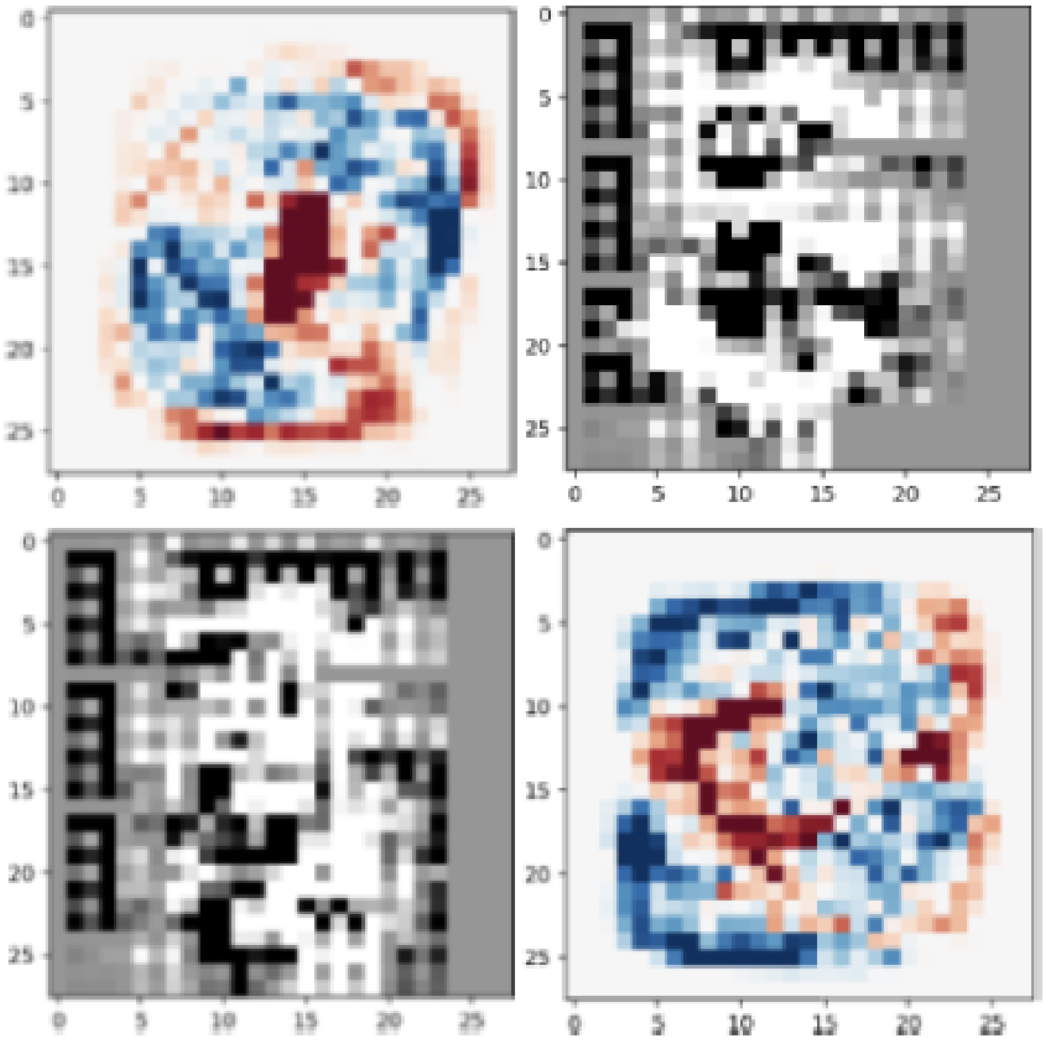
ИТМО

И. Ю. Попов, А. Я. Бучаев, Д. А. Есипов

ВАЛИДАЦИЯ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА



Санкт-Петербург
2023

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

УНИВЕРСИТЕТ ИТМО

И. Ю. Попов, А. Я. Бучаев, Д. А. Есипов

ВАЛИДАЦИЯ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

ЛАБОРАТОРНЫЙ ПРАКТИКУМ

РЕКОМЕНДОВАНО К ИСПОЛЬЗОВАНИЮ В УНИВЕРСИТЕТЕ ИТМО
по направлениям подготовки
10.04.01, 23.04.03, 11.04.03
в качестве учебно-методического пособия для реализации основных профессиональных образовательных программ высшего образования бакалавриата

ИТМО

Санкт-Петербург
2023

Попов И. Ю., Бучаев А. Я., Есипов Д. А., Валидация систем искусственного интеллекта – СПб: Университет ИТМО, 2023. – 29 с.

Рецензент(ы):

Изложены основы современных концепций валидации и верификации систем искусственного интеллекта (ИИ), таких как валидация обучающей выборки, анализ интерпретации моделей ИИ, верификация моделей нейронных сетей. Особое внимание уделено интерпретации моделей нейронных сетей, что позволит изучить поведение модели и принципы принятия решений классификации. Издание может быть использовано в качестве учебного пособия для реализации основных профессиональных образовательных программ высшего образования магистратуры по направлениям 10.04.01 – Информационная безопасность, 23.04.03 – Эксплуатация транспортно-технологических машин и комплексов, 11.04.03 – Конструирование и технология электронных средств.

ИТМО

Университет ИТМО – ведущий вуз России в области информационных и фотонных технологий, один из немногих российских вузов, получивших в 2009 году статус национального исследовательского университета. С 2013 года Университет ИТМО – участник программы повышения конкурентоспособности российских университетов среди ведущих мировых научно-образовательных центров, известной как проект «5 в 100». Цель Университета ИТМО – становление исследовательского университета мирового уровня, предпринимательского по типу, ориентированного на интернационализацию всех направлений деятельности.

