Научная статья

Original article

УДК 004.89

ПОИСК ОПТИМАЛЬНОЙ МОДЕЛИ ПРЕДВАРИТЕЛЬНО ОБУЧЕННОЙ НЕЙРОННОЙ СЕТИ ДЛЯ РЕШЕНИЯ ЗАДАЧИ ИДЕНТИФИКАЦИИ ЛИЦ

SEARCH FOR THE OPTIMAL MODEL OF A PRELIMINARY TRAINED NEURAL NETWORK FOR SOLVING THE PROBLEM OF FACE IDENTIFICATION

Сиденко Никита Евгеньевич, студент 2 курса магистратуры по направлению подготовки «Информатика и вычислительная техника» (Искусственный интеллект), ФГБОУ ВО «Иркутский национальный исследовательский технический университет» (664074 Россия, г. Иркутск, ул. Лермонтова, д. 83)

Sidenko Nikita Evgenevich, 2nd year Master's student in the field of study «Informatics and Computer Engineering» (Artificial Intelligence), Federal State Budget Educational Institution of Higher Education «Irkutsk National Research Technical University» (83 Lermontova st., Irkutsk, 664074 Russia)

Аннотация. В статье рассматриваются предварительно обученные модели сверточных сетей для решения задачи идентификации лиц. Проведен анализ эффективности наиболее распространенных предварительно обученных моделей. Для сравнения моделей использовался подход сиамской нейронной сети. Качество предсказания оценивалось по евклидовому расстоянию между двумя векторами. В статье представлены описание моделей и их сравнительные

характеристики. Практическая значимость проведенного исследования заключается в том, что полученные результаты могут быть использованы в проектах с идентификаций лиц.

Annotation. The article considers pre-developed CNN-models for solving face identification problems. An analysis of the effectiveness of the popular models was carried out. The Siamese neural network approach is used to compare models. Evaluation of the quality of forecasting by the Euclidean distance between two vectors. The article presents descriptions of models and their comparative characteristics. This study provides new insights into face recognition projects.

Ключевые слова: методы машинного обучения, сиамские нейронные сети, сверточные нейронные сети, предварительные обученные модели, сравнительный анализ.

Keywords: machine learning methods, Siamese neural networks, convolutional neural networks, pretrained models, benchmarking.

Введение

Для эффективного решения задач машинного зрения при малом количестве данных для обучения модели используют предварительно обученные сверточные сети [1]. Данный подход позволяет выделить признаки из входных данных, для последующего обучения полносвязной сети. На сегодняшний день предлагается множество предобученных моделей от разных сообществ и крупных производителей, таких как Google или Facebook, при этом нужно понимать, что некоторые модели могут быть избыточными или наоборот неэффективными при решении определенных задач. Поэтому важно уметь произвести анализ необходимой модели и принять решение о её применимости к конкретной задаче. В данной статье представлен анализ предварительно обученных моделей для идентификации лиц.

Предварительно обученные сверточные сети для распознавания лиц

На сегодняшний день существует несколько популярных предобученных CNN-моделей для распознавания лиц [2], а именно: VGG-Face, Facenet,

OpenFace, DeppFace. Эти модели были обучены для классификации изображений лиц на крупномасштабном наборе данных.

VGG-Face

VGG-Face разработана исследователями из группы Visual Geometry Group в Оксфорде. VGG-Face достаточно глубокая сверточная сеть, она имеет 22 последовательных слоя.

Facenet

Facenet разработана исследовательской группой Google. Facenet является сетью, которая преобразовывает изображения лиц в компактное евклидовое пространство, где расстояния напрямую соответствуют мере сходства лиц. Как только это пространство создано, такие задачи, как распознавание лиц, проверка и кластеризация, могут быть легко реализованы с использованием стандартных методов с встраиванием FaceNet в качестве векторов признаков. FaceNet использует особенную функцию потерь TripletLoss, она минимизирует расстояние между изображением с похожей внешностью. Модель построена на Inception Resnet. Resnet – остаточные нейронные сети, имеющие соединения для быстрого доступа, данная сеть разработана для сокращения потерь при оптимизации CNN с глубокими слоями. Inception Resnet — это сверточная нейронная архитектура, основанная на семействе архитектур Inception, но включающая в себя подходы Resnet.

