Санкт-Петербургский государственный университет

**Кафедра информационных систем**

**Кисляков Владислав Сергеевич**

**Магистерская диссертация**

**Идентификация человека по биометрическим данным**

Направление 01.04.02

Прикладная математика и информатика

Магистерская программа Математическое моделирование в задачах естествознания

Научный руководитель,

доктор физ.-мат. наук,

профессор

Матросов Александр Васильевич

Санкт-Петербург

2017

Содержание

Введение 3

Постановка задачи 4

Глава 1. Теоретическое обоснование 5

* 1. Нейронные сети 5
  2. Свёрточные нейронные сети 6
  3. Методы предварительной обработки изображений 9
  4. Оценка точности сети 10
  5. Полный цикл идентификации 12

Глава 2. Описание исследуемых сетей 13

2.1 VGG 13

2.2 GoogleNet 13

2.3 Inception v3 15

2.4 ResNet 15

Глава 3. Методика исследования 17

3.1 Описание набора данных и подготовка его к использованию 17

3.2 Обработка ответов сети 17

Глава 4. Результаты анализа сетей 19

4.1 Результаты 19

4.2 Вывод 23

Заключение 24

Список литературы 25

**Введение**

Сверточные нейронные сети были наиболее успешным подходом к проблеме классификации изображений за последнее десятилетие. В последнее время несколько предварительно обученных сетей были опубликованы и стали открытыми. Все эти сети прошли обучение по большому количеству изображений лиц, собранных с онлайн-ресурсов, таких как the Internet Movie Database и результаты поиска Google.

Применение сверточных сетей не ограничивается решением задачи классификации. Так же им нашли применение в задачах сегментации изображений, что косвенно является задачей классификации, и идентификации человека по изображению лица [1].

Задача идентификации человека по биометрическим данным актуальна в последние годы. Распознавание личности по изображению лица является одним из наиболее удобных способов идентификации, так как не требует близкого контакта человека с прибором и может проводиться без ведома объекта анализа, что бывает полезно для различных охранных структур. Так же существуют методы идентификации по отпечаткам пальцев, голосу, радужке глаза и анализам биоматериалов. Однако, они не обладают в полной мере указанными выше преимуществами [4][16].

Системы идентификации личности часто используются для проверки допуска человека к различным данным или объектам. Для систем, применяемых подобным образом важной характеристикой, является сведение к минимуму вероятности ложноположительных ответов, т.е. предоставление допуска лицам, немеющим такового.

В данной работе будут проанализированы предобученные свёрточные сети, находящиеся в открытом доступе.

Простановка задачи

Решаемая в данной работе задача заключается в проведении сравнения точности работы и количества ложноположительных ответов свёрточных сетей, находящихся в открытом доступе, применяемых в задаче идентификации человека по изображению лица, и выявлении сети, показавшей лучшие результаты.

Для решения задачи требуется:

* подобрать подходящий набор данных, из которого будут составлены пары, ожидающие положительного (пара изображений принадлежит одному человек) и негативного (пара изображений принадлежит разным личностям) результата;
* провести предварительную обработку собранных изображений для передачи сети;
* передать сетям обработанные изображения и получить данные для дальнейшего анализа;
* проанализировать полученные данные для выявления сети с лучшими результатами.

**Глава 1. Теоретическое обоснование.**

* 1. **Нейронные сети**

Нейронные сети — это вычислительные модели, основанные на биологии мозга [1]. Нейронные сети состоят из набора узлов, которые представляют искусственные нейроны. Узлы связаны направленными ребрами. Эти ребра представляют синапсы биологической нейронной сети. Каждый нейрон имеет функцию активации *f*. Обозначим индекс узла, где начинается ребро как *i* и узла, где заканчивается ребро как *j*. Каждое ребро связано с весом , который соответствует направленному ребру между двумя узлами. Значение нейрона рассчитывается по взвешенной сумме значений входных узлов. Этот расчет в [1] описывается как входящая активация, обозначается как и рассчитывается как в формуле 1. Во-вторых, функция активации для этой взвешенной суммы вычисляется как в формуле 2. Пример простого искусственного нейрона, также известного как персептрон, показан на рисунке 1.

Формула 1

Формула 2

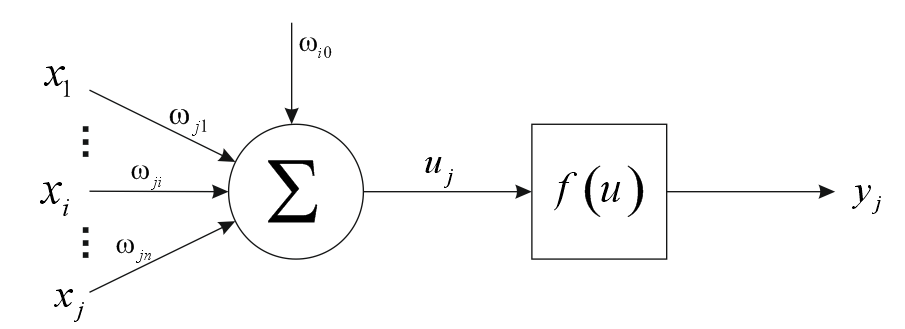


Рисунок 1 Модель нейрона

Нейроны могут формировать слои. Слой – это группа одинаковых по строению нейронов с различными весами, получающая некоторый n-мерный вектор, где n должно соответствовать числу синапсов в каждом нейроне слоя и на выходы отдает вектор размерности, соответствующий числу нейронов в слое. Нейроны слоя не связаны друг с другом [1].

Многослойная сеть прямого распространения объединяет в себе последовательно расположенные слои. Сигнал передается по сети только в прямом направлении, слева направо, от слоя к слою. На рисунке 2 изображен на архитектура многослойной сети с тремя слоями. Выходные нейроны составляют выходной слой сети. Остальные нейроны относятся к скрытым слоям. Скрытые слои не являются частью входа или выхода сети. Первый скрытый слой получает входной сигнал. Результирующий сигнал первого скрытого слоя, в свою очередь, поступает на следующий скрытый слой, и т.д., до конца сети [3].

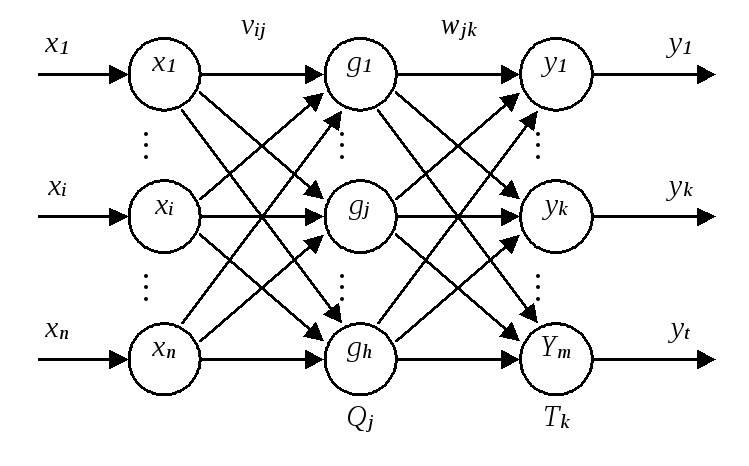


Рисунок 2 Архитектура многослойной сети

* 1. **Свёрточные нейронные сети**

Свёрточные сети состоят из различных типов слоев: слои свёртки, слои субдискретизации и слои, представленные в предыдущем параграфе. Пример архитектуры сверточной сети с наглядным расположением слоев представлен на рисунке 3 [5].