Оглавление

Введение

Системы искусственного интеллекта (ИИ) имеют широкое применение: их повсеместно встречают в экономике, финансах, промышленности, логистике, медицине, кинематографе и даже изобразительном искусстве, психология и т.д. ИИ используются в системах автоматизированного управления, системах принятия решений, системах, связанных с обработкой естественного языка, распознаванием образов, речи, анализом данных, логистикой и других практически значимых задач обработки данных. Однако ошибки при решении таких задач могут привести к тяжким последствиям, связанным с угрозой для жизни и здоровья людей, серьезному экономическому и экологическому ущербу. Поэтому доверие к системам ИИ является важнейшим условием, определяющим возможность применения этих систем. Наличие практических навыков работы по валидации систем ИИ на сегодняшний день является большим преимуществом среди конкурентов, выгодно выделяющим для работодателей на рынке вакансий специалиста, работающего в области информационных технологий. В лабораторном практикуме объединены лабораторные работы по анализу, валидации и визуализации обучающей выборки, оценке модели нейронной сети и составлению ее паспорта. Большинство лабораторных работ данного практикума рассчитаны на создание обучающимися некоторого программного обеспечения на языке программирования Python.

Лабораторный практикум предназначен для студентов магистратуры, обучающихся по направлениям 10.04.01 – Информационная безопасность, 23.04.03 – Эксплуатация транспортно-технологических машин и комплексов, 11.04.03 – Конструирование и технология электронных средств.

Все задания могут быть выполнены в бесплатном облачном сервисе Google Colaboratory или любой другой среде разработки, поддерживающей Python. Выбор IDE не влияет на ход выполнения лабораторных работ.

Постановка каждой задачи лабораторного практикума предельно коротка, однако для понимания задачи необходимо знать материал соответствующего раздела курса «Валидация систем искусственного интеллекта», а для написания вспомогательных скриптов – уметь программировать на Python. При необходимости более глубокого изучения какой-либо темы, обучающиеся могут обратиться к литературе, представленной в практикуме после каждой лабораторной работы.

Программа лабораторного практикума

Практикум включает четыре лабораторные работы.

Лабораторная работа 1. Валидация обучающего датасета

Цель работы – изучение основных принципов валидации обучающего датасета; получение навыков анализа и валидации обучающей выборки; получение навыков визуализации полученных данных в различных графических форматах. Работа состоит из четырех частей, помимо составления краткой сводки полученной информации и оформления отчета по лабораторной работе. В первой части описывается представленный датасет, во второй части производится фильтрация его аномальных элементов, в третьей части оценивается репрезентативность данных, в четвертой части визуализируются полученные результаты.

Лабораторная работа 2. Валидация модели нейронной сети классификации изображений

Цель работы – изучение основных принципов валидации целевой модели; получение навыков анализа и описания модели. Работа состоит из четырех частей, помимо составления краткой сводки полученной информации и оформления отчета по лабораторной работе. В первой части описывается архитектура модели нейронной сети классификации изображений согласно варианту, во второй части производится верификация модели с помощью предоставленного датасета, в третьей части осуществляется проверка представленного программного кода, в четвертой части оценивается качество работы модели.

Лабораторная работа 3. Интерпретация модели нейронной сети классификации изображений

Цель работы – получение навыков исследования механизма принятия решений моделей классификации изображений и формирования представления об интерпретации обрабатываемых моделью данных. Работа состоит из трех частей, помимо составления краткой сводки полученной информации и оформления отчета по лабораторной работе. В первой части исследуются инструменты интерпретации и визуализации элементов нейронной сети, во второй части производится интерпретация весов пикселей модели в виде изображения, в третьей части исследуется интерпретация весов пикселей на предмет аномалий.

Лабораторная работа 4. Паспорт модели нейронной сети классификации изображений. Заключающая характеристика

Цель работы – получение навыков сводного анализа полученных результатов; написание заключающей характеристики и составление паспорта модели. Работа состоит из трех частей, которые объединяют результаты всех предыдущих работ, помимо оформления отчета по

лабораторной работе. В первой части составляется паспорт модели. Во второй части пишется заключающая характеристика. В третьей части проводится защита результатов.

Требования к оформлению отчета по лабораторным работам

Общие требования

1) Отчет выполняется в виде самодостаточного документа. Материал, изложенный в отчете, должен быть понятен без дополнительных комментариев со стороны исполнителей.
2) Отчет выполняется в виде текстового документа в соответствии с ГОСТ 2.105-95 и представляется в электронном виде.
3) В отчёте должен(-ы) быть полностью представлен(-ы) вспомогательный(-ые) скрипт(-ы) на Python, которые были написаны для выполнения задач лабораторной работы.
4) Листы отчета должны быть пронумерованы, кроме первого листа, который считается титульным.
5) Если отчет содержит большое количество листов, рекомендуется добавлять лист с содержанием отчета (разделы и номера листов).

Содержание отчета

Отчет должен содержать следующие разделы:

1) Титульный лист;
2) Тема, цель и задачи работы;
3) Содержательная часть;
4) Выводы.

Примеры выполнения лабораторных работ №1 и №2 представлены в Приложениях 1 и 2.

Тема, цель и задачи работы

В данном разделе должна быть сформулирована тема и цель работы. Их необходимо скопировать из данного лабораторного практикума. Также должны быть сформулированы задачи, решаемые в процессе выполнения лабораторной работы для достижения поставленной цели.

Содержательная часть

В данном разделе должны быть представлены и описаны шаги для достижения поставленной цели. Для этого содержательная часть условно делится на теоретическую и экспериментальную.

Теоретическая часть отчета по лабораторной работе должна включать в себя всю необходимую информацию о предметной области, к которой относятся описание работы программных компонентов, используемых при ее выполнении, и т.п. Данный раздел не является обязательным.

Экспериментальная часть лабораторной работы является обязательной и не может быть опущена. В ней отражаются персональные результаты обучающегося, полученные им при выполнении лабораторной работы. Этими результатами являются листинги написанного программного кода, таблицы и изображения, полученные в процессе решения сформулированных задач, а также характеристики и аналитические заключения. Исходные данные, использованные обучающимся для практической проверки реализованных в ходе практической работы программных компонентов, также должны быть представлены.

