

ПАВЛОВ
Максим Павлович

бакалавриат, Петрозаводский государственный университет
(Петрозаводск, Российская Федерация)
maksim pavlov 2003@list.ru

ПОДГОТОВКА ДАННЫХ ДЛЯ МАШИННОГО ОБУЧЕНИЯ РАСПОЗНАВАНИЯ ОБЪЕКТОВ И ИХ КЛЮЧЕВЫХ ТОЧЕК

Научный руководитель:

Марахтанов Алексей Георгиевич

Статья поступила: 01.04.2023;

Принята к публикации: 02.04.2023;

Опубликована: 15.04.2023

Аннотация. Современные методы искусственного интеллекта требуют тщательной выборки данных для достижения наилучшей точности. В статье рассматриваются подходы, связанные с улучшением набора данных машинного обучения распознавания объектов и ключевых точек. Автор формирует методику их подготовки и работы с ними.

Ключевые слова: подготовка данных, распознавание объектов, апробация датасета, контроль качества данных, аннотирование и маркировка данных

Для цитирования: Павлов М. П. Подготовка данных для машинного обучения распознавания объектов и их ключевых точек // StudArctic Forum. 2023. Т. 8, № 1. С. 9—14.

ВВЕДЕНИЕ

Задача распознавания объектов и распознавания на них определенного набора точек – одна из ведущих в компьютерном зрении. Для её решения существует целый набор различных алгоритмов и реализаций. Но перед их применением всегда возникает задача подготовки данных для обучения. В настоящее время многие исследователи и предприятия имеют ограниченный доступ к подготовленным данным: их недостаточно [Willemink]. И, как правило, они не удовлетворяют все потребности исследователей, так как каждая практическая задача компьютерного зрения уникальна. Подготовка данных – дорогостоящий и трудоёмкий процесс, результатом которого являются обученные алгоритмы с ограниченной сферой применения. Появляется всё больше компаний, предоставляющих услуги по их созданию, но они ещё не могут покрыть весь спрос исследователей. В данной работе описана методика по формированию наборов данных для распознавания объектов и их ключевых точек, основанная на практическом опыте Центра искусственного интеллекта ПетрГУ и включающая многие из современных подходов и стратегий. Методика была опробована на практике: сформирован датасет фотографий радужной форели, и сравнены показатели при обучении нейронной сети на первоначальном наборе данных (без использования некоторых из предложенных методов), и после использования всех подходов по улучшению.

МЕТОДИКА ПО ПОДГОТОВКЕ ДАННЫХ ДЛЯ РАСПОЗНАВАНИЯ ОБЪЕКТОВ И ИХ КЛЮЧЕВЫХ ТОЧЕК

Перед использованием изображений для разработки алгоритма машинного обучения необходимо предпринять определённые шаги (Рисунок 1). Как правило, всё начинается с анализа условий, в которых планируется использовать нейронную сеть. Собранные данные должны им полностью соответствовать и предусматривать все возможные сценарии. Также необходимо разработать алгоритм получения изображений объекта исследований, согласовать возможные правовые и этические нюансы. Одной из самых сложных задач является получение доступа к данным, например, к записям с городских камер видеонаблюдения. Идеальным вариантом является сотрудничество разработчиков систем видеонаблюдения с частными и государственными компаниями. Также всегда есть вариант сбора данных собственными силами, с возможным поиском изображений из открытых интернет-источников.

Все собранные данные хранятся либо в локальном хранилище (при этом обычно создается RAID-массив), либо в облачном. Преимуществами локального хранения являются безопасность и доступность данных для обучения: снижаются временные затраты на подачу изображений непосредственно нейронной сети. Но при этом ограничивается обмен данными между другими группами исследователей. Облачное хранилище решает данную проблему, но оно обладает своими недостатками: меньше уровень безопасности, существует постоянная необходимость в высокоскоростном соединении.



