

УДК 004.032.26

АНАЛИЗ ЭФФЕКТИВНОСТИ ПРОГРАММНОГО МОДУЛЯ РАСПОЗНАВАНИЯ ЛИЦ НА БАЗЕ СИАМСКОЙ СВЁРТОЧНОЙ НЕЙРОННОЙ СЕТИ FACENET

Климбасов А.А., студент гр.950501

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Перцев Д.Ю. – канд. техн. наук

Аннотация. В данной работе проведен анализ эффективности программного модуля распознавания лиц на базе сиамской свёрточной нейронной сети FaceNet. Выявлены его достоинства и недостатки.

Ключевые слова. Глубокая нейронная сеть, идентификация лиц, FaceNet, самообучение нейронной сети.

Введение. Систематический рост мощностей вычислительных машин, наблюдаемый в последние десятилетия, открывает новые возможности в области анализа и обработки больших данных. Задачи, требующие автоматизации, обладают все менее структурированными данными, а построение математической модели процессов решения представляется все менее тривиальной задачей. Например, в области распознавания и идентификации лица человека данный факт дал толчок к развитию применения нейронных сетей. Благодаря возможности обучения, нейронные сети способны находить сложные зависимости между входными и выходными данными, выявление которых в виде математического отношения методами ручного анализа представляется, как минимум, сложно выполнимым.

Распознавание лиц – это сложный процесс, плохо поддающийся ручной алгоритмизации при условии достижения высокой точности. Распространенным методом решения данной задачи являются свёрточные нейронные сети глубокого обучения.

В случае стандартной классификации модель обучена определять ряд классов, каждый из которых является лицом, которое нейронная сеть может идентифицировать. При этом в процессе распознавания вычисляется распределение вероятностей по всем классам.

Минусом применения классических свёрточных нейронных сетей является необходимость переобучения первых слоёв сети, отвечающих за классификацию при добавлении нового лица. Переобучение модели требует времени и мощностей. Также повышается сложность реализации доступности сервисов, использующих подобные модели. Требование доступности сервиса ставит задачу одновременного обучения модели и её использования для обработки данных, что повышает сложность. Также такие архитектуры требуют большого количества данных для обучения для обеспечения конкурентного уровня точности.

Вышеописанные проблемы не распространяются на сиамскую архитектуру. Основная идея состоит в использовании двух идентичных моделей с одинаковыми весами, реализующих извлечение векторов признаков лиц из изображений в качестве встраиваемых структур, и функции сравнения векторов для вычисления меры схожести. Выходные вектора признаков встроены моделей поддаются сравнению простыми алгоритмами наподобие евклидова расстояния.

Главное преимущество модели с сиамской архитектурой – их способность к обучению на одиночном изображении (англ. One-Shot learning).

Конвейер предварительной обработки изображения. Для улучшения точности работы сети требуется предварительная обработка изображения. Такие факторы, как освещение, резкость изображения, угол поворота лица сильно влияют на точность предсказания. Как следствие требуется создать конвейер предобработки изображения для минимизации влияния вышеописанных факторов на работу модуля.

Корректировка яркости. Освещение – сильно влияющий на качество распознавания лица фактор. Тени на лице и не выровненная яркость могут вносить существенные неточности как при распознавании, так и в процессе добавления лица в базу данных, что является еще более уязвимой частью модуля идентификации. Для устранения дефектов освещения применяется гамма-коррекция с автоматическим определением гамма-коэффициента следующего вида:

$$\gamma = \log(\text{mid} \times 255) / \log(\text{mean}) \quad (1)$$

Корректировка контраста. Для корректного извлечения признаков лица требуется высокая контрастность изображения. Повышение контраста изображения происходит при помощи метода

адаптивной ограниченной нормализации гистограммы (англ. Contrast Limited Adaptive Histogram Equalization) [1].

Обнаружение лица. Лицо человека в общем случае не единственный объект на фотографии. Кроме того, возможно наличие нескольких лиц, в результате чего требуется проведение части процедуры предобработки и распознавания для каждого лица в отдельности. Обнаружение лица происходит на основе ориентиров лица, полученных при помощи модели `shape_predictor_5_face_landmarks` [2].

Выравнивание лица. Выравнивание лица происходит на основе евклидовых преобразований на основе ориентиров, полученных при помощи вышеупомянутой модели. Основная суть преобразования – изменить угол поворота лица относительно горизонтали.

