

Санкт-Петербургский государственный университет

***ТИМОФЕЕВ Александр Валентинович***  
**Выпускная квалификационная работа**  
***Распознавание лиц с окклюзией***

Направление 02.04.02

«Фундаментальная информатика и информационные технологии»

ООП «Цифровые технологии и системы»

Научный руководитель:  
доцент кафедры КММС,  
к.т.н Гришкин Валерий  
Михайлович

Рецензент:  
Космачев Валентин Михайлович

Санкт-Петербург

2022

# СОДЕРЖАНИЕ

Введение.....	3
Постановка задачи .....	5
Обзор литературы .....	6
Глава 1. Наборы данных.....	8
1.1 Labeled Faces in the Wild.....	8
1.2 Masked LFW.....	8
Глава 2. Свёрточные нейронные сети.....	10
2.1 Свёрточный слой.....	10
2.2 Слой субдискретизации.....	10
2.3 Слой активации .....	11
2.4 Residual block.....	11
2.5 Inception module.....	12
2.6 Depthwise Separable Convolution.....	13
Глава 3. Архитектуры свёрточных нейронных сетей .....	14
3.1 AlexNet .....	14
3.2 VGG16 .....	15
3.3 ResNet.....	16
Глава 4. Модель распознавания лиц с окклюзией .....	19
4.1 Используемые инструменты.....	19
4.2 Построение модели.....	19
4.3 Оценка результатов работы модели.....	22
Выводы.....	24
Заключение .....	25
Список использованных источников .....	26

## Введение

Актуальность задачи распознавания человека не вызывает сомнений. Системы распознавания используют для контроля доступа на территорию, различного контроля перемещений и посещений. Также распознавание может использоваться для обеспечения безопасности и предотвращения совершения противоправных действий.

Распознавание лиц является одним из способов идентификации человека, наравне с распознаванием отпечатков пальцев, глаз и вен. Но в отличие от этих способов распознавание лица является наиболее не навязчивым способом идентификации. Также распознавание при помощи лица можно делать на расстоянии, что позволяет распознать человека без его непосредственного участия в процессе распознавания.

Как задача, распознавание лиц развивается уже несколько десятилетий, на ранних этапах распознавание строилось на простых изображениях, где была одинаковая освещенность, четкое изображение и было полностью видно лицо человека. В реальных условиях такие системы, часто показывали результаты намного хуже, чем при тестировании их на подготовленных наборах данных. Благодаря этому появились новые наборы задач, которые бы решали часть возникающих проблем при распознавании в реальных условиях. Такими задачами являются: распознавание лица в различных позах, при различном освещении, при низком качестве изображения, при различных эмоциях и окклюзии.

Распознавание лиц с окклюзией является одной из самых сложных задач, так как на изображении отсутствует информация о скрытых частях лица. Также на практике очень сложно собрать базу данных лиц, которая бы покрывала все возможные окклюзии, что делает данную задачу еще более не тривиальной.

На сегодняшний день нейронные сети чаще всего используют для распознавания лиц, так как при помощи них был достигнут наибольший прогресс в распознавании.

## **Постановка задачи**

В данной работе стоит задача распознавания лиц с окклюзией. Входными данными модели распознавания является база данных лиц, которые алгоритм сможет распознать. Предполагается, что детектирование лица с окклюзией уже произошло и лицо было центрировано. На вход подается изображение с единственным центрированным лицом. В результате должно происходить распознавание лица и должно возвращаться имя человека на изображении.

## Обзор литературы

В статье [1] приведен обзор современных техник детектирования и распознавания лиц с окклюзией. Авторы рассматривают три различных метода распознавания лиц. Метод извлечения надежных признаков окклюзии (occlusion robust feature extraction) состоит в поиске наименее пострадавших от окклюзии признаков лица и проведении распознавания на основе этих признаков. В методе распознавания лица со знанием окклюзии (occlusion aware face recognition) предполагается, что известно, где находится окклюзия, можно ее отбросить и работать только с видимыми частями. Стратегия метода реконструкции лица на основе восстановления окклюзии (occlusion recovery based face recognition) состоит в том, чтобы восстановить лицо без окклюзии и в дальнейшем использовать обычные алгоритмы распознавания лиц. Также в статье приводятся сравнительные оценки работы рассматриваемых подходов.

В статье [2] проводится сравнение несколько подходов, применимых для распознавания лиц с окклюзией. Рассматриваемыми методами являются: метод главных компонент, анализ основных компонент ядра, обобщенный дискриминационный анализ, метод опорных векторов и сеть радиально-базисных функций. Авторы сравнивают результаты этих подходов, для различных изображений, где скрыты наиболее важные для распознавания лица элементы, такие как глаза, брови, рот. Полученные в данной работе результаты показывают, что для подавляющего большинства методов основными элементами распознавания лица являются область глаз и бровей. В результате экспериментов лучше всего в распознавании себя показали метод опорных векторов и сеть радиально-базисных функций, а хуже всего себя показали метод главных компонент и анализ основных компонент ядра.

В статье [3] авторы предлагают решение задачи распознавания лиц с окклюзией при помощи сверточной нейронной сети. В исследовании предложена архитектура нейронной сети MaskNet основанной на ResNet для решения поставленной задачи и проведено сравнения результатов с

результатами других нейронных сетей. Архитектура сверточной нейронной сети основана на идее, чтобы присваивать более высокие значения весам, активируемым незакрытыми частями лица и более низкие значения тем, которые активируются скрытыми частями лица. Таким образом авторы данной работы добились точности распознавания в 91,6%.

