НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
май–июнь 2023 Том 23 № 3 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
May–June 2023 Vol. 23 No 3 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

doi: 10.17586/2226-1494-2023-23-3-506-518
УДК 004.032.26

# Детекция ключевых точек лица с помощью капсульных нейронных сетей

**Антон Александрович Бойцев[1]✉, Дмитрий Геннадьевич Волчек[2], Егор Николаевич Магазенков[3], Максим Кириллович Неваев[4], Алексей Андреевич Романов[5]**

[1,2,3,5] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация
[4] ЗАО «Центр финансовых технологий», Санкт-Петербург, 191002, Российская Федерация

[1] boitsevanton@gmail.com✉, https://orcid.org/0000-0002-3374-8256
[2] dvolchek@itmo.ru, https://orcid.org/0000-0002-0310-1654
[3] egormaga04@mail.ru, https://orcid.org/0000-0002-7563-0846
[4] m.nevaev@alumni.nsu.ru, https://orcid.org/0000-0002-9000-7841
[5] romanov@itmo.ru, https://orcid.org/0000-0002-6991-464X

**Аннотация**

**Введение.** Исследована актуальная и востребованная задача детекции ключевых точек лица. Рассмотрены существующие подходы к решению задачи детекции ключевых точек лица, часто классифицируемые на параметрические и непараметрические. Определен наиболее качественный на сегодняшний день подход, основанный на методах глубокого обучения. Предложено два решения: капсульная сеть с динамической маршрутизацией и глубокая капсульная сеть. В качестве данных для проведения эксперимента выбраны 10 000 сгенерированных лиц из базы сайта Kaggle, размеченных с помощью фреймворка MediaPipe. **Метод.** Предложен метод использования капсульных архитектур нейронных сетей для решения задачи детекции ключевых точек лица. Метод включает в себя использование сегментации по распознанным с помощью фреймворка MediaPipe ключевым точкам лица. Для построения сетки лица применена триангуляция Делоне. Предложена архитектура глубокой капсульной сети с учетом семантической сегментации. **Основные результаты.** На основе размеченных данных выполнены эксперименты по детекции ключевых точек с помощью разработанных капсульных нейронных сетей. По результатам тестирования получены значения функции потерь 2,5–2,9 и точности 0,87–0,9. **Обсуждение.** Предложенная архитектура может быть использована в технологиях по сопоставлению геометрий сеток лица реального человека и трехмерной модели. Архитектура может найти применение в исследованиях капсульных нейронных сетей в области обработки и анализа изображений.

**Ключевые слова**

капсульные нейронные сети, детекция ключевых точек лица, распознавание изображений лиц, нейросети

**Ссылка для цитирования:** Бойцев А.А., Волчек Д.Г., Магазенков Е.Н., Неваев М.К., Романов А.А. Детекция ключевых точек лица с помощью капсульных нейронных сетей // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23, № 3. С. 506–518. doi: 10.17586/2226-1494-2023-23-3-506-518

# Facial keypoints detection using capsule neural networks

**Anton A. Boitsev[1]✉, Dmitry G. Volchek[2], Egor N. Magazenkov[3], Maxim K. Nevaev[4], Aleksei A. Romanov[5]**

[1,2,3,5] ITMO University, Saint Petersburg, 197101, Russian Federation
[4] ZAO “Center of Financial Technologies”, Saint Petersburg, 191002, Russian Federation

[1] boitsevanton@gmail.com✉, https://orcid.org/0000-0002-3374-8256
[2] dvolchek@itmo.ru, https://orcid.org/0000-0002-0310-1654
[3] egormaga04@mail.ru, https://orcid.org/0000-0002-7563-0846
[4] m.nevaev@alumni.nsu.ru, https://orcid.org/0000-0002-9000-7841
[5] romanov@itmo.ru, https://orcid.org/0000-0002-6991-464X

**Abstract**
The problem of detecting key points of the face is investigated. This problem is quite relevant and important. The existing approaches of solving this problem, which are usually divided into parametric and nonparametric methods, are considered. As a result of the study, it was concluded that, nowadays, the most qualitative results are demonstrated by approaches based on deep learning methods. Two solutions are proposed: a capsule network with dynamic routing and a deep capsule network. The data for the experiments are 10,000 generated faces taken from Kaggle, marked up using MediaPipe. A method of using capsule architectures in neural networks to solve the problem of detecting key points of the face is proposed. The method includes the use of segmentation based on the key points of the face recognized using MediaPipe. Delaunay triangulation was used to build the face mesh. The architecture of a deep capsule network considering semantic segmentation was proposed. Based on the marked-up data, experiments on the detection of key points using the developed capsule neural networks were performed. According to the test results, the loss function reached values in range 2.50–2.90, the accuracy reached values in range 0.87–0.9. The proposed architecture can be used in technologies for comparing the geometry of the face grid of a real person with the geometry of the face grid of a three-dimensional model as well as in further studies of capsule neural networks by researchers in the field of image processing and analysis.

