Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Московский государственный технический университет имени Н. Э. Баумана (национальный исследовательский университет)»
(МГТУ им. Н. Э. Баумана)

ФАКУЛЬТЕТ «Информатика и системы управления»
КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

Научно-исследовательская работа

на тему:

«Методы идентификации объектов с целью их дальнейшего отслеживания»

Студент: **Слиняков М.Л.**
Группа: **ИУ7-54Б**
Научный руководитель: **Тассов К.Л.**

Москва, 2024

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Системы видеонаблюдения и анализа видеоинформации стали важной частью современной технологии осуществления безопасности. Одной из актуальных задач становится автоматизированная идентификация объектов на видеопотоках и их дальнейшее отслеживание. Решение данной задачи особенно востребована в системах, связанных с безопасностью, мониторингом общественных мест, дорог. В условиях наблюдения объектов в динамической среде одной лишь идентификации объекта оказывается недостаточно. Возникает необходимость не только идентифицировать объект, но и прогнозировать его возможное местоположение на последующих кадрах или на других видеокамерах. Такой подход позволяет продолжить отслеживание объекта, даже если он временно исчезает из поля зрения одной камеры, но может вскоре появиться в зоне видимости другой. Следовательно, эффективные методы идентификации и прогнозирования траекторий движения объектов оказываются крайне значимыми для повышения надежности и точности систем наблюдения.

Целью данной работы является анализ существующих методов идентификации объектов на видео, а так же методов для их идентификации. Для достижения поставленной цели необходимо выполнить следующие задачи:

1. Изучить современные методы идентификации объектов на изображениях и видео.

2. Изучить современные методы отслеживания объектов на изображениях и видео.

3. Рассмотреть алгоритмы отслеживания объектов, включая подходы, основанные на вероятностных методах, сверточных нейронных сетях.

4. Проанализировать применимость скрытых марковских моделей (HMM) для идентификации и предсказания положения объектов в видеопотоках.

1 Обзор предметной области

Задачей распознавания объекта называется построение алгоритма, который вычисляет некоторые характеристики это объекта по его наблюдаемым свойствам. Работать этот алгоритм должен не только для объектов, предьявленных заранее, но и для объектов, которые заранее представлены не были (рисунок 1.1). Задачей обучения является построение таких алгоритмов по имеющемуся набору объектов.

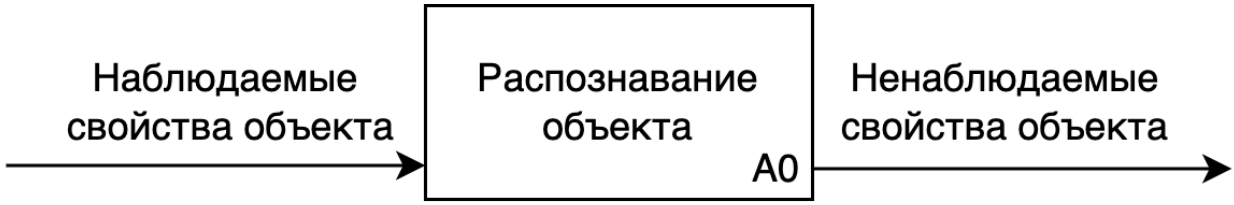

Рисунок 1.1 – Схема алгоритма, решающего задачу распознавания объектов в формате idef0

Под наблюдаемыми свойствами объекта принимаются значения векторов свойств, образующих некоторое пространство свойств X. Аналогично, результаты распознавания являются результаты векторов в пространстве ответов Y. Исходя из этого, алгоритм решающий задачу распознавания объекта осуществляет некоторое отображения $X \to Y$.

Можно привести такой пример:

Пусть имеется некоторый обучающий набор из n объектов с известными наблюдаемыми признаками из X и известными ненаблюдаемыми признаками из Y (формула 1.1).

$$M = ((x_1, y_1), ..., (x_n, y_n) : x_i \in X, y_i \in Y) \tag{1.1}$$

В качестве результата алгоритм распознавания для некоторого объекта O с наблюдаемыми свойствами x будет выдавать результат y_i такой, что наиболее близким значением наблюдаемых свойств к x (согласно какой-то метрике) будет значение наблюдаемых свойств x_i.