Рис. 1. Схема подготовки данных для видеоаналитики объекта и его ключевых точек

После сбора данных об объектах их необходимо структурировать - разбить на соответствующие подмножества, описывающие все возможные ситуации положения объекта и внешней среды. Тщательное разбиение на подмножества пригодится в дальнейшем при улучшении собранного набора данных на основе самосовершенствования модели. Перед работой с данными они обязательно должны пройти предварительную проверку. Качество и объём изображений зависит от предметной области и задачи исследования. В данных часто есть определенные изъяны, поэтому рекомендуется, чтобы одну и ту же аннотацию проверило несколько экспертов с целью свести человеческий фактор и индивидуальные особенности разметки каждого аннотатора к минимуму. Затем из каждого изображения выделяется вся необходимая для обучения нейронной сети информация. Как правило, это прямоугольные области, каждая из которых содержит один из объектов изучения, и определённый набор его ключевых точек. Полученные аннотации сохраняются в определённом формате, предусмотренном алгоритмом загрузки данных выбранной нейронной сети для обучения, и связываются с исходными изображениями. Одними из популярных форматов для распознавания объектов является COCO Json и Pascal VOC [Renu]. Рекомендуется подготавливать аннотации именно в них, так как в дальнейшем можно будет конвертировать их в любой другой формат. Также отметим, что, как правило, процесс подготовки данных – итеративный. После сбора небольшого, первоначального набора подготовленных аннотаций и изображений можно их включать в обучение нейронной сети с последующим добавлением или исключением данных исходя из оценки самосовершенствования обученной модели, её показателей на тестовом наборе данных после обучения. Такой подход часто называется активным обучением модели [Способы обеспечения качества данных для машинного обучения], и это обеспечивает улучшение не только обученных моделей, но и самого набора данных. Однако может потребоваться большое количество запусков, чтобы выявить то подмножество данных, которые вносят отрицательный вклад в обучение.

В последнее время наблюдается тенденция к увеличению не только количества данных при обучении, но и их качества [Willemink]. Так как успешность и точность распознавания как объекта, так и его ключевых точек напрямую зависит от качества и разнообразия данных, то стоит придерживаться всех современных и продуктивных стратегий по формированию набора данных. И только пройдя все этапы формирования набора данных, можно раскрыть весь потенциал точности предсказаний модели.

ПРАКТИЧЕСКИЙ ПРИМЕР - СОСТАВЛЕНИЕ ДАТАСЕТА РАДУЖНОЙ ФОРЕЛИ

В последнее время технологии машинного обучения всё чаще используются в рыбопромышленной отрасли - возникает необходимость в формировании набора данных для изучения рыбы [Марахтанов]. Радужная форель – один из многочисленных представителей семейства лососёвых. Её активно вылавливают в естественной среде обитания, выращивают в искусственных садках рыбных хозяйств. И для решения определённых прикладных задач рыбоводческих хозяйств, а также наблюдения за особями необходимы данные о радужной форели: её местоположение в кадре и совокупность определенных ключевых точек. В результате нашего анализа объекта исследований были выделены следующие 8 ключевых точек радужной форели (Рисунок 2): mouth, eye, dorsal fin, adipose fin, tail, anal fin, ventral fin, pectoral fin. Они являются ключевыми для дальнейших исследований. Например, с их помощью можно оценить массу особи [Царёв].