# Глава 1. Наборы данных

## 1.1 Labeled Faces in the Wild

Одной из самых популярных баз данных для распознавания лиц является Labeled Faces in the Wild (LFW) [4]. В данной базе 13 233 изображения 5 749 человек, собранных из интернета, которые были обнаружены и центрированы детектором лиц Виолы-Джонса. Также 1 680 из изображенных людей имеют две или более разных фотографии в наборе данных. На рис. 1 представлены примеры изображений из набора LFW.



Рис. 1. Примеры лиц, представленных в LFW [4]

## 1.2 Masked LFW

Набор данных Masked LFW [5] построен на основе базы данных Cross-Age LFW, при помощи автоматизированного инструмента создания лиц в масках из лиц без масок. Также для генерации использовались различные шаблоны масок, охватывающие большинство распространённых стилей, появляющихся в повседневной жизни, для получения различных эффектов генерации. Данный набор данных часто используется для сравнения



различных алгоритмов распознавания лиц с окклюзией. На рис. 2 представлены примеры изображений из набора Masked LFW.



Рис. 2. Примеры лиц из MLFW [5]

## Глава 2. Свёрточные нейронные сети

Наилучшие результаты в области распознавания лиц показывают свёрточные нейронные сети. Свёрточная нейронная сеть состоит из разных видов слоев, идущих друг за другом. Существует огромное количество различных архитектур нейронных сетей, но принципы их построения и слои из которых они состоят остаются прежними.

### 2.1 Свёрточный слой

Свёрточный является основным слоем нейронной сети и представляет из себя результат применения операции свёртки к выходам предыдущего слоя. Ядро свёртки «скользит» по исходным данным, выполняя поэлементное умножение и затем суммирование результатов произведения, а также запись в один пиксель выхода. На рис. 3 представлен результат операции свёртки.

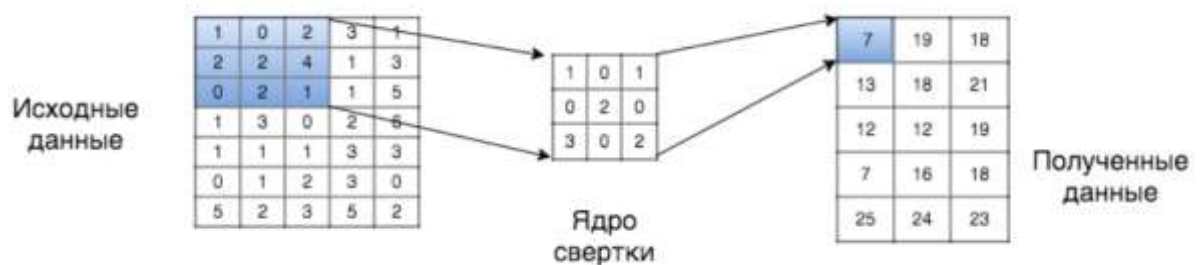


Рис 3. Операция свёртки

### 2.2 Слой субдискретизации

Слой субдискретизации служит для снижения размерности входного изображения. При его применении исходное изображение делится на равные блоки размером  $n \times m$  и для каждого из этих блоков вычисляется некоторая функция. Чаще всего такой функцией является максимум или взвешенное среднее. На рис. 4 изображен пример операции субдискретизации.



Рис. 4. Операция субдискретизации

### 2.3 Слой активации

Слой активации представляет собой поэлементное применение функции активации к входному изображению, где функция активации – некая несложная нелинейная функция. Пример применения функции активации изображен на рис. 5.



Рис. 5. Функция активации

### 2.4 Residual block

При обучении глубоких нейронных сетей часто можно столкнуться с серьёзной проблемой в виде исчезающего градиента. Эта проблема возникает потому, что при дифференцировании по цепочке до последних слоев нейронной сети доходит очень маленькая величина градиента из-за

многократного умножения на маленькие величины. Для борьбы с данной проблемой был придуман residual block [6]. Принцип его работы заключается в том, чтобы взять пару слоёв и добавить проходящую мимо них дополнительную связь. На рис. 6 изображено устройство residual block.

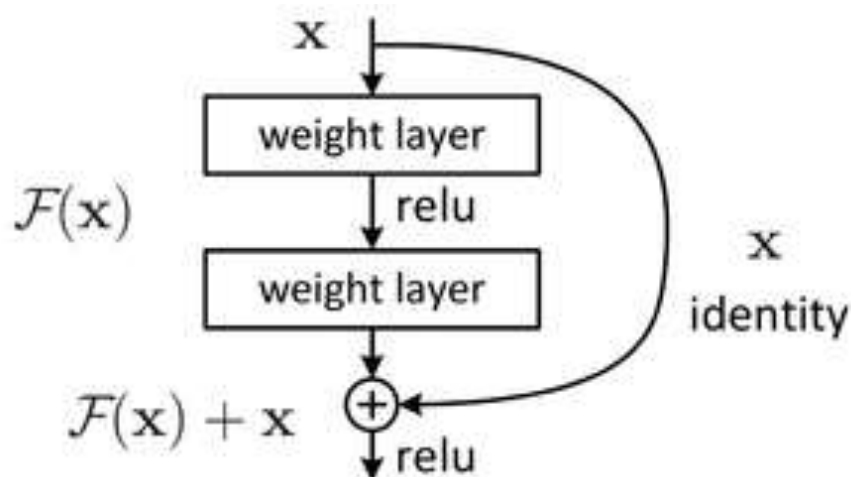


Рис. 6. Устройство residual block

## 2.5 Inception module

При работе со свёрточными нейронными сетями, часто возникает вопрос выбора размера ядра свертки. Для решения данного вопроса была предложена архитектура Inception module [7]. Идея этого подхода заключается в использовании сразу нескольких слоев свертки с различными размерами ядер и последующей конкатенацией их. Но такой подход приводит к увеличению операций для активации одного слоя. Авторы решили эту проблему, добавив дополнительно к каждому фильтру слой свертки  $1 \times 1$ , который снижает размерность поступающего сигнала. На рис. 7 изображено устройство Inception module.