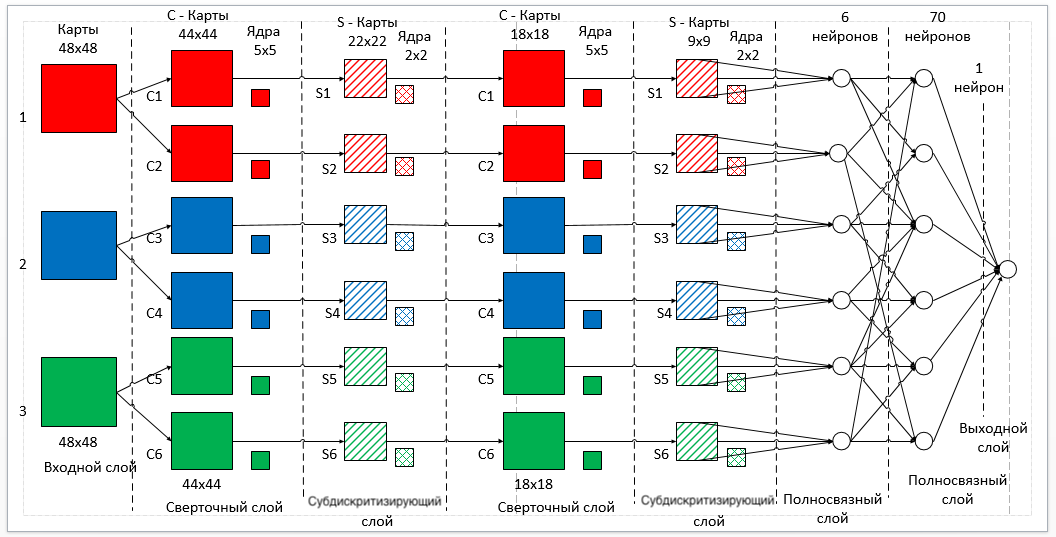


Рисунок 3 Пример архитектуры свёрточной сети

Свёрточный слой является набором карт признаков. Каждая карта имеет ядро или иначе фильтр. Количество карт определяется требованиями к решаемой задаче, при большом количестве карт, повышается качество распознавания до определенного предела, а далее происходит снижение и повышается количество вычисляемых параметров. Исходя из анализа научных статей [4], в большинстве случаев предлагается брать соотношение один к двум, то есть каждая карта предыдущего слоя связана с двумя картами сверточного слоя, в соответствии с рисунком 4.

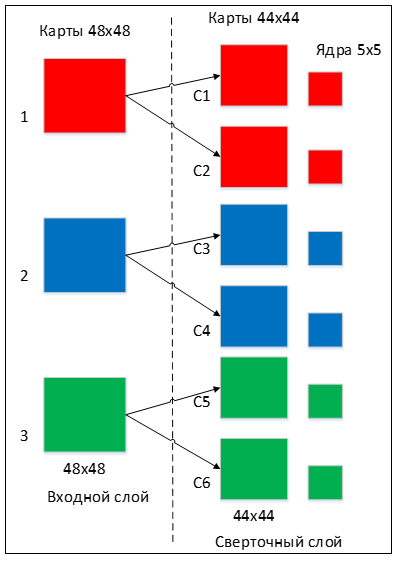
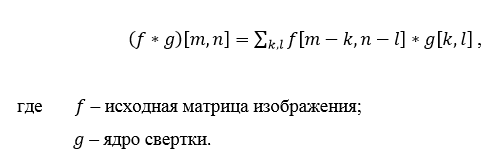


Рисунок 4 Карты сверточного слоя

Ядро является фильтром, который проходит по карте и выявляет заданные в нем признаки. Так, если обучение проводилось на базе данных лиц, тогда ядро могло в процессе обучения выдавать наибольший сигнал в области глаза, рта, брови или носа [9].

Ядро выступает системой разделяемых весов, это одна из основных особенностей сверточных сетей. Многослойные сети имеют много связей между нейронами, что замедляет процесс обучения. Сверточные нейросети в свою очередь за счет общих весов дают возможность уменьшить число связей и позволяют идентифицировать один и тот же признак по всей области изображения [6].

Веса ядер изначально заданы случайным образом, чаще всего в диапазоне [-0.5..0.5]. Ядро проходит по карте и выполняет свертку по формуле 3, где f – исходная матрица изображения, g – ядро свертки, получая новую карту как показано на рисунке 5.



Формула

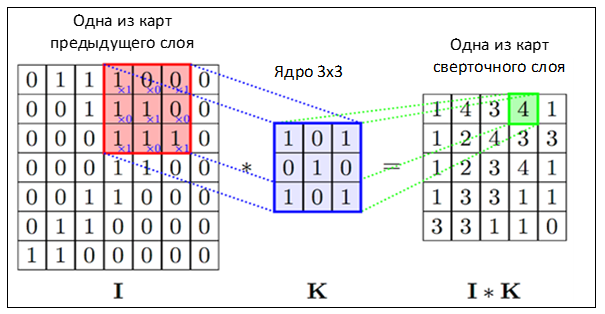
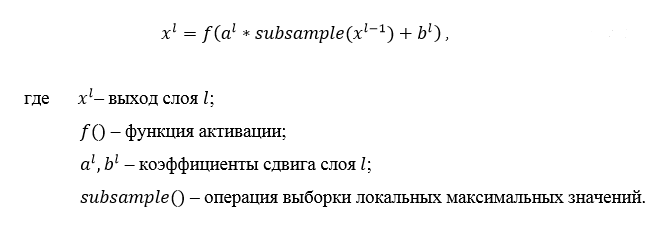


Рисунок 5 Операция свертки

Слой субдискретизации выполняет сжатие изображения проходя по нему фильтром 2x2 и выбирая максимальное значение из попавших в фильтр как показано на рисунке 6, тем самым формируя новое изображение или как в нашем случае карту. Операция можно описать формулой 4, где f – функция активации, a и b -коэффициенты сдвига слоя, subsample – операция выборки локальных максимальных значений.



Формула

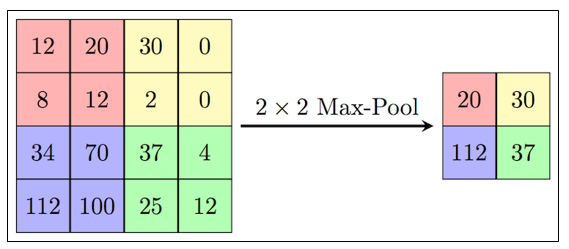


Рисунок 6 Операция субдискретизации

* 1. **Методы предварительной обработки изображений**

Существуют различные способы поиска лиц на изображении, но в данной работе используется метод названный - гистограмма направленных градиентов или сокращенно HOG. Данный метод применим не только для поиска лиц, но и для других объектов. Его суть состоит с в том, чтобы на изображении, к черно-белому виду, заметить каждый пиксель стрелкой направления в котором изображение становится темнее или, иначе говоря, направление градиента. Однако, стрелка на каждом пикселе дает избыточную информацию, с которой трудно работать, поэтому используется подход разбиение изображения на небольшие квадраты, в которых подсчитывается количество каждого направления и выбирается стрелка с максимальным количеством совпадений. Для поиска лица на таком изображении достаточно найти участок, который более всего похож на известную HOG-структуру, полученную из группы лиц, представленную на рисунке 7 [10].

Так же для более качественной идентификации человека по лицу требуется привести каждое изображение к эталонной форме, где глаза, рот и нос будут находиться применено в одинаковых точках. Для этого можно использовать следующий метод. Выделяется 68 специфических точек (меток) (рисунок 8), имеющихся на каждом лице, и имея представление, где находятся глаза, рот и нос, требуется провести только простые преобразования масштабирования и вращения, не приводящие к искажениям изображения, т.е. аффинные преобразования [11]. Пример таких манипуляций приведен на рисунке 9.

