

МИНИСТЕРСТВО ВНУТРЕННИХ ДЕЛ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ КАЗЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ МОСКОВСКИЙ УНИВЕРСИТЕТ МВД РОССИИ ИМЕНИ В.Я.КИКОТЯ

**КОМАНДА ВАСИЛИСК**

**ВЫБОР АЛГОРИТМА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

**для задачи интеллектуального поиска предметов антиквариата   
на открытых торговых площадках**

Москва 2021

Исходя из постановки задачи необходимо решить вопрос схожести двух изображений. Так как входные данные приложения являются изображениями, а они могут быть представлены на вход нейронной сети, как многомерный массив чисел в узлах равномерной сетки с произвольным числом осей – тензор, решение подобной задачи невозможно без выделения вторичных параметров изображений и обработки данных с сеточной топологией. Под вторичными параметрами понимается результат некоторого преобразования изображения для выделения наиболее значимых характеристик объектов, присутствующих на нем [1]. Вторичные параметры играют определяющую роль в задачах, связанных с определением соответствия изображений некоторых предметов, так как первичные признаки, такие как цвет пикселя, не могут являться определяющим фактором для решения задачи идентификации графического объекта.

Анализ литературы показал, что в настоящий момент наиболее эффективно вторичные параметры выделяют сверточные нейронные сети (СНС) – особый вид искусственных нейронных сетей, созданных для обработки данных, представленных сеточной топологией [2, 3]. Состояние нейрона сети без свертки определяется следующей формулой:

, 

где n – число входов нейрона;  – значение i-го входа нейрона;  – вес i-го синапса. А аксона по формуле:

, 

где f – некоторая функция, которая называется активационной, на входе, для решения данной задачи была выбрана функция гиперболического тангенса, которая имеет следующий вид:

. 

На выходе – функция softmax, которая имеет следующий вид:

, 

где большая *K* указывает на количество классов.

СНС представляют собой нейронные сети, в которых вместо общей операции умножения на матрицу, в одном или более слоев используется операция математической свертки. Они, в свою очередь, являются одним из алгоритмов искусственного интеллекта. Отклик первого слоя СНС (как и последующих сверточных слоев) для одного канала входного изображения описывается следующей математической моделью:

, 

где m и n – высота и ширина матрицы соответственно; F – значение ядра свертки; b – смещение: H и W – высота и ширина выхода соответственно описываются следующим образом:

, 

, 

где и  – ширина и высота входа соответственно;  и  – размеры фильтра;  и  – смещение;  и  – заполнение нулями (padding).

В целях решения задачи схожести двух изображений СНС эффективнее традиционной нейронной сети за счет того, что входное изображение может содержать большое количество пикселей (тысячи или миллионы), но при этом небольшие значимые вторичные признаки, обнаружить которые возможно, используя ядро, охватывающее десятки пикселей. Схема сверточного слоя (СL) представлена на рисунке 1.