Так же необходимо формализовать постановку задачи обучения. Пусть имеется пространство наблюдаемых свойств X, пространство ненаблюдаемых

свойств Y, пространство алгоритмов распознавания $A : X \to Y$, пространство вероятностных мер P на $X \times Y$, функция штрафа L (формула 1.2), обучающий набор M.

$$L(a(x), y, x) = \begin{cases} 0, & \text{если } a(x) = y, \\ 1, & \text{если } a(x) \neq y. \end{cases} \tag{1.2}$$

Требуется по этим данным, построить алгоритм $a \in A$, при котором математическое ожидание штрафа минимально по некоторому распределению $\pi \in P$ (формула 1.3).

$$E_p(f) = \int_{x,y} E(a(x), y, x)\, d\pi(x, y) \to \min_a \tag{1.3}$$

По методу Монте-Карло, можно приблизить математическое ожидание штрафа (формула 1.4).

$$E(a, M) = \frac{1}{N} \sum_{i=1}^{N} E(f(x_i, y_i, x_i) \to \min_a . \tag{1.4}$$

Приведенный способ называется способом минимизации ошибки обучения. В нем есть такой недостаток: в результате обучения был составлен некоторый распознающий алгоритм a такой, что $a(x_i) = y_i$, и он имеет малую ошибку на обучающем наборе и большую ошибку на случайных значениях. Такая ситуация называется переобучением.

В качестве примера алгоритма распознавания можно привести распознающие деревья. Для распознаваемого объекта проводится некоторая конечная цепочка сравнений значений его наблюдаемых свойств с некоторыми значениями. Обучение дерева заключается в составлении его структуры, значений и операций для сравнения и ответов в каждом листе. Пусть имеем дерево с одним листом ($f(x) = r$). При квадратичной ошибке минимум по r достигается в среднем арифметическом ответе из всего обучающего набора (формула 1.5).

$$r = \frac{1}{N} \sum_{i=1} N y_i \tag{1.5}$$

2 Анализ последовательностей и Марковские модели

Во многих реальных задачах размерность пространства наблюдаемых свойств не совсем естественна. Чаще всего такие объекты представляются ввиде какого-то структурированного набора из более элементарных объектов. Эти элементарные объекты уже можно закодировать вектором свойств. Изображение представленно в виде набора пикселей, которые в свою очередь могут кодироваться некоторым вектором наблюдаемых свойств.

Вероятностной моделью последовательностей в пространстве X называется последовательность совместных распределений вероятности $p^k(x_1, ..., x_k)$ (формула 2.1).

$$p^k(x_1, ..., x_k) = p_{k|k-1}(x_k|x_1, ..., x_{k-1}) \cdot p^{k-1}(x_1, ..., x_{k-1}) = \\ = ... = p_{k|k-1}(x_k|x_1, ..., x_{k-1})...p_{2|1}(x_2|x_1) \cdot p^1(x_1) \quad (2.1)$$

В модели Маркова условные вероятности $p_{k|k-1}$ зависят от фиксированного числа величин, имеющими больший коэффициент (наиболее близкие к k). То есть вероятность следующей величины зависит от вероятностей n предыдущих величин. На рисунках 2.1, 2.2 приведены схемы зависимости вероятностей случайных величин (в данном случае свойств), для моделей Маркова разной степени. В круге представлены вероятности, стрелки определяют направление зависимости вероятности свойства (к зависимому).