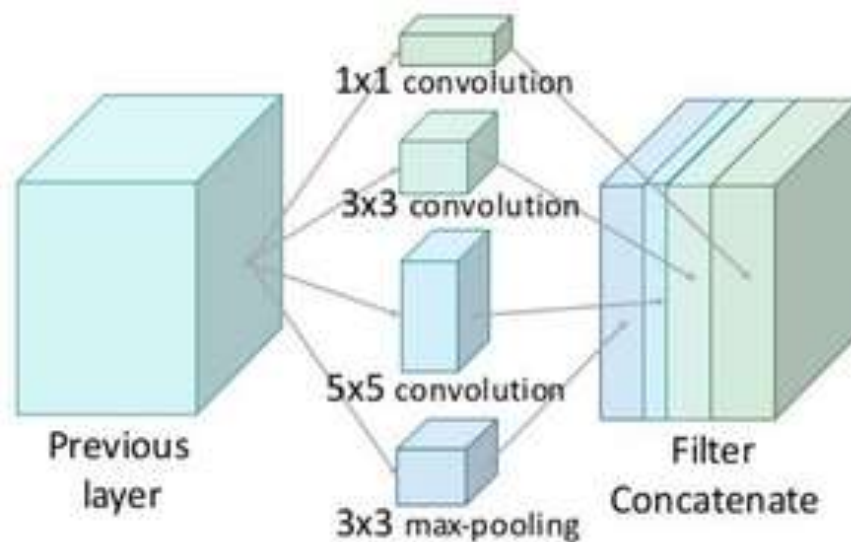


Рис. 7. Устройство Inception module

## 2.6 Depthwise Separable Convolution

Идеологическим продолжением Inception Module является идея Depthwise Separable Convolution [8]. Идея данного подхода заключается в том, чтобы на первом этапе над входным изображением производилась свёртка с ядром  $1 \times 1$  и затем к каждому из каналов применялась свёртка с ядром  $3 \times 3$ . Такая архитектура позволяет сделать нейронную сеть более компактной. На рис. 8 изображено устройство блока Depthwise Separable Convolution.

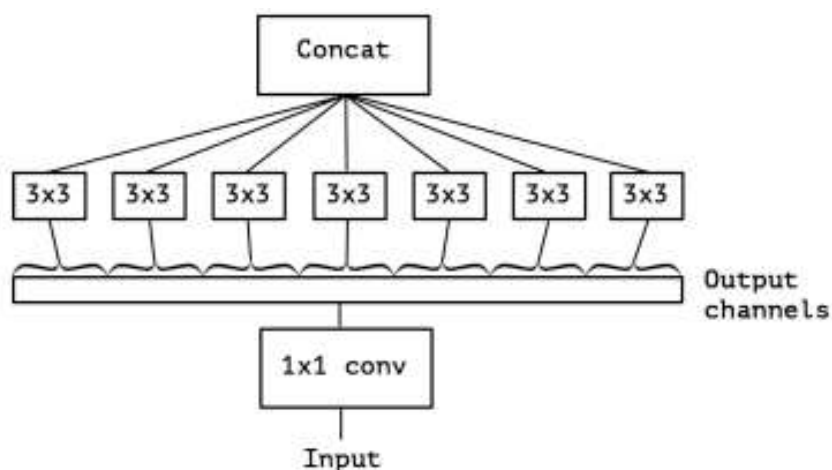


Рис. 8. Устройство блока Depthwise Separable Convolution

## Глава 3. Архитектуры свёрточных нейронных сетей

Все выше описанные слои можно комбинировать друг с другом для создания и улучшения свёрточной нейронной сети. Но на практике создать свою архитектуру нейронной сети, которая бы удовлетворяла нашим потребностям с точки зрения качества работы достаточно сложно. Поэтому рассмотрим наиболее популярные и часто используемые архитектуры для распознавания изображений.

### 3.1 AlexNet

AlexNet [9] – свёрточная нейронная сеть, которая оказала огромное влияние на алгоритмы распознавания изображений. Архитектура нейронной сети представлена на рис. 9. Как видно на изображении, она состоит из восьми уровней. Пять слоев являются свёрточными, оставшиеся три являются полносвязными. Выходные данные пропускаются через функцию softmax, которая формирует распределение 1000 меток класса. Вместо функции гиперболического тангенса (tanh), которая была стандартом в то время, в качестве функции активации была взята функция ReLU.

$$ReLU(x) = \max(0, x)$$

Выбор данной функции обусловлен простотой ее вычисления, а также простотой вычисления ее производной, что дает прирост к скорости обучения нейронной сети. Также, благодаря тому, что сеть разделена на несколько частей, можно проводить обучение нейронной сети на нескольких графических процессорах, что дает прирост к скорости обучения.

