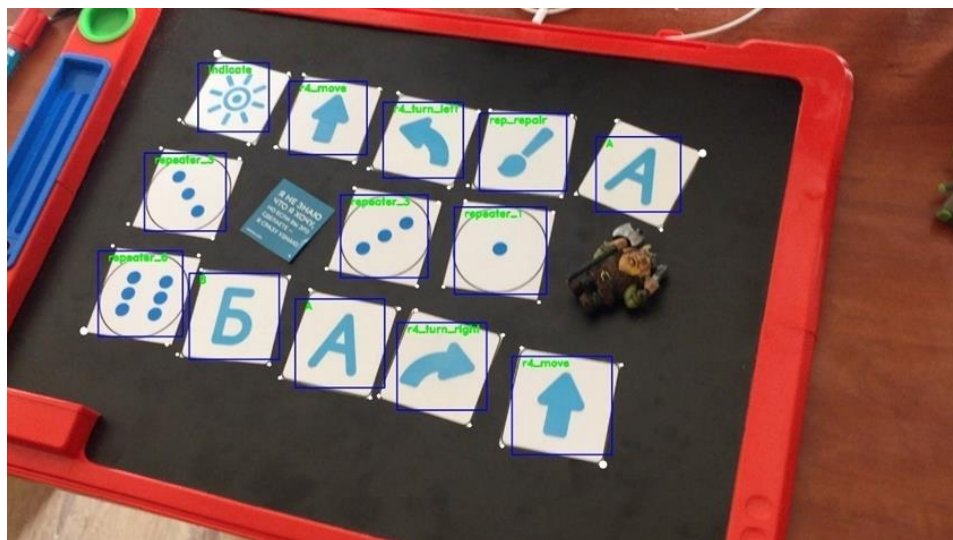


Рис. 1. Начальное изображение (разметка по 4 точкам)



Рис. 2. Искаженное изображение (разметка по 4 точкам)

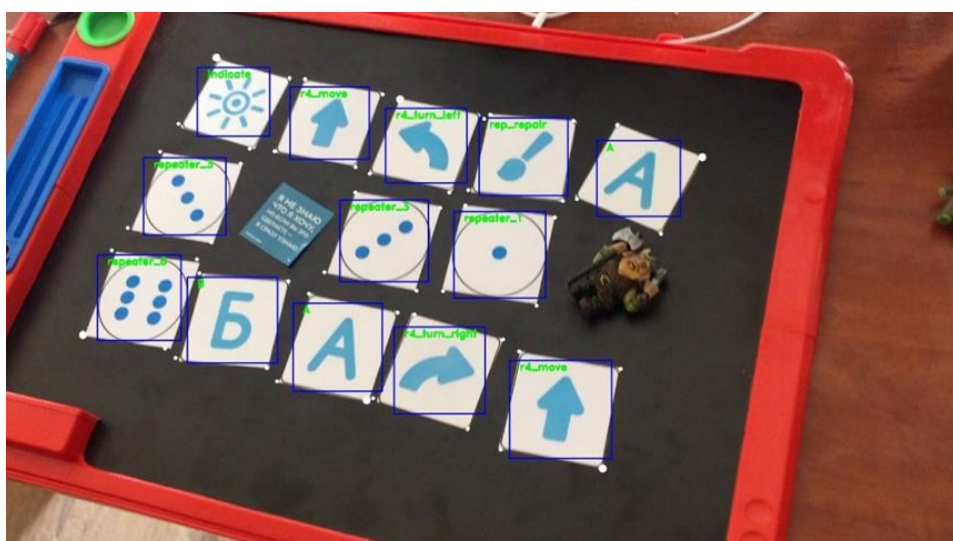


Рис. 3. Начальное изображение (разметка по 6 точкам)

8. Object Localization and Detection. URL: https://leonardoaraujosantos.gitbooks.io/artificial-intelligence/content/object_localization_and_detection.html (дата обращения: 20.11.2018).