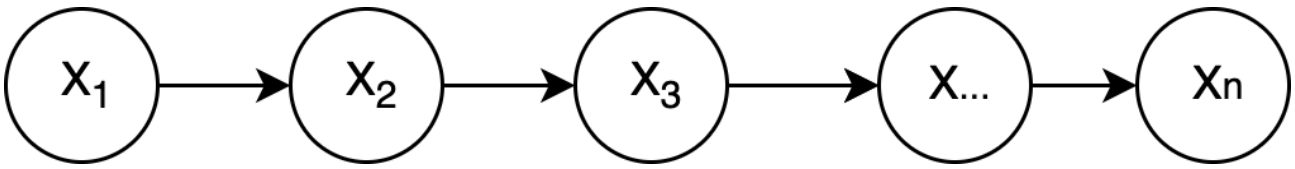

Рисунок 2.1 – Зависимости случайных свойств в модели Маркова, при n = 1

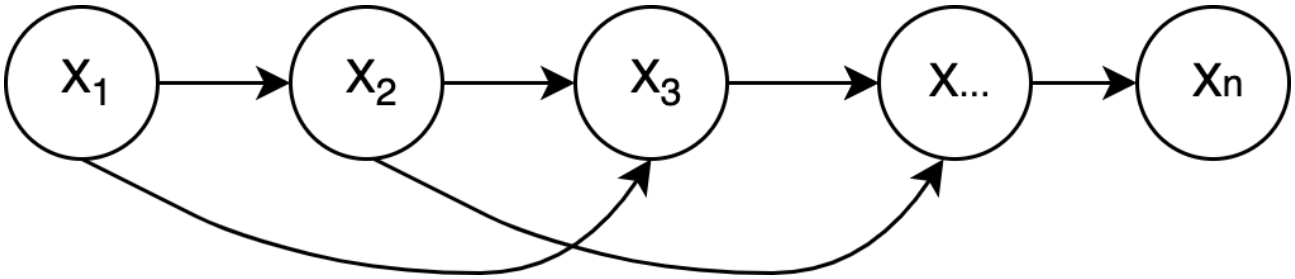

Рисунок 2.2 – Зависимости случайных свойств в модели Маркова, при n = 2

На рисунке 2.3 представлена схема зависимости вероятности величин (в данном случае и наблюдаемых, и ненаблюдаемых свойств) в скрытой модели Маркова (HMM). В ней текущее состояние системы (ненаблюдаемые свойства) неизвестно напрямую, а вместо него видно лишь определенные состояния (наблюдаемые свойства), которые зависят от скрытых состояний. С каждым скрытым состоянием связана вероятность наблюдения определенного открытого состояния. При работе со скрытой моделью Маркова приходится решать задачу нахождения скрытых состояний на основе видимых состояний. Эту задачу как раз решает алгоритм распознавания $A : X \to Y$.

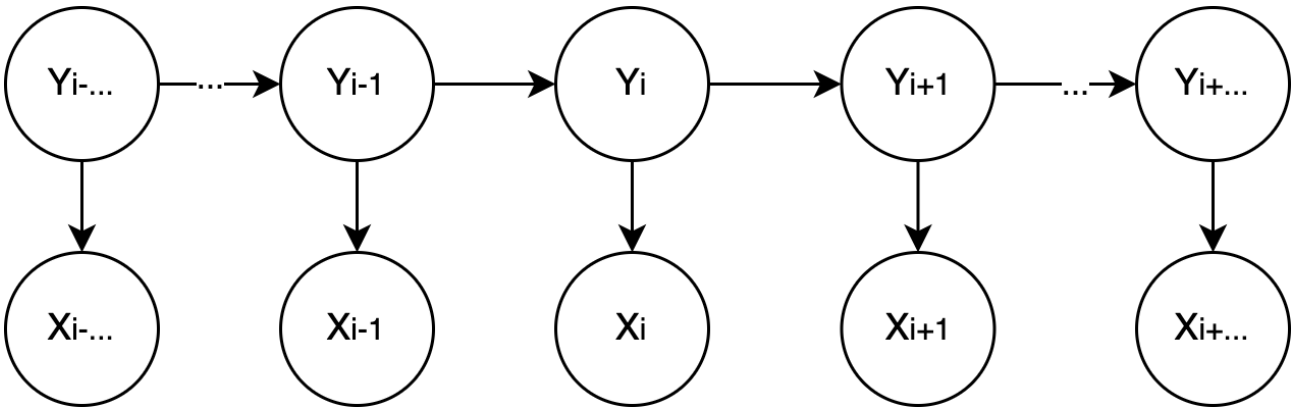

Рисунок 2.3 – Зависимости случайных свойств в модели Маркова, при n = 2

Таким образом, обычная марковская модель применима в задачах отслеживания некоторой последовательности действий. Это как раз подходит для отслеживания объекта и его действий после идентификации. Скрытая марковская модель может быть полезна, если приходится решать задачу распознавания. В задаче распознавания объекта на видеопотоках, скрытые состояния могут представлять собой сам класс объекта или что-то, что его идентифицирует, а так же его позицию (относительно того, на какой камере его

