

Генерация датасетов для учебных задач компьютерного зрения

А.Н. Вильданов

Уфимский университет науки и технологий, Нефтекамский филиал, Нефтекамск

Аннотация: В настоящее время существует множество причин, почему обучение технологиям искусственного интеллекта (ИИ) может быть важным для сегодняшних студентов. Поэтому дисциплины, связанные с ИИ, вузы активно включают в образовательные программы бакалавриата и магистратуры. Важно научить студентов понимать, как работают такие технологии, и как их можно использовать для решения различных задач. В свою очередь, обучение немислимо без демонстрации примеров решения различных задач. Важным же этапом в решении задачи машинного обучения вообще, в т.ч. задачи компьютерного зрения, является этап формирования обучающей выборки. Поэтому возникла идея написания программы, которая способна была бы генерировать датасеты по разнообразным тематикам для задач компьютерного зрения. Формат данных генерируемой выборки для обучения train.csv общепринят и выглядит следующим образом: каждая строка представляет собой описание одного изображения; первый столбец содержит метки классов, к которым принадлежит изображение; оставшиеся столбцы содержат пиксельные значения изображения, например, в виде плоского вектора, где каждое значение соответствует яркости соответствующего пикселя на изображении. Полученные датасеты можно использовать для организации проектной деятельности учащихся по искусственному интеллекту.

Ключевые слова: искусственный интеллект, машинное обучение, компьютерное зрение, нейронная сеть, обучающая выборка, датасет, C#, пиксель, субпиксельная обработка изображений, организация проектной деятельности учащихся.

Введение

С развитием технологий искусственного интеллекта (ИИ) возникла необходимость в обучении студентов в данной области науки и техники. Такое обучение важно для студентов в настоящее время, поскольку ИИ все больше внедряется в различные сферы жизни и работы. Их изучение поможет студентам понимать, как работают такие технологии, и как их можно использовать для решения различных задач. Это может быть полезно для будущей карьеры в области новейших технологий или любой другой отрасли, которая использует ИИ.

Сегодня искусственный интеллект обозначен, как одна из перспективных научных тенденций в развитии экономики Российской Федерации [1].



Далее, обучение технологиям ИИ может помочь студентам развивать навыки аналитического мышления и решения проблем, которые могут быть применимы в любой отрасли. Использование ИИ для обработки больших объемов данных может помочь студентам научиться эффективно анализировать информацию и выявлять тенденции.

Кроме того, не менее важным является то обстоятельство, что обучение технологиям ИИ может помочь студентам понимать возникающие сегодня этические и социальные вопросы, связанные с использованием ИИ. Студенты могут изучить вопросы, связанные с безопасностью данных, защитой личной информации и прозрачностью алгоритмов, что может быть полезно для формирования их собственных взглядов и мнений на эти вопросы.

В целом, обучение технологиям искусственного интеллекта может быть полезно для студентов, независимо от их будущей карьеры. Это поможет им развивать востребованные навыки.

Важная область искусственного интеллекта – это технологии компьютерного зрения. Она имеет множество применений в различных отраслях, таких, как медицина, производство, автомобильная промышленность, розничная торговля, безопасность и т.д. Одним из основных применений компьютерного зрения является автоматическое распознавание образов и классификация объектов. Это может быть полезно в различных задачах, например, в автоматической проверке качества продуктов на производственной линии, в автоматическом распознавании лиц на камерах видеонаблюдения, в диагностике заболеваний на медицинских изображениях и т.д.

Кроме того, компьютерное зрение может быть использовано для автоматического анализа изображений и видео, например, в анализе сцен в автомобилях с автопилотом, в определении трафика на дорогах, в анализе поведения клиентов в магазинах и т.д. Активно развиваются различные

методы распознавания изображений и системы, построенные на их основе, с успехом решают такие задачи, как идентификация отпечатков пальцев, роговицы глаза, анализ аэрокосмических снимков, обнаружение подделок и т.д. [2]. Ещё одним изобретением стала система мониторинга космического пространства, назначение которой состоит в наблюдении за космическими территориями, космическим оборудованием и получаемыми данными об образующемся космическом мусоре [3].