**Keywords**
capsule neural networks, detection of key points of the face, face image recognition, neural networks

**For citation:** Boitsev A.A., Volchek D.G., Magazenkov E.N., Nevaev M.K., Romanov A.A. Facial keypoints detection using capsule neural networks. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 3, pp. 506–518 (in Russian). doi: 10.17586/2226-1494-2023-23-3-506-518

## Введение

Задача детекции ключевых точек лица — одна из задач компьютерного зрения. Детекция ключевых точек лица находит множество применений, начиная от создания анимационных фильмов и заканчивая сложными системами биометрии и компьютерной безопасности. Существуют различные хорошо изученные подходы к решению задачи детекции ключевых точек лица, однако, появление новых архитектур нейронных сетей дает серьезный качественный скачок в этой области как с точки зрения скорости, так и точности. Тем самым рассматриваемая задача остается актуальной и по сей день. Отметим, что искусственные нейронные сети на сегодняшний день де-факто стали популярным аппаратом для решения разного рода задач. Исследователи находят все больше областей применения и развития нейронных сетей [1–3]. Каждый год разрабатывают новые архитектуры нейронных сетей, усовершенствуют алгоритмы оптимизации, за счет которых и происходит «обучение», и многое другое. Не секрет, что одна из самых востребованных областей применения нейронных сетей в настоящее время — область компьютерного зрения. Это связано с широким развитием и внедрением технологий компьютерного зрения в медицине, робототехнике, видеонаблюдении, компьютерной графике, создании автономных автомобилей.

При решении задач компьютерного зрения активно применяются сверточные нейронные сети. Особо отметим следующие яркие отраслевые представители последних: AlexNet, ResNet, YOLO, и др. [4–6]. Несмотря на активное развитие отраслевых решений, сохраняется проблема решения задач, в которых встречаются следующие особенности в данных: повороты одних и тех же объектов, разная степень освещенности, изменение взаимного расположения в пространстве отдельных частей объектов. В работе [7] предложена архитектура капсульных нейронных сетей, способная расширить область задач, поддающихся решению с привлечением аппарата искусственных нейронных сетей. Положительные результаты достигнуты с помощью использования слоев, в которых применены так называемые «капсулы», которые, кроме привычных активационных карт, используют дополнительную структуру (информацию) для учета взаимосвязей отдельных частей объекта друг с другом. Еще одно преимущество капсульной архитектуры — требуется меньший объем обучающей выборки в силу большего объема извлекаемых из данных информации.

## Постановка задачи

Капсульные нейронные сети обладают значительным потенциалом по работе с двумерными изображениями и извлечению из них трехмерных признаков [8]. Появляющиеся возможности вызывают огромный интерес в исследованиях областей применения данной архитектуры, особенно применительно к области компьютерной графики. В настоящее время компьютерная графика затрагивает такие аспекты нашей жизни как: фильмы, игры, компьютерное моделирование реальных физических систем; тренажеры для водителей, пилотов и военных; трехмерное моделирование органов человека для детекции аномалий; моделирование протезов; наглядные обучающие материалы, и многое другое. Несмотря на серьезное развитие вычислительных устройств и программного обеспечения по работе с трехмерным моделированием, сам процесс часто может требовать использования дорогих технических решений и высококвалифицированных специалистов. Перечисленные аспекты еще раз мотивируют интерес изучения применимости архитектуры капсульных нейронных сетей к трехмерному моделированию, а также к анализу выгод и проблем.

Основная причина выполнения настоящей работы — факт, что большинство современных решений распознавания ключевых точек лица используются в коммерческих проектах. В связи с этим исследователи в своих работах не предоставляют подробную архитектуру своего решения и размеченные данные, на которых проводились исследования. Цель работы — разработка решения, сопоставимого по качеству с аналогами, но

полностью открытого и прозрачного для других исследователей.

### Достижения в области детекции ключевых точек лиц

Определение ключевых точек лица является актуальной и востребованной задачей на протяжении последних десятилетий. Процесс детекции ключевых точек лица (Facial Feature Point Detection, FFPD) обычно рассматривается как задача обучения с учителем, для решения которой используются наборы данных, размеченные вручную. Существует широкий спектр методов, позволяющих выявлять ключевые точки на изображениях человеческих лиц. В работе [9] разделены существующие подходы на две глобальные категории: параметрические (в которых данные подчинены известным вероятностным распределениям с настраиваемыми параметрами) и непараметрические (которые непосредственно строят зависимости) (рис. 1).

FFPD-методы по критерию категоризации «модель формы» (Categorization criterion: Shape model) делятся на:

— методы на основе параметрической модели формы (Parametric Shape Model-based Methods);
— методы на основе непараметрических моделей формы (Non-parametric Shape Model-based Methods).

Методы на основе параметрической модели формы по критерию категоризации «модель внешности» (Categorization criterion: Appearance model) делятся на:

— локальные методы на основе частей (Local Part-based Methods);
— холистические методы (Holistic Methods).

Методы на основе непараметрической модели формы по критерию категоризации «связь между формой и внешним видом» (Categorization criterion: Connection between Shape and Appearance) делятся на:

— образцовые методы (Exemplar-based Methods);
— методы на основе графических моделей (Graphical Model-based Methods);
— методы каскадной регрессии (Cascaded Regression-based Methods);
— методы, основанные на глубоком обучении (Deep Learning-based Methods).

Параметрические модели основаны на двух предположениях: части лица людей (глаза, нос, губы, и т. д.) представляют собой локальные области, в рамках границ которых находятся ключевые точки конкретного лица (Local Part-based Methods); целое лицо рассматривается, как отдельный признак, игнорируя отдельные черты, такие как глаза, рот, нос и т. д. (Holistic Methods). На рис. 2 показаны ключевые точки для 600 изображений лиц (черные крестики); красными точками выделены средние значения для соответствующих ключевых точек. Видно, что ключевые точки действительно образуют некоторые области-кластеры, привычные для описания русским языком: глаза, брови, контуры лица, и др.

Примером холистического метода служит преобразование пикселов изображения лица в векторы признаков (рис. 3) или использование метода главных компонент для разложения изображения на «собственные лица» (eigenfaces) (рис. 4) для последующего их анализа [10].

Непараметрические методы основаны только лишь на данных, не предполагают построение статистических моделей и представлены широким набором различных подходов. Образцовые методы используют каждое изображение в качестве отдельного признака, при этом образцовые изображения содержат полные (все, встречающиеся при тесте) признаки.