можно увидеть). Открытыми состояниями могут быть наблюдаемые свойства объекта, такие как форма, цвет, движения. Эти наблюдения зависят от скрытых состояний и могут быть неполными, но модель может интерпретировать их в зависимости от переходных вероятностей. Модель также может учитывать, что объект на основе текущего состояния (например скорости) будет двигаться плавно (или наоборот) и появится на определенной камере. Это может быть особенно полезно в ситуациях, когда видеокамеры не перекрывают области видимости друг друга и объект в данный момент находится вне поле зрения.

3 Сверточные нейронные сети

Сверточная нейронная сеть состоит из следующих слоев: сверточные слои, субдискретизирующие слои, слои перцептрона. Задача сверточных и субдискретизирующих слоев состоит в том, чтобы формировать входной вектор наблюдаемых свойств, который будет передан перцептрону. В контексте обработки изображений, свертка — это процесс, при котором фильтр (или ядро) применяется к изображению, чтобы извлечь из него наблюдаемые признаки. Это делается путем перемещения фильтра по изображению и вычисления взвешенной суммы пикселей, которые фильтр охватывает. Для положения фильтра вычисляется сумма 3.1, где N, M - размеры фильтра, p - пиксель изображения, w - вес. Эта сумма образует новый пиксель, который затем становится частью выходного изображения.

$$S = \sum_{i=0,j=0}^{i=N,j=M} p_{ij} \cdot w_{ij} \tag{3.1}$$

Свертка позволяет извлекать такие наблюдаемые признаки с изображения как края и текстуры. Каждый фильтр может извлекать какой-то один наблюдаемый признак, поэтому для формирования вектора наблюдаемых признаков необходимо иметь несколько фильтров, каждый из которых будет извлекать конкретный признак. На рисунке 3.1 представлена схема прохода фильтра размером 3x3 по изображению 5x5.

