МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное

образовательное учреждение высшего образования

**«Челябинский государственный университет»**

**(ФГБОУ ВО «ЧелГУ»)**

Математический факультет

Кафедра компьютерной безопасности и прикладной алгебры

**ДИПЛОМНАЯ РАБОТА**

Реализация системы распознавания 3D лиц с использованием нейронных сетей

|  |  |
| --- | --- |
| ДОПУСТИТЬ К ЗАЩИТЕ  Заведующий кафедрой,  Кандидат физико-математических наук,  Доцент  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Ручай А.Н.    «\_\_\_» \_\_\_\_\_\_\_\_\_ 2020 г. | Выполнил студент  Токарев Игорь Вячеславович  группы МК-601  очной формы обучения  направления подготовки 10.05.01 компьютерная безопасность  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  «\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_ 2020 г. |

|  |  |
| --- | --- |
| Нормоконтроль  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Шалагинов Л.В.    «\_\_\_» \_\_\_\_\_\_\_\_\_ 2020 г. | Научный руководитель  Ручай Алексей Николаевич  Заведующий кафедры КБиПА  Кандидат физико-математических наук  Доцент  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  «\_\_\_» \_\_\_\_\_\_\_\_\_ 2020 г. |

Челябинск

2020

Оглавление

[Введение 3](#_Toc31948395)

[Теоретическая часть 4](#_Toc31948396)

[Наборы данных. 4](#_Toc31948397)

[Задача идентификации по 3D изображению лица. 9](#_Toc31948398)

[Нейронные сети. 11](#_Toc31948399)

[Практическая часть 14](#_Toc31948400)

[Выполнение процесса идентификации 14](#_Toc31948401)

# Введение

Задача идентификации по биометрическим признакам сегодня имеет большое значение в задачах компьютерной безопасности. Особенно, важным является при разграничении доступа к важным хранилищам информации или субъектам, которые должны быть доступны лишь ограниченному кругу лиц.

Целью данной работы является создание системы идентификации по 3D изображению лица человека.

Для реализации системы были поставлены следующие задачи:

1. Сбор данных;
2. Обработка полученных данных;
3. Дообучение нейронной сети на данных;
4. Обработка результата.

# Теоретическая часть

## Наборы данных

Для реализации систем компьютерного зрения на основе первостепенной задачей является сбор подходящих наборов данных. Для применения в поставленной задаче были выбраны следующие наборы:

1. 3DTEC
2. BU-3DFE
3. Bosphorus
4. FaceWarehouse

Опишем каждый из наборов.

**3DTEC** (<http://i14s50.anthropomatik.kit.edu/download/BEFIT_Notre_Dame_3DTEC_v2.pdf>)

В базе данных 3DTEC имеется 440 изображения в формате ABS.

Чтобы иметь возможность работать с изображениями, сначала необходимо изменить формат каждого изображения на PLY. Что было получено с использованием языка программирования Python и библиотеки NumPy.

Данная база данных была разработана для использования в задачах по распознаванию близнецов.

3DTEC имеет собственную лицензию (<https://cse.nd.edu/labs/cvrl/data-sets/und-license-agreement-3d-tec>).

**BU-3DFE** (<http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html>)

В данной базе располагается 2500 изображений 3D лиц более чем 100 человек, которые были созданы для задач распознавания и анимации.

Соотношение изображений мужчин и женщин составляет 56% к 44%. Возраст актёров от 18 до 70 лет. Содержит множество рас:

* Европейцы;
* Африканцы;
* Среднеазиатская;
* Индийцы.

Каждый субъект исполнил 7 различных выражений лица перед 3D-сканером

Изображены следующие виды эмоций:

* Счастье
* Отвращение
* Страх
* Злость
* Удивление
* Грусть
* Нейтральная

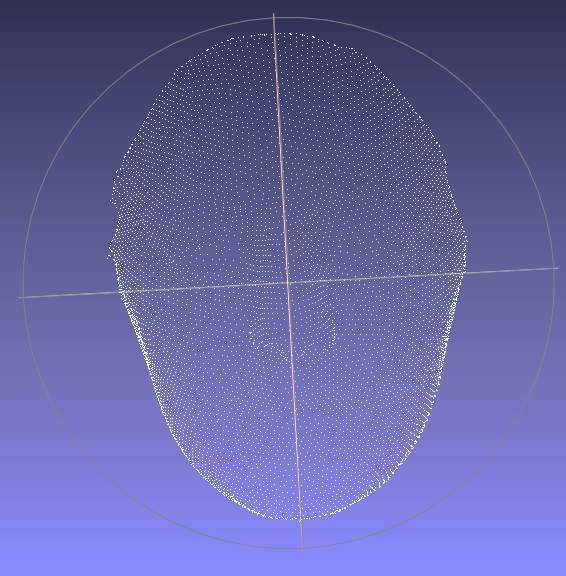
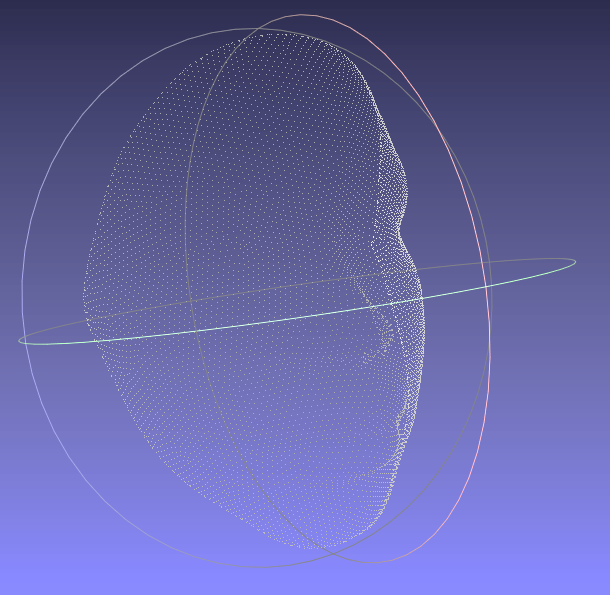


Рис 4. Пример облаков точек, полученных из файлов 3D-изображений библиотеки 3B-FU.

**FaceWarehouse** (<http://kunzhou.net/zjugaps/facewarehouse/>)

Cao Chen, Yanlin Weng, Shun Zhou, Yiying Tong, Kun Zhou: "FaceWarehouse: a 3D Facial Expression Database for Visual Computing", IEEE Transactions on Visualization and Computer Graphics, 20(3): 413-425, 2014

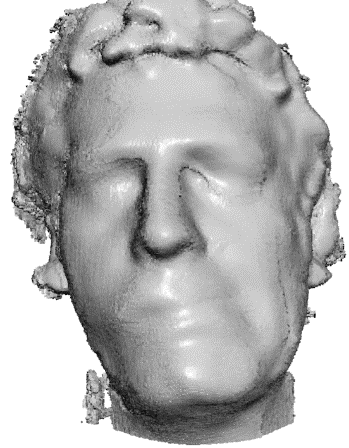
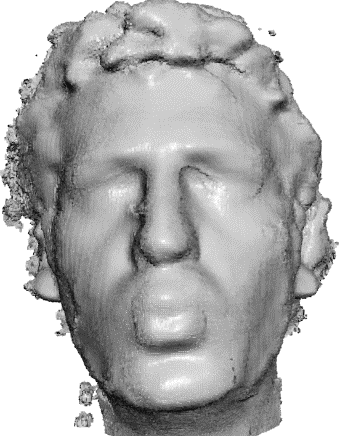
База данных 3D изображений лиц с различными выражениями для создания приложений. Содержит изображения 150 субъектов в возрасте от 7 до 80 лет разных этнических групп. Для получения данных использовалась RGBD камера Kinect.

Данные для каждого человека хранятся в отдельной директории. В каждой директории "Tester\_xx" хранятся RGBD данные, для каждого изображения отмечены 74 особые точки, созданы меши, также представлены некоторые дополнительные данные.

* Tester\_1: folder name with person 1
  + TrainingPose
    - pose\_0.pose - pose\_20.pose: данные от Kinect RGBD камеры
    - pose\_0.obj - pose\_20.obj:реконструированные меши.
    - pose\_0.land - pose\_20.land: отмеченные 2D точки для первого кадра в директории.
  + Blendshape
    - shape.bs: двоичный файл с выражением blendshape
    - shape\_0.obj - shape\_46.obj: выражения blendshapes

Примеры возможных применений:

* перенос эмоций.
* создание анимаций изображений в реальном времени.
* создание новых эмоций для отдельного персонажа используя совмещение имеющихся



**Bosphorus**

База данных BOSPHORUS специально создана чтобы проводить исследования в области 3D и 2D обработки изображений лиц людей, а также 3D реконструкции лиц. База состоит из изображений 105 человек, и всего 4666 изображений 3D облаков. В базе содержатся не только нормализованные изображения, но также:

1. Выражение эмоций – до 35 для отдельного человека.
2. Подсчитанные лицевые движения.
3. Третья часть базы создана с привлечением профессиональных актёров.