OpenFace

В создании OpenFace принимали участие: Брэндон Амосом, Бартош Людвичюк и Махадев Сатьянараян. Данная разработка была поддержана Национальным научным фондом (NSF). Дополнительную поддержку оказали корпорация Intel, Google, Vodafone, NVIDIA и семейный фонд Конклина Кистлера. Несмотря на большее количество слоёв и связей OpenFace является лёгкой моделью для распознавания лиц по сравнению с остальными популярными моделями. Модель построена на Inception Resnet.

DeepFace

DeepFace разработана исследовательской группой Facebook. Данная CNN использует 8 уровневую сверточную нейронную сеть с более чем 120 миллионами параметров. В обучении сети использовалось 4 миллиона изображений 4000 человек. Процесс работы модели включает в себя этап обработки, а именно детекцию лица и выравнивание.

Сравнение предварительно обученных моделей

Для сравнения предварительно обученных моделей были выбраны следующие критерии: размер модели, вывод модели, затраченное время на предсказание, точность предсказания (Таблица 1).

Таблица 1 – Критерии для сравнительного анализа предварительно обученных моделей

№	Критерий для оценки	Способ оценивания
1.	Размер модели	Объем занимаемого места на дисковом пространстве
2.	Вывод модели	Визуализация выходного вектора
3.	Затраченное время на предсказание	Время работы скрипта для сравнения 36-ти фотографий.
4.	Точность предсказания	Сравнение выходных векторов

Все тесты производились на одном компьютере с процессором ryzen 5 4500u, в среде разработки PyCharm.

Точность прогнозирования

Чтобы получить предсказание для фотографии человека сравнивается тестовая фотография с оригинальной. Оригинальная фотография обычно хранится в базе данных и является эталонной для идентификации определенного человека. Две фотографии получают предсказания от одной модели нейронной сети и их результаты сравниваются, такой подход называется сиамские сети [2]. Для сравнения результата используется евклидово расстояние.

Загрузка моделей

Для загрузки моделей была использована библиотека Deepface. Deepface является легкой библиотекой для распознавания лиц и анализа лицевых атрибутов, таких как: возраст, пол, эмоции и раса. Данная библиотека имеет в своём составе предварительно обученные модели VGG-Face, Google FaceNet, OpenFace и Facebook DeepFace. Также библиотека имеет инструменты для детекции и обработки лиц.

Входные данные

Для анализа предсказаний было выбрано 6 фотографий 2-х человек. Входные данные представлены на рисунке 1. Для каждой фотографии была применена детекция лица и обработка с помощью метода detectFace из библиотеки deepface. Пример фотографии после детекции и обработки отображен на рисунке 2. Следует отметить, что detectFace также произвёл выравнивание лица, что чрезвычайно важно для некоторых моделей.