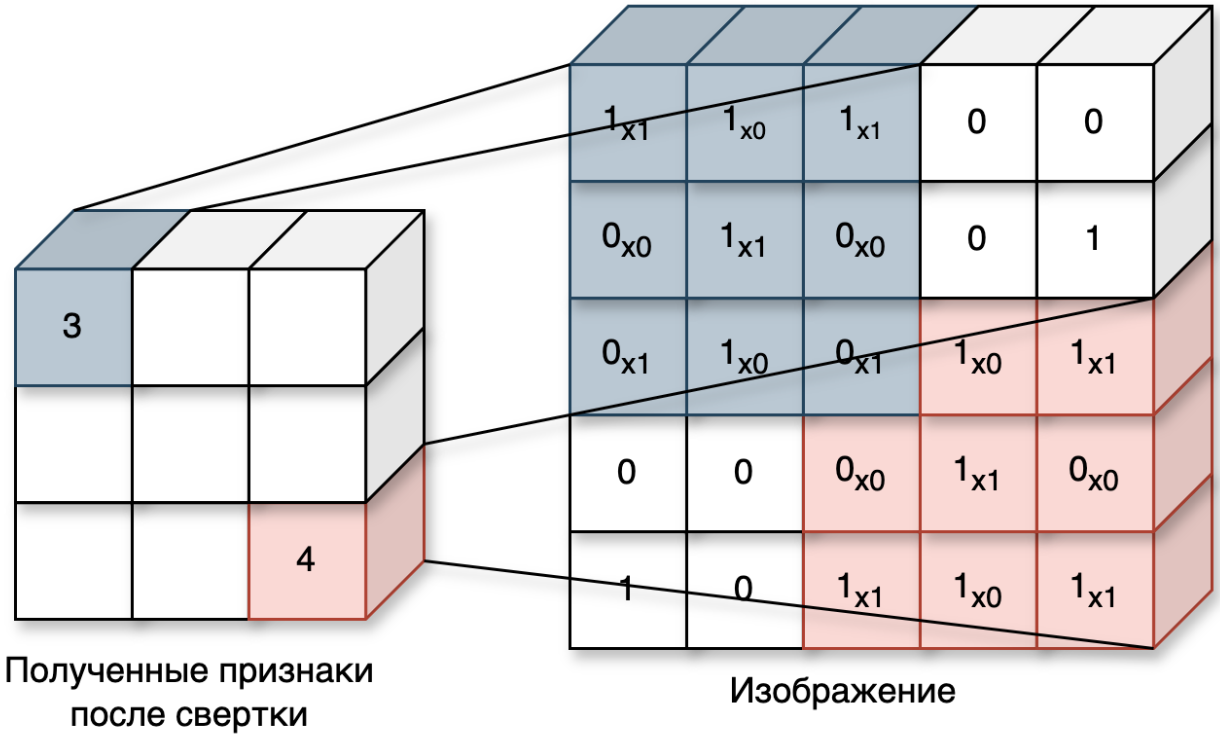

Рисунок 3.1 – Схема прохода фильтра размеров 3x3 по изображению 5x5

В представленной схеме есть недостаток. Крайние пиксели изображения никогда не оказываются в центре ядра, так как тогда ядру будет неоткуда брать информацию из пикселей рядом с крайним вне изображения. Эту проблему решает технология padding. Ее суть заключается в том, чтобы прибавить к изображению ложные пиксели нулевого значения. На рисунке 3.2 представлена схема прохода фильтра по изображению с ложными пикселями нулевого значения.