Данные по выполненной лабораторной работе можно предоставить отдельно, в том числе в виде ссылки на Интернет-ресурс. При изучении ранее выполненных работ обучающимся рекомендуется самостоятельно переписать всю программу в том виде, в каком она была написана, с целью лучшего понимания алгоритма программы.

Если обучающийся не завершил цикл выполнения программы, должно быть проведено ее окончательное редактирование. В экспериментальную часть отчета по лабораторной работе необходимо включить пояснения и комментарии к написанному программному коду (если написание программного кода является одной из задач лабораторной работы).

Выводы

Выводы должны быть кратко изложены в виде списка и отражать наиболее важные аспекты лабораторной работы. В конце отчета студент должен привести список использованной при подготовке отчета и процитированной литературы (в том числе ссылки на стандарты, руководства пользователя и прочую техническую документацию). Текст отчета по лабораторной работе может являться планом ответа при защите лабораторной работы.

Правила оформления текста отчета по лабораторной работе

Отчет должен быть представлен в формате PDF. Поля текста могут быть выбраны произвольно, например, 25 мм по всем сторонам печатного листа. Рекомендуется при оформлении отчета использовать шрифт чёрного цвета (например, Times New Roman) высотой 14 пунктов с полуторным интервалом между строками. Обязательным требованием при оформлении текста отчёта по лабораторной работе является его выравнивание по ширине страницы и использование автоматических переносов. На рисунках в отчете

могут быть представлены блок-схемы, графики, снимки экранов виртуального окружения, а также прочая графическая информация. Рисунки должны быть четкими и легкочитаемыми.

Листинги программного кода должны оформляться как скриншоты окна редактора, либо непосредственно в тексте отчета. В обоих случаях должны быть соблюдены следующие правила:

1) Должны использоваться моноширинные шрифты;
2) Должны быть пронумерованы строки программы (нумерация должна начинаться с единицы);
3) Должна быть включена «подсветка» синтаксических конструкций используемого языка программирования.

Исходные коды должны быть снабжены комментариями. На рисунки и листинги обязательно должны присутствовать ссылки в тексте, например: «На рисунке 1 изображен ...» или «Исходный код программы представлен в листинге 2 ...». Листинги, таблицы и рисунки должны быть снабжены подписями, отражающими их содержание.

Защита лабораторной работы

Во время защиты лабораторной работы обучающемуся следует ясно и чётко изложить тему, цель работы и основные решённые задачи, а также ответить на дополнительные вопросы, заданные преподавателем в процессе защиты отчёта. Критерии, по которым оценивается защита лабораторной работы, представлены в таблице 1.

Таблица 1 – Критерии оценивания защиты лабораторной работы

№ п/п	**Критерий**	**Оценка (уровень)**		
		Высокий	**Средний**	**Низкий**
1	Корректность оформления отчета	2	1,5	1
2	Качество устного представления результатов работы	2	1,5	1
3	Понимание (воспроизведение) выполненных команд	2	1,5	1
4	Правильность и полнота аналитического заключения достигнутого результата	2	1,5	1
5	Качество ответов на дополнительные вопросы	2	1,5	1
Итого баллов:		**10**	**7,5**	**5**

Лабораторная работа № 1

Валидация обучающего датасета

Цель работы:

Изучение основных принципов валидации обучающего датасета; получение навыков анализа и валидации обучающей выборки; получение навыков визуализации полученных данных в различных графических форматах.

Задачи:

1) Описать представленный датасет;
2) Произвести фильтрацию аномальных элементов датасета;
3) Оценить репрезентативность данных;
4) Визуализировать полученные результаты;
5) Составить краткую сводку полученной информации.

Ход работы:

Для выполнения данной работы необходимо скачать набор данных согласно варианту (https://disk.yandex.ru/d/YJiSKAt5c8AQvg) и ознакомиться с прилагаемым датасетом, а затем написать его характеристику.

Необходимо написать вспомогательные скрипты на языке программирования Python, которые решают перечисленные задачи.

Под фильтрацией подразумевается поиск и исключение аномальных элементов (например, дубликатов и тд.), приветствуются автоматизированные методы фильтрации.

Репрезентативность модели отражает полноту представления данных и влияет на качество классификации модели. Необходимо произвести качественную и количественную оценку репрезентативности.

Для визуализации полученных данных рекомендуется использование программных инструментов (например, библиотеку matplotlib).

Список литературы:

1) V. Varkarakis and P. Corcoran, "Dataset Cleaning — A Cross Validation Methodology for Large Facial Datasets using Face Recognition," 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), Athlone, Ireland, 2020, pp. 1-6, doi: 10.1109/QoMEX48832.2020.9123123.
2) Пылов, П. А. Фундаментальные типы кросс-валидации для оценки качества моделей машинного и глубокого обучения / П. А. Пылов, О. А. Ивина // Россия молодая : Сборник материалов XIII Всероссийской научно-практической конференции с международным участием, Кемерово, 20–23 апреля 2021 года / Редколлегия: К.С. Костиков (отв. ред.) [и др.]. – Кемерово: Кузбасский государственный технический университет имени Т.Ф. Горбачева, 2021. – С. 31520.1-31520.7. – EDN NYULSQ.
3) AltexSoft, Data Quality Management: Roles, Processes, Tools [Электронный ресурс] .- https://www.altexsoft.com/blog/data-quality-management-and-tools/ (дата обращения: 23.01.2023)
4) Хабр, О важности датасета и о том, как сделать его лучше. Опыт нашей компании [Электронный ресурс] .- https://habr.com/ru/post/678808/ (дата обращения: 23.01.2023)
5) GeekBrains, Датасет: виды, применение, набор лучших [Электронный ресурс] .- https://gb.ru/blog/dataset/ (дата обращения: 23.01.2023)
6) Loginom, Репрезентативность выборочных данных [Электронный ресурс] .- https://loginom.ru/blog/representativity (дата обращения: 23.01.2023)

Лабораторная работа № 2

Валидация модели нейронной сети классификации изображений

Цель работы

Изучение основных принципов валидации целевой модели; получение навыков анализа и описания модели.