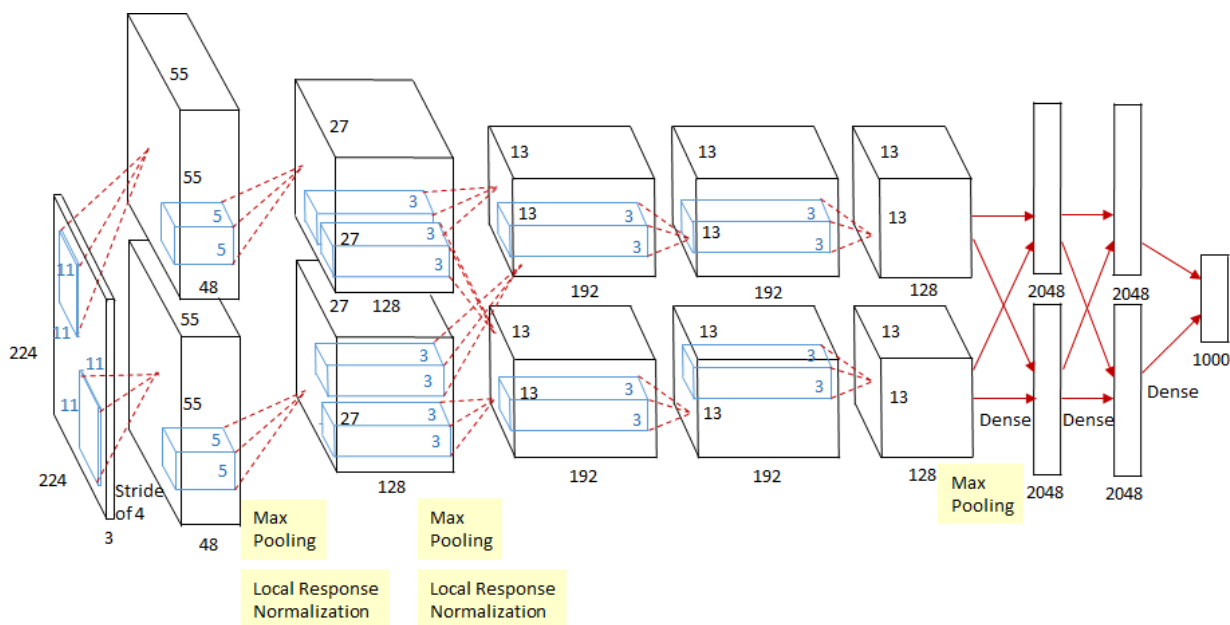


Рис. 9. Архитектура AlexNet [9]

## 3.2 VGG16

Улучшением AlexNet, является модель свёрточной нейронной сети VGG16 [10]. Архитектура нейронной сети представлена на рис. 10. В данной сети заменены фильтры в первом и втором свёрточных слоях размерами 11x11 и 5x5 на несколько фильтров размером 3x3, идущих друг за другом. Данная замена позволяет уменьшить количество обучаемых переменных, что означает более быстрое обучение. Таким образом, архитектура представляет собой множество фильтров с маленьким рецептивным полем размером 3x3.



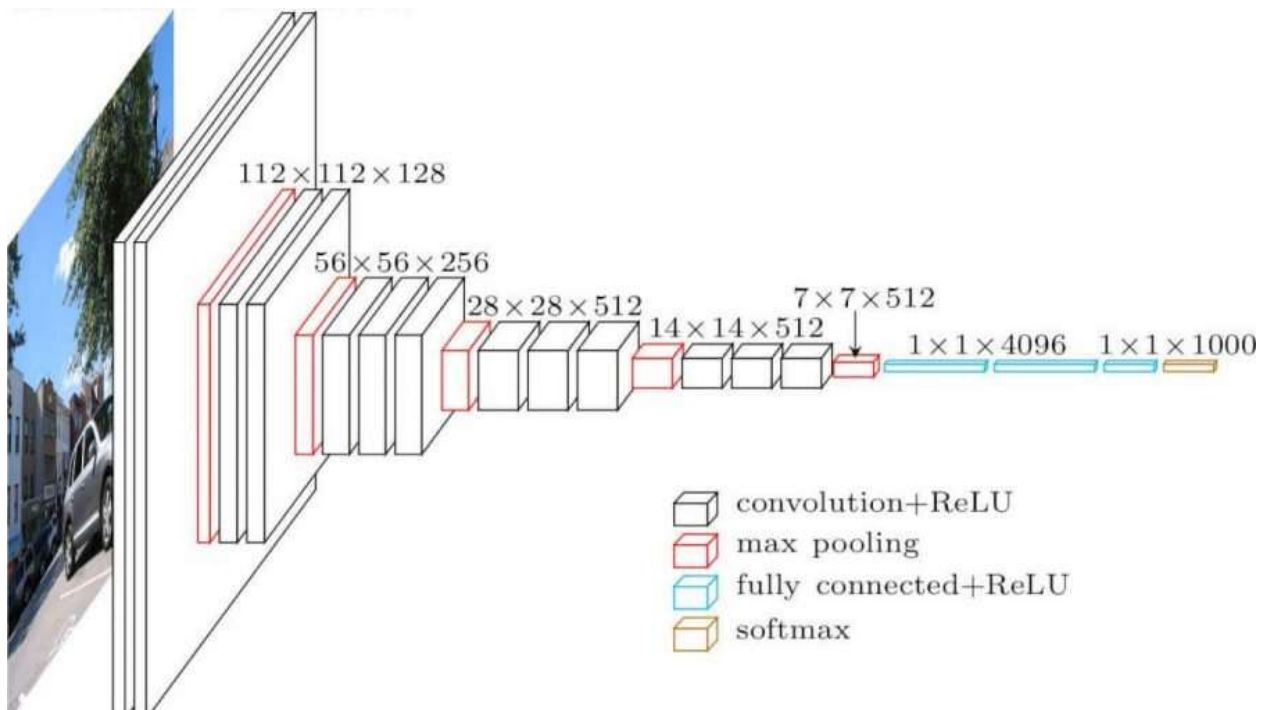


Рис. 10. Архитектура VGG16 [10]

### 3.3 ResNet

В последние годы нейронные сети становятся все более глубокими, так как чем больше слоев у нейронной сети, тем более сложную функцию она может представить. Но при простом добавлении слоев возникнет проблема затухания градиента, описанная выше. Благодаря этому можно получить такой результат, что при увеличении слоев нейронной сети, ошибка распознавания будет увеличиваться. На рис. 11 представлено сравнение ошибки для 20-слойной и 50-слойной нейронных сетей при обучении и на тестовых данных.