В 2015 году была разработана модель YOLO (You Only Look Once) - это нейронная сеть глубокого обучения для задачи обнаружения объектов в изображениях. Она быстро стала очень популярной благодаря своей способности обнаруживать объекты на изображении за один проход, что делает ее быстрой и эффективной [4].

YOLO использует сверточные нейронные сети для извлечения признаков из изображения и прогнозирования координат ограничивающей рамки (bounding box) для каждого объекта на изображении, а также вероятности принадлежности каждого объекта к каждому из заданных классов.

Это позволяет быстро и точно обнаруживать объекты различных классов, таких как люди, автомобили, животные и другие, и широко применяется в различных областях, включая компьютерное зрение, автономные системы и робототехнику.

Таким образом, компьютерное зрение является очень важной областью искусственного интеллекта, которая имеет множество применений в различных отраслях и может помочь улучшить эффективность и точность выполнения различных процессов и решения задач. И поэтому процесс изучения технологий искусственного интеллекта немислим без обучения решению задач компьютерного зрения.

Важным этапом в решении задачи машинного обучения вообще, в т.ч. задачи компьютерного зрения, является этап формирования обучающей выборки [5], или датасета. При этом на первый план выходит необходимость обеспечения репрезентативности обучающей выборки. Она имеет три аспекта [5]:

- достаточность — число обучающих примеров должно быть достаточным для обучения;
- разнообразие;
- равномерность представления классов

Существует множество источников, где можно найти датасеты для тренировки решения задач компьютерного зрения. Некоторые из наиболее популярных источников включают [6]:

- ImageNet: ImageNet является одним из наиболее широко используемых датасетов для обучения нейронных сетей компьютерного зрения. Он содержит более миллиона изображений, размеченных на более чем 1 тысячу классов.

- COCO: COCO (Common Objects in Context) - это еще один популярный датасет, который содержит изображения с разметкой объектов и сцен.

- CIFAR: CIFAR (Canadian Institute for Advanced Research) - это набор изображений, который содержит 10 классов изображений, каждый из которых имеет 50 000 образцов.

- MNIST: MNIST (Modified National Institute of Standards and Technology) - это набор изображений рукописных цифр, который широко используется для обучения нейронных сетей распознаванию рукописного текста.

- Open Images Dataset: Open Images Dataset - это набор изображений, содержащий более 9 миллионов изображений и 6000 категорий.

Кроме того, существует множество других датасетов, специализированных на различных задачах, таких как распознавание лиц, детектирование объектов, распознавание жестов и т.д. Важно выбрать датасет, который наилучшим образом соответствует задаче, которую предстоит решить.

При проведении занятий в учебной группе желательно обеспечить разнообразие задач для большей активности каждого учащегося и выполнение им самостоятельной работы [7]. Известно, что самостоятельная работа студентов способствует более эффективному овладению материалом, стимулирует познавательные и профессиональные интересы, развивает творческую активность и инициативу, способствует росту мотивации обучения [8]. Поэтому возникла идея написания программы, которая способна была бы генерировать датасеты по разнообразным тематикам.

Как это можно сделать? В работе [9] для обучения нейронной сети по распознаванию рукописных букв предложено искусственно генерировать изображения букв для датасета путем их рисования с помощью различных шрифтов на языке C# (исходный код можно найти в репозитории GitHub [9] в папке DatasetGenerator). В этом языке имеется класс `Graphics`, который предоставляет метод `DrawString` для рисования объектов. С его помощью для рисования и сохранения изображения буквы «А» можно использовать, например, такой код [10]:

```
char symb = 'A';  
var res = new Bitmap(image_width, image_height);  
var g = Graphics.FromImage(res);  
g.DrawString(symb.ToString(), drawFont, drawBrush, x, y,  
drawFormat);  
res.Save(path);
```

Следуя [9], чтобы подготовить требуемый датасет, будем генерировать изображения, выполнив следующие условия:

- буквы изображений будем записывать различными шрифтами;
- буквы будем записывать разными размерами. Главное, чтобы буквы были различимыми и не выходили за размеры изображения;
- будем наклонять буквы на некоторый угол (для аугментации данных).