Различные уровни поворота лица.

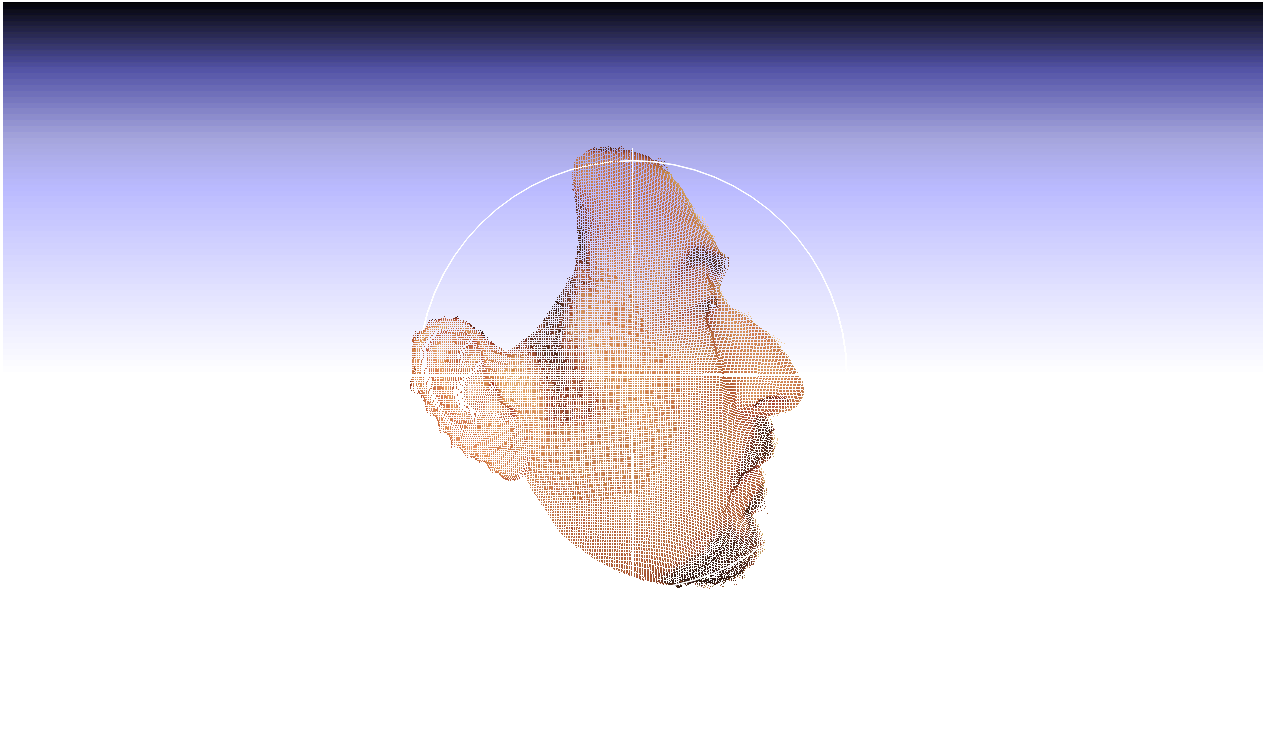


Рис 1. Пример изображения bs002\_YR\_L90\_0.ply до процедуры поворота.