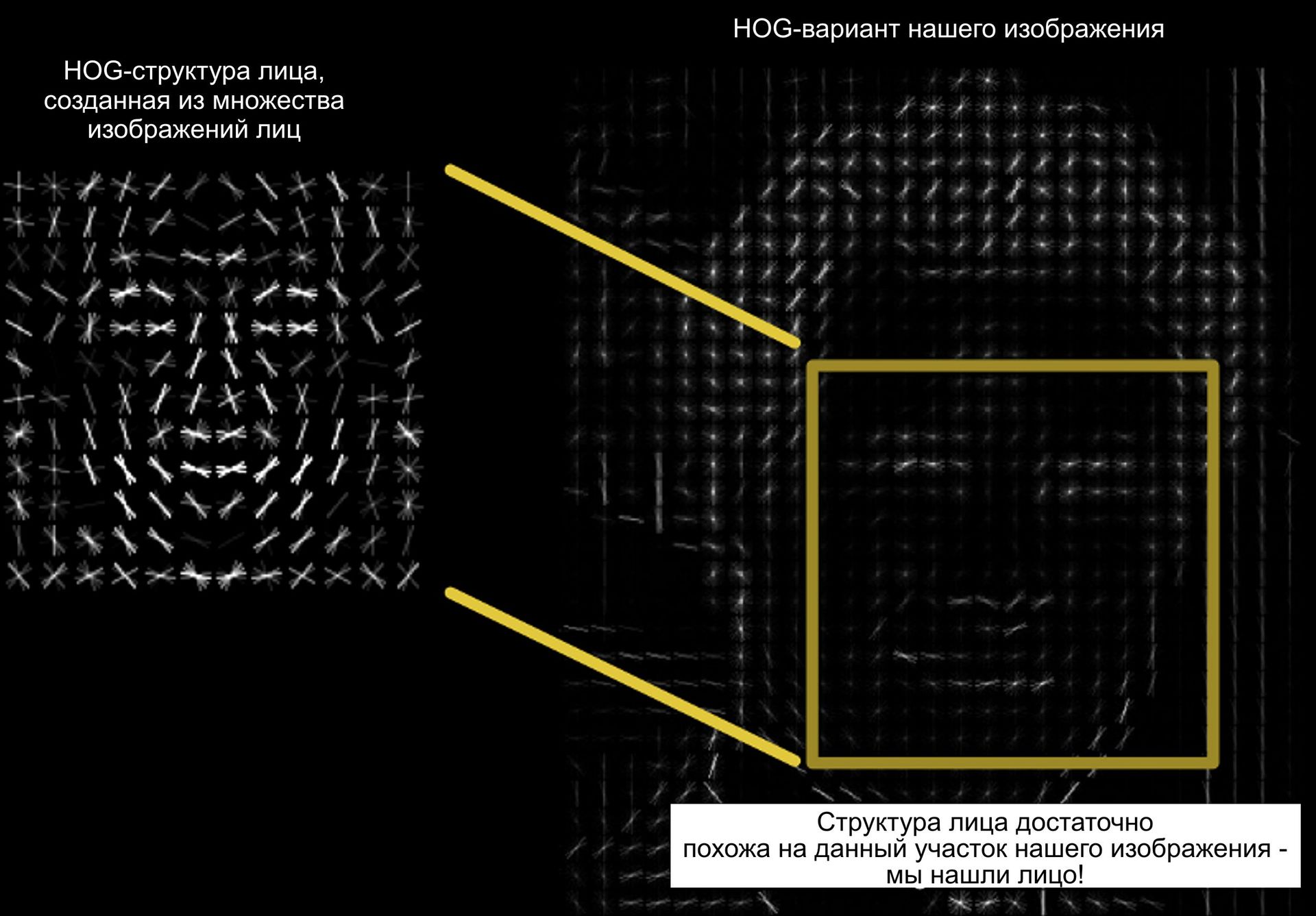


Рисунок 7. HOG-структура лица, созданная из множества изображений лиц

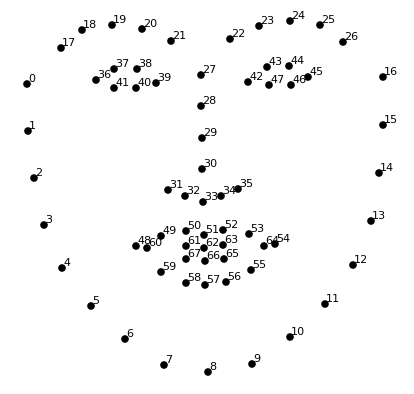


Рисунок 8. 68 антропометрических точек



Рисунок 9. Процесс выравнивания лица на изображении

* 1. **Оценка точности сети**

Простейший способ оценить качество идентификации — это определить долю ответов, на которых сеть дала правильное решений, представлена формулой 5, где где, P – количество правильных решений, а N – размер выборки.

Формула 5

Однако, при сильных сдвигах классов она не предоставляет корректную точность. В таких случаях разумно использовать точность (precision) и полноту (recall). В таблице 1 содержится информация сколько раз система приняла верное и сколько раз неверное решение идентификации. А именно: TP — истиноположительное решение; TN — истиноотрицательное решение; FP — ложноположительное решение; FN — ложноотрицательное решение [14].



Таблица 1

Точность сети идентификации людей– это доля пар изображений, действительно принадлежащих данному одному человек относительно всех пар, которые система посчитала позитивными вариантами. Полнота сети идентификации людей – это доля пар изображений, принадлежащих классу относительно всех пар в являющихся позитивными. Точность рассчитывается по формуле 6, а полнота по формуле 7.

Формула 6

Формула 7

Точность и полноту также можно использовать в качестве базиса для F-метрики. Она является гармоническим средним между точностью и полнотой. Она стремится к нулю, если точность или полнота стремится к нулю [14]. Рассчитывается по формуле 8.

Формула 8

* 1. **Полный цикл идентификации**

Распознавание лиц можно рассматривать как часть конвейера обнаружения-выравнивания-распознавания [2]. Каждая фаза в конвейере генерирует входные данные для следующей фазы, как показано на рисунке 10.

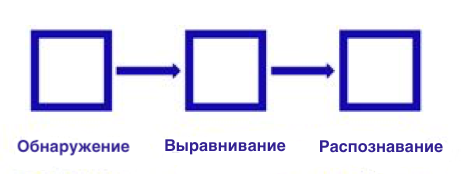


Рисунок 10 Обнаружение-выравнивание-распознавание конвейера. Каждая фаза генерирует вход для следующей фазы. Изображение из [3]

На этапе обнаружения положения лиц на изображении находятся. Обычно самое большое лицо, которое можно найти — это лицо, идентифицированное по этому изображению. Прямоугольная ограничительная рамка самого большого найденного лица затем обрезается по изображению. На этапе выравнивания обрезанное изображение преобразуется в каноническую позу, которая обычно поворачивает и переводит изображение так, чтобы глаза были на уровне, а нос - в центре изображения. Наконец, фаза распознавания включает в себя сопоставление изображения с известной идентификационной информацией или сравнение двух изображений с предположением, принадлежат ли они одному человеку [2][13].

**Глава 2. Описание исследуемых сетей**

**2.1 VGG**

VGG Net — модель свёрточной сети, представленная в [4]. В данной сети отказались от использования ядер размером более, чем 3×3. В [4] показали, что слои с ядрами 7×7 эквивалентены трём слоям с ядрами 3×3, причём это позволяет использовать на 55% меньше параметров. Так же слои с ядрами 5×5 эквивалентны двум слоям с фильтром 3×3, которые экономят 22% параметров сети. Пример подобной декомпозиции представлен на рисунке 11 [5].

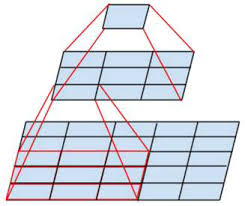


Рисунок 11. Визуальное представление декомпозиции 5x5 в 2 3х3

На соревновании ILSVRC 2014 ансамбль из двух VGG Net получил top-5 ошибку 7,3 %. Хотя данная модель и не победила в соревновании, из-за её простоты она используется в более сложных сетях, предназначенных для детектирования предметов семантической сегментации или маскирования объектов [5].

**2.2 GoogleNet**

GoogleNet продукт компании Google, другое название Inception. В 2014 году сеть заняла первое место на конкурсе ILSVRC, она немного превзошла по точности сетьVGG и в три раза превзошла сеть AlexNet. Архитектура представлена на рисунке 12 [15].

Архитектура сети имеет три особенности. Первой особенностью является то, что сеть имеет не последовательную структуру, а разветвленную в виде ациклического графа. Под ацикличным имеется в виду, что нет обратной связи как в рекуррентных сетях. Вторая особенность — это наличие у сети нескольких выходов. Последний выход используется непосредственно для задач классификации, а промежуточные выходы служат для решения проблемы затухания градиента. И ключевой особенностью является использование Inception блоков, так же представленных на рисунке 12 и описанный далее по тексту**.**

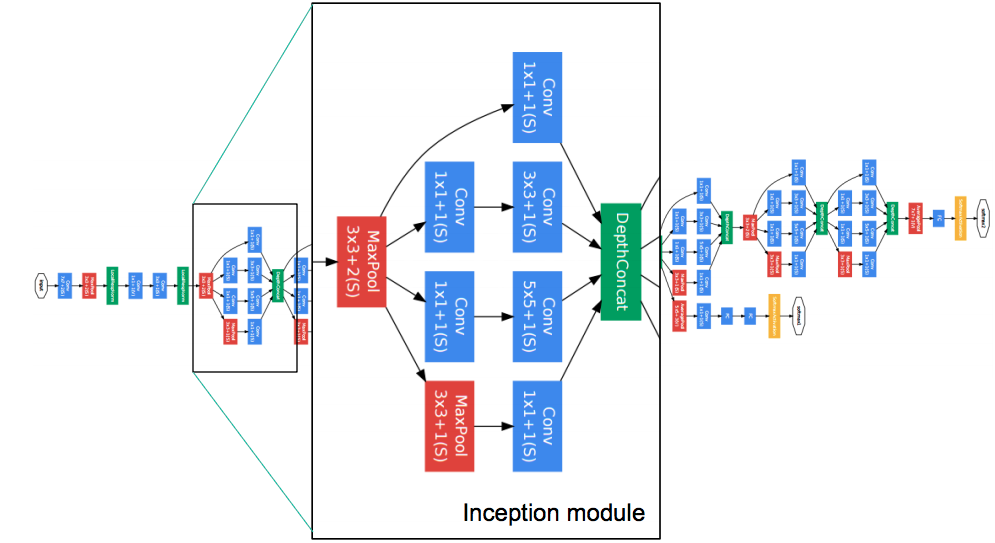


Рисунок 12. Архитектура GoogleNet

Авторы GoogleNet отталкивались от того, что после каждого слоя необходимо сделать выбор каким будет следующий слой: слой свёрткой с ядром 3×3, 5×5, 1×1 или слой субдискретизации. Все представленные слои полезны по-своему. Например, ядро 1×1 выявляет корреляцию между каналами, фильтры большего размера реагируют на более глобальные признаки, а слой субдискретизации позволяет уменьшить размерность без больших потерь информации. Вместо выбора одного конкретного слоя, используются все слои сразу, параллельно друг другу, а затем полученные результаты объединяются в один. Во избежание роста числа вычисляемых параметров, перед каждым слоем свёртки используется свёртка 1×1, которая уменьшает число карт признаков. Такой блок слоёв назвали модулем Inception [19][20]. Также в GoogLeNet отказались от использования полносвязного слоя в конце сети, используя вместо него слой Average Pooling, благодаря чему сильно сократилось число вычисляемых параметров в сети. Таким образом, GoogLeNet, состоящая из более чем ста базовых слоёв, имеет почти в 12 раз меньше параметров, чем AlexNet (≈ 7 миллионов против 138 миллионов) [17][18].