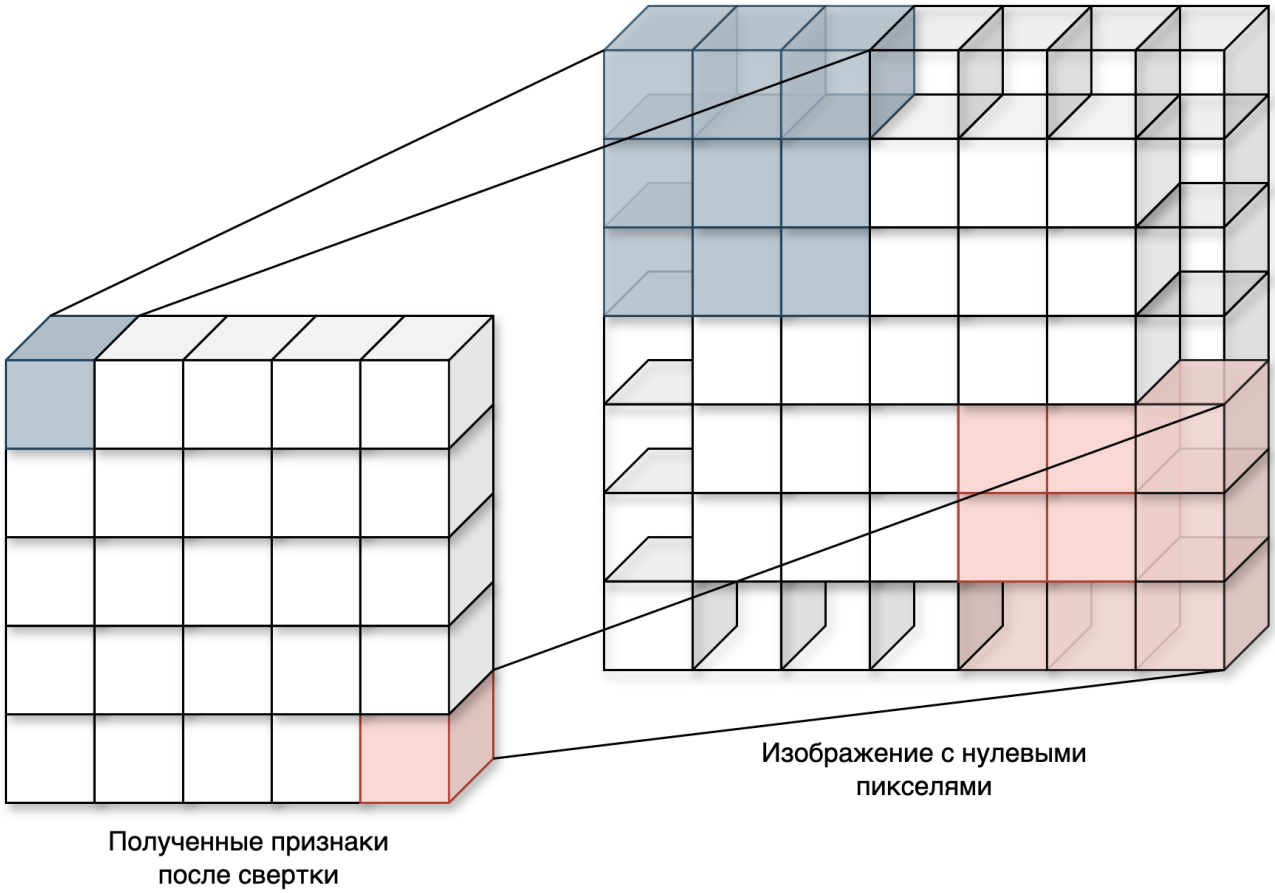

Рисунок 3.2 – Схема прохода фильтра по изображению с ложными пикселями

Субдискретизирующие слои уменьшает размерность матриц, полученных на этапе свертки. На этом этапе фильтр ”скользит”вдоль матрицы, полученной на этапе свертки и выполняет либо усреднение (average pooling) или выбор максимального (max pooling) из сканируемой области.

Таким образом, сверточные слои можно располагать друг за другом, формируя иерархию признаков - от низкоуровневых (края и текстуры), которые выделяются в начальных слоях, до высокоуровневых (формы), которые выделяются в более глубоких слоях. Подвыборочные (субдискретизирующие) слои обычно следуют за одним или несколькими сверточными слоями и являются промежуточными шагами для уменьшения размерности, подготавливая данные к более глубоким сверточным слоям или к полносвязанным слоям.

Полносвязный слой (или перцептрон) занимается классификацией признаков, полученных от подвыборочных слоев. Каждый подвыборочный слой связан с одним нейроном полносвязного слоя. Значение нейрона

вычисляется по формуле 3.2, где X это вектор подготовленных наблюдаемых свойств, переданных нейрону, размерностью X, w - веса, b - коэффициент сдвига слоя, f - функция активации.

$$Y = f(\sum_{i}^{N} X_i \cdot w_i + b) \tag{3.2}$$

4 Распознавание лиц

В предыдущей главе были представлены сверточные нейронные сети, с помощью которых можно решать задачу распознавания изображения. Для этого необходимо иметь ядра (фильтры) для выявления наблюдаемых признаков из изображения. Получить их можно в результате обучения. В начале обучения значения весов ядра случайны, в процессе обучения они корректируются, чтобы уменьшить ошибку предсказания. Изображение проходит через сеть, ядра свёртки создают карты признаков путём свертки, в конце прохода сеть вычисляет ошибку предсказания. Модель использует алгоритм обратного распространения для корректировки весов в ядрах на основе ошибки. Вес фильтров, которые лучше помогают отличать черты лица, настраиваются так, чтобы усилить эти признаки, а менее значимые фильтры корректируются или игнорируются. В ходе обучения сеть автоматически подстраивает фильтры, чтобы они лучше распознавали конкретные черты лица, такие как глаза, нос и рот. На начальных слоях СНС учится выделять простые признаки, такие как линии и углы, что помогает определить границы черт лица. На более глубоких уровнях сети активируются более сложные признаки, которые представляют собой комбинации линий и текстур, характерных для форм глаз, носа и рта.

Для решения задачи распознавания лиц необходимо составить архитектуру нейронной сети.

