«Национальный исследовательский университет

«Высшая школа экономики»

Групповой проект

**«Классификация изображений,**

**Датасет на основе скриншотов/моделей видеоигры»**

Выполнили:

*Баранова Дарья*

*Павлов Сергей*

*Петренко Денис*

*Озеров Максим*

Содержание

**Введение1**

Цели и задачи работы2

**Сбор данных1**

Введите название главы (уровень 2)2

Введите название главы (уровень 3)3

**Построение baseline модели4**

Введите название главы (уровень 2)5

Введите название главы (уровень 3)6

**Улучшение модели4**

Введите название главы (уровень 2)5

Введите название главы (уровень 3)6

**Заключение4**

**Литература4**

**Приложение4**

1. **Введение**
   1. **Цели и задачи работы**

Целью работы ставится:

* изучение подходов к обработке изображений и выделению из них ключевых признаков;
* изучение способов классификации изображений с помощью классических подходов машинного обучения;
* постановка и решение собственной задачи классификации изображений методами классического машинного обучения (без использования нейронных сетей).

В качестве собственной задачи была выбрана задача классификации изображений со скриншотами моделей оборудования из игры «[Космические рейнджеры 2: Доминаторы](https://ru.wikipedia.org/wiki/%D0%9A%D0%BE%D1%81%D0%BC%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B5_%D1%80%D0%B5%D0%B9%D0%BD%D0%B4%D0%B6%D0%B5%D1%80%D1%8B_2:_%D0%94%D0%BE%D0%BC%D0%B8%D0%BD%D0%B0%D1%82%D0%BE%D1%80%D1%8B)».

Для достижения цели были выделены следующие задачи работы:

1. Собрать и обработать (при необходимости) датасет необходимого размера;
2. Выбрать метрики качества, с помощью которых будет оцениваться качество будущей модели;
3. Разработать baseline метод для классификации изображений, а также спроектировать и обучить модель;
4. Сформулировать и обосновать возможные способы улучшить полученное качество классификации и реализовать их. Оценить, насколько улучшилось качество классификации;
5. Подвести итоги и описать наилучшую полученную модель для решения поставленной задачи.

1. **Сбор данных**

В качестве изображений для классификации были выбраны статические рейнджеры оборудования в космическом корабле игрока из игры «Космические рейнджеры 2: Доминаторы».

Сбор данных был произведён вручную при помощи созданной фанатами утилиты [https://rangers.fandom.com/ru/wiki/ResEditor], предназначенной для извлечения графических элементов из дистрибутива игры. Такая возможность позволила не пользоваться внутриигровыми скриншотами, которые содержали бы в себе лишние элементы фона и, возможно, были бы хуже исходников по качеству.

В исходниках игры каждое изображение оборудования представляет собой циклический анимированный ряд из ~100 кадров. Поскольку для задачи необходимы статические изображения, для каждой модели оборудования был вручную выбран наиболее репрезентативный (на глаз) кадр.

В итоге каждое исходное изображение обладает следующими характеристиками:

* формат: PNG
* размер: 75 \* 75 pixel
* фон: прозрачный

Исходная выборка содержит 9 классов оборудования со следующим распределением:

****

Каждый класс представляет собой множество элементов, внешне сильно различающихся между собой. Например, класс «Оружие» состоит из 15 различных видов орудийный инструментов различной формы, размера и цвета. Изображения элементов классов представлены в приложении.

Далее исходная выборка была аугментирована разными искажениями для получения 700 изображений со следующим распределением:

****

Аугментация проводилась с помощью пайплайна из случайных изменений:

* Поворот (*rotate*) от -8 до +8 градусов с вероятностью 0.5;
* Изменение масштаба (*zoom*) с коэффициентом от 0.8 до 1.01 с вероятностью 0.5;
* Отражение по горизонтали (*flip\_left\_right*) с вероятностью 0.5;
* Случайное искажение (*random\_distortion*) с вероятностью 0.7 (пример искажения представлен на рисунке ниже).