Задачи:

1) Описать архитектуру модели нейронной сети классификации изображений согласно варианту;
2) Произвести верификацию модели с помощью предоставленного датасета;
3) Осуществить проверку представленного программного кода;
4) Оценить качество работы модели;
5) Составить краткую сводку полученной информации.

Ход работы:

Для выполнения данной работы необходимо скачать необходимые файлы согласно варианту (https://disk.yandex.ru/d/YJiSKAt5c8AQvg), выполнить описание модели согласно варианту. Для получения детальной характеристики модели можно анализировать прилагаемый исходный код или использовать вспомогательные методы библиотек (например, tensorflow позволяет получить структуру модели и графическое представление в виде блоков).

Проведение верификации подразумевает использование прилагаемого датасета.

Проверка программного кода предполагает анализ построения модели, формирования выборок и т.д.

Под оценкой модели подразумевается качественная оценка.

Список литературы:

1) Пылов, П. А. Фундаментальные типы кросс-валидации для оценки качества моделей машинного и глубокого обучения / П. А. Пылов, О. А. Ивина // Россия молодая : Сборник материалов XIII Всероссийской научно-практической конференции с международным участием, Кемерово, 20–23 апреля 2021 года / Редколлегия: К.С. Костиков (отв. ред.) [и др.]. – Кемерово: Кузбасский государственный технический университет имени Т.Ф. Горбачева, 2021. – С. 31520.1-31520.7. – EDN NYULSQ.
2) OSP, Валидация автономных систем [Электронный ресурс] .- https://www.osp.ru/os/2019/04/13055222 (дата обращения: 23.01.2023)
3) Хабр, Валидация моделей машинного обучения [Электронный ресурс] .- https://habr.com/ru/company/glowbyte/blog/569970/ (дата обращения: 23.01.2023)
4) Atlassian, Различные виды тестирования ПО [Электронный ресурс] .- https://www.atlassian.com/ru/continuous-delivery/software-testing/types-of-software-testing (дата обращения: 23.01.2023)

Лабораторная работа № 3

Интерпретация модели нейронной сети классификации изображений

Цель работы:

Получение навыков исследования механизма принятия решений моделей классификации изображений; формирования представления об интерпретации обрабатываемых моделью данных.

Задачи:

1) Исследовать инструменты интерпретации и визуализации элементов нейронной сети;
2) Произвести интерпретацию весов пикселей модели в виде изображения;
3) Исследовать интерпретацию весов пикселей на предмет аномалий;
4) Составить краткую сводку полученной информации.

Ход работы:

Для выполнения лабораторной работы возможно использование следующих инструментов:

1) tensorboard;
2) sklearn.

В ходе выполнения работы требуется скачать необходимые файлы согласно варианту (https://disk.yandex.ru/d/YJiSKAt5c8AQvg), проанализировать скрытые слои модели и произвести интерпретацию весов для получения информативных визуализаций, с помощью которых будет произведен анализ для определения фокусировки модели на отдельных элементах изображения, выявления аномалий и неявных признаков классификации. Вспомогательные скрипты, выполняющие перечисленные задачи, необходимо писать на языке Python.

Инструмент tensorboard предоставляет краткую информацию о модели на основе использованных логов (https://www.tensorflow.org/tensorboard/image_summaries?hl=ru).

При обучении модели с помощью библиотеки sklearn возможна визуализация отдельных представлений классов.

Для ручного анализа и визуализации интерпретации весов возможно использование метода закрашивания окна. Суть метода заключается в модификации анализируемого изображения путем закрашивания небольшой области, затем модифицированное изображение анализируется нейронной сетью и вычисляется степень принадлежности к определенному классу (пример изображен на рисунке 1). Данный процесс

является итеративным и должен покрывать все области изображения для получения информативной интерпретации. При закрашивании наиболее чувствительных участков изображения степень принадлежности падает, что говорит о мере влияния данного участка на принятие решения классификации.