Например, в работе [11] описанный подход использован для решения задачи детекции лица. Процесс обнаружения детектором лиц на основе образцов состоит в следующем: каждый образец обрабатывает тестовое изображение для получения карты достоверности (рис. 5, *a*); карты достоверности, полученные на уровне образцов, суммируются для построения окончательной карты достоверности (рис. 5, *b*). Далее становится возможным найти лица, ссылаясь на пики на окончательной карте достоверности (рис. 5, *c*).

Методы FFPD, работающие на базе графической модели, основаны на древовидной структуре и на марковских случайных полях. В методах на основе древовидной структуры каждая точка черты лица рассматривается как узел, а совокупность точек — как

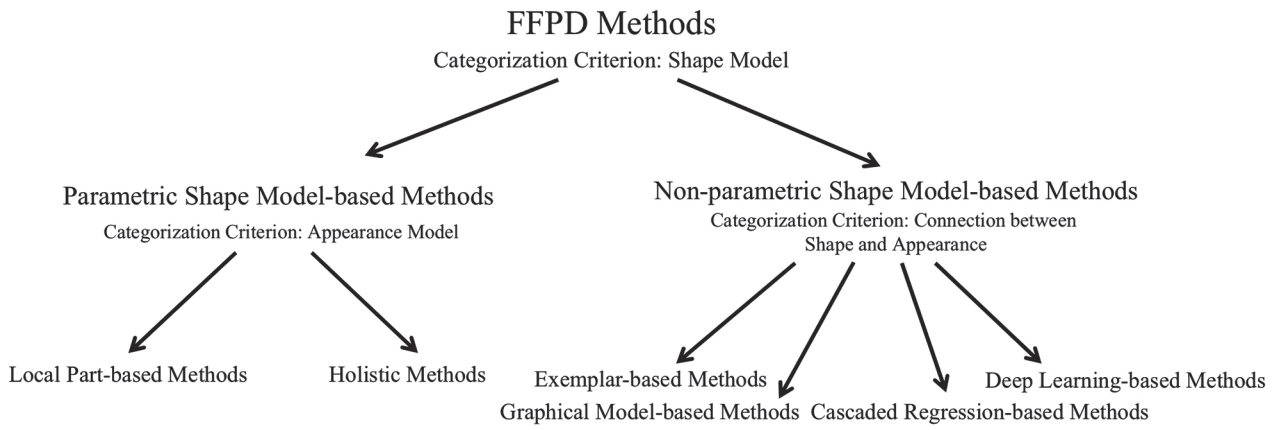



*Рис. 1*. Классификация методов FFPD [9]
*Fig. 1*. Classification of FFPD methods [9]

*Рис. 2*. Распределение ключевых точек для 600 изображений [9]
*Fig. 2*. Key points distribution for 600 images [9]

дерево. Расположение характерных точек лица можно оптимально определить с помощью динамического программирования. Наиболее популярный метод основан на каскадной регрессии. Этот метод подразумевает обучение серии регрессоров каскадным образом. При этом каждый из регрессоров уточняет значения предыдущих до получения истинных значений (рис. 6).

Развитие непараметрических методов во многом произошло благодаря применению методов на основе глубокого обучения. В работе [12] впервые рассмотрена комбинация сверточного подхода и принципа каскада регрессоров. Дальнейшее развитие, как, например, в работе [13], позволило кроме ключевых точек распознать и какие-то дополнительные детали. В [14] разработаны оптимизационные алгоритмы, предназначенные для решения задач компьютерного зрения, впоследствии усовершенствованные при помощи рекуррентных ней-

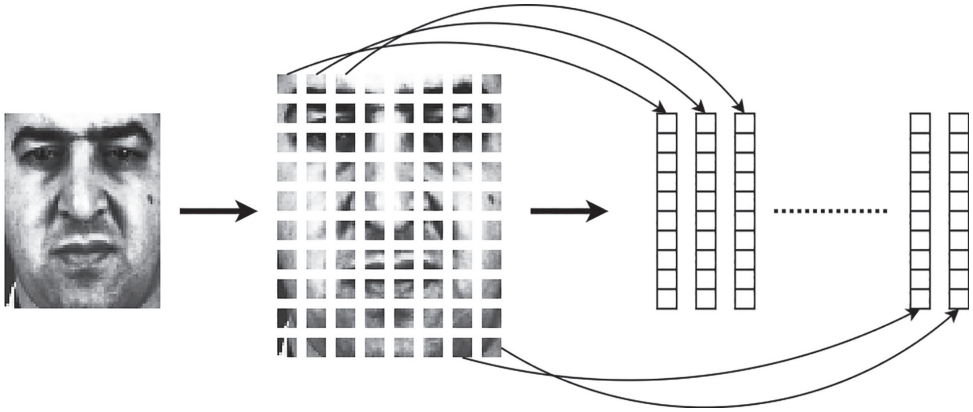


*Рис. 3*. Лицо представлено небольшим количеством признаков [10]
*Fig. 3*. Face is represented by a sufficiently small number of features [10]

*Рис. 4*. Собственные лица [10]
*Fig. 4*. The eigenfaces [10]