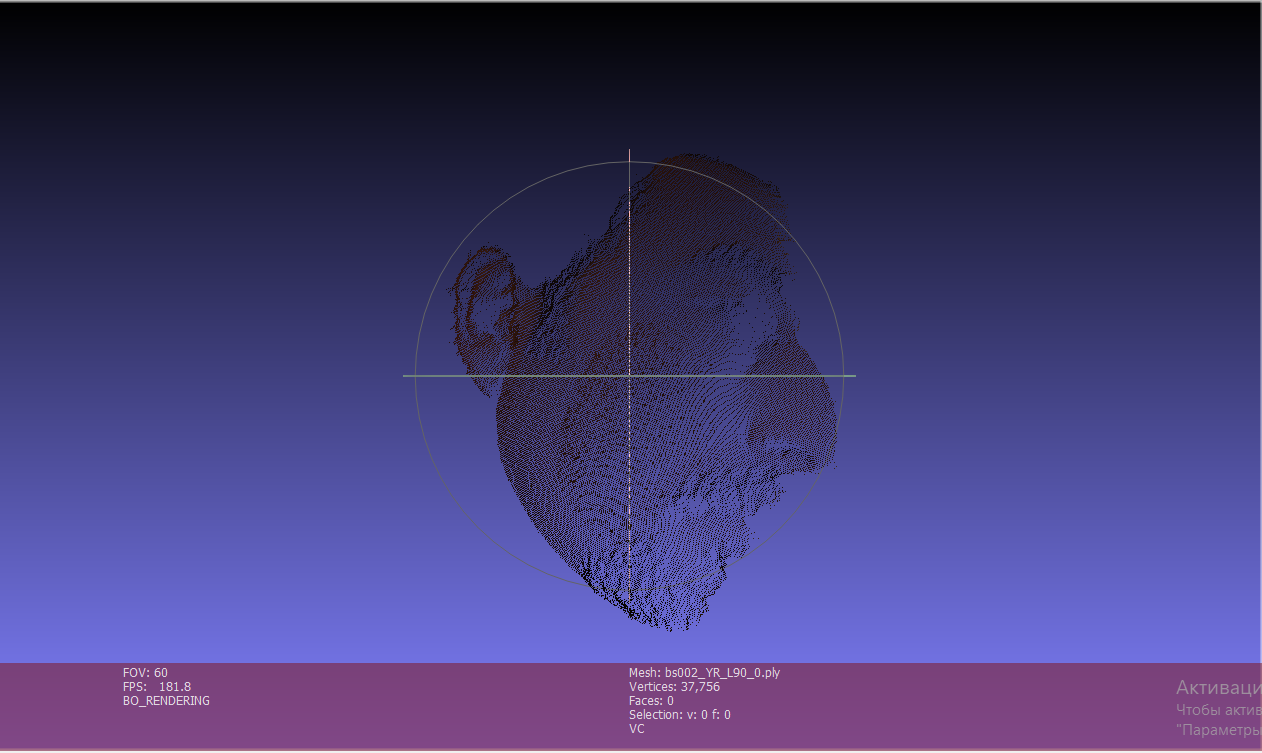


Рис 2. Пример изображения bs002\_YR\_L90\_0.ply после процедуры поворота.