Кроме того, по сравнению с работой [9], в предлагаемой работе проделаны следующие улучшения:

- при повороте буквы могут сильно сместиться от центра и даже выйти за границы рисунка. Этот недостаток исправлен путем расчета ограничивающей рамки для буквы и последующего сдвига буквы в центр картинки;

- можно генерировать изображения не только русских букв, но и другие популярные символы, например, знаки зодиака, греческие буквы, пиктограммы и т.д.;

- в задачах на компьютерное зрение для увеличения скорости обработки данных картинки обычно кодируются в один файл для обучения `train.csv` (что также реализовано в программе). В этом файле есть заголовочная строка – метка и номера пикселей. Далее идут строки данных. Каждая строка кодирует одну картинку:

```
label,pixel0,pixel1,pixel2,pixel3,pixel4,pixel5,...  
5, 0, 0, 0, 200, 250, 220,...  
2, 0, 0, 120, 170, 200, 180,...
```

- кроме того, часть данных (20%) содержит ответы и генерируется отдельно для валидации (файл `validate.csv`);

– можно указать (для проведения конкурса), что часть данных не содержит ответов. Тогда генерируется отдельно тестовые данные (файл test.csv) и ответы к нему (solution.csv).

Интерфейс программы достаточно простой (рис. 1). Имеется возможность выбора готовых вариантов (рис. 2).