**2.3 Inception v3**

В [24] авторы развили идею Inception (или иначе GoogleNet) в Inception-v3. В данной модели они развили идею декомпозиции фильтров как в VGG, предложив декомпозировать ядро N×N двумя последовательными ядрами 1×N и N×1.

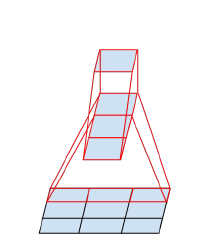


Рисунок . Декомпозиция ядра

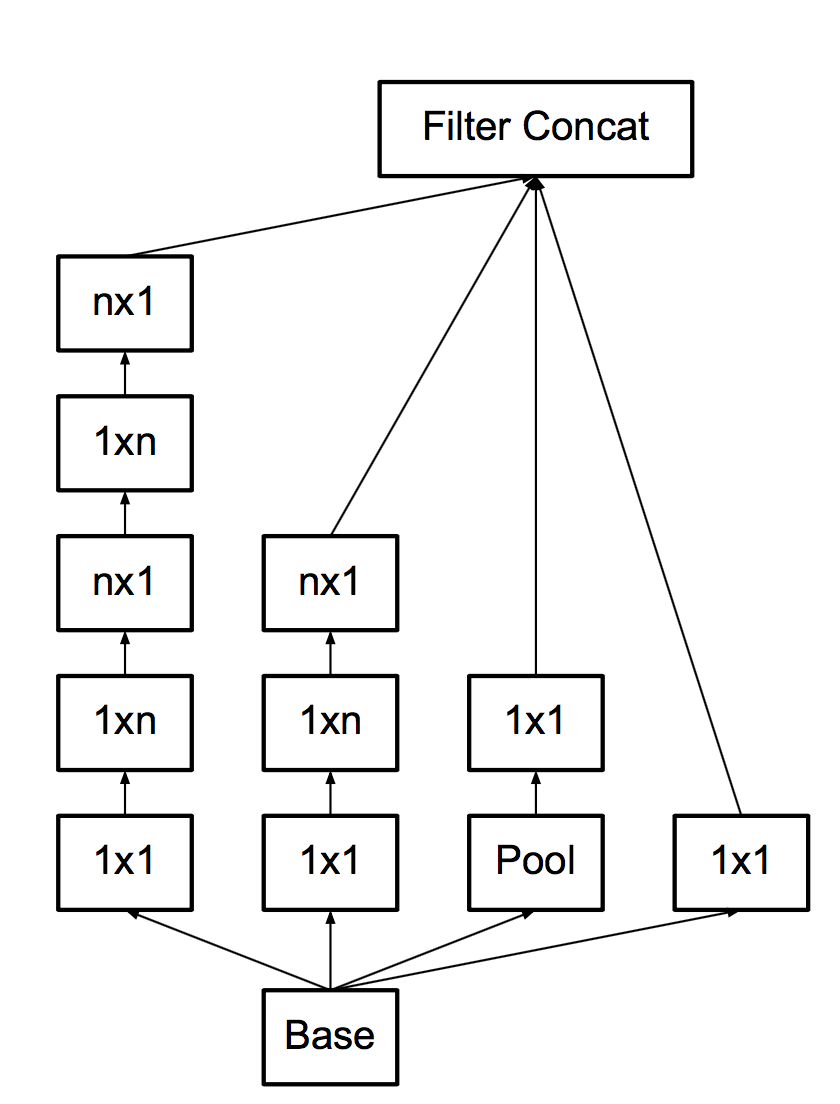


Рисунок . Inception блок с декомпозированными ядрами

**2.4 ResNet**

Авторы ResNet отметили, что с увеличением количества слоёв качество работы свёрточной сети начинает снижаться. Одновременное падение точности на валидационном и тренировочном множествах наталкивает на мысль, что проблема заключается не в переобучении. Предположили, что если свёрточная нейронная сеть достигла своего предела точности на каком-то слое, тогда все следующие слои должны будут выродиться в тождественное преобразование, но из-за сложности обучения глубоких сетей этого не происходит [21]. Для того чтобы исключить подобное, предложили ввести пропускающие соединения, представленное на рисунке 15.

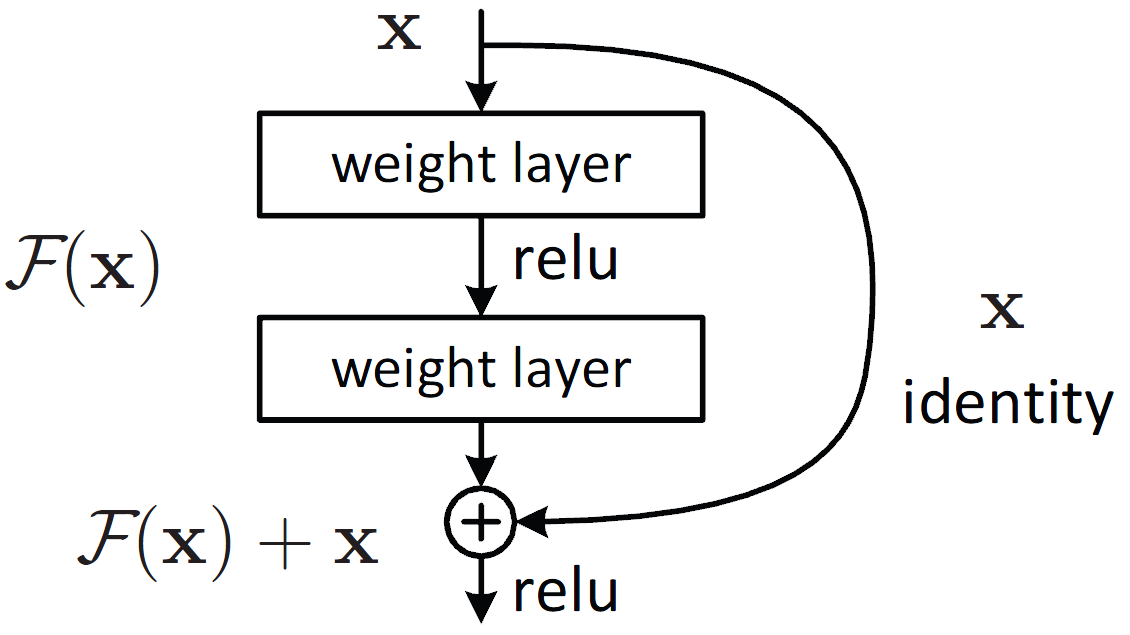


Рисунок 15. Пропускающие соединение

Добавив пропускающие соединения, как показано на рисунке 15, сеть учится остаточной функции, которая затем складывается с тождественным преобразованием. Анализ показал, что глубокие остаточные нейронные сети можно считать ансамблем, состоящим из более мелких остаточных нейронных сетей, чья эффективная глубина увеличивается в процессе обучения [12].