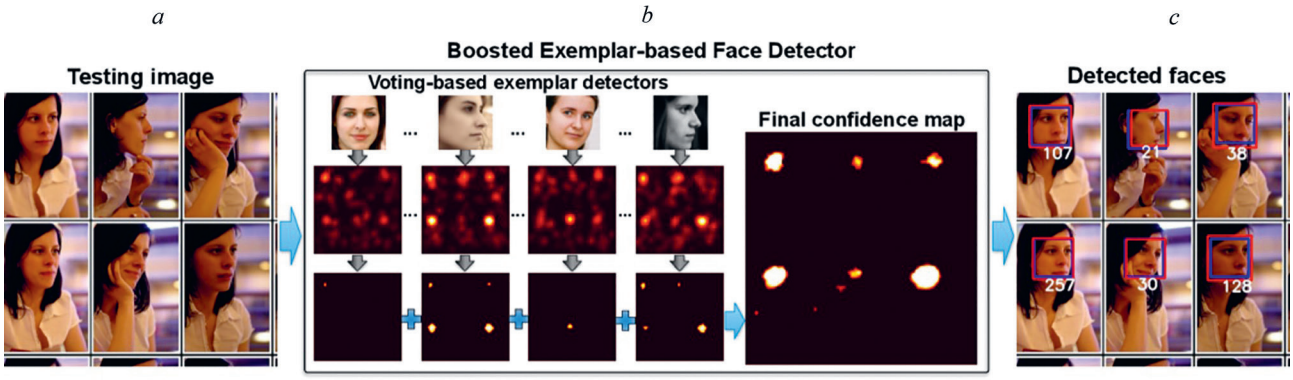



*Рис. 5*. Процесс обнаружения детектором лиц на основе образцов: тестовые изображения (*a*); карты достоверности на уровне образцов (*b*); поиск лиц по окончательной карте достоверности (*c*)

*Fig. 5*. The face-detection process: testing image (*a*); exemplar-level confidence maps (*b*); detection of faces based on the final confidence map (*c*)

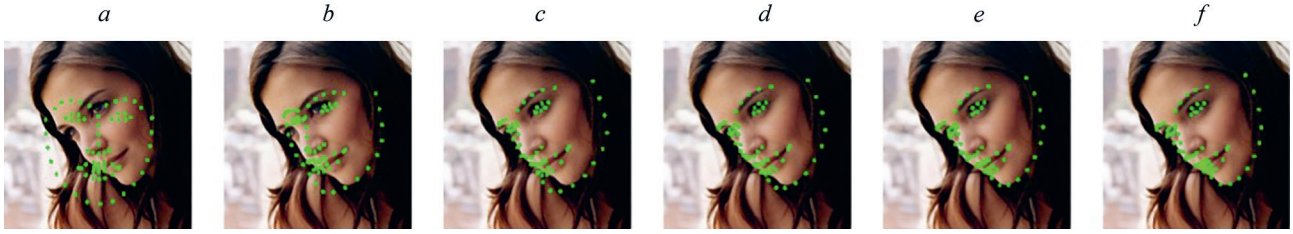



*Рис. 6*. Результат прогнозирования для набора данных 300W с использованием технологии обработки изображений cGPRT. Оценка контура лица инициализируется и итеративно обновляется через каскад деревьев регрессии: начальная оценка контура (*a*), оценки контура на разных этапах cGPRT (*b–f*)

*Fig. 6*. The result for 300W dataset using cGPRT technology. The face contour is initialized and updated using the regression trees cascade: initialization (*a*), contour correction using cGPRT (*b–f*)

ронных сетей. В работах [15, 16] описаны предсказания позы и ключевых точек лица в 3D и использовано сопоставление трехмерных морфируемых моделей с двумерным изображением лица. В работе [12] сопоставляются плотные морфируемые модели с помощью каскада сверточных сетей. В [13] итеративно сопоставляется морфируемая модель одной сверточной сетью, расширенной с помощью дополнительных каналов, с формами признаков на каждой итерации. Самое быстрое современное решение обработки изображений предложено в работе [17] от инженеров корпорации Google. Решение позволило на мобильном графическом процессоре обнаруживать 468 трехмерных точек лица со скоростью до 1000 кадров/с (рис. 7). Самые впечатляющие на данный момент результаты представлены в работе [18] от авторов из компании Microsoft. Предложенное решение способно с производительностью более 150 кадров/с на одном ядре CPU предсказывать плотную сетку лица из более 700 трехмерных точек с последующей трехмерной реконструкцией (рис. 8).

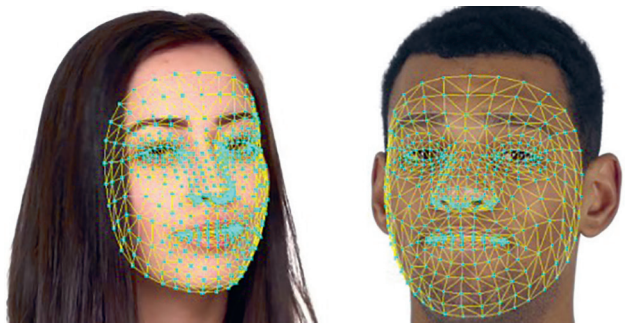


*Рис. 7*. Примеры предсказания сетки лица [17]

*Fig. 7*. Examples of facial keypoints prediction [17]

### Мотивация введения капсульных архитектур

Так как сверточные архитектуры имеют недостатки, то для их устранения разработаны капсульные архитектуры. Рассмотрим основные недостатки.