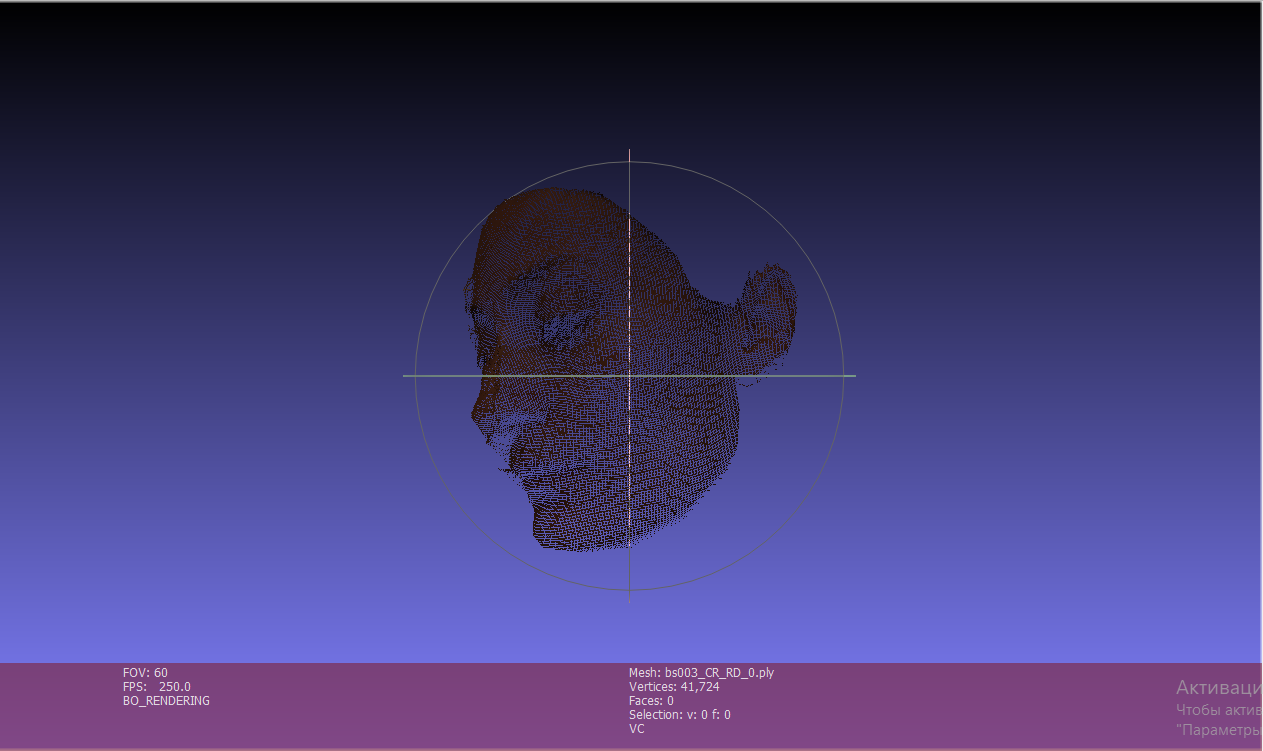


Рис 3. Пример изображения bs003\_CR\_RD\_0.ply после процедуры поворота.

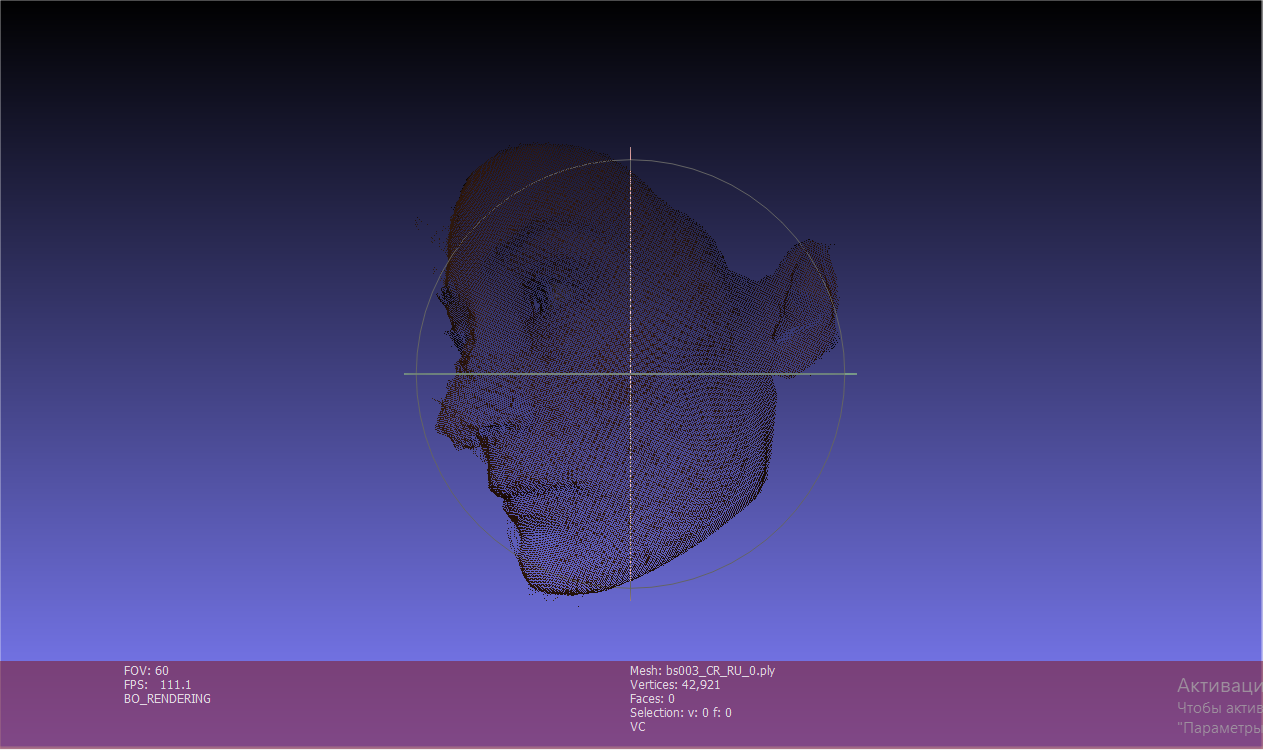


Рис 4. Пример изображения bs003\_CR\_RD\_0.ply после процедуры поворота.