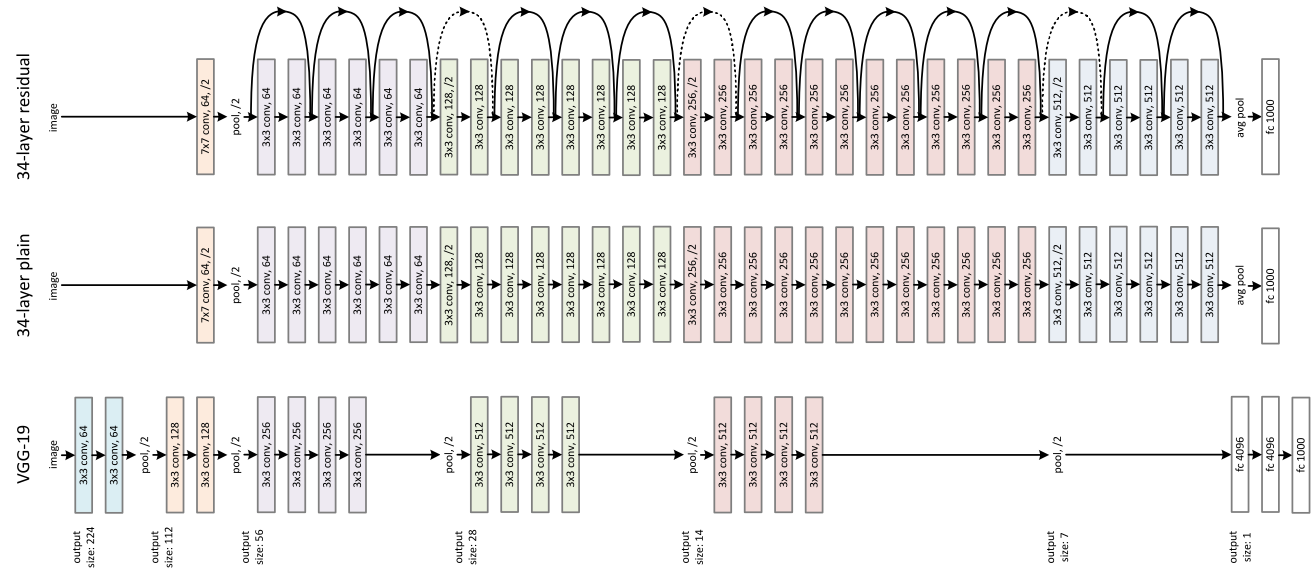


Рисунок . Сравнение архитектур ResNet и VGG

**Глава 3. Методика исследования**

**3.1 Описание набора данных и подготовка его к использованию**

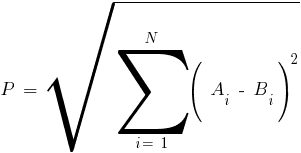
В качестве набора данных, используемого при анализе, выбрана база данных FEI Face Database. Это база данных лиц, собранная в Бразилии. Она содержит набор изображений лиц, сделанных в период с июня 2005 года по март 2006 года в Лаборатории искусственного интеллекта FEI в Сан-Бернардо-ду-Кампу, Сан-Паулу. Есть 14 изображений для каждого из 200 человек, в общей сложности 2800 изображений. Все изображения являются красочными и взяты на белом однородном фоне в вертикальном фронтальном положении с поворотом профиля примерно до 180 градусов. Масштаб может варьироваться примерно на 10%, а исходный размер каждого изображения составляет 640x480 пикселей. Все лица в основном представлены студентами и сотрудниками FEI в возрасте от 19 до 40 лет с отчетливой внешностью, прической и украшениями. Количество представленных мужских и женских объектов одинаково и равно 100 [23].

Поскольку количество позитивных пар в представленном наборе данных много меньше, чем негативных было принято решение снизить их количество путем ограничения их количества и подбора случайным образом. Таким образом количество позитивных пар равно 18200, а негативных 100000.

Для поиска лица и его границ на изображении использовался метод, описанный в параграфе 1.3- гистограмма направленных градиентов [7][8]. Далее для выравнивания лица использовались методы оценки антропометрических точек и аффинных преобразований, описанные в том же параграфе 1.3. Так же необходимо выполнить изменение размена изображения в соответствии с требованиями первого слоя сети.

**3.2 Обработка ответов сети**

Передав сети изображение, мы получим на выходе n-мерный вектор [9]. Соответственно для пары изображений мы получим два вектора. Чтобы решить какие изображения являются позитивными, а какие негативными можно вычислить евклидово расстояние между двумя полученными векторами, по формуле 9, и чем ближе полученное значение к нулю, тем больше вероятность что изображения принадлежат одному человеку. Соответственно, чем выше это расстояние, тем выше вероятность что на изображениях представлены разные личности [17].



Формула

Однако, для дальнейшего исследования будет удобней перевести полученные расстояния на процентное соотношение, где за 100% принимается минимальное из имеющихся расстояний, а за ноль соответственно максимальное. Таким образом далее будем вместо расстояния использовать процентную вероятность принадлежности пары изображений одному человеку.

Получив вероятности принадлежности для всех пар изображений, можно построить гистограммы полученных результатов выделив позитивные и негативные пары разными цветами для наглядности. На них будет наглядно видно, как распределились вероятности для каждого типа пар.

Чтобы найти оптимальное пороговое значение процентной вероятности принадлежности необходимо так же провести оценку точности сети. Самым простым вариантом оценки точности сети будет найти частное количества правильных ответов и общего количества пар. Однако, поскольку количество позитивных пар много меньше количества негативных пар данная оценка не сможет дать качественной оценки. Поэтому будет использована F-мера, описанная в параграфе 1.4. Обе эти оценки будут отражены на графиках для наглядности.

**Глава 4. Результаты анализа сетей**

**4.1 Результаты**

Ниже для каждой из рассматриваемых сетей представлены гистограммы с распределением пар изображений в зависимости от процентной вероятности принадлежать пары одному человеку. Оранжевым цветом изображены негативные пары, а синим позитивные.

Так же для каждой сети представлен график оценки точности их работы в зависимости от порогового значения принятия решения о принадлежности пары одному человеку. Оранжевым цветом изображён график F-меры, а голубым частное количества правильных ответов и общего количества пар.

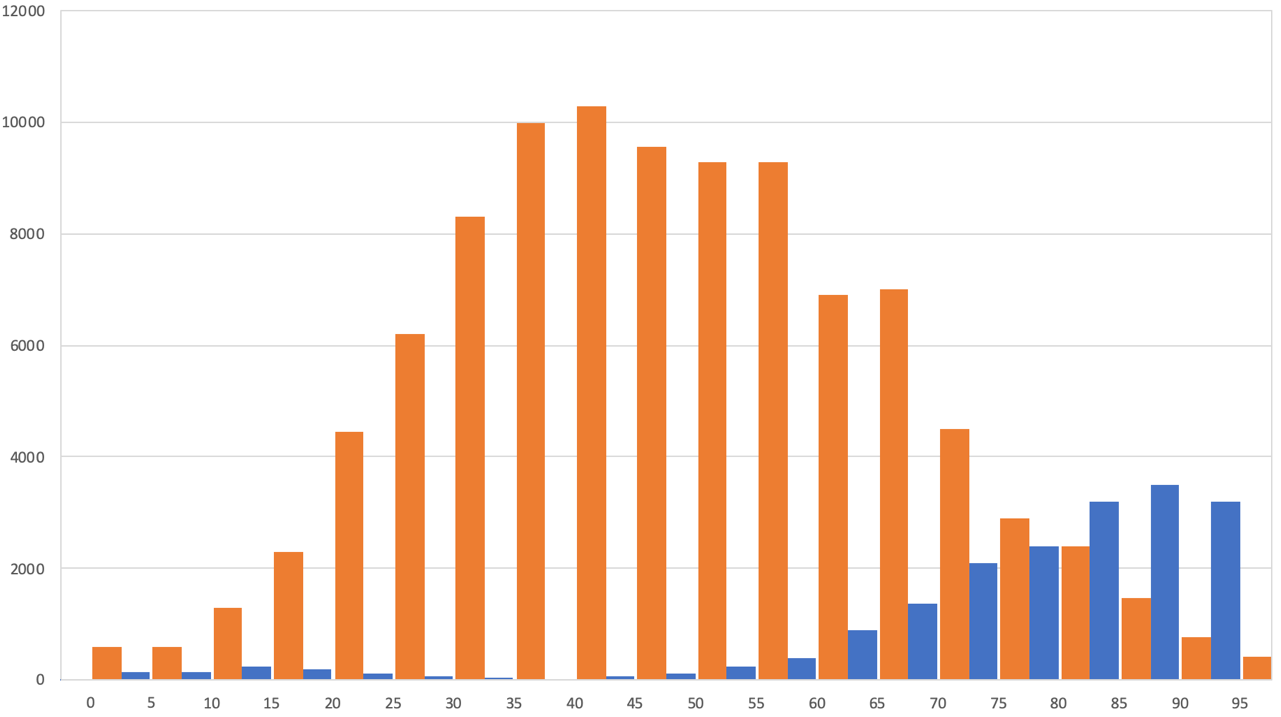
По положению максимального значения графика F-меры можно определить оптимальное значение порога принятия решения о принадлежности пары одному человеку и далее оценить количество ложноположительных ответов по гистограмме.