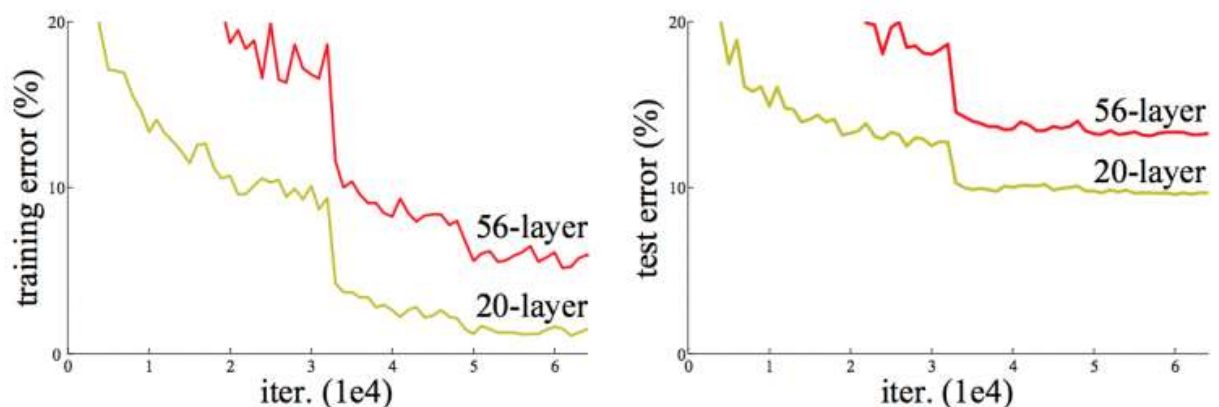


Рис. 11. Сравнение 56-слойной и 20-слойной нейронных сетей [6]



Для того чтобы решить эту задачу, исследователи компании Microsoft представили residual block, рассмотренный выше. Таким образом можно из более простой нейронной сети, например, VGG16 сделать более глубокую нейронную сеть ResNet [6], добавляя дополнительные слои, а также быстрые соединения между слоями (residual block). На рис. 12 представлен пример преобразования VGG16 в ResNet.

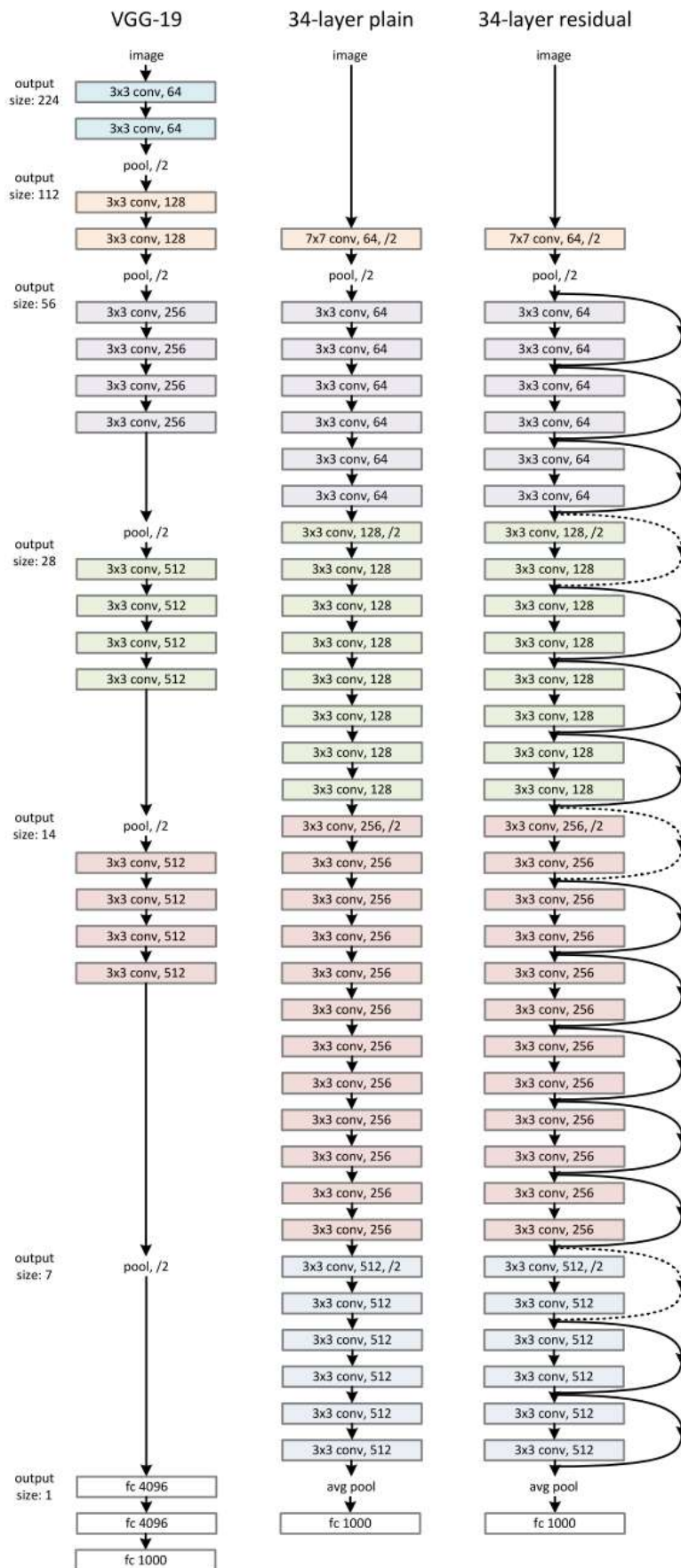


Рис. 12. Пример построения ResNet из VGG16 [6]

## Глава 4. Модель распознавания лиц с окклюзией

### 4.1 Используемые инструменты

Обучение нейронной сети производилось при помощи языка Python 3 в онлайн сервисе Google Colab, который предоставляет возможность удаленного запуска Jupiter Notebook на удаленных машинах компании Google. Для реализации и обучения нейронных сетей использовалась библиотека PyTorch [11].