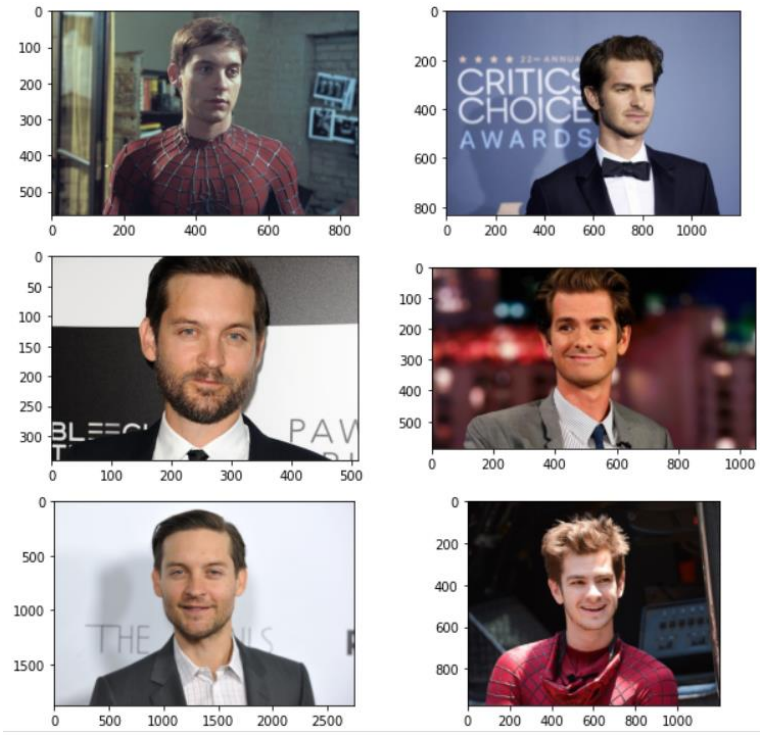


Рисунок 1 – Входные данные

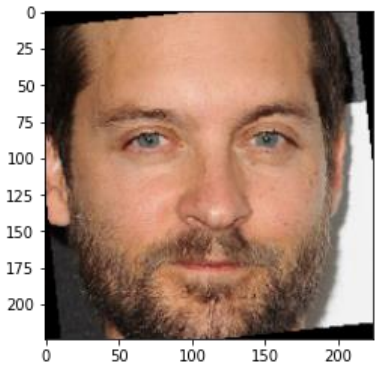


Рисунок 2 – обработанная фотография.

В итоге был создан входной датасет для работы с моделями сверточных нейронных сетей.

VGG-Face

Модель данной сверточной нейронной сети имеет размер 580 мб.

Модель VGG-Face имеет на выходе вектор из 2622 значений. На вход сети подана фотография face1.jpg, визуальное представление выхода отображено на рисунке 3.

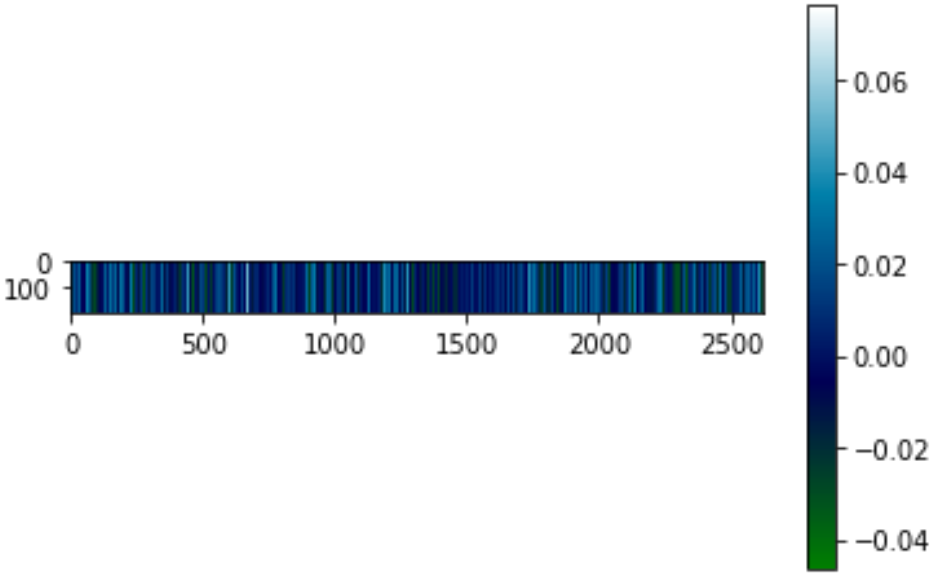


Рисунок 3 – Вывод модели VGG-Face

Время работы при сравнении 6-ти фотографий в цикле: при выполнении 36-ти итераций затраченное время составило 59 секунд.