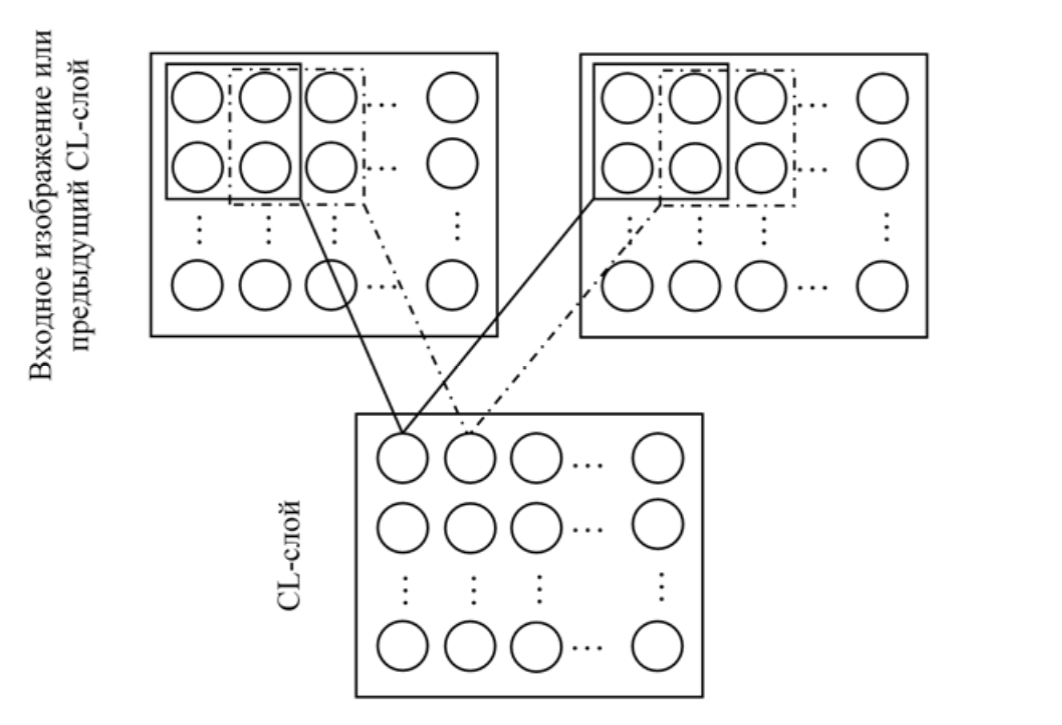


Рисунок 1 – Схема CL

Нейроны СL имеют структуру, представленную на рисунке 2.

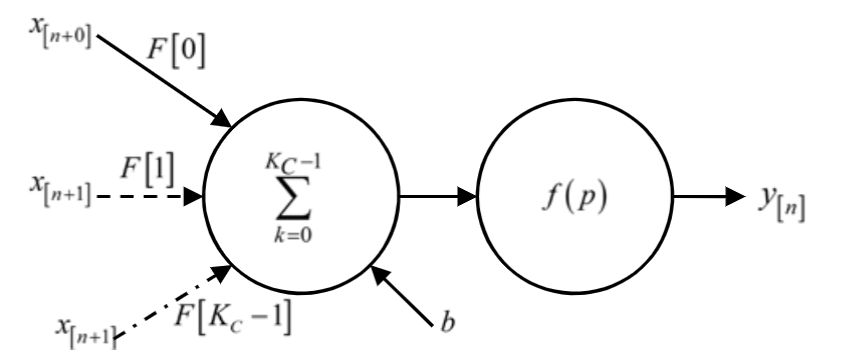


Рисунок 2 – Структура нейрона (CL)

На рисунке 2  – общее количество нейронов или элементов входного вектора, входящих в рецептивное поле n-го нейрона СL;  – настраиваемые веса нейрона; b – смещение n-го нейрона, причём b и – одни и те же для всей карты СL;  – входные данные для n-го нейрона СL .

С другой стороны, возможно использовать классические подходы, такие как метод Виолы-Джонса, основанный на суммировании пикселей из прямоугольных регионов изображения [4], однако несмотря на то, что метод чувствителен к вертикальным и горизонтальным линиям на изображениях, результат поиска более груб и не обеспечивает приемлемую скорость и эффективность работы (рис. 3).



Рисунок 3 – Эффективность обнаружения границ

Правое изображение получено вычитанием из каждого пикселя исходного изображения значения пикселя слева от него. Результатом получается сила всех вертикальных границ во входном изображении, что необходимо для обнаружения объектов. Высота обоих изображений 280 пикселей. Ширина входного изображения 320 пикселей, а выходного – 319. Это преобразование также описывается сверткой с ядром, содержащим два элемента, оно требует 319 × 280 × 3 = 267 960 операций с плавающей точкой (два умножения и одно сложение на каждый выходной пиксель). Если то же самое преобразование выполнять путем перемножения матриц, то потребуется 320 × 280 × 319 × 280, т. е. больше восьми миллиардов элементов матрицы, так что с точки зрения потребления памяти свертка эффективнее такого преобразования в четыре миллиарда раз. Даже при прямом перемножении матриц необходимо выполнить свыше 16 миллиардов операций с плавающей точкой, в этом случае свертка примерно в 60 000 раз эффективнее. Необходимо отметить, что большинство элементов матрицы было бы равно нулю. В матрице без применения свертки было бы 2 × 319 × 280 = 178 640 элементов. Свертка – чрезвычайно эффективный способ описания преобразований, в которых одно и то же линейное преобразование многократно применяется к небольшим участкам изображения.

Таким образом, решение выбрать в качестве алгоритма выделения вторичных параметров аппарат СНС является наиболее целесообразным в рамках решаемой задачи, так как необходимо хранить меньшее количество параметров, что снижает требования модели к объему памяти и повышает ее статистическую эффективность. Также, для вычисления отклика требуется меньше операций.

При решении вопроса выбора конкретной архитектуры параметризатора, были проанализированы различные современные решения [5–7]. При наложении ограничений на эффективность (максимальная точность при минимальном количестве параметров) была выбрана архитектура остаточной 50-слойной СНС для классификации изображений ResNet-50 [6], отвечающей высокой точностью при небольшом количестве вычислений, что показано на графике (рис.4), и содержащей 60 миллионов весовых коэффициентов – параметров, присвоенных каждому входу нейрона на которые перемножаются входные данные, поданные на нейрон.