### 4.2 Построение модели

В процессе работы за основу была взята архитектура ResNet152, так как она представляет собой наиболее продвинутую архитектуру из представленных для распознавания изображений. В ходе работы была обучена предобученная модель на наборах данных LFW и Masked LFW. Точность распознавания данной модели составила 0,93 и 0,75 соответственно. На рис. 13, 14 представлены изменения функции потерь и точности во время обучения для соответствующих баз данных.

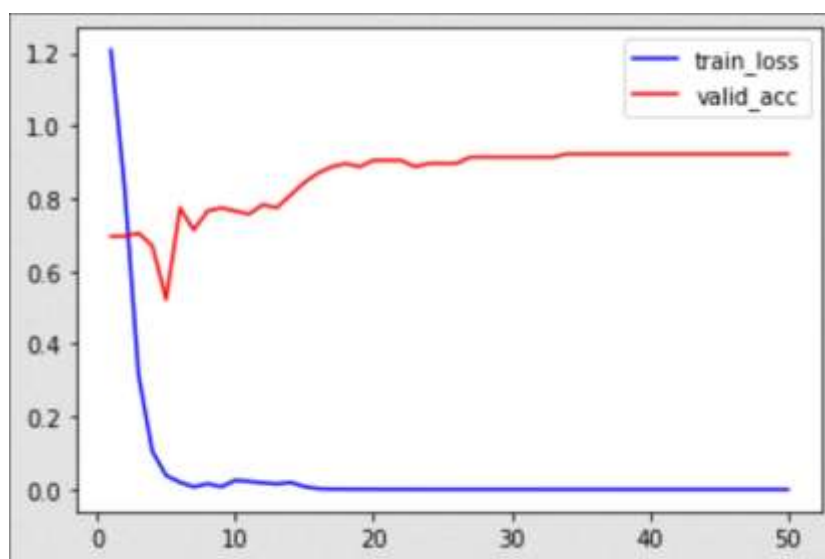
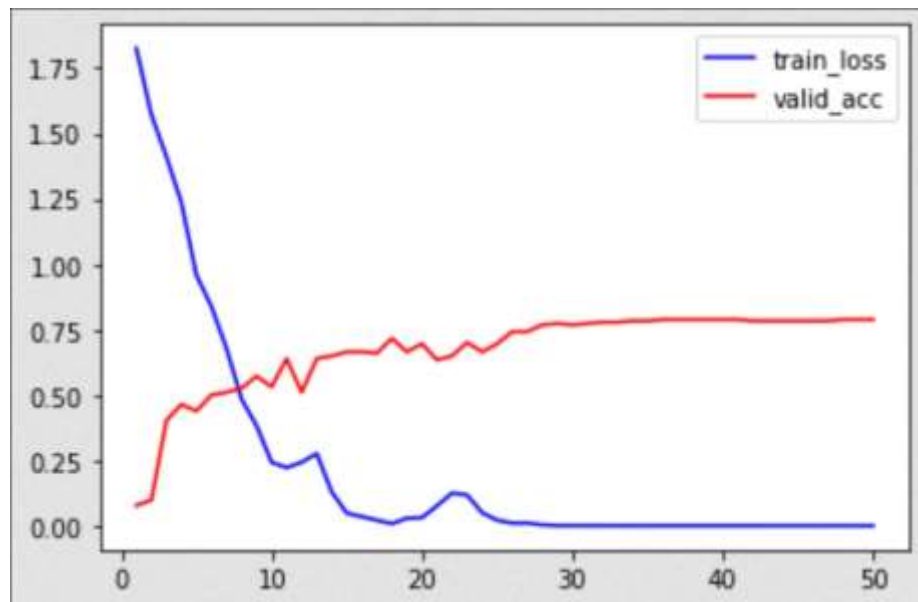
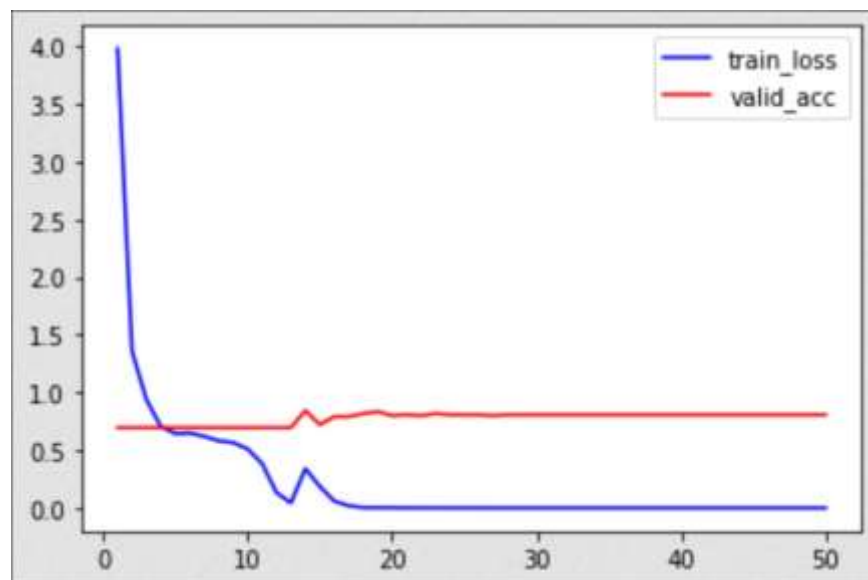