Для тестирования точности выводились результаты сравнения в матрицу, где пересечение столбца и строки показывает евклидово расстояние между двумя фотографиями.

Результат сравнения фотографий представлен в таблице 2.

Таблица 2 – Итоги сравнения, модель VGG-Face

	Tobi 1	Tobi 2	Tobi 3	Andrew 1	Andrew 2	Andrew 3
Tobi 1	0.000	0.543	0.505	0.804	0.799	0.826
Tobi 2	0.543	0.000	0.498	0.702	0.615	0.765
Tobi 3	0.505	0.498	0.000	0.837	0.827	0.902
Andrew 1	0.804	0.702	0.837	0.000	0.533	0.541
Andrew 2	0.799	0.615	0.827	0.533	0.000	0.468
Andrew 3	0.826	0.765	0.902	0.541	0.468	0.000

Из таблицы 2 видно, что средний порог между значениями составляет 0.579, при этом разница в евклидовом расстоянии между разными людьми имеет минимальное значение в 13.3 процента. Данный результат вполне приемлем для идентификации лица, но при таких значениях возможны ложные предсказывания. Для повышения точности можно использовать другой метод обработки входной фотографии или использовать классификатор после сверточной сети.

Facenet

Модель данной сверточной нейронной сети имеет размер 92.2 мб.

Модель Facenet имеет на выходе вектор из 512-ти значений. На вход сети подана фотография face1.jpg, визуальное представление выхода отображено на рисунке 4.

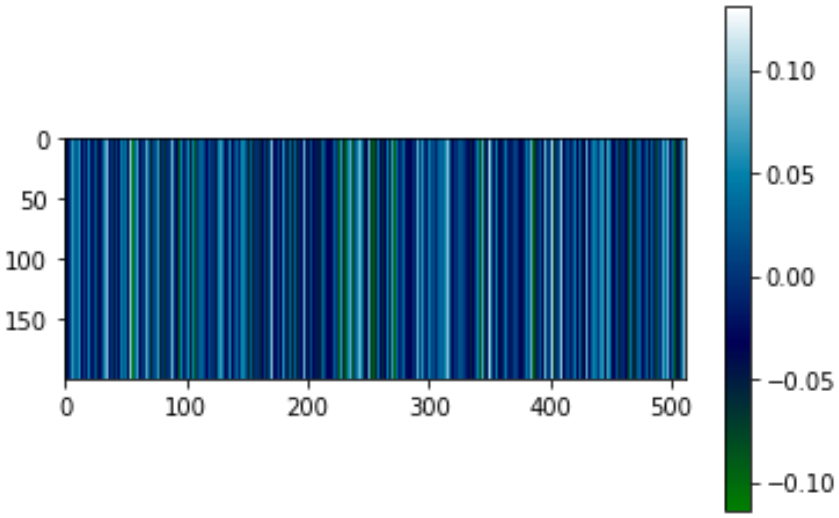


Рисунок 4 – Вывод модели Facenet

Время работы при сравнении 6 фотографий в цикле: при выполнении 36-ти итераций затраченное время составило 18 секунд.

Результат с таблицей точности представлен в таблице 3.

Таблица 3 – Итоги сравнения, модель Facenet

	Tobi 1	Tobi 2	Tobi 3	Andrew 1	Andrew 2	Andrew 3
Tobi 1	0.000	0.764	0.628	1.271	1.336	1.232
Tobi 2	0.764	0.000	0.768	1.297	1.362	1.283
Tobi 3	0.628	0.768	0.000	1.334	1.387	1.289
Andrew 1	1.271	1.297	1.334	0.000	0.717	0.694
Andrew 2	1.336	1.362	1.387	0.712	0.000	0.695
Andrew 3	1.232	1.283	1.289	0.694	0.695	0.000

Из таблицы 3 видно, что средний порог между значениями составляет 1, при этом разница в евклидовом расстоянии между разными людьми имеет минимальное значение в 65 процента. Данный результат достаточно высокий, что позволяет использовать данную модель без классификатора.