По результатам предварительного исследования представленных наборов, можно заключить, что наиболее подходящим является Bosphorus, так как обладает изначально впечатляющей вариативностью изображений и аугментированностью.

Приведём сводную таблицу с характеристиками каждого из наборов:

|  |  |  |  |
| --- | --- | --- | --- |
| Набор | Уровень аугментированности | Количество изображений | Необходимый уровень предобработки |
| Bosphorus | Высокий | 2622 | Низкий |
| 3DTEC | Низкий | 440 | Высокий |
| BU-3DFE | Средний | 5000 | Средний |
| FaceWarehouse | Высокий | 2850 | Высокий |

**VGG Face 2**

В качестве исходного набора данных для протестированных архитектур нейронных сетей используется VGG Face 2 (<http://www.robots.ox.ac.uk/~vgg/data/vgg_face2/index.html>).

<http://www.robots.ox.ac.uk/~vgg/publications/2018/Cao18/cao18.pdf>

VGG Face 2 собран научной группой Оксфордского Университета (University of Oxford). Набор содержит более 3.3 миллионов изображений лиц более чем 9000 субъектов. Размеры изображений варьируются. Всем изображениями соответствуют различные методы аугментации, такие как:

* различные положения лица,
* присутствие посторонних лиц,
* использование дополнительных средств сокрытия элементов лица (очки, шляпы и т.д.),
* изменение фона,
* различие тона, яркости изображения.

** **

*Рис. 3 Примеры изображения лица субъекта в наборе данных VGG Face 2.*

** ** 

*Рис. 3 Примеры изображения лица субъекта в наборе данных VGG Face 2.*

## Задача идентификации по 3D изображению лица.

Назовём режимом идентификации следующий эксперимент. Создадим два набора данных – Gallery и Probe.

Стандартным способом проведения идентификации является расположения в Gallery нормализованных изображений, а в Probe произвольного набора изображений.

На выходе нейронной сети получается набор признаков. Набор признаков представляется в виде вектора, длина которого определяется конкретной архитектурой нейронной сети. Например:

* для сети VGG16 набор признаков слоя fc7 имеет длину 4096;
* для сети ResNet-50 вектор признаков слоя classifier имеет длину 8631.

Для того, чтобы сравнивать похожесть изображений в задаче идентификации мы используем косинусное расстояние, вычисляемое для двух векторов признаков, получаемых на выходе нейронной сети.

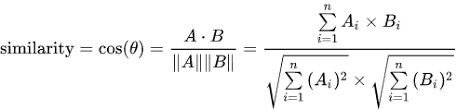


Рис 1. Косинусное расстояние между векторами A и B.

Подобный подход позволяет освободиться от создания классификатора для фиксированного набора изображений и вычислять сходство между произвольным количеством субъектов идентификации.

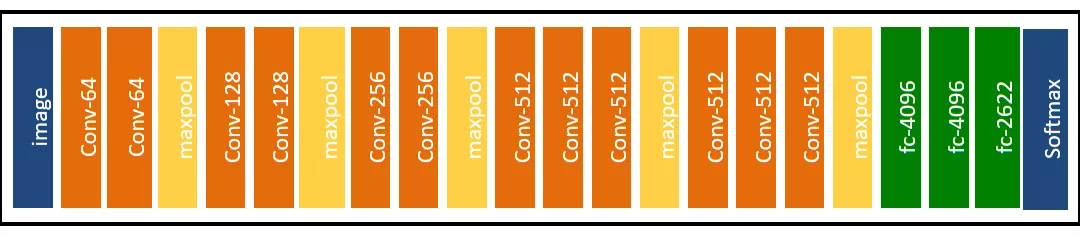
В нормализованной группе находились изображения нормальные изображения для каждого из объектов идентификации. Затем в 105 экспериментах для каждого отдельного объекта выбирались оставшиеся, кроме нормального, изображения и проводилось их сравнение нейронной сетью. На выходе нейронной сети мы получаем вероятности совпадения изображения из нормализованной и тестовой групп.

*.*

## Нейронные сети.

В предыдущих работах была исследована нейронная сеть с VGG16 обученнная на наборе данных VGGFace. (<https://www.robots.ox.ac.uk/~vgg/data/vgg_face/>)

Данная нейронная сеть является сверхглубокой. Имеет 22 слоя, включая полносвязные и свёрточные слои. На вход сети подаётся изображение размерности 224\*224 пикселей.