Рис. 13. Изменение функции потерь и точности во время обучения (LFW)

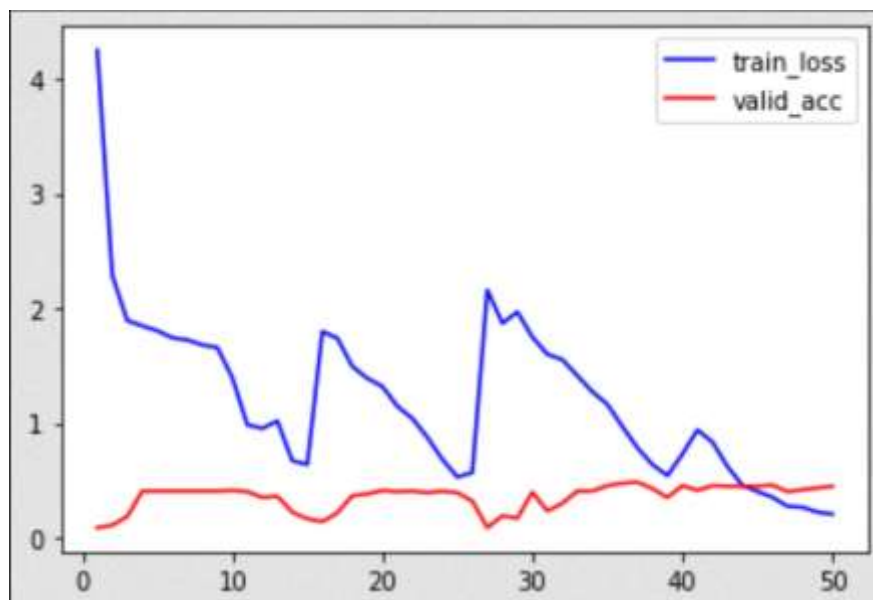


**Рис. 14.** Изменение функции потерь и точности во время обучения (MLFW)

Также, модель на основе архитектуры ResNet152 была обучена с нуля. На рис. 15, 16 представлены изменения функции потерь и точности во время обучения, полученные в ходе работы данной модели. Точность распознавания на наборах данных LFW и MLFW составила 0,83 и 0,44 соответственно.



**Рис. 15.** Изменение функции потерь и точности во время обучения (LFW)



**Рис. 16.** Изменение функции потерь и точности во время обучения (MLFW)

На основе полученных результатов было принято решение, что улучшение модели будет происходить на предобученной нейронной сети, так как там была достигнута большая точность распознавания. К предобученной модели добавлены три слоя свертки 3x3 идущих друг за другом, дальше идет линейное преобразование над входящими данными и за ним идет функция dropout, которая «исключает» случайно выбранные нейроны из сети, для избежания проблемы переобучения. На рис. 17 представлена архитектура полученной нейронной сети.

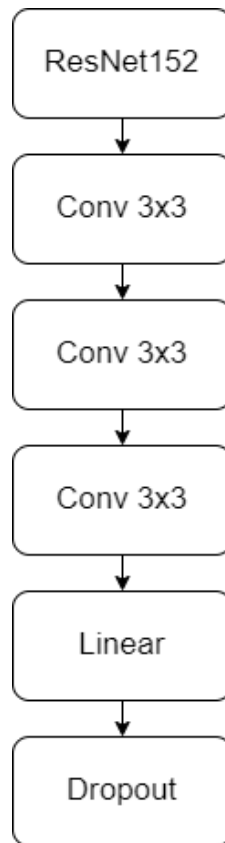
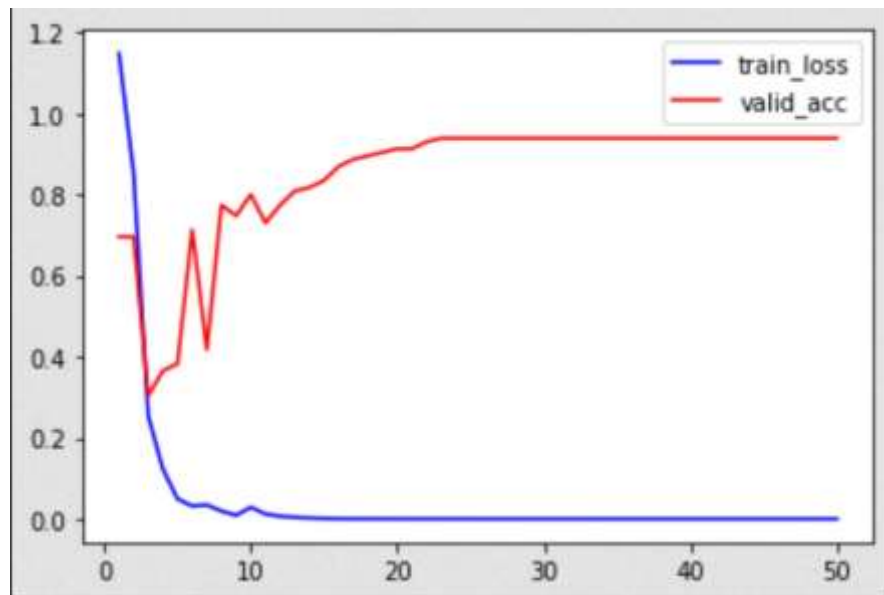