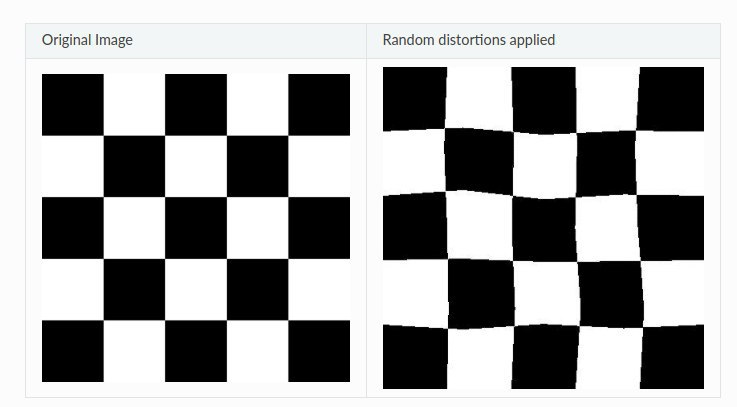


Рис. Пример случайного искажения (слева – оригинальное изображение, справа - искаженное)

Количество сгенерированных изображений выбиралось исходя из двух предпосылок:

1. Чем больше обучающая выборка будет, тем лучшего качества можно будет достичь при обучении модели;
2. Слишком большое количество данных может повлечь слишком долгое обучение (подробнее это описано в разделе построения baseline модели).

**3. Построение baseline модели**

**3.1 Предпосылки**

Изображения разных классов не сильно различаются между собой. Можно сказать, что даже для человеческого глаза изображения разных классов трудно различимы, так как все модели игры имеют довольно разнообразный внешний вид.

Исходя из этого, было принято решение строить baseline на основе техники Bag-of-words для computer vision [2].

**3.2 Описание метода**

Для извлечения признаков из изображений использовалось масштабно-инвариантное преобразование функций SIFT (Scale Invariant Feature Transform). <https://medium.com/data-breach/introduction-to-sift-scale-invariant-feature-transform-65d7f3a72d40>

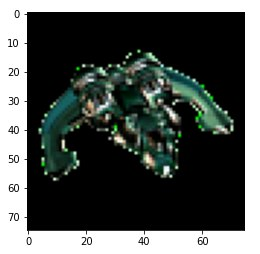
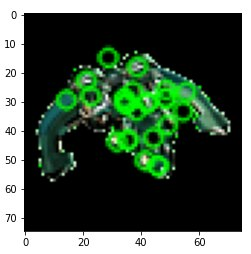
 

Рис.

После этого на основе выделенных дескрипторов каждого изображения строится соответствующее векторное пространство. Из множества дескрипторов для изображений получается набор векторов в этом пространстве с соответствующими координатами. Однако так как векторов много и нужно как-то выделить «ключевые точки», которые и будут представлять собой слова в словаре, нужно все это пространство кластеризовать. Проводится эта процедра с помощью K-Means алгоритма, который данное векторное пространство разбивает на K кластеров – опорных точек пространства. Эти опорные точки представляют собой «словарь» в представлении «Bag-of-words».

Следующим шагом исходные изображения описывались с помощью полученных опорных точек. Для каждого изображения определяется какие опорные точки лучше всего его описывают (какие кластеры находятся ближе всего к каждому из векторов его векторного представления). Таким образом, каждое изображение мы представляем как набор «слов из словаря».

После этого для каждого такого набора слов можно построить гистограмму. Таким образом, для каждого изображения мы получим по одной гистограмме. Именно их мы и будет классифицировать с помощью метода опорных векторов.

Описанный метод классификации накладывает значительное ограничение на объем обучающей выборки. K-Means алгоритм в процессе работы вычисляет матрицу расстояний, из-за этого время обучения на больших объемах данных может сильно возрастать.

**3.3 Полученное качество**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Размер обучающей выборки | Размер тестовой выборки | Количество кластеров |
| Качество на train |  |  |  |
| Качество на test |  |  |  |

**4. Улучшение модели**

**Заключение**

**Литература**

1. Ссылка на утилиту https://rangers.fandom.com/ru/wiki/ResEditor
2. <https://en.wikipedia.org/wiki/Bag-of-words_model_in_computer_vision>
3. <https://customers.pyimagesearch.com/the-bag-of-visual-words-model/>
4. <https://machinelearningknowledge.ai/image-classification-using-bag-of-visual-words-model/>

**Приложение**

Изображения классифицируемых моделей класса «CARGOHOOK»



Изображения классифицируемых моделей класса «DEFGENERATOR»



Изображения классифицируемых моделей класса «ENGINE»



Изображения классифицируемых моделей класса «FUELTANKS»



Изображения классифицируемых моделей класса «HULL»



Изображения классифицируемых моделей класса «RADAR»



Изображения классифицируемых моделей класса «REPAIRROBOT»



Изображения классифицируемых моделей класса «SCANER»



Изображения классифицируемых моделей класса «WEAPON»