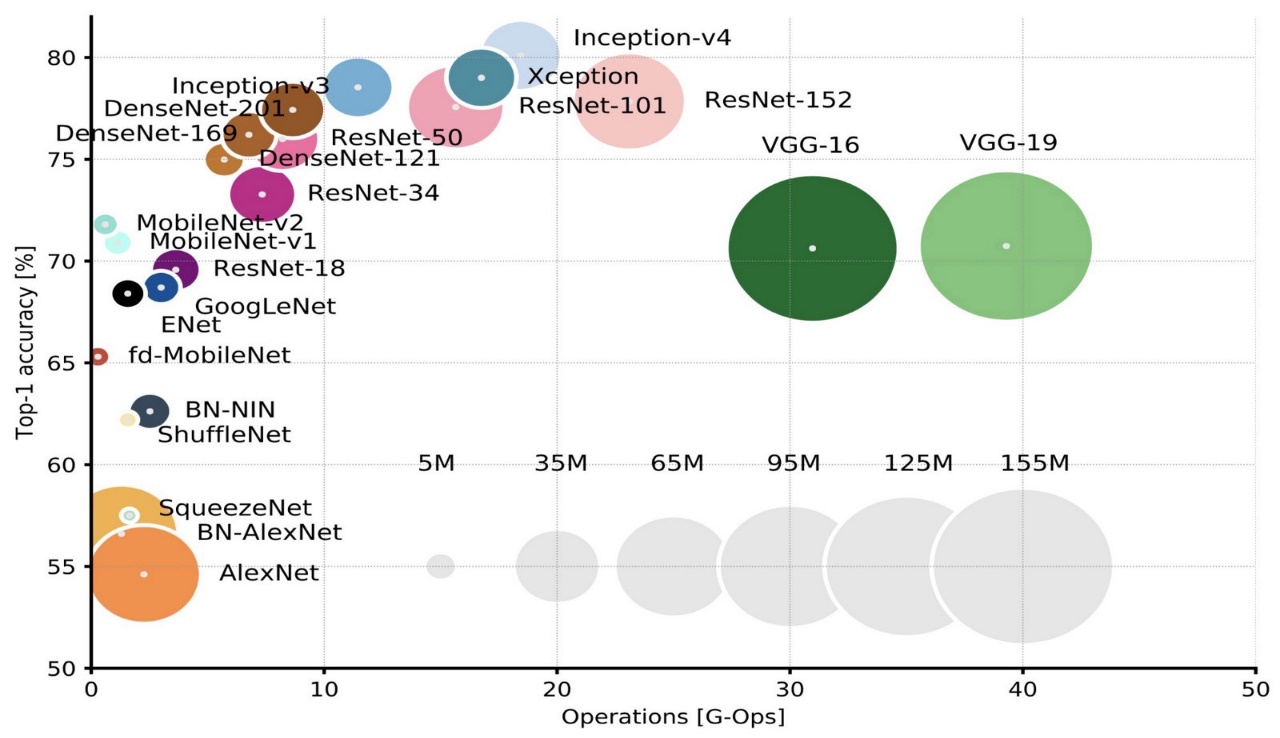


Рисунок 4 – График сравнения эффективности СНС

В виду того, что сети подобного размера обучить за время проведения хакатона не представляется возможным было принято решение использовать подход fine-tune [2]. Он позволил, корректируя слои нейронной сети, которые имеют наиболее абстрактные представления и производя дообучение только нескольких слоев, уменьшить шанс переобучения и сделать текущую модель ResNet-50 ещё более подходящей к решению задачи определения схожести двух изображений. При использовании подхода fine-tune, загружались весовые коэффициенты, используемые в соревновании Imagenet [8], после чего отсекалась выходная часть СНС (содержащая полносвязные [2] слои). На выход СНС добавлялся слой average pooling [6], таким образом при выделении вторичных параметров изображение (рис. 5) преобразовывалось в вектор размерностью .



Рисунок 5 – Исходное изображение

Если отложить по оси абсцисс вторичные параметры из вектора, а по оси ординат веса, то получится визуализированное представление вторичных параметров изображения (рис. 6). Тем самым задача решения вопроса схожести двух изображений была сведена к задаче определения близости двух векторов в *n*-мерном пространстве.