OpenFace

Размер модели данной сверточной нейронной сети составляет всего 15.3 мб.

Модель OpenFace имеет на выходе вектор из 128 значений. На вход сети подана фотография face1.jpg, визуальное представление выхода отображено на рисунке 5.

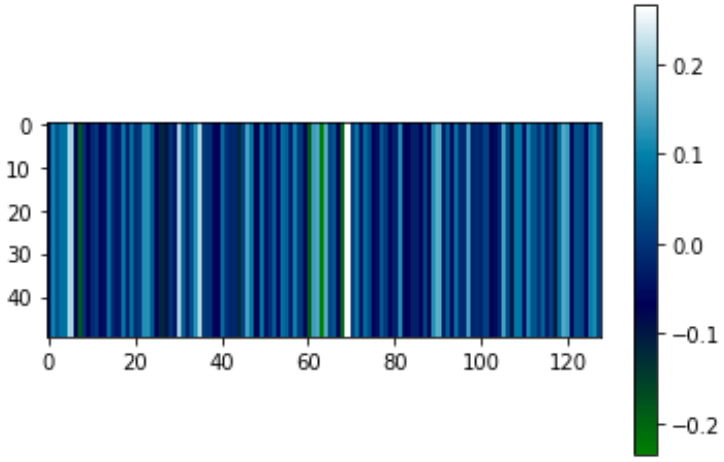


Рисунок 5 – Вывод модели OpenFace

Время работы при сравнении 6 фотографий в цикле: при выполнении 36-ти итераций затраченное время составило 9 секунд.

Результат с таблицей точности представлен в таблице 4.

Таблица 4 – Итоги сравнения, модель OpenFace

	Tobi 1	Tobi 2	Tobi 3	Andrew 1	Andrew 2	Andrew 3
Tobi 1	0.000	0.676	0.540	0.869	0.859	0.952
Tobi 2	0.676	0.000	0.394	0.960	0.855	0.978
Tobi 3	0.540	0.394	0.000	0.992	0.831	0.983
Andrew 1	0.869	0.960	0.992	0.000	0.639	0.682
Andrew 2	0.859	0.855	0.831	0.639	0.000	0.489
Andrew 3	0.952	0.978	0.983	0.682	0.489	0.000

Из таблицы 4 видно, что средний порог между значениями составляет 0.75, при этом разница в евклидовом расстоянии между разными людьми имеет минимальное значение в 22.9 процента. Данный результат можно считать приемлемым, учитывая, что модель имеет небольшой вес 15.3 мб. Но для повышения порогового значений предсказания требуется выравнивания лица на фотографии.

DeepFace

Модель данной сверточной нейронной сети имеет размер 511 мб.

Модель DeepFace имеет на выходе вектор из 4096 значений. На вход сети подана фотография face1.jpg, визуальное представление выхода отображено на рисунке 6.