**VGG**

Оптимальное значение порога принятия решения: 75%.

Точность: 71,4%.

Количество ложно положительных ответов: 7980.

****

Гистограмма 1  
 Распределение сравниваемых пар изображений по вероятности их принадлежности одному человеку  
(ось Х – промежутки вероятности в процентах с шагом 5%, ось Y – количество пар попавших в промежуток,  
 негативные пары - оранжевый, позитивные пары - синий)

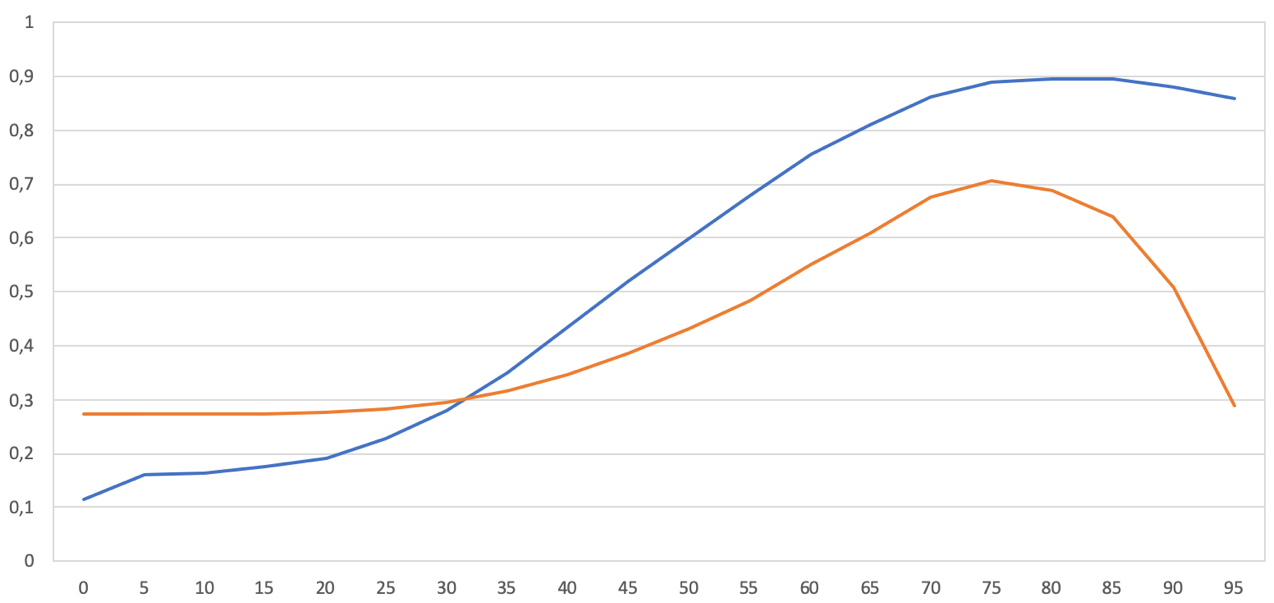
****

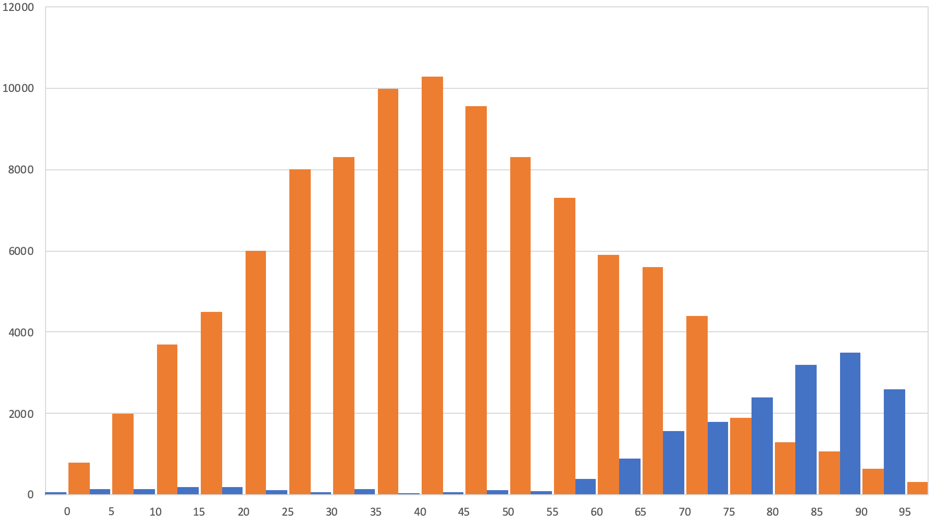
График 1  
Оценка точкости ответов сети в зависимоти от порога принятия решения  
(Ось Х – порог принятия решения идентичности личностей в паре изображений,  
Ось Y – значение меры точности,   
F-мера оранжевый, кол-во проавильных ответов деленное на все множество -синий)

**GoogleNet**

Оптимальное значение порога принятия решения: 75%.

Точность: 74,4%.

Количество ложно положительных ответов: 6251.



Гистограмма 2  
Распределение сравниваемых пар изображений по вероятности их принадлежности одному человеку  
(ось Х – промежутки вероятности в процентах с шагом 5%, ось Y – количество пар попавших в промежуток,  
 негативные пары - оранжевый, позитивные пары - синий)

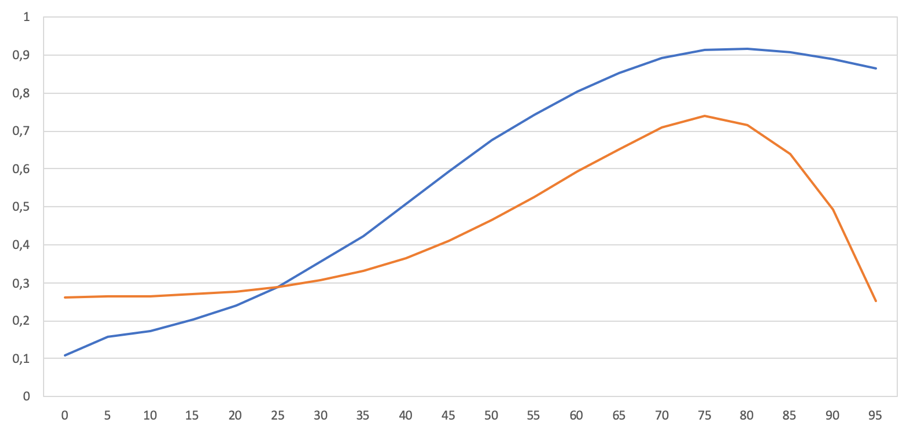


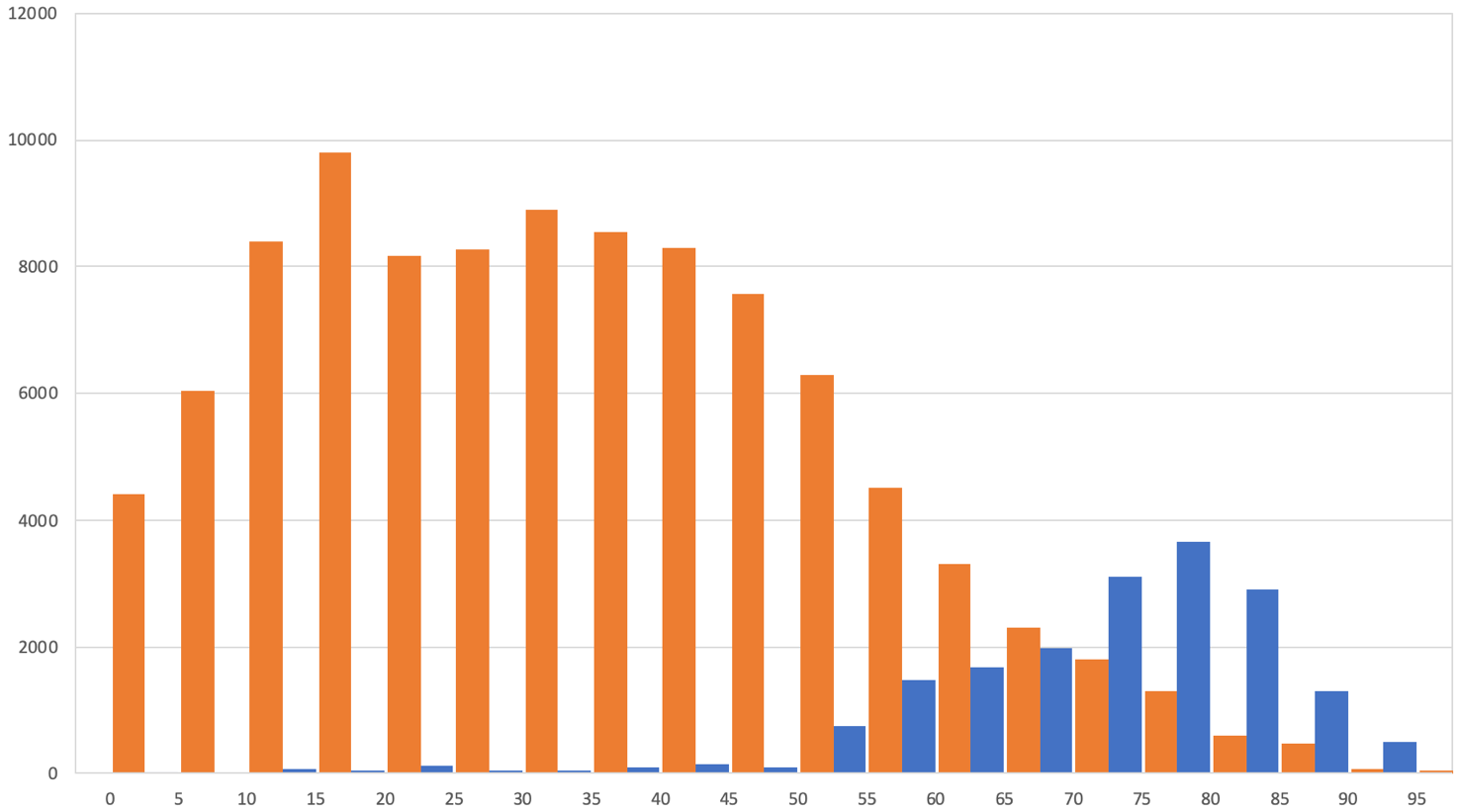
График 2  
Оценка точкости ответов сети в зависимоти от порога принятия решения  
(Ось Х – порог принятия решения идентичности личностей в паре изображений,  
Ось Y – значение меры точности,   
F-мера оранжевый, кол-во проавильных ответов деленное на все множество -синий)