Архитектура ResNet также является сверхглубокой нейронной сетью. Различные модификации могут включать различные наборы слоёв. Отличием ResNet от других свёрточных нейронных сетей является наличие специальных Residual слоёв.

Рис. 4. Архитектура нейронной сети VGG-Face.

В данной работе используется архитектура ResNet-50. Д

Схематиченое изображение Residual слоя изображено на рисунке Рис. 2. В задачах классификации изображений архитектуры ResNet неоднократно доказывали свою эффективность и показывая state-of-art результаты.

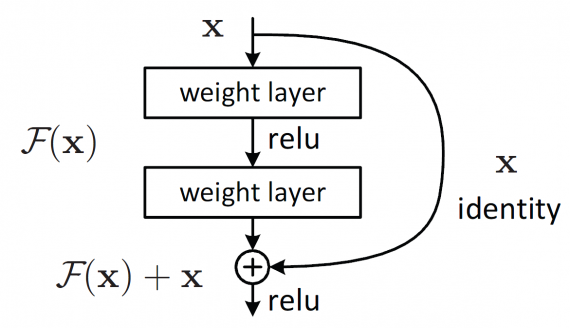


Рис. 2. Схематичное изображение Residual блока. Такой блок состоит из суммы вектора данных X и вектора обработанных данных преобразованием F(X).

Использование ResNet-50 было определено тем, что существуют несколько свободно распространяемых предобученных на VGG Face 2 нейронных сетей данной архитектуры.

Описание предобученных архитектур нейронных сетей ResNet-50.

(<https://github.com/ox-vgg/vgg_face2>)

| **Архитектура** | **Размер выходного слоя** | **TAR@FAR = 0.001** | **TAR@FAR = 0.01** | **Фреймворк** |
| --- | --- | --- | --- | --- |
| ResNet-50 | 8631 | 0.891 | 0.947 | Caffe, PyTorch |
| ResNet-50-256D | 256 | 0.898 | 0.956 | Caffe, PyTorch |
| ResNet-50-128D | 128 | 0.904 | 0.956 | Caffe, PyTorch |

Каждая из представленных выше моделей имеет 228 слоёв. На рисунке Рис 4 изображена архитектура нейронной сети ResNet-50. За исключением того, что на последнем слое имеется 8631 выход, вместо 1000.

Различия заключаются лишь в количестве нейронов выходного слоя.

Для реализации системы выбрана сеть с наибольшим количеством параметров последнего слоя для обеспечения возможности наиболее точного распознавания.

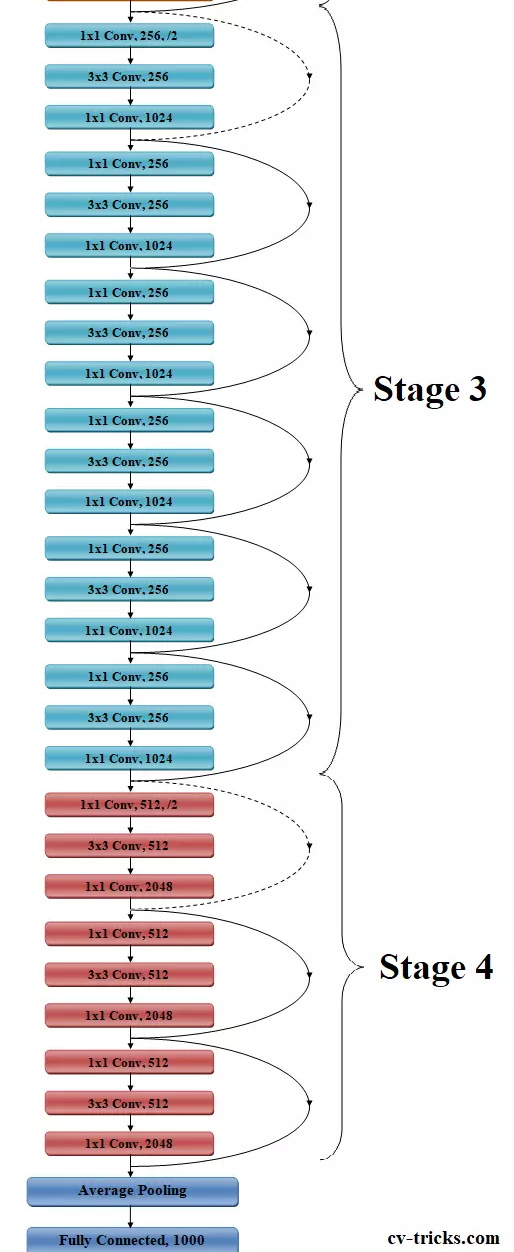
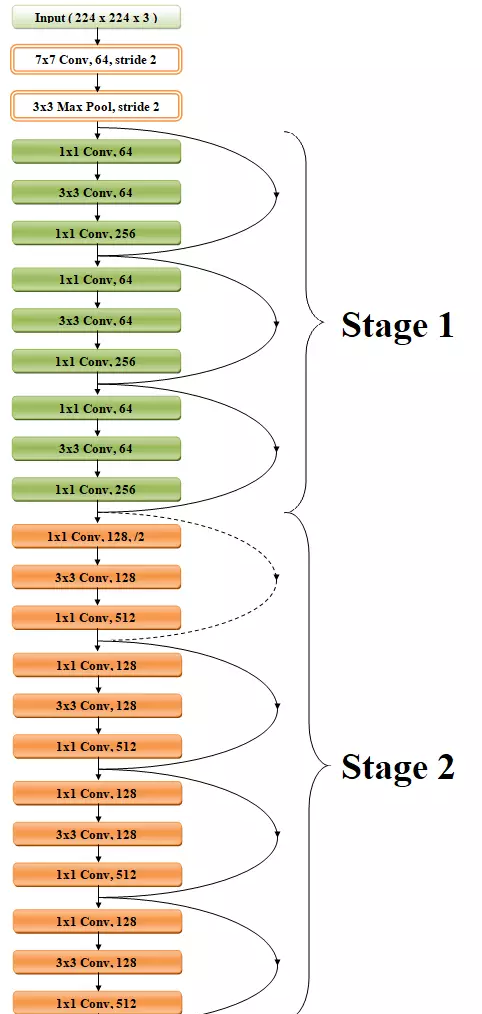