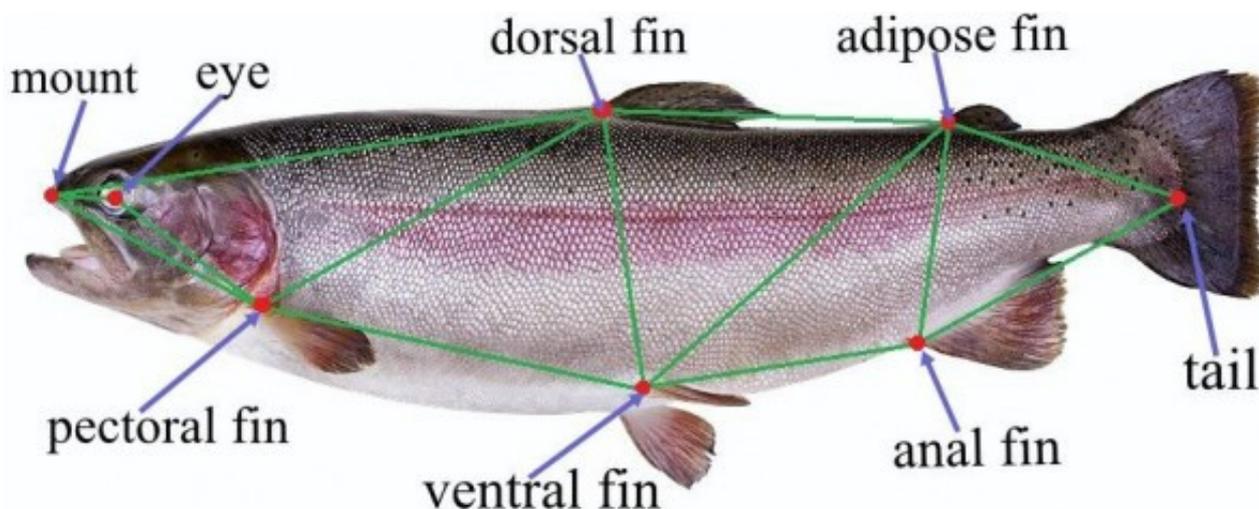


Рис. 2. Ключевые точки радужной форели

Разметка фотографий проходила с помощью популярной библиотеки COCO Annotator [COCO Annotator], развёрнутой в отдельном кластере, чтобы обеспечить возможность разметки с разных устройств и для нескольких экспертов. Перед подготовкой данных был сформирован набор инструкций, регулирующий многие из возможных событий и вопросов при аннотации, которому придерживались все аннотаторы. Например, если видно одновременно оба глаза и/или оба плавника pectoral fin или ventral fin (Рисунок 3), то размечалась только та сторона рыбы, которая лучше всего видна в кадре. В случае если видно одинаково хорошо все элементы (eyes, pectoral fin, ventral fin) аннотировалась правая сторона относительно рыбы. Если рыба смотрит прямо в кадр, то mouth перепроверялся несколько раз другими экспертами, так как данная точка - геометрическая середина верхней губы рыбы, и её неверное местоположение может изменить предсказание всех точек такого типа. В результате в подготовленном наборе данных всегда была размечена лишь одна сторона рыбы – левая или правая, которую было видно в кадре лучше всего. Такой подход помогает в дальнейшем распознавать только одну сторону рыбы с соответствующими ключевыми точками, без смешивания точек с разных сторон.

При подготовке данных также важно аннотировать все объекты в кадре, это может быть

критично для некоторых из современных детекторов объектов - например, для YOLO [Redmon], который мы использовали при решении задачи. Если не разметить объект, который видно в кадре, то нейронная сеть в процессе обучения будет получать «штрафы» при его распознавании, что может ухудшить весь процесс обучения. Но если отметить что-то непонятное и излишнее, то можно научить детектировать лишнее. Исходя из этого, размечались все рыбы в кадре: если эксперту было понятно, что это рыба, и он мог точно увидеть на ней хотя бы одну из ключевых точек. На каждой отмеченной рыбе ставились все возможные ключевые и скрытые точки. Точка является скрытой, если перекрыта в кадре другим объектом, но по видимой части можно с большой вероятностью понять её фактическое местонахождение. Для дальнейшей работы со скрытыми точками использовалась система флагов видимости каждой точки, предусматриваемая форматом аннотаций COCO Json [Renu]: 0 – точку не видно в кадре, 1 – точка является скрытой, 2 – точка видна в кадре. В дальнейшем такая система флагов была важна при обучении модели для создания возможностей предсказаний не только видимых, но и скрытых точек. При этом мы использовали контроль качества с последующей проверкой двумя экспертами каждой сформированной аннотации.