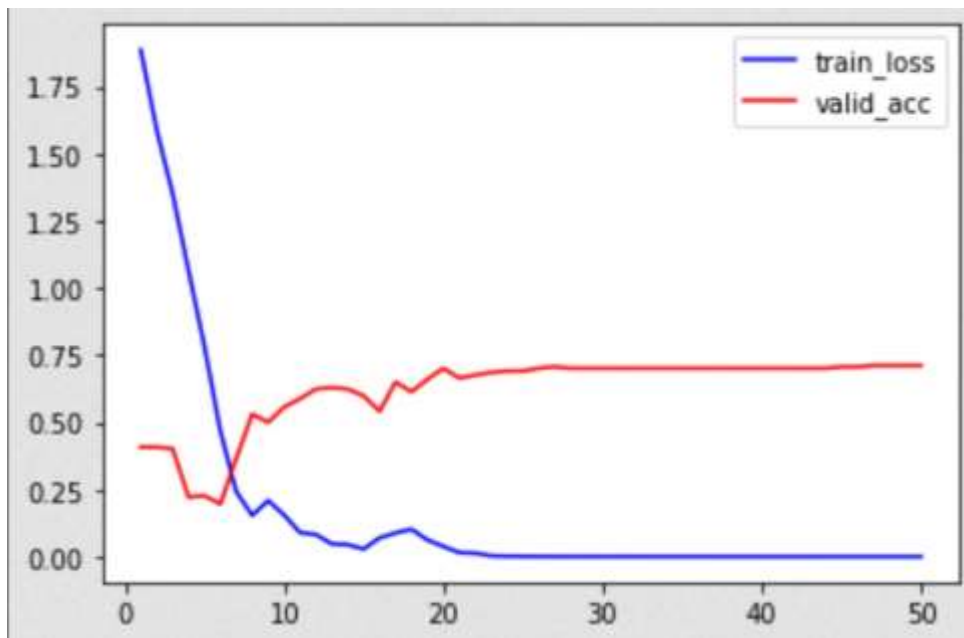
Рис. 17. Архитектура улучшенной модели

### 4.3 Оценка результатов работы модели

Модифицированная модель достигла точности распознавания 0,94 на LFW и 0,77 на Masked LFW, что позволяет сказать, что улучшение модели позволило увеличить точность распознавания. На рис. 18, 19 представлены изменения функции потерь и точности во время обучения, полученные в ходе работы улучшенной модели.



**Рис. 18.** Изменение функции потерь и точности во время обучения (LFW)



**Рис. 19.** Изменение функции потерь и точности во время обучения (MLFW)

## **Выводы**

В результате данной работы были получены следующие результаты:

- были обучены предобученная и обученная с нуля модели ResNet152, на основе полученных результатов была выбрана модель для улучшения;
- была разработана модификация ResNet152 с добавлением новых слоев;
- была проведена оценка полученных результатов распознавания.



## **Заключение**

В ходе данной работы были выбраны и подготовлены наборы изображений лиц с окклюзией и без. Также были рассмотрены современные подходы к распознаванию изображений и была выбрана модель для улучшения. На основе архитектуры ResNet152 была построена модификация с добавлением новых слоев. В итоге построенная модель позволила улучшить точность распознавания лиц с окклюзией и без относительно сети ResNet.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. D. Zeng, R. Veldhuis, and L. Spreewers. A survey of face recognition techniques under occlusion. IET Biometrics, pages 1–23, 2020, doi: arXiv:2006.11366
2. B. Jozer, F. Matej, O. Luboš, O. Miloš and P. Jarmila, "Face recognition under partial occlusion and noise," Eurocon 2013, 2013, pp. 2072-2079, doi: 10.1109/EUROCON.2013.6625266.
3. W. Wan and J. Chen, "Occlusion robust face recognition based on mask learning," 2017 IEEE International Conference on Image Processing (ICIP), 2017, pp. 3795-3799, doi: 10.1109/ICIP.2017.8296992.
4. Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments"
5. Chengrui Wang, Han Fang, Yaoyao Zhong, Weihong Deng, "MLFW: A Database for Face Recognition on Masked Faces", doi: arXiv:2109.05804
6. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, "Deep Residual Learning for Image Recognition", doi: arXiv:1512.03385
7. Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, "Going Deeper with Convolutions", doi: arXiv:1409.4842
8. François Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions", doi: arXiv:1610.02357
9. Alex Krizhevsky, "One weird trick for parallelizing convolutional neural networks", doi: arXiv:1404.5997
10. Karen Simonyan, Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", doi: arXiv:1409.1556
11. Документация библиотеки PyTorch [Электронный ресурс], // URL: <https://pytorch.org> (дата обращения 28.05.2022).

12. Y. Li, J. Zeng, S. Shan and X. Chen, "Occlusion Aware Facial Expression Recognition Using CNN With Attention Mechanism," in *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2439-2450, May 2019, doi: 10.1109/TIP.2018.2886767.

13. K. V. Arya and Anukriti, "Face recognition under partial occlusion using HMM and Face Edge Length Model," 2014 9th International Conference on Industrial and Information Systems (ICIIS), 2014, pp. 1-6, doi: 10.1109/ICIINFS.2014.7036574.

14. L. Mao, F. Sheng and T. Zhang, "Face Occlusion Recognition With Deep Learning in Security Framework for the IoT," in *IEEE Access*, vol. 7, pp. 174531-174540, 2019, doi: 10.1109/ACCESS.2019.2956980.