При объединении блоков предобработки учитывается возможность улучшения показателей последующего блока за счёт предыдущего. Текущая последовательность конвейера: обнаружение лица, экстрагирование лица с достаточной для дальнейших преобразований областью, корректировка яркости, корректировка контраста, выравнивание лица, экстрагирование лица в соответствии с требованиями модели распознавания.

Архитектура сиамской нейронной сети изображена на рисунке 1. Архитектура данной нейронной сети состоит из блока извлечения и блока сравнения признаков. В качестве встраиваемой свёрточной сети для извлечения признаков используется реализация FaceNet [3] со встроенной моделью Inception [4].

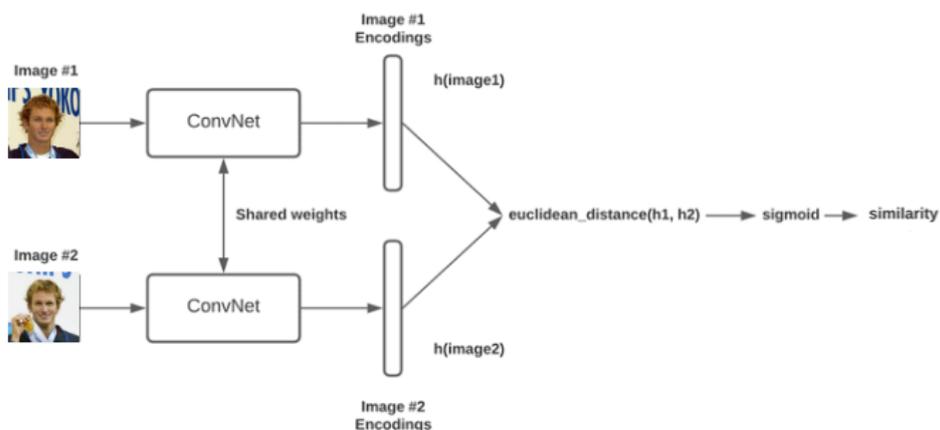


Рисунок 1 – Архитектура нейронной сети

В качестве входных данных в нейронную сеть поступает изображение размером 96 x 96 x 3.

Блок 1: Извлечение признаков (англ. Feature Extractor). В данном блоке к изображению применяется структура нейронной сети FaceNet с использованием вложенной сети Inception. Сеть FaceNet состоит из входного плотно связанного слоя размерностью 96 x 96 x 3. Далее следует блок сети Inception, описанный ниже. Выход Inception подаётся на слой нормализации, возвращающий вектор признаков. В процессе тренировки также следует функция потерь (англ. Triplet Loss) [4].

Архитектура FaceNet приведена на рисунке 2.

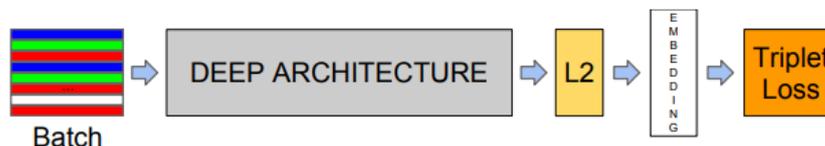


Рисунок 2 – Архитектура встроенной сети FaceNet

Сеть Inception состоит из 9-ти связанных последовательно повторяющихся блоков. Каждый блок представляет собой структуру, состоящую из 4 параллельных последовательностей обработки входных данных. Первая последовательность состоит из операции свёртки 1x1. Вторая – из операции свёртки 1x1 с последующей свёрткой 3x3. Третья – из операции свёртки 1x1 с последующей свёрткой 5x5. Четвёртая – из операции субдискретизации (англ. Max Pooling) 3x3 с последующей свёрткой 1x1. Выходные данные последовательностей конкатенируются в один выходной вектор, являющийся входным по отношению к вышестоящему блоку.

Структура блока изображена на рисунке 3.