Все данные подготавливались и апробировались в рамках программно-аппаратного комплекса для управления жизненным циклом садка аквакультуры «Fish Grow Platform» компании «Интернет-бизнес-системы» [FishGrowPlatform. Сервисы для рыбоводов]. Сформированный датасет для обучения содержит 7662 изображения, 19371 особей радужной форели, 14891 из которых имеют ключевые точки. Он размечался в несколько проходов с системой проверки самосовершенствования модели нейронной сети. Также тестирование модели после обучения проходило на трёх специально подготовленных видео, приближенных к реальным условиям: множество рыб в ограниченном бассейне; одиночная особь в аквариуме; надводные манипуляции с рыбой. Такой подход обеспечивает полноценную предварительную оценку модели в приближенных к реальным условиям в системе видеоаналитики объектов.

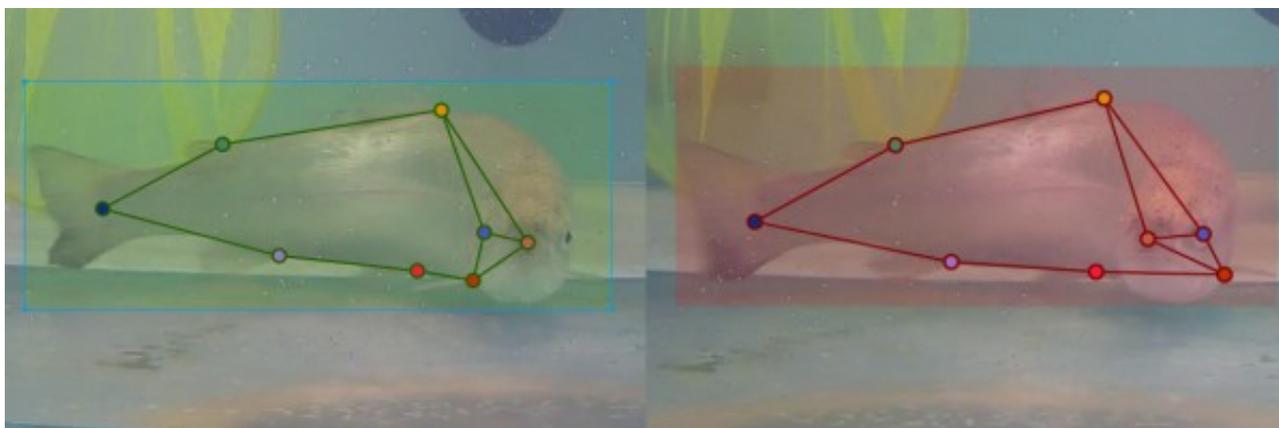


Рис. 3. Спорные случаи разметки, регламентированные правилами, для распознавания только одной из сторон рыбы

Рассмотрим конкретный практический пример, основанный на опыте создания набора данных радужной форели. Проанализируем обучение одной и той же нейронной сети на двух разных наборах данных. В первом случае не применялись:

- 1) итеративное изменение набора данных на основании самосовершенствования модели, улучшения её показателей после каждого обучения;
- 2) контроль разметки несколькими экспертами.

Описанный ранее датасет содержал 9997 изображений, 29064 особи радужной форели, 17685 из которых с ключевыми и скрытыми точками. Обученная на нём модель достигала: 0.89 mAP 0.5; 0.63 mAP 0.5:0.95; среднюю погрешность предсказания координат точек 4-6% относительно размеров особи в кадре.