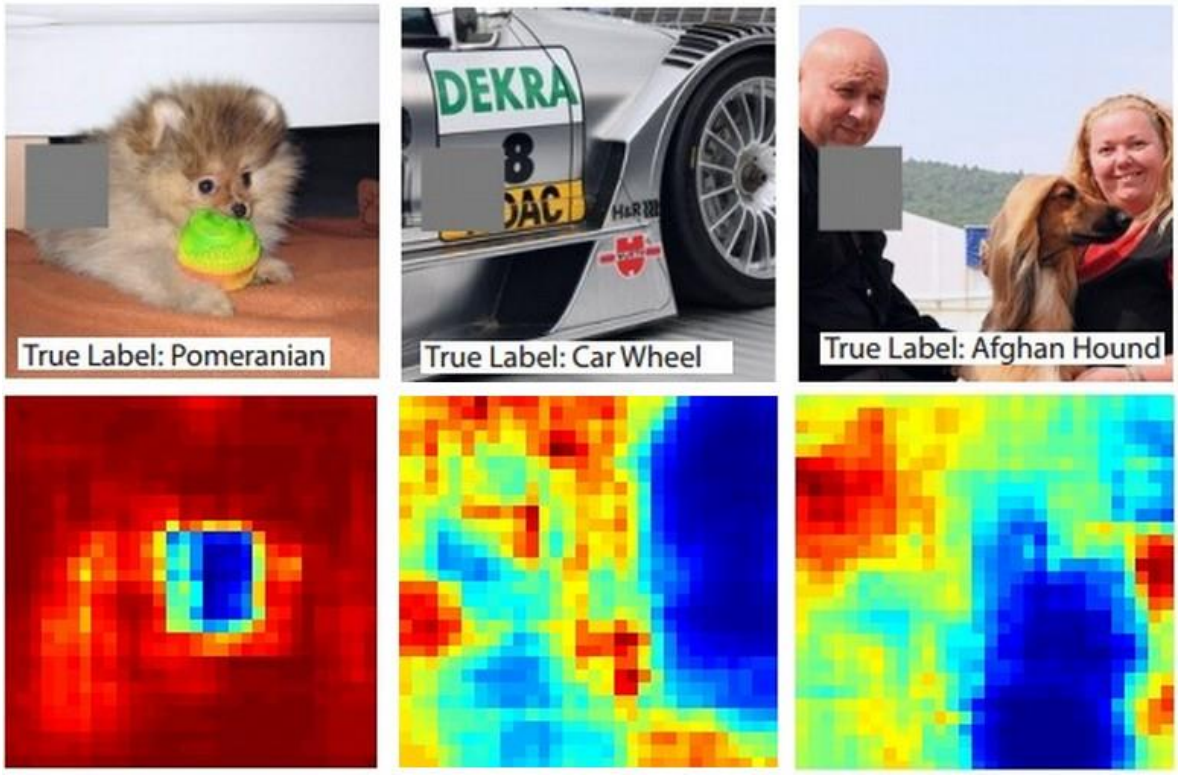


Рисунок 1 – Выявление чувствительных участков изображений методом закрашивания

Список литературы:

1) Zeiler, M.D., Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8689. Springer, Cham. https://doi.org/10.1007/978-3-319-10590-1_53
2) Owen Shen, "Interpretability in ML: A Broad Overview", The Gradient, 2020.
3) Хабр, «Сделать красиво». Визуализация обучения с Tensorboard от Google [Электронный ресурс] .- https://habr.com/ru/post/349338/ (дата обращения: 23.01.2023)
4) NewTechAudit, Визуализация архитектуры и отдельных блоков нейросети с помощью Netron [Электронный ресурс] .- https://newtechaudit.ru/netron/ (дата обращения: 23.01.2023)
5) Хабр, Визуализация процесса обучения нейронной сети средствами TensorFlowKit [Электронный ресурс] .- https://habr.com/ru/post/342934/ (дата обращения: 23.01.2023)
6) StevenRush, Визуализация того, чему обучается сверточная нейронная сеть [Электронный ресурс] .- https://stevenrush.github.io/understanding-cnn/ (дата обращения: 23.01.2023)

Лабораторная работа № 4

Паспорт модели нейронной сети классификации изображений. Заключающая характеристика

Цель работы:

Получение навыков сводного анализа полученных результатов; написание заключающей характеристики и составление паспорта модели.

Задачи:

1) Составить паспорт модели;
2) Написать заключающую характеристику;
3) Провести защиту результатов.

Ход работы:

Для выполнения лабораторной работы необходимо подготовить паспорт модели и итоговое заключение. Данные документы имеют свободный формат, требуется информативное заключение, краткая сводка по всем полученным результатам. Подразумевается использование предыдущих работ для формирования паспорта и заключения. В заключении необходимо представить результат валидации модели.

Процесс защиты представляет собой оценку полученных результатов и анализ написанных скриптов.

Приложение 1

Пример выполнения Лабораторной работы №1

Отчет по лабораторной работе №1

По дисциплине: «Валидация систем искусственного интеллекта»

На тему: «Валидация обучающего датасета»

Выполнил(-а):

Магистрант гр. N42101с Иванов И. И.

_____

Проверил(-а):

Доцент ФБИТ, к.т.н. Попов И. Ю.

_____

Санкт-Петербург
2023

Тема работы: валидация обучающего датасета.

Цель работы: изучение основных принципов валидации обучающего датасета; получение навыков анализа и валидации обучающей выборки; получение навыков визуализации полученных данных в различных графических форматах.

Задачи:

1) Описать представленный датасет;
2) Произвести фильтрацию аномальных элементов датасета;
3) Оценить репрезентативность данных;
4) Визуализировать полученные результаты;
5) Составить краткую сводку полученной информации.

Описание датасета

Представленный датасет является набором изображений. Данный набор содержит в себе совокупности пяти различных классов фруктов: яблоки, бананы, груши, манго и апельсины.

Суммарно в датасете 5000 изображений, каждый класс представляется 1000 экземплярами соответствующего фрукта.

Каждый экземпляр представляет собой изображение размером 30x30 пикселей в цветовом пространстве RGB.

Фильтрация данных датасета

Для обработки датасета был написан программный код на языке Python (представлен на рисунке 2), который реализует фильтрацию данных, а именно удаление аномальных экземпляров:

1) Дубликаты;
2) Изображения нестандартного размера;
3) Изображения иного формата;
4) Изображения в серых тонах. Такая характеристика была выбрана, так как считается, что цветовая компонента важна для классификации фруктов, следовательно, изображения в оттенках серого можно считать выбросами в выборке.

```python
from PIL import Image, ImageDraw
from utils import check_size, check_color, check_format, load_dataset



def validate_image(file_name, width, height):
    size = check_size(file_name)
    image_format = check_format(width, height)
    color = check_color(file_name)

    return size and image_format and color

def check_dataset(dataset):
    for i in range(len(dataset)):
        if not(validate_image(el)):
            del dataset[i]

def main():
    data = load_dataset('fruits')
    check_dataset(data)

if __name__ == '__main__':
    main()
```

Рисунок 2 – Фильтрация датасета

Оценка репрезентативности данных

В рамках выполнения лабораторной работы необходимо произвести качественную и количественную оценку репрезентативности данных.

При качественной оценки важным фактором является наличие в выборке всех представленных классов с равной вероятностью выбора того или иного класса. С помощью инструментов визуализации на языке Python была получена круговая диаграмма, показывающая относительное содержание экземпляров исследуемых классов (представлена на рисунке 3).