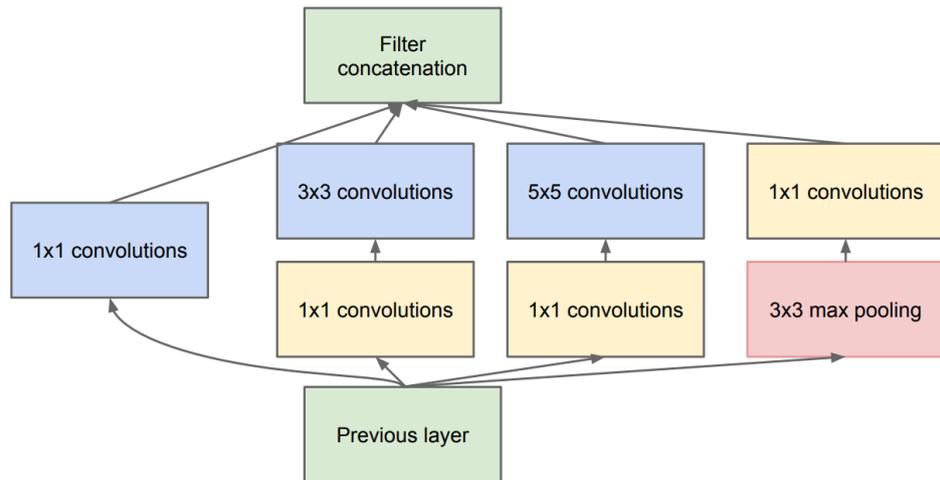


Рисунок 3 – Блок Inception с уменьшением размерности

Выходные данные представляют собой 128 байтовый вектор признаков.

Блок 2: Компаратор признаков. Выполняет вычисление евклидова расстояния между двумя векторами признаков. Представляет собой простую одноименную математическую функцию.

Тестирование нейронной сети.

Тестирование нейронной сети производилось на датасете Labeled Faces in the Wild [5] с использованием метрик точности (англ. precision) и полноты (англ. recall). В качестве порога срабатывания для выходного слоя нейронной сети использовались значения 0.8, 0,7, 0,68, 0,65, 0,5. Значения метрик сведены в таблицу 1.

Таблица 1 – Результаты измерения качества работы нейронной сети на основе выбранных метрик.

Метрика	Порог срабатывания				
	0,8	0,7	0,68	0,65	0,6
Precision	0,772	0,812	0,841	0,906	0,953
Recall	0,813	0,515	0,524	0,468	0,346

Анализ сравниваемых пар изображений, ошибочно распознанных сетью, показал, что сеть плохо справляется с изображениями с повернутыми относительно оси фокуса объектива камеры лицами. Предложено решение в виде добавления в конвейер предобработки выравнивания лиц на основе аффинных преобразований.

На основе полученных результатов можно сделать вывод, что нейронная сеть в совокупности с конвейером предобработки изображений с достаточной точностью выполняет распознавание в рамках области идентификации лиц.

Список использованных источников:

1. Contrast-limited adaptive histogram equalization [Electronic Resource] / Cinvestav. – Mode of access: <https://www.tamps.cinvestav.mx/~wgomez/material/AID/CLAHE.pdf>. – Date of access: 08.04.2023.
2. Shape Predictor 5 Face Landmarks model [Electronic Resource] / Dlib. – Mode of access: http://dlib.net/files/shape_predictor_5_face_landmarks.dat.bz2. – Date of access: 08.04.2021.
3. FaceNet: A Unified Embedding for Face Recognition and Clustering [Electronic Resource] / ArXiv. – Mode of access: <https://arxiv.org/pdf/1503.03832.pdf>. – Date of access: 08.04.2023.
4. Going Deeper with Convolutions [Electronic Resource] / THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL. – Mode of access: <https://www.cs.unc.edu/~wliu/papers/GoogLeNet.pdf>. – Date of access: 08.04.2023.
5. Labeled Faces in the Wild / university of massachusetts. – Mode of access: <http://vis-www.cs.umass.edu/lfw>. – Date of access: 10.04.2023

EFFICIENCY ANALYSIS OF THE SOFTWARE MODULE FOR FACE RECOGNITION BASED ON THE SIAMESE CONVOLUTIONAL NEURAL NETWORK FACENET

Klimbasau A.A.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Pertsau D. – PhD in Computer Science

Annotation. In the paper, we analyzed the effectiveness of the face recognition software module based on the Siamese convolutional neural network FaceNet. Its advantages and disadvantages are revealed.

Keywords. Deep neural network, face identification, FaceNet, neural network self-learning.