При использовании же всех описанных методов набор данных для обучения стал меньше на 24% и составлял 7662 изображений, 19371 особей радужной форели, 14891 из

которых содержат ключевые и скрытые точки. На нём нейронная сеть показала следующие результаты: 0.93 mAP 0.5 (+4,49%); 0.7 mAP 0.5:0.95 (+11,1%); среднюю погрешность предсказания координат точек 1-2% относительно размеров особи в кадре.

Полученная разница между показателями показывает, что для улучшения точности предсказаний одной и той же модели нейронной сети стоит начать именно с улучшения набора данных, на котором она обучается.

ЗАКЛЮЧЕНИЕ

На подготовку данных для машинного обучения может уходить подавляющее большинство времени, требуемого для решения задач компьютерного зрения. Ещё на этапе анализа объекта, его разметки, могут возникнуть необходимые предположения и гипотезы, которые будут полезны и на этапе машинного обучения. Создание отлаженного механизма по сбору и обработке данных – крайне кропотливый процесс, требующий точности, концентрации и усидчивости всех аннотаторов. Даже на относительно небольшом наборе данных можно достичь лучших результатов, если более тщательно подойти к подготовке датасета. Несмотря на все сложности, формирование наборов данных для машинного обучения является той необходимой и ключевой частью, которая не только даёт возможность обучения моделей искусственного интеллекта в целом, но и помогает совершенствовать и расширять сферу их применения.

СПИСОК ЛИТЕРАТУРЫ

Марахтанов А. Г. Fish Grow Platform // Цифровые технологии в образовании, науке, обществе : материалы XV всероссийской научно-практической конференции. Петрозаводск, 2022. С. 56—59. URL: <https://it2022.petrso.ru/doc/it2022.pdf?t=1671521859> (дата обращения: 31.03.2023).

Способы обеспечения качества данных для машинного обучения. URL: <https://habr.com/ru/post/588266/>. (дата обращения: 31.03.2023).

Царёв Н. В. Оценка массы рыбы на основе её линейных размеров, измеренных методами видеоанализа // Цифровые технологии в образовании, науке, обществе : материалы XV всероссийской научно-практической конференции. Петрозаводск, 2022. С. 112—115. URL: <https://it2022.petrso.ru/doc/it2022.pdf?t=1671521859>. (дата обращения: 31.03.2023).

COCO Annotator. URL: <https://github.com/jsbroks/coco-annotator?ref=madewithvuejs.com> (дата обращения: 31.03.2023).

Fish Grow Platform. Сервисы для рыбоводов. URL: <https://inbisyst.ru/solutions/product/fish-grow-platform> (дата обращения: 31.03.2023).

Preparing Medical Imaging Data for Machine Learning / M. J. Willemink, A. K. Wojciech, H. Cailin [et al.] // Radiology. 2020. Vol. 1, No. 295. P. 4—15.

Renу К. COCO and Pascal VOC data format for Object detection // Medium [Электронный ресурс]. Режим доступа: <https://towardsdatascience.com/coco-data-format-for-object-detection-a4c5eaf518c5> (дата обращения 31.03.2023)

You only look once: Unified, real-time object detection / J. Redmon, S. Divvala, R. Girshick [et al.] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. P. 779-788.

Maksim P. PAVLOV

bachelor's degree, Petrozavodsk State University
(Petrozavodsk, Russian Federation)
maksim pavlov 2003@list.ru

PREPARING DATA FOR MACHINE LEARNING OBJECT AND KEY POINTS RECOGNITION

Scientific adviser:

Aleksey G. Marakhtanov
Paper submitted on: 04/01/2023;
Accepted on: 04/02/2023;
Published on: 15/04/2023.

Abstract. Modern methods of artificial intelligence require careful selection of information to obtain the best accuracy. The article considers approaches to improve the data set of machine learning objects and key points recognition. The author forms a methodic of preparing and working with these.

Keywords: data preparation, object recognition, dataset approbation, data quality control, data annotation and labeling

For citation: Pavlov M. P. Preparing data for machine learning object and key points recognition. *StudArctic Forum*. 2022; 8(1): 9—14.