**Inception v3**

Оптимальное значение порога принятия решения: 66,4%.

Точность: 76,2%.

Количество ложно положительных ответов: 6582.



Гистограмма 3  
Распределение сравниваемых пар изображений по вероятности их принадлежности одному человеку  
(ось Х – промежутки вероятности в процентах с шагом 5%, ось Y – количество пар попавших в промежуток,  
 негативные пары - оранжевый, позитивные пары - синий)

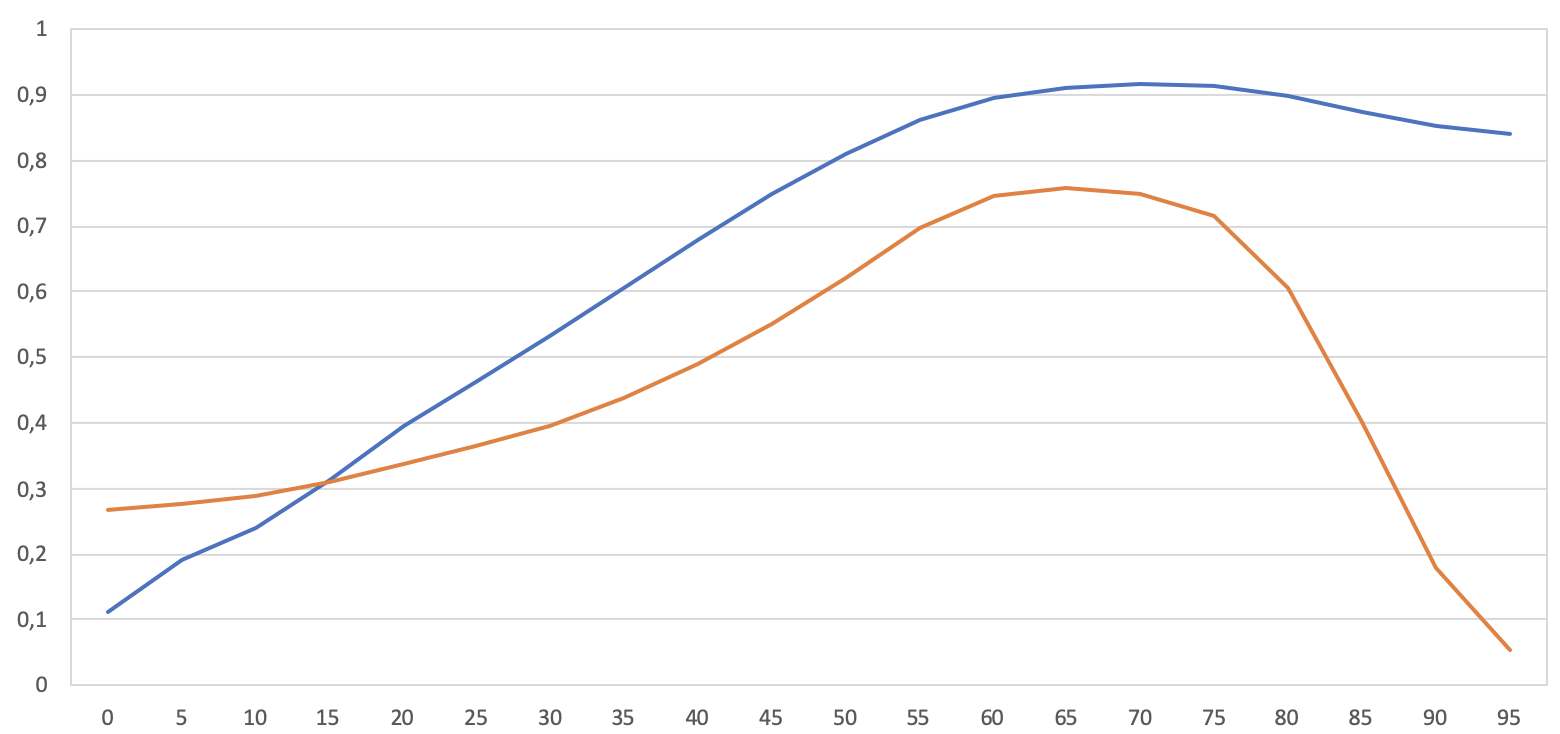


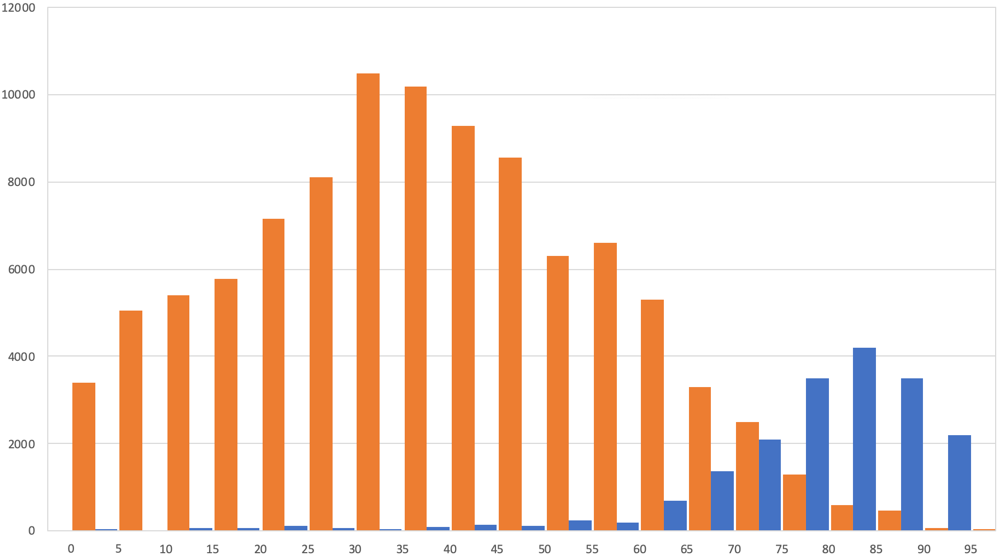
График 3  
Оценка точкости ответов сети в зависимоти от порога принятия решения  
(Ось Х – порог принятия решения идентичности личностей в паре изображений,  
Ось Y – значение меры точности,   
F-мера оранжевый, кол-во проавильных ответов деленное на все множество -синий)

**ResNet**

Оптимальное значение порога принятия решения: 74%.

Точность: 84,3%.

Количество ложно положительных ответов: 2482.