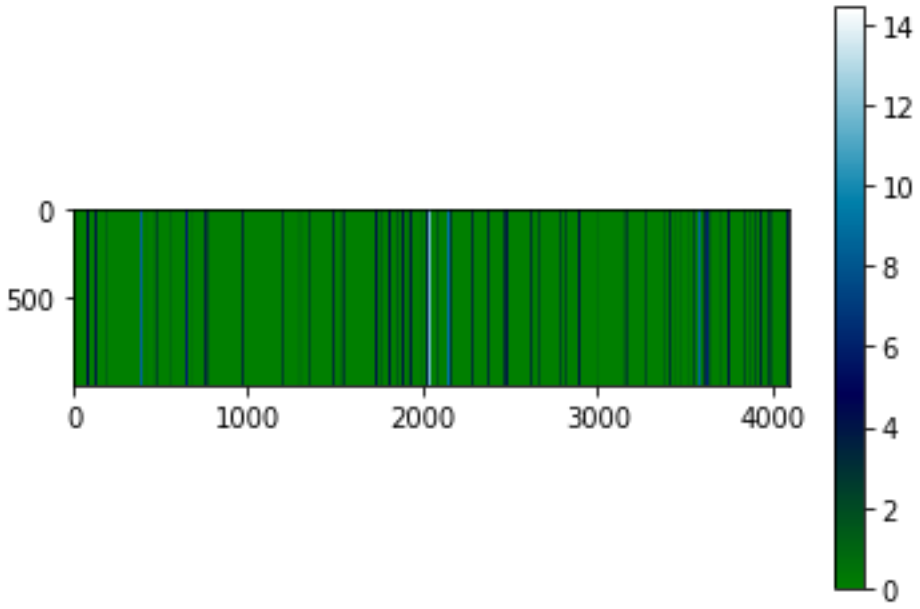


Рисунок 6 – Вывод модели DeepFace

Время работы при сравнении 6 фотографий в цикле: при выполнении 36 итераций затраченное время составило 13 секунд. Хороший показатель времени при таком размере сети.

Результат с таблицей точности представлен в таблице 5.

Таблица 5 – Итоги сравнения, модель DeepFace

	Tobi 1	Tobi 2	Tobi 3	Andrew 1	Andrew 2	Andrew 3
Tobi 1	0.000	0.624	0.562	0.670	0.622	0.728
Tobi 2	0.624	0.000	0.640	0.694	0.680	0.773
Tobi 3	0.562	0.640	0.000	0.679	0.662	0.719
Andrew 1	0.670	0.694	0.679	0.000	0.595	0.661
Andrew 2	0.622	0.680	0.662	0.595	0.000	0.599
Andrew 3	0.728	0.773	0.719	0.661	0.559	0.000

Данная модель имеет ложные предсказывания. Если мы исключим ложные предсказания из таблицы 5 (желтые метки), то получим следующие результаты: средний порог между значениями составляет 0.615, при этом разница в евклидовом расстоянии между разными людьми имеет минимальное значение в 3.4 процента. Данная сеть не подходит под модель сиамской сети в данной конфигурации, возможно необходимо сменить подход в предварительной обработке фотографий.

Итоги сравнения

В таблице 6 представлен итог тестирования 4-х моделей преобученных сверточных сетей.

Таблица 6 – Параметры

	VGG-Face	Facenet	OpenFace	DeeopFace
Размер	580 mb	92.2 mb	15.3 mb	512 mb
Вектор	2622	512	128	4096
Время	59 сек.	18 сек.	9 сек.	13 сек.
Расстояние	13.3 %	65 %	22.9 %	3.4 %

В таблице в строках указаны следующие параметры:

- Размер – размер сети в мегабайтах;
- Вектор – размер вектора;
- Время – время выполнения 36 итераций предсказания;
- Расстояние – оценка качества предсказания в разнице между двумя ближними оценками (самое высокое значение в предсказании одного человека и самое низкое значение в предсказании разных людей).

Таким образом, из анализа результатов тестирования следует, что оптимальным выбором для задачи определения лиц являются модели Facenet и OpenFace. VGG-Face и DeepFace показывают не лучшие результаты в решении нашей задачи. Для вариантов, где мы ограничены в ресурсах, например мобильное приложение со встроенной моделью нейронной сети, лучше использовать OpenFace. В остальных случаях хорошим выбором будет Facenet.

Заключение

В результате проведенных исследований выявлено, что оптимальной моделью для идентификации лиц является FaceNet.

При применении предварительно обученных моделей было произведено сравнение эффективности моделей по выбранным критериям: размер модели, размер выходного вектора, скорость работы, точность предсказания.