Рис. 4 Архитектура нейронной сети ResNet-50.

# Практическая часть

## Выполнение процесса идентификации

Конфигурация тестовой системы: CPU = intel i3 9100F, GPU = NVIDIA RTX 2060, RAM = 16 GB.

Транслирование наборов данных из различных форматов в облако точек.

В связи с тем, что программные системы применяют разные форматы хранения 3D изображений, вначале их следует приводить в один формат данных.

Для этих целей разработан набор скриптов. На последнем шаге приводящие данные в формат PLY. Применяется язык программирования Python версии 3.7 с библиотеками для научных вычислений (NumPy, SciKit-learn) и реконструирования 3D-изображений (PCL).

**Шаг 1.** Применение набора скриптов для преобразования данных в формате PLY из других форматов, таких как (OBJ, WRL, ABS).

**Шаг 2.** Преобразование PLY облаков точек в двоичный формат NPY, содержащий представление 2D проекции облака точек на плоскость XY.

Для этого используются файлы Preprocessing.py и XYProjection.py

Для каждого изображения происходит вычитание значения средних параметров значения красной компоненты (R), зелёной компоненты (G), синей компоненты (B) значения цвета.

На вход нейронной сети передаются данные в двоичном формате, получаемые из файлов облаков точек PLY.

**Шаг 3.** Выполнение идентификации с помощью файла Recognition.py

* Считываются значения из всех бинарных данных, представленных в директориях Probe и Gallery, разделяющие в соответствующие группы.
* Данные для каждого изображения подаются на вход нейронной сети.
* Выходы нейронных сети обеих групп сравниваются друг с другом с использованием меры косинусного расстояния.
* Наиболее близкое к единице значение выдаётся за результирующее для данного файла из Gallery.

## Реализация процесса дообучения.

Использование нейронной сети с фремворком глубинного обучения Caffe предполагает возможность реализации процесса дообучения (transfer learning) модели.

Учитывая, что дообучение производится для улучшения результатов предсказания в задаче идентификации (в которой нейронной сети нужно уметь правильно соотносить изображения одного человека к одному классу), производилась следующая разметка данных.

Из каждого набора данных для изображений одного объекта идентификации присваивался уникальный идентификатор. Идентификаторы не пересекаются.

# Заключение

В процессе выполнения дипломной работы был выработан план и реализованы компоненты системы распознавания 3D лиц людей с помощью нейронной сети. Удалось выполнить сбор данных из открытых источников, провести предварительную обработку части данных, а также адаптировать для применения в обучении нейронной сети.

Возможно продолжение работы в виде более полного обучения нейронной сети.