Архитектура **LeNet-5** была создана для распознавания рукописных цифр, однако она может быть адаптирована для распознавания лиц. Ее структура состоит в следующем:

- Входное изображение.
- Сверточный слой с 6 фильтрами 5x5.
- Подвыборочный слой Max Pooling 2x2.
- Сверточный слой с 16 фильтрами 5x5.
- Подвыборочный слой Max Pooling 2x2.
- Полносвязный слой с 120 нейронами.
- Выходной слой Softmax для классификации на основе признаков лица.

Архитектура **VGG-16** подходит для более крупных датасетов и достигает точности 92.7%. Входному слою подаются RGB изображения размером 224x224 пикселей. Далее изображения проходят через сверточные слои. Размерность ядер в этих слоях - 3x3. В одной из конфигураций используется сверточный фильтр размера 1x1, который может быть представлен как линейная трансформация входных каналов (с последующей нелинейностью). Сверточный шаг фиксируется на значении 1 пиксель. Пространственное дополнение (padding) входа сверточного слоя выбирается таким образом, чтобы пространственное разрешение сохранялось после свертки, то есть дополнение равно 1 для 3x3 сверточных слоев. Подвыборка осуществляется при помощи пяти max-pooling слоев, которые следуют за одним из сверточных слоев (не все сверточные слои имеют последующие max-pooling). Операция max-pooling выполняется с ядром 2x2 пикселей с шагом 2. После свертки идвут два полносвязных слоя по 4096 нейронов.

- Входное RGB изображение 224x224.
- Два сверточных слоя с 64 фильтрами 3x3.
- Подвыборочный слой Max Pooling 2x2 и шагом 2.
- Два сверточных слоя с 128 фильтрами 3x3.
- Подвыборочный слой Max Pooling 2x2 и шагом 2.
- Два сверточных слоя с 256 фильтрами 3x3.
- Подвыборочный слой Max Pooling 2x2 и шагом 2.
- Два сверточных слоя с 256 фильтрами 3x3.
- Подвыборочный слой Max Pooling 2x2 и шагом 2.
- Два сета из трех сверточных слоев с 512 фильтрами 3x3.
- Подвыборочный слой Max Pooling 2x2 и шагом 2.
- 2 полносвязных слоя с 4096 нейронами.
- Выходной слой Softmax для классификации на основе признаков лица.

FaceNet использует глубокое обучение для получения компактного и эффективного представления лиц, называемого эмбеддингом лица. В отличие от традиционных подходов, которые классифицируют изображение лица, FaceNet проецирует лица в многомерное пространство, где расстояния между лицами указывают на их схожесть. Модель позволяет с высокой точностью решать задачи верификации и идентификации лиц. В данной технологии сверточная нейронная сеть используется для получения вектора фиксированной длины. Для определение схожести полученных векторов, FaceNet использует триплетную функцию потерь то есть при обучении перед СНС также стоит задача минизации ошибки, однако функцией определения ошибки в данном случае будет триплетная функция 4.1, где A - эталонный вектор свойств человека (anchor), P - вектор свойств этого же человека, полученный с другой фотографии (positive), N - вектор свойств другого человека (negative), d - функция растояния, a - добавочный член, чтобы P и N не оказались на одинаковом расстоянии от A.

$$loss = max(d(A, P) - d(A, N) + \alpha \tag{4.1}$$

Для успешного обучения необходимо правильно подбирать триплеты. Существуют так называемые сложные триплеты (hard-triplets), где A и P сильно отличаются друг от друга, A и N близко друг к другу. Также для первичного обучения необходимо использовать легкие триплеты (easy-triplets), где схожесть между A и P, как и различие между A и N, очевидна.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

[1] Мерков А.Б. Распознавание образов. Построение и обучение вероятностных моделей. Москва: ЛЕНАНД, 2014.

[2] Калиновский И.А. Методы нейросетевого детектирования лиц в видеопотоке сверхвысокого разрешения. Национальный исследовательский Томский государственный университет, 2016.

[3] Patrik KAMENCAY Miroslav BENCO Tomas MIZDOS Roman RADIL. A New Method for Face Recognition Using Convolutional Neural Network // ADVANCES IN ELECTRICAL AND ELECTRONIC ENGINEERING. 2017. Т. 15, № 4. С. 663–672.