1. **Угасание признаков при использовании субдискретизации.** Несмотря на то, что Max Pooling и Average Pooling слои призваны уменьшать объем вычислений и увеличивать скорость обучения сети, они же приводят к потере информации при осуществлении прямого прохода. Например, на рис. 9 на вход подается карта признака, после этого к ней применяются алгоритмы Max Pooling и Average Pooling. Видно, что после обработки изображений алгоритмы приводят к угасанию сигнала. С учетом того, что первые слои сверточных сетей состоят из подобных кривых, происходит серьезная потеря информации из данных.
2. **Неспособность к извлечению информации о пространственных отношениях между объектами.**

*Рис. 8*. Примеры предсказания сетки лица и последующая 3D-реконструкция [18]
*Fig. 8*. Examples of facial keypoints prediction and 3-D reconstruction [18]

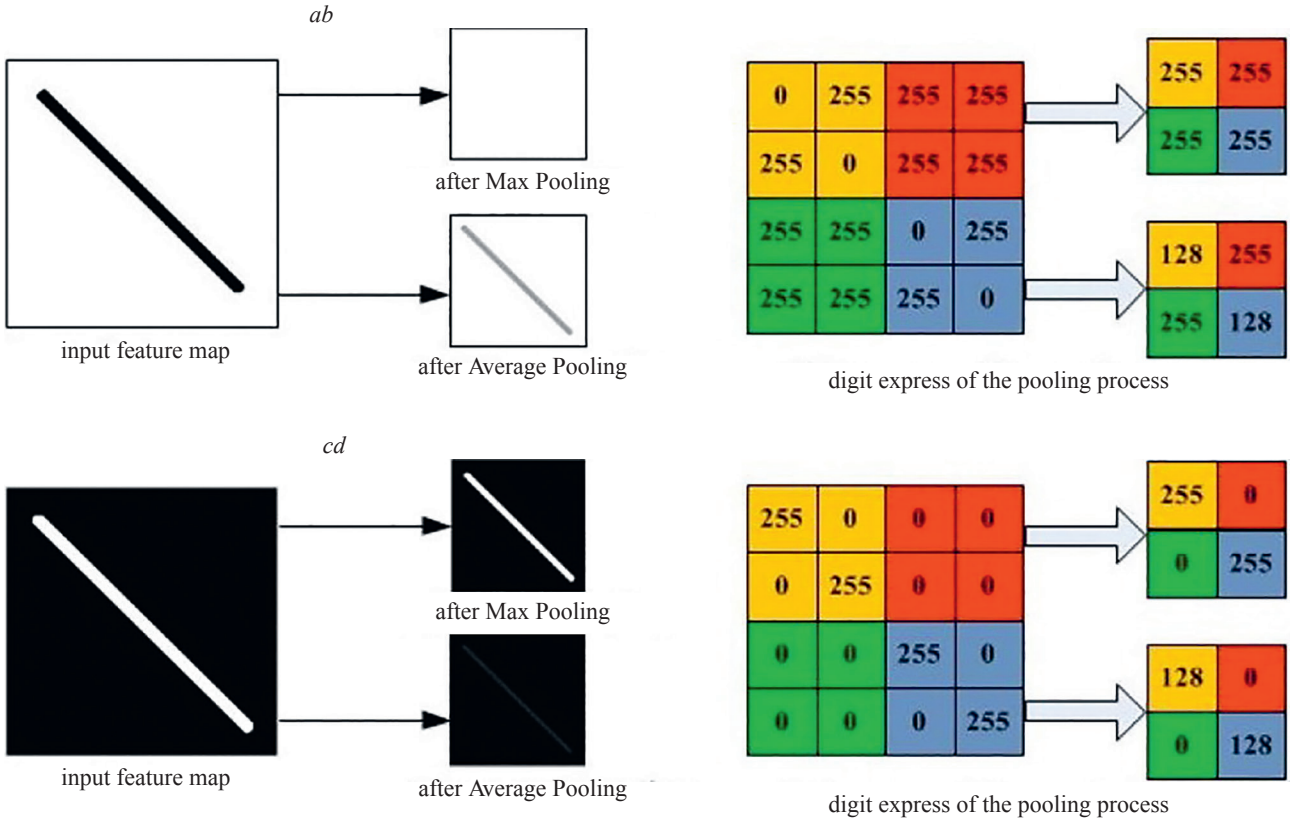



*Рис. 9*. Примеры затухания информации в слоях субдискретизации [7] для алгоритмов Max Pooling (*a*, *b*) и Average Pooling (*c*, *d*), в графическом (*a*, *c*) и цифровом (*b*, *d*) режимах
*Fig. 9*. Examples of information vanishing in pooling layers [7] in case of Max Pooling (*a*, *b*) and Average Pooling (*c*, *d*). In graphical (*a*, *c*) and digital (*b*, *d*) cases, correspondingly

Операция свертки, используемая в классических сверточных архитектурах, способна обнаруживать в данных определенные признаки, но не способна отслеживать их взаимное расположение друг относительно друга. Это может приводить (и приводит) к ошибкам и неточностям в работе моделей [19]. Пример показан на рис. 10. Сверточная нейронная сеть в обоих случаях классифицирует объект как лицо, но второй случай лицом не является.

3. **Отсутствие свойства пространственной инвариантности.** Для формирования представления о том, как объект выглядит «с разных сторон», свер-

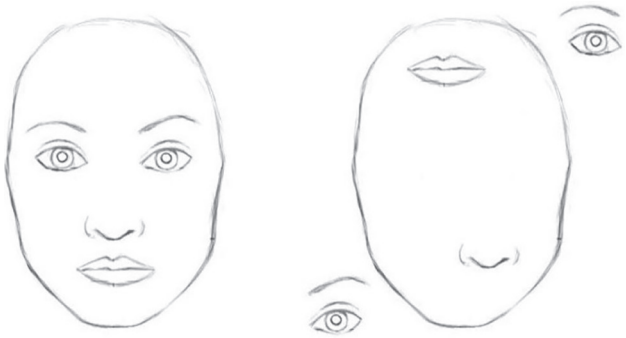


*Рис. 10*. Пример ошибочного определения наличия лица [19]
*Fig. 10*. An example of incorrect face detection [19]

точным архитектурам обычно необходимо огромное количество данных. Это приводит к необходимости сбора большого объема данных одной природы, и к длительному обучению модели. То, что понятно человеку, оказывается совершенной новинкой для классических сверток. На рис. 11 изображена статуя Свободы, которая легко узнаваема с разных углов обзора и с разным уровнем освещения для человека. Но для распознавания данного объекта сверточной нейронной сетью при обучении потребуется своя подвыборка данных для каждого из вариантов изображений.