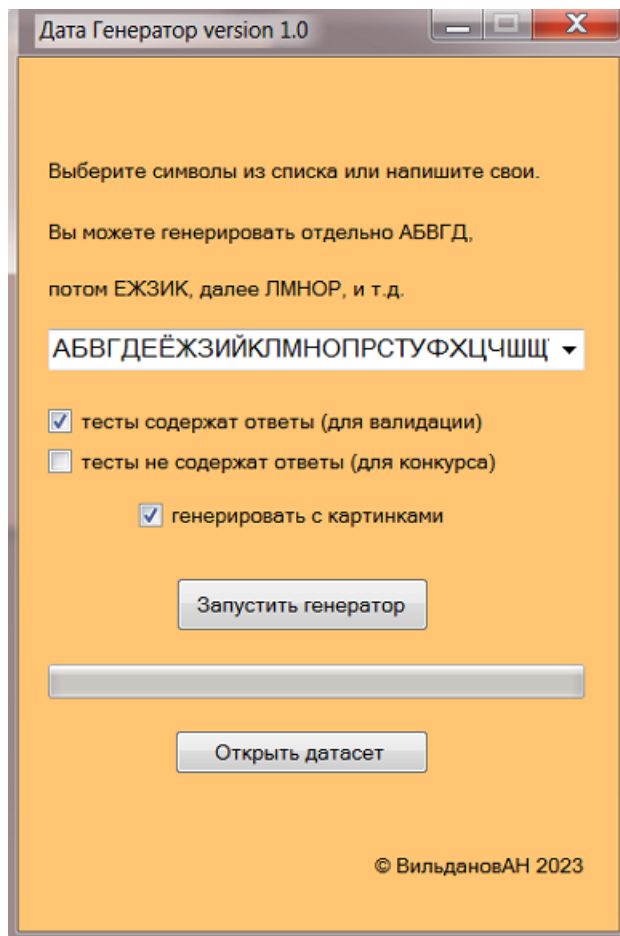


Рис. 1. – Интерфейс программы

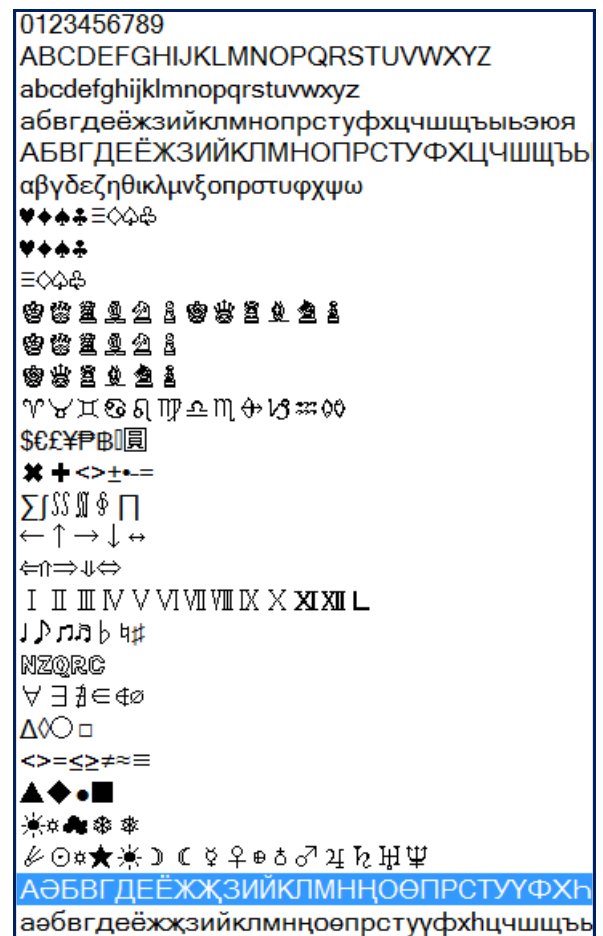


Рис. 2. – Варианты символов изображений

С помощью этой программы преподаватель может легко подготовить различные варианты заданий по компьютерному зрению. Например, построить и обучить модель сверточной нейронной сети для задачи распознавания шахматных фигур, для задачи распознавания обозначений

знаков зодиака, распознавания математических операторов, изображений греческих букв, карточных мастей и т.д.

Приложение выложено на Github. GitHub – это веб-платформа, которая позволяет разработчикам хранить код и другие типы программных проектов, делиться ими и совместно работать над ними. Ознакомиться с демонстрацией возможностей генератора датасетов для учебных задач компьютерного зрения можно по адресу gevaraweb.github.io/aidatagen/. Это приложение работает на компьютере под управлением ОС Windows 7 и выше, с установленным NET Framework 4.6+.

Приложение является бесплатным. Любой преподаватель может его использовать с учащимися на уроках или на внеурочных занятиях.

Литература

1. Элайджа Н. Цифровая Россия-2024. Как новые технологии изменят жизнь россиян // Российское информационное агентство. 2019. URL: tass.ru/nacionalnyye-proekty/6469993.

2. Крысова И.В. Экспертная система для автоматизации процесса классификации деталей по ЕСКД // Инженерный вестник Дона. 2013. № 4. URL: ivdon.ru/ru/magazine/archive/n4y2013/1888.

3. Боев С. ПАО МАК «Вымпел». Система мониторинга космического пространства // МАК ВЫМПЕЛ. 2022. URL: macvimpel.ru/upload/catalog/rucatalog-2022.pdf.

4. Hao, L., Jinguo, Zh., Xiaohu, J., Xizhou Zh., Hongsheng, L., Chun Y., Xiaohua, W., Yu Q., Xiaogang W., Wenhai W., Jifeng D. Uni-Perceiver v2: A Generalist Model for Large-Scale Vision and Vision-Language Tasks//arXiv preprint arXiv:2211.09808. 2022. URL: arxiv.org/pdf/2211.09808.pdf.

5. Галушка В.В., Фатхи В.А. Формирование обучающей выборки при использовании искусственных нейронных сетей в задачах поиска ошибок баз

данных // Инженерный вестник Дона. 2013. № 2. URL: ivdon.ru/magazine/archive/n2y2013/1597.