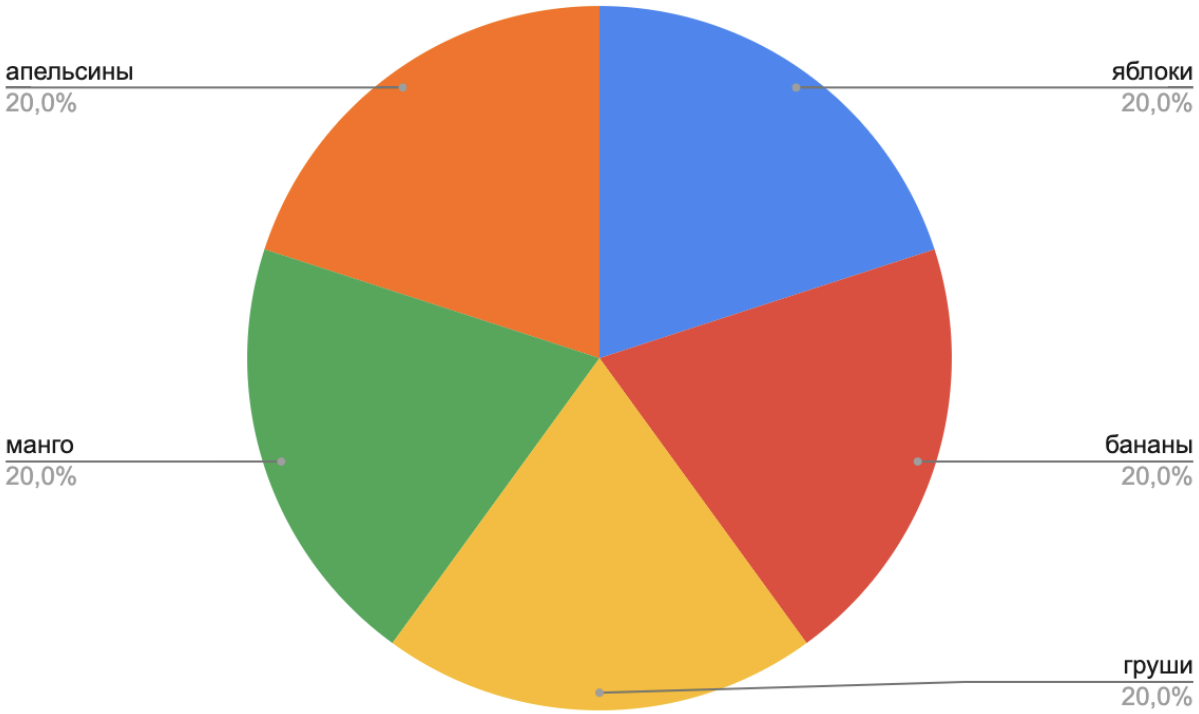


Рисунок 3 – Оценка качественной репрезентативности классов выборки

Количественная оценка вычисляется с помощью формулы 1:

$$n = \frac{t^2 * p * q}{\Delta^2}, \quad (1)$$

где t – доверительный коэффициент, показывающий, какова вероятность того, что размеры показателя не будут выходить за границы предельной ошибки;

p – доля единиц наблюдения, обладающих изучаемым признаком;

$q = 1 - p$ — доля единиц наблюдения, не обладающих изучаемым признаков;

Δ — допустимая ошибка выборки.

Выводы

В ходе выполнения работы был описан и охарактеризован датасет согласно варианту. Данная совокупность является репрезентативной, также была проведена фильтрация данных в соответствии с предполагаемыми задачами и целями при использовании данного датасета.

Приложение 2

Пример выполнения Лабораторной работы №2

Отчет по лабораторной работе №2

По дисциплине: «Валидация систем искусственного интеллекта»

На тему: «Валидация модели нейронной сети классификации изображений»

Выполнил(-а):

Магистрант гр. N42101с Иванов И. И.

_____

Проверил(-а):

Доцент ФБИТ, к.т.н. Попов И. Ю.

_____

Санкт-Петербург
2023

Тема работы: валидация модели нейронной сети классификации изображений.

Цель работы: изучение основных принципов валидации целевой модели; получение навыков анализа и описания модели.

Задачи:

1) Описать архитектуру модели нейронной сети классификации изображений согласно варианту;
2) Произвести верификацию модели с помощью предоставленного датасета;
3) Осуществить проверку представленного программного кода;
4) Оценить качество работы модели;
5) Составить краткую сводку полученной информации.

Описание архитектуры согласно варианту

Данная модель предназначена для классификации изображений фруктов. На вход модели подаются изображения, выходом является предсказанный класс входного изображения – принадлежность к какому-либо виду фруктов. Представленная модель нейронной сети классификации изображений имеет архитектуру, показанную на рисунке 4. Данная модель имеет несколько слоев:

1) Входной слой Rescaling – используется для изменения размера входных изображений;
2) Слой Conv2D – блок свертки с 16 фильтрами и ядром свертки размером 3;
3) Слой maxpooling2d – слой производит более жесткую светку изображения сразу уменьшая размерность каждого слоя, например, в 2 раза (представлено на рисунке 5);
4) Слой Conv2D – блок свертки с 32 фильтрами и ядром свертки размером 3;
5) Слой maxpooling2d;
6) Слой Conv2D – блок свертки с 64 фильтрами и ядром свертки размером 3;
7) Слой maxpooling2d;
8) Слой Flatten – переводит полученный тензор к вектору;
9) Слой Flatten;
10) Слой Dense – полносвязный слой размерностью 128;
11) Выходной слой Dense – полносвязный слой размерностью, которая равна количеству классов.