Перечисленные проблемы возможно исключить при помощи капсульных архитектур. Работа [7] является одной из первых, в которой рассмотрены идеи капсульных архитектур.

### Модификация капсульной архитектуры DeepCaps

Рассмотрим модификацию архитектуры, предназначенной для решения поставленной цели работы и по совместительству являющейся одной из самых популярных на данный момент — архитектуры DeepCaps [20]. Архитектура состоит из одного классического сверточного слоя, 16 капсульных слоев (четыре из которых — слои со Skip Connections), а также одного полносвязного слоя, используемого для определения координат ключевых точек лица (решения задачи регрессии относительно координат этих точек).

Модификация архитектуры DeepCaps состоит в следующем.

1. В работе [20] выполнено тестирование архитектуры DeepCaps в задаче классификации на наборах данных MNIST, FashionMNIST и SVHN. Размеры входных данных (ширина × высота в пикселах, число каналов) для черно-белых изображений равны (64 × 64, 1), для цветных изображений — (64 × 64, 3). В настоящей работе использован размер данных (128 × 128, 3). Изменение размера связано с тем, что для данных меньшего размера значительно уменьшалось качество работы обученной модели, а для данных большего размера — возрастала вычислительная сложность при обучении и использовании модели, а также ухудшалась способность модели обобщать признаки.
2. На вход предлагаемой архитектуры поступали не только исходные изображения, но и изображения, полученные в результате их семантической сегментации. Сегментация получена при помощи заранее предобученной сети U-net. Такой подход позволил извлечь дополнительные признаки из исходного изображения и повысить способность сети к обобщению.
3. В результате сеть на вход получила данные, размер которых равен (128 × 128, 6) — склейка исходных трехканальных изображений и их сегментированных трехканальных копий.
4. В исходной архитектуре, решающей задачу классификации, выходы нейронов выходного слоя, количество которых равно количеству классов, нормировались, чтобы на выходе было получено совместное

*Рис. 11*. Примеры видов изображений при распознавании статуи Свободы [19]
*Fig. 11*. Examples of images types when recognizing the Statue of Liberty [19]

вероятностное распределение (сеть может определять несколько объектов на изображении и относить их к разным классам). В настоящей работе такой подход не годится: для каждой точки необходимо предсказывать сразу три координаты, причем связь предсказаний в рамках одного контекста оказывается крайне нежелательной. Это означает, что нерационально использовать три капсулы для каждой точки, поэтому задействовано по капсуле на точку, т. е. 468 капсул, каждая из которых выдает три координаты.

5. В качестве функции потерь берется евклидово расстояние между истинными и предсказанными координатами соответствующих точек. В качестве функции, показывающей точность предсказаний, выбрана индикаторная функция (весовая), которая за каждую попавшую в 0,01 шаровую окрестность точку добавляла 1/468 к точности.

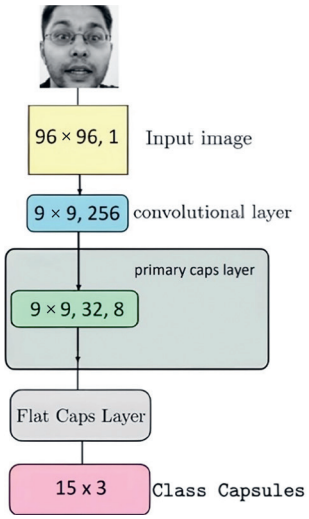



*Рис. 12*. Архитектура для детекции 15 точек лица
*Fig. 12*. 15 face-points detection architecture

### Полученные результаты

Проверка работоспособности капсульной нейронной сети выполнена в два этапа. Это связано с тем, что детекция трехмерной сетки лица является сложной задачей. Исходя из этого, сначала проведено предсказание небольшого количества ключевых точек лица в пространстве 2D. Для этого выполнена разработка простой архитектуры (рис. 12) на основании архитектуры, описанной в работе [7]. На данном этапе использован датасет, состоящий из изображений в градациях серого лиц людей, размеченный 15 ключевыми точками лица: уголки губ, центры верхней и нижней губ, нос, зрачки, уголки глаз, начало и конец бровных дуг. Примеры данных представлены на рис. 13, *a*. Решение оптими-