Гистограмма 4  
Распределение сравниваемых пар изображений по вероятности их принадлежности одному человеку  
(ось Х – промежутки вероятности в процентах с шагом 5%, ось Y – количество пар попавших в промежуток,  
 негативные пары - оранжевый, позитивные пары - синий)

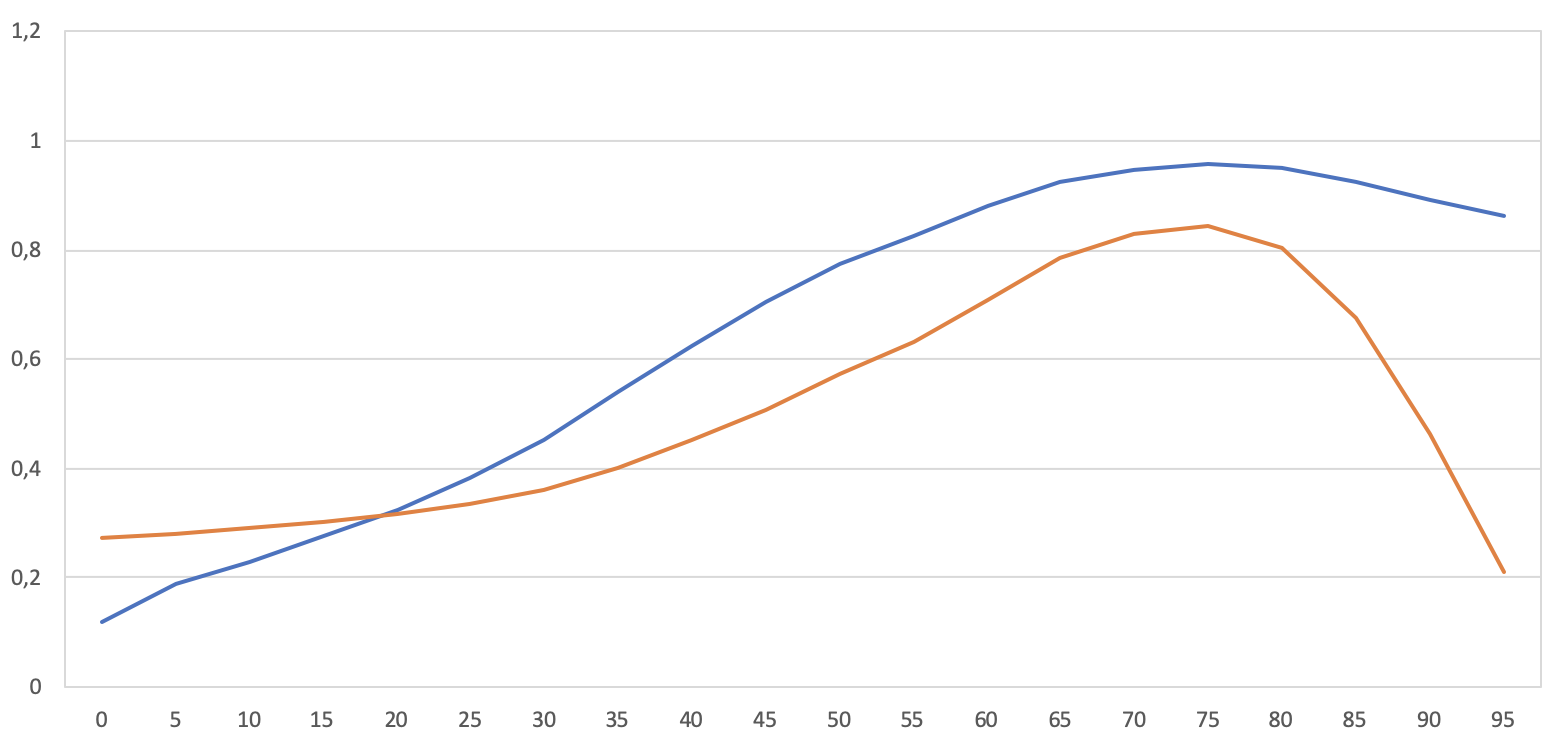


График 4  
Оценка точкости ответов сети в зависимоти от порога принятия решения  
(Ось Х – порог принятия решения идентичности личностей в паре изображений,  
Ось Y – значение меры точности,   
F-мера оранжевый, кол-во проавильных ответов деленное на все множество -синий)

**4.2 Вывод**

Лучшие результаты среди представленных архитектур показала сеть ResNet с результатом общей точности 84,3% и наименьшим количеством ложноположительных ответов (2482) при пороге принятия решения 74% достаточно близком к порогам остальных сетей.

Таким образом можно сделать вывод о правильности архитектурных решений, принятых в сети ResNet, но всё-таки её точность далека от систем, использующих коалиции из нескольких сетей. Следует отметить, что она успешно используется в подобных системах. Например, ICARE\_FACE\_V1, Uniview Technology и QINIU ATLAB - FaceX V1, имеющие в своем составе ResNet, показали высокие результаты в идентификации человека по лицу получив точность более 99% в тестах ресурса MegaFace [22].

Заключение

В результате исследования проведен сравнительный анализ работы находящихся в открытом доступе свёрточных сетей, применяемых в задаче идентификации человека по изображению лица, и выявить сеть, показавшую лучшие результаты точности решений и количества ложноположительных ответов. Ей оказалась сеть ResNet c наилучшим относительно других сетей результатом точности и меньшим количеством ложно положительных ошибок. Остальные сети также показали качественные результаты, однако количество ложно положительных ошибок полученных ими почти в три раза выше чем у ResNet, а этот показатель крайне важен для систем идентификации человека.

По полученным результатам так же можно сделать выводы о удачности архитектурных решений, применяемых в рассматриваемых сетях, и выбрать сети для включения их в коалицию для получения более качественных результатов. Так, например, VGG имея наиболее примитивную архитектуру (относительно рассматриваемых сетей) показала худший результат, Inception v3 являясь продолжением идей VGG и GoogleNet превзошла их, а ResNet, использующая пропускающие соединения, проказала лучший результат.

**Список литературы**

1. Lipton, Z. C., Berkowitz, J., and Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019.
2. Huang GB, Jain V, Learned-Miller E. Unsupervised Joint Alignment of Complex Images. 2007 IEEE 11th International Conference on Computer Vision. 2007. pp. 1–8. doi:1​0.1109/ICCV .2007.4408858
3. D. Stansbury. Derivation: Derivatives for Common Neural Network Activation Functions. In: The Clever Machine [Internet].
4. K. Simonyan, A. Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint* arXiv:1409.1556.
5. Сикорский О.С., МГТУ им. Н.Э. Баумана, Обзор свёрточных нейронных сетей для задачи классификации изображений.
6. V. Kazemi, J. Sullivan. One Millisecond Face Alignment with an Ensemble of Regression Trees. https://pdfs.semanticscholar.org/d78b/ 6a5b0dcaa81b1faea5fb0000045a62513567.pdf
7. DLib Library. http://dlib.net/
8. OpenCV Library. http://opencv.org/
9. PyTorch https://pytorch.org/
10. A. Krizhevsky, I. Sutskever, G.E. Hinton. ImageNet Classification with  
    Deep Convolutional Neural Networks. https://papers.nips.cc/paper/ 4824-imagenet-classification-with-deep-convolutional-neural-networks. pdf
11. R. G. Cinbis, J. J. Verbeek, and C. Schmid. Unsupervised metric learning for face identification in TV video. In Proc. ICCV, pages 1559–1566, 2011.
12. Matthew D Zeiler, Rob Fergus. 2014. Visualizing and understanding convolutional networks. *European conference on computer vision*, pp. 818-833. Springer International Publishing.
13. Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich. 2015. Going deeper with convolutions. *The IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-9
14. Geoffrey Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, Ruslan R Salakhutdinov. 2012. Improving neural networks by preventing co- adaptation of feature detectors. *arXiv preprint* arXiv:1207.0580
15. Ross Girshick. 2015. Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440-1448.
16. Lee H, Grosse R, Ranganath R, Ng AY. Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations. Proceedings of the 26th Annual International Conference on Machine Learning. New York, NY, USA: ACM; 2009. pp. 609–616.
17. Review: DenseNet — Dense Convolutional Network (Image Classification) <https://towardsdatascience.com/review-densenet-image-classification-b6631a8ef803>
18. Siebert Looije, University of Groningen, Pre-trained Deep Convolutional Neural Networks for Face Recognition
19. Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). How transferable are features in deep neural networks? In Advances in neural information processing systems, pages 3320–3328.
20. Qian, N. (1999). On the momentum term in gradient descent learning algorithms. Neural networks, 12(1):145–151.
21. Guo, Y., Zhang, L., Hu, Y., He, X., and Gao, J. (2016). Ms-celeb-1M: challenge of recognizing one million celebrities in the real world. Electronic Imaging, 2016(11):1–6.
22. MegaFace <http://megaface.cs.washington.edu/results/facescrub.html>.
23. FEI Face Database <https://fei.edu.br/~cet/facedatabase.html>
24. Sergey Ioffe, Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167