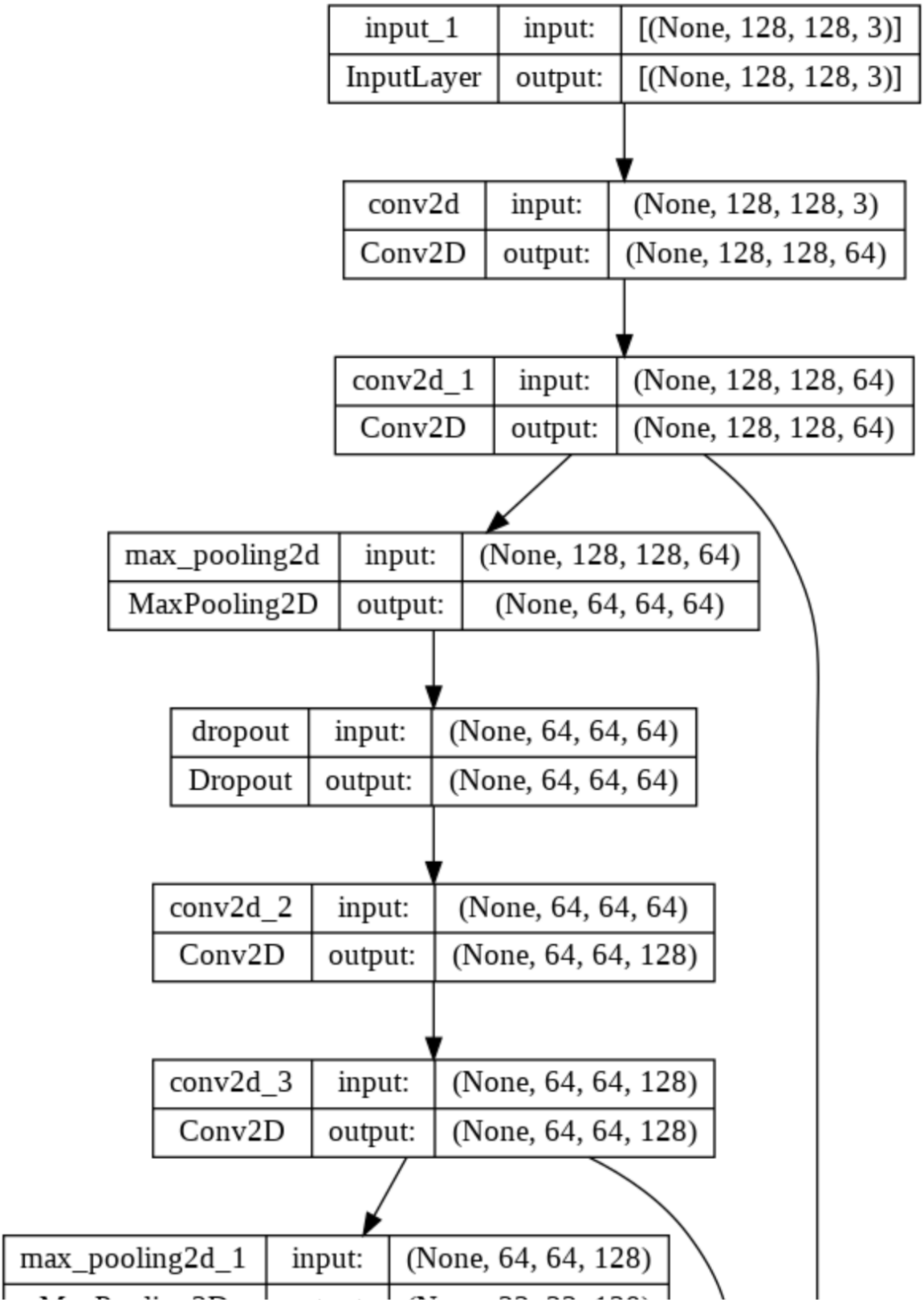


Рисунок 4 – Визуализация архитектуры модели

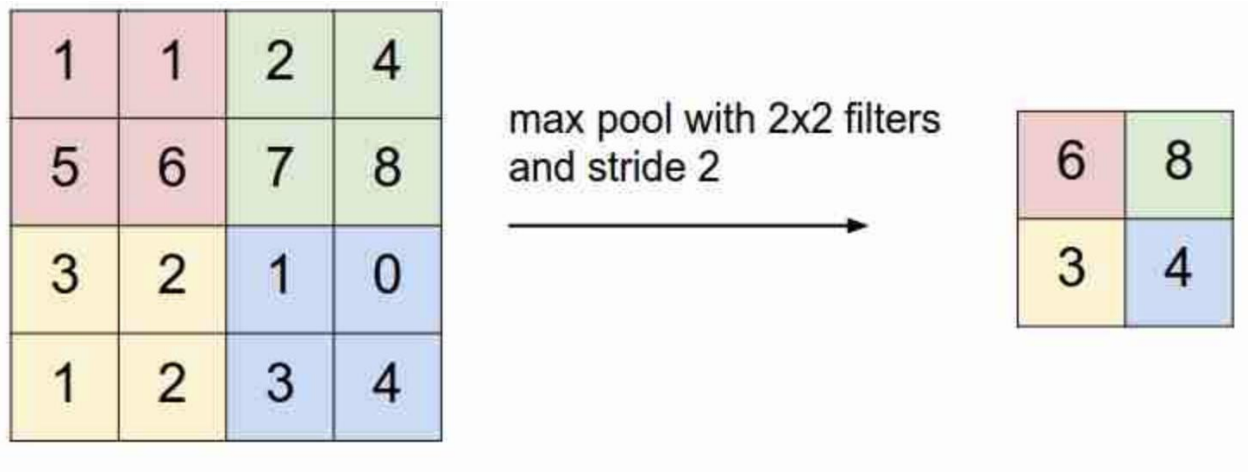


Рисунок 5 – Преобразование MaxPooling

Верификация модели нейронной сети

Для произведения верификации модели нужно загрузить несколько изображений, принадлежащих к различным классам, для проверки работоспособности модели (представлено на рисунке 6).