Сравнительный анализ моделей по указанным критериям позволил выявить две оптимальные модели - Facenet и OpenFace. Модели показали хорошую скорость работы и точность идентификации фотографий. Модель OpenFace имеет маленький размер сети (15.3 мб) и требует выравнивание фотографий, данная модель подходит для размещения на мобильных устройствах, где есть ограничение ресурса, в остальных случаях подойдёт модель Facenet. Данные модели можно использовать в сиамских сетях для идентификации лиц. Для повышения точности работы сети необходимо рассмотреть разные варианты обработки входного изображения. Результаты проведенного исследования могут быть использованы в проектах, где необходимо встроить модуль идентификации лиц.

Литература

1. Глубокое обучение на Python / Ф. Шолле - СПБ.: Питер, 2018. – 400 с.
2. A Gentle Introduction to Face Recognition in Deep Learning [Электронный ресурс]. – Режим доступа: https://sefiks.com/2020/05/01/a-gentle-introduction-to-face-recognition-in-deep-learning/ - (дата обращения: 15.12.21)
3. keras-vggface [Электронный ресурс]. – Режим доступа: https://github.com/rcmalli/keras-vggface - (дата обращения: 17.12.21)
4. OpenFace [Электронный ресурс]. – Режим доступа: https://vggfacehttps://cmusatyalab.github.io/openface/ - (дата обращения: 17.12.21)
5. FaceNet: A Unified Embedding for Face Recognition and Clustering [Электронный ресурс]. – Режим доступа: https://arxiv.org/abs/1503.03832 - (дата обращения: 17.12.21)

6. DeepFace: Closing the Gap to Human-Level Performance in Face Verification [Электронный ресурс]. – Режим доступа: https://research.facebook.com/publications/deepface-closing-the-gap-to-human-level-performance-in-face-verification/ - (дата обращения: 17.12.21)
7. Deepface [Электронный ресурс]. – Режим доступа: https://github.com/serengil/deepface - (дата обращения: 1.12.21)

References

1. Глубокое обучение на Python / Ф. Шолле - СПБ.: Питер, 2018. – 400 с.
2. A Gentle Introduction to Face Recognition in Deep Learning [Electronic resource]. – Access mode: https://sefiks.com/2020/05/01/a-gentle-introduction-to-face-recognition-in-deep-learning/ - (date of access: 15.12.21)
3. keras-vggface [Electronic resource]. – Access mode: https://github.com/rcmalli/keras-vggface - (date of access: 17.12.21)
4. OpenFace [Электронный ресурс]. – Access mode: https://vggfacehttps://cmusatyalab.github.io/openface/ - (date of access: 17.12.21)
5. FaceNet: A Unified Embedding for Face Recognition and Clustering [Electronic resource]. – Access mode: https://arxiv.org/abs/1503.03832 - (date of access: 17.12.21)
6. DeepFace: Closing the Gap to Human-Level Performance in Face Verification [Electronic resource]. – Access mode: https://research.facebook.com/publications/deepface-closing-the-gap-to-human-level-performance-in-face-verification/ - (date of access: 17.12.21)
7. Deepface [Electronic resource]. – Access mode: https://github.com/serengil/deepface - (date of access: 1.12.21)

© Сиденко Н.Е., 2022 Научно-образовательный журнал для студентов и преподавателей «StudNet» №4/2022.

Для цитирования: Сиденко Н.Е. ПОИСК ОПТИМАЛЬНОЙ МОДЕЛИ ПРЕДВАРИТЕЛЬНО ОБУЧЕННОЙ НЕЙРОННОЙ СЕТИ ДЛЯ РЕШЕНИЯ ЗАДАЧИ ИДЕНТИФИКАЦИИ ЛИЦ// Научно-образовательный журнал для студентов и преподавателей «StudNet» №4/2022.