6. Krizhevsky A. Learning multiple layers of features from tiny images. 2009. URL: cs.toronto.edu/~kriz/learning-features-2009-TR.pdf.

7. Вильданов А.Н. О генерации многовариантных задач по теории игр // Достижения и приложения современной информатики, математики и физики: Материалы VI Всероссийской научно-практической заочной конференции, Нефтекамск, 01 ноября 2017 года. Нефтекамск: Башкирский государственный университет. 2017. С. 87 - 94.

8. Гарифуллин Р.И., Девяткин Е.М. Электронный комплекс виртуальных лабораторных установок по механике и молекулярной физике // Сборник научных статей международной молодежной школы-семинара «Ломоносовские чтения на Алтае»: в 6 частях, Барнаул, 05–08 ноября 2013 года / Под редакцией Родионова Е.Д. Барнаул: Алтайский государственный университет, 2013. С. 309-311.

9. Amialchenia Andrei. Handwriting_Letter_Recognition_ML_NET // github.com URL: github.com/flash2048/Handwriting_Letter_Recognition_ML_NET (дата обращения: 20.03.2023).

10. Graphics.DrawString метод // learn.microsoft.com / URL: learn.microsoft.com/ru-ru/dotnet/api/system.drawing.graphics.drawstring (дата обращения: 20.03.2023).

References

1. E`lajdzha N. Cifrovaya Rossiya-2024. Kak novy`e texnologii izmenyat zhizn` rossiyan [Digital Russia-2024. How new technologies will change the lives of Russians]. Rossijskoe informacionnoe agentstvo. 2019. URL: tass.ru/nacionalnye-proekty/6469993.

2. Krysova I.V. Inzhenernyj vestnik Dona, 2013, №4. URL: ivdon.ru/ru/magazine/archive/n4y2013/1888.
 3. Boev S. PAO MAK «Vympel». Sistema monitoringa kosmicheskogo prostranstva [Vympel. Space monitoring system]. MAK VYMPEL 2022. URL: macvypel.ru/upload/catalog/ru-catalog-2022.pdf.
 4. Hao, L., Jinguo, Zh., Xiaohu, J., Xizhou Zh., Hongsheng, L., Chun Y., Xiaohua, W., Yu Q., Xiaogang W., Wenhai W., Jifeng D. Uni-Perceiver v2: A Generalist Model for Large-Scale Vision and Vision-Language Tasks. arXiv preprint arXiv:2211.09808. 2022. URL: arxiv.org/pdf/2211.09808.pdf.
 5. Galushka V.V., Fathi V.A. Inzhenernyj vestnik Dona, 2013. № 2. URL: ivdon.ru/magazine/archive/n2y2013/1597.
 6. Krizhevsky A. Learning multiple layers of features from tiny images. 2009. URL: cs.toronto.edu/~kriz/learning-features-2009-TR.pdf.
 7. Vildanov A.N. Dostizheniya i prilozheniya sovremennoj informatiki, matematiki i fiziki: Materialy VI Vserossijskoj nauchno-prakticheskoy zaochnoj konferencii. Neftekams, 2017. pp. 87 - 94.
 8. Garifullin R.I., Devyatkin E.M. Sbornik nauchnyh statej mezhdunarodnoj molodezhnoj shkoly-seminara «Lomonosovskie chteniya na Altae»: v 6 chastyah, Barnaul, 05–08 noyabrya 2013 goda. Barnaul, 2013. pp. 309-311.
 9. Amialchenia Andrei. Handwriting_Letter_Recognition_ML_NET. github.com URL: github.com/flash2048/Handwriting_Letter_Recognition_ML_NET (date assessed: 20.03.2023).
 10. Graphics. DrawString Method. URL: learn.microsoft.com/en-us/dotnet/api/system.drawing.graphics.drawstring.
-