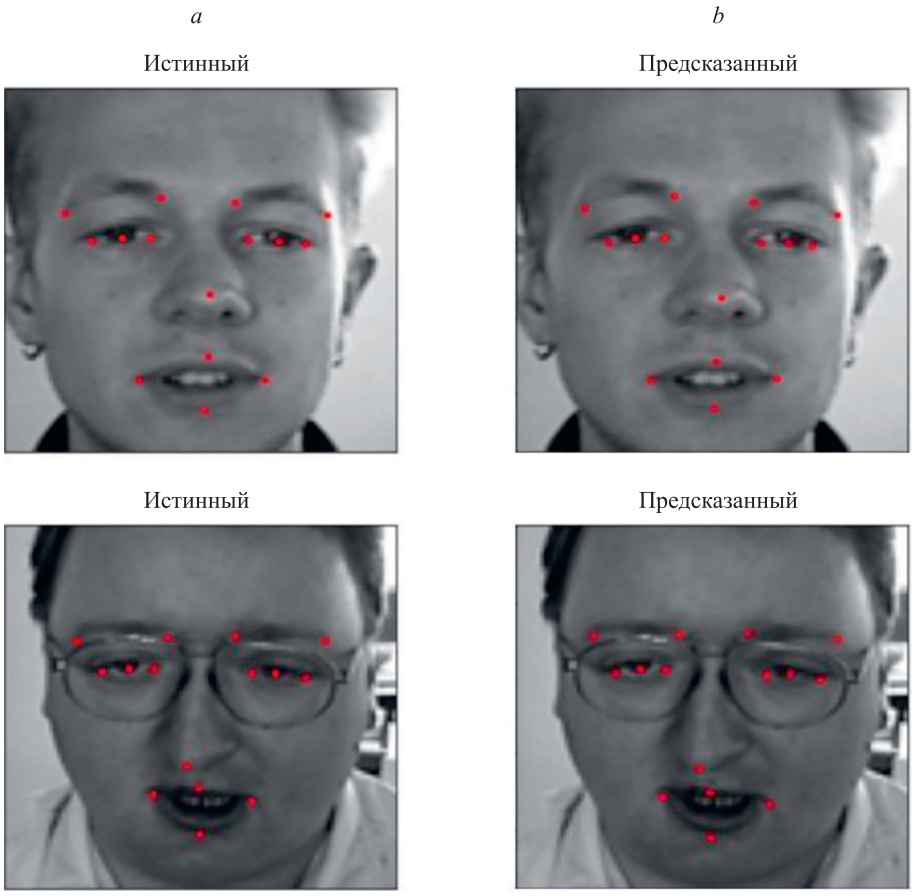



*Рис. 13*. Примеры работы обученной модели: истинное (*a*) и предсказанное (*b*) изображения
*Fig. 13*. Examples of model application: initial image (*a*) and model's prediction (*b*)

зационной задачи для модели выполнено на видеокарте GeForce RTX 3070. Обучение заняло 20 эпох, а общее время обучения — 54 мин. Значение функции потерь на тестовой выборке равно 0,2618, а функция точности на тестовой выборке достигла значения в среднем 0,8601. Данные значения не являются показательными, поэтому продемонстрируем результаты работы обученной модели на случайных данных из тестовой выборки (рис. 13, *b*) — расположение красных точек на предсказанных изображениях сопоставимо с точками на исходных изображениях.

Выполним проверку работоспособности архитектуры с помощью распознавания трехмерной сетки лица с относительно большим числом ключевых точек (468 точек). Используем данные, состоящие из 10 000 сгенерированных лиц (рис. 14, *a*), размеченные с помощью фреймворка MediaPipe (рис. 14, *b*).

На основе полученных данных осуществим сегментирование по распознанным с помощью MediaPipe ключевым точкам лица. Для построения сетки лица используем триангуляцию Делоне. Сегментацию проведем по отдельным сегментам: овал лица, левый и правый глаза, нос и рот (рис. 14, *c*).

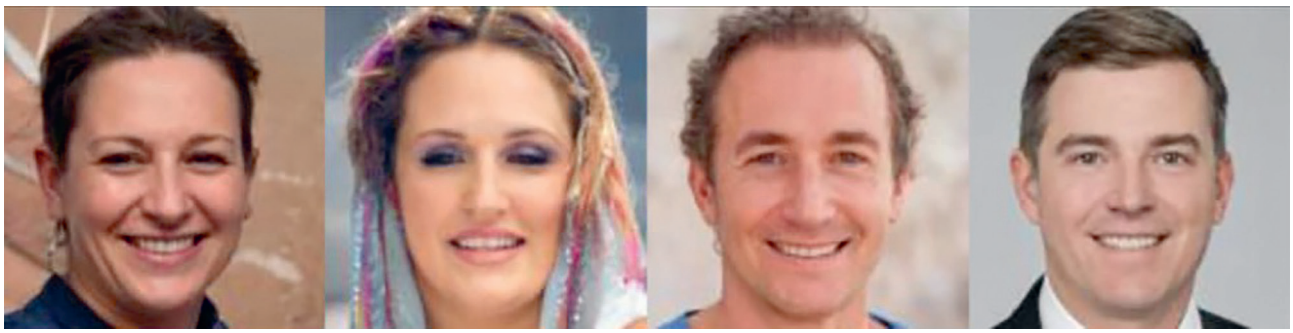
Полученную сегментацию используем в процессе обучения нейронной сети U-net [21] для решения задачи семантической сегментации (рис. 15).

Используем вторую нейронную архитектуру (DeepCaps) глубоких капсульных сетей. На рис. 16 показана схема разработанной архитектуры.

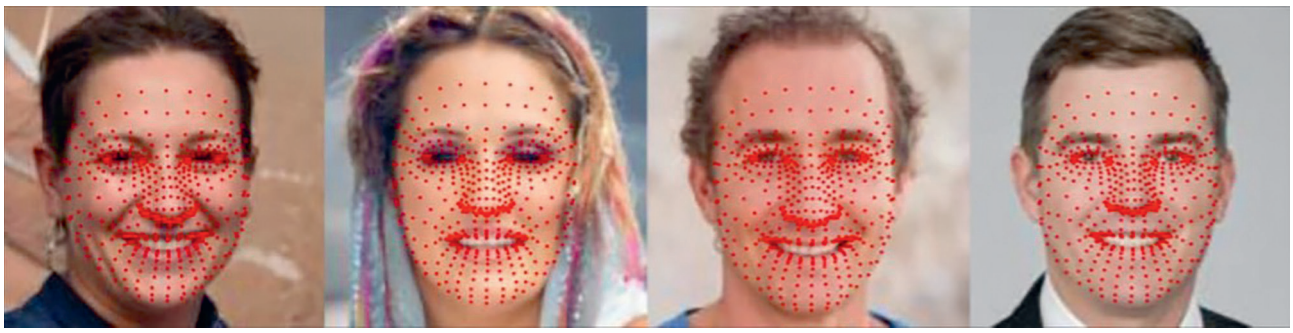
Предлагаемая архитектура с учетом семантической сегментации и предсказания сетки лица показана на рис. 17.