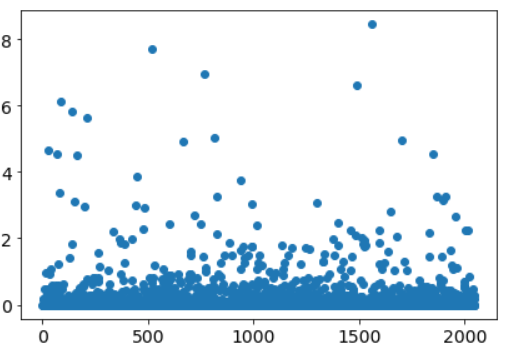


Рисунок 6 – График визуализации вторичных параметров изображения

Решение такой задачи при классическом подходе решается нахождением евклидова расстояния, вычисляемого по теореме Пифагора, или косинусного расстояния через скалярное произведение векторов и косинуса угла между ними [3], исходя из анализа литературы, данный подход также обладает достаточно низкой эффективностью по сравнению с искусственными нейронными сетями. В качестве инструмента сравнения двух векторов была выбрана искусственная нейронная сеть – многослойный перцептрон, решающая задачи классификации, в которой каждый нейрон связан со всеми нейронами предыдущего слоя.

Задача разметки данных решена в реализуемом парсере. Аннотированные изображения необходимы в качестве входных данных для обучения нейронной сети. Так как парсинг происходит по отдельным категориям антиквариата, то название этой категории и конкретного изображения и будут являться разметкой.

Также, в качестве основных подходов обучения и настройки параметров следует отметить использование аугментации [9], которая хорошо показывает себя в задачах классификации изображений. Аугментация – это увеличение выборки данных для обучения через модификацию существующих данных. Таким образом, выборка может увеличиться в десятки раз, а процесс обучения становится более эффективным, так как один и тот же объект на разных изображениях может быть представлен по-разному, что значительно повышает точность интеллектуального поиска схожих изображений.

Задача аугментации решена использованием библиотеки imgaug [10], включающей в себя такие типы функций, как операции со светом, геометрические операции и операции с объектами. В данном случае аугментация производилась посредством увеличения изображений, вырезания части, изменения перспективы и ротации (поворота) изображения (рис. 7).



Рисунок 7 – Пример аугментации изображений

Исходя из проведенного анализа, в качестве алгоритма выделения вторичных параметров был выбран аппарат СНС с архитектурой остаточной 50-слойной нейросети для классификации изображений ResNet-50. Также было принято решение использовать подход fine-tune, загружая весовые коэффициенты, используемые в соревновании Imagenet, отсекая выходную часть СНС и добавляя на выход слой average pooling, преобразовывая изображения в вектор. Задачу определения схожести двух изображений по их векторам было выбрано решить, используя многослойный перцептрон. Для решения задачи аугментации, выбрана библиотека imgaug.

Литература

1. Гонсалес Р., Вудс Р. Цифровая обработка изображений. Москва: Техносфера, 2012. 1104 с.
2. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение. Москва: «ДМК Пресс», 2017. 652 с.
3. Вьюгин В.В. Математические основы теории машинного обучения и прогнозирования. Москва. 2013. 387 с.
4. Viola P., Jones M. Rapid Object Detection using a Boosted Cascade of Simple Features // proccedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2001. pp. 511-518.
5. Chollet F. [1610.02357] Xception: Deep Learning with Depthwise Separable Convolutions // arXiv.org e-Print archive. 2016. URL: http://arxiv.org/abs/1610.02357 (дата обращения: 16.04.2021).
6. Larsson G., Maire M., Shakhnarovich G. [1605.07648] FractalNet: Ultra-Deep Neural Networks without Residuals // arXiv.org e-Print archive. 2017. URL: https://arxiv.org/abs/1605.07648 (дата обращения: 17.04.2021).
7. Szegedy C., Ioffe S., Vanhoucke V., Alemi A.A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning // International Conference on Learning Representations (ICLR) Workshop, 2016. С. 375-387.
8. ImageNet Large Scale Visual Recognition Competition (ILSVRC) [Электронный ресурс] // ImageNet: [сайт]. URL: http://www.image-net.org/challenges/LSVRC/ (дата обращения: 17.04.2021).
9. Wong S., Gatt A., Stamatescu V., McDonnell M. [1609.08764] Understanding data augmentation for classification: when to warp? // arXiv.org e-Print archive. 2016. URL: https://arxiv.org/abs/1609.08764 (дата обращения: 17.04.2021).
10. Jung A.B., Wada K., Crall J., Tanaka S., Graving J., Reinders C., Yadav S., Banerjee J., Vecsei G., Kraft A. и др. imgaug // github. 2020. URL: https://github.com/aleju/imgaug (дата обращения: 26.04.2021).