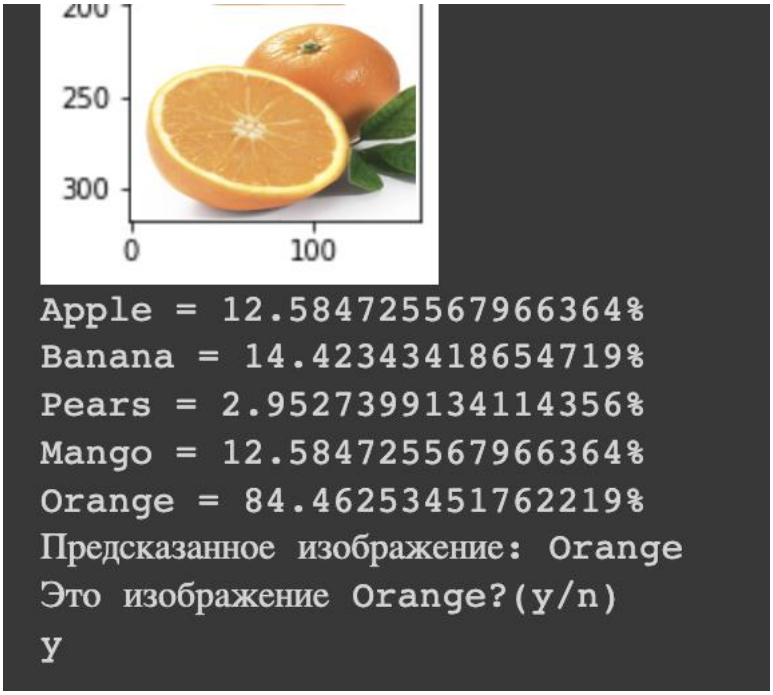


Рисунок 6 – Верификация модели

После тестирования модели на нескольких различных образцах, можно сделать вывод, что модель действительно проводит классификацию фруктов, изображенных на картинках.

Проверка программного кода

Исследуемый программный код рассматриваемой модели имеет несколько недостатков:

1) Наличие дублированного слоя Flatten. Так как слой Flatten переводит тензор в вектор, то дублирование этого слоя в данном случае не нужно;
2) Неверные пропорции разделения обучающей и валидационной выборки.

Оценка качества обучения и работы модели

Была произведена модификация представленного программного кода с помощью библиотеки для визуализации matplotlib, которая предоставляет инструменты визуализации. График обучения модели представлен на рисунке 7. Как видно, ошибка в процессе проверки модели на валидационной выборке растет, это связано с выше перечисленными недостатками.

Рисунок 7 – График обучения и валидации модели

После исправления недостатков программного кода были получены результаты, представленные на рисунке 8.

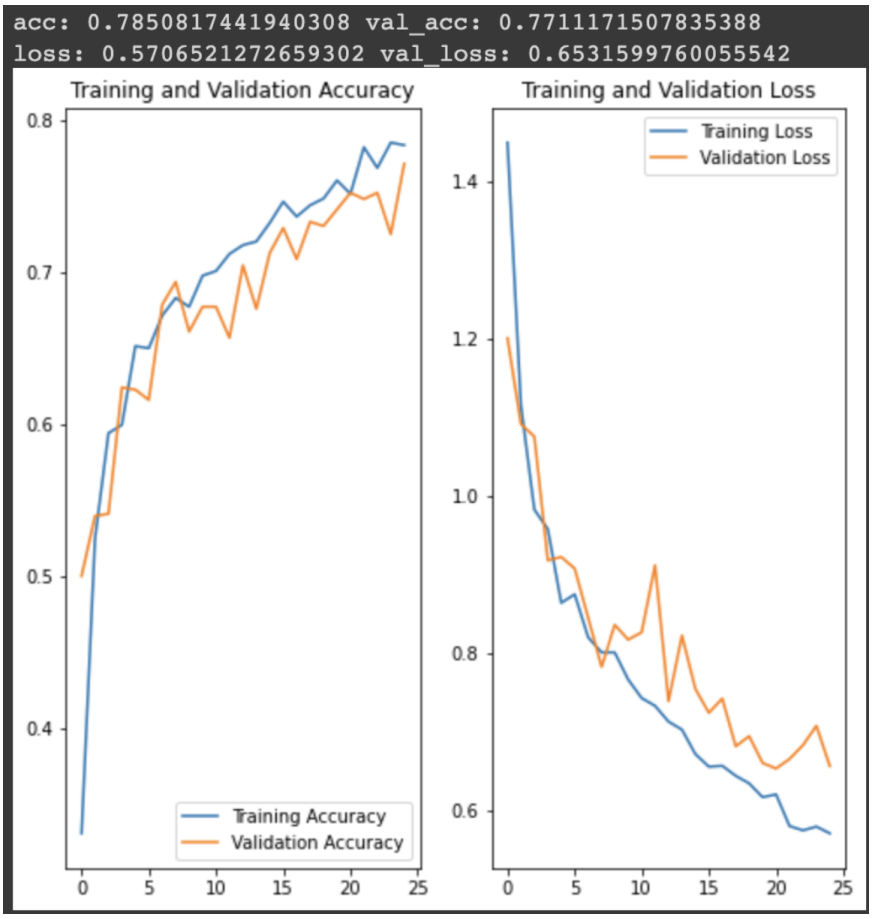


Рисунок 8 – График обучения и валидации после корректировки

Выводы

В ходе выполнения работы была описана модель нейронной сети классификации изображений, ее архитектура, были проанализированы используемые слои. Также был произведен анализ программного кода обучения модели, исправлены отмеченные недостатки и проведена верификация модели.

Попов Илья Юрьевич
Бучаев Абдулхамид Яхьяевич
Есипов Дмитрий Андреевич

Валидация систем искусственного интеллекта

Лабораторный практикум

В авторской редакции
Редакционно-издательский отдел Университета ИТМО
Зав. РИО Н.Ф. Гусарова
Подписано к печати
Заказ №
Тираж
Отпечатано на ризографе

Редакционно-издательский отдел
Университета ИТМО
197101, Санкт-Петербург, Кронверкский пр., 49