Обучение заняло 20 эпох, всего в выборке было 10 000 изображений, 1000 из них была отложена на тест. По результатам обучения и тестирования функция потерь во время обучения варьировалась в пределах 2,30–2,60, а точность — 0,9–0,94. Во время тестирования функция потерь достигла значений 2,50–2,90, а точность — 0,87–0,90. Пример работы обученной сети на тестовых данных представлен на рис. 18.

*a*



*b*



*c*

*Рис. 14*. Примеры: лиц из набора данных (*a*); разметки фреймворка MediaPipe (*b*) и двумерной сегментации (*c*)

*Fig. 14*. Examples: faces from the dataset (*a*); MediaPipe application (*b*) and 2D segmentation (*c*)

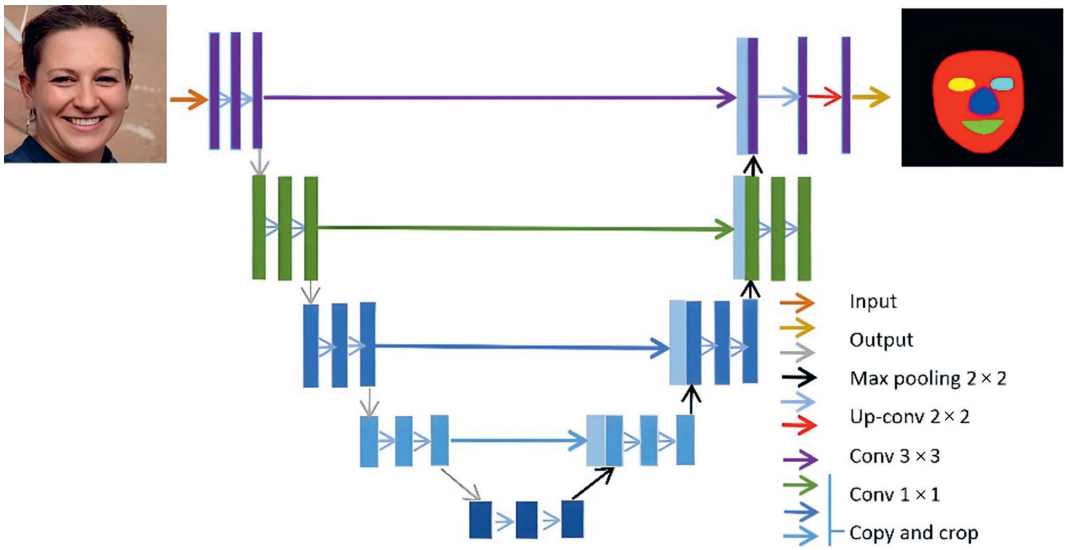



*Рис. 15*. Архитектура нейронной сети U-net [21]
*Fig. 15*. U-net architecture [21]

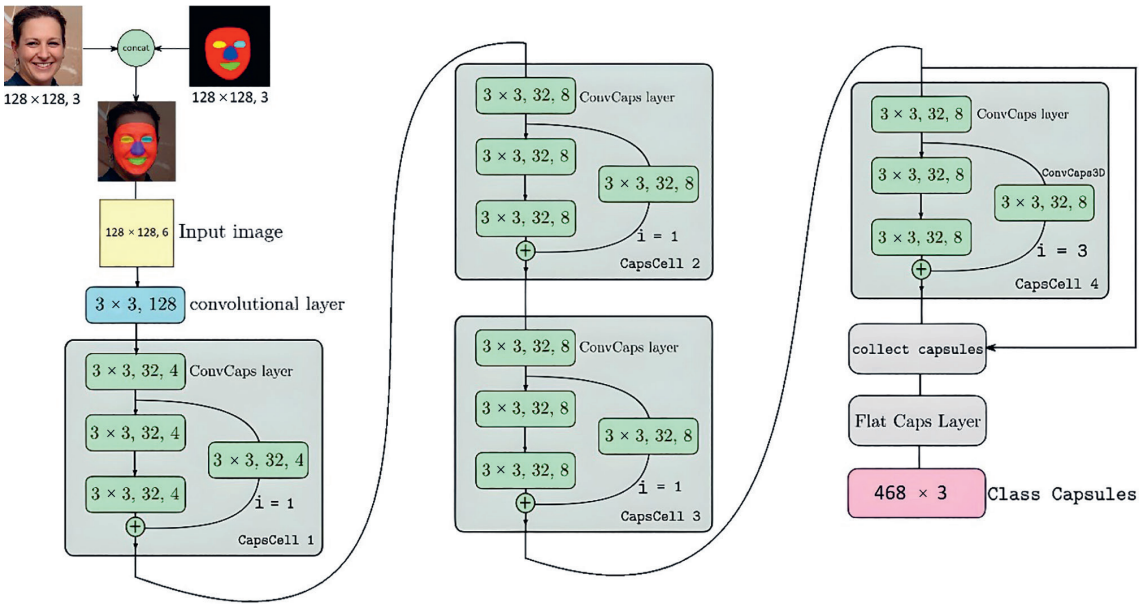



*Рис. 16*. Схема разработанной архитектуры предсказания сетки лица глубокой капсульной сетью
*Fig. 16*. Developed deep capsule network architecture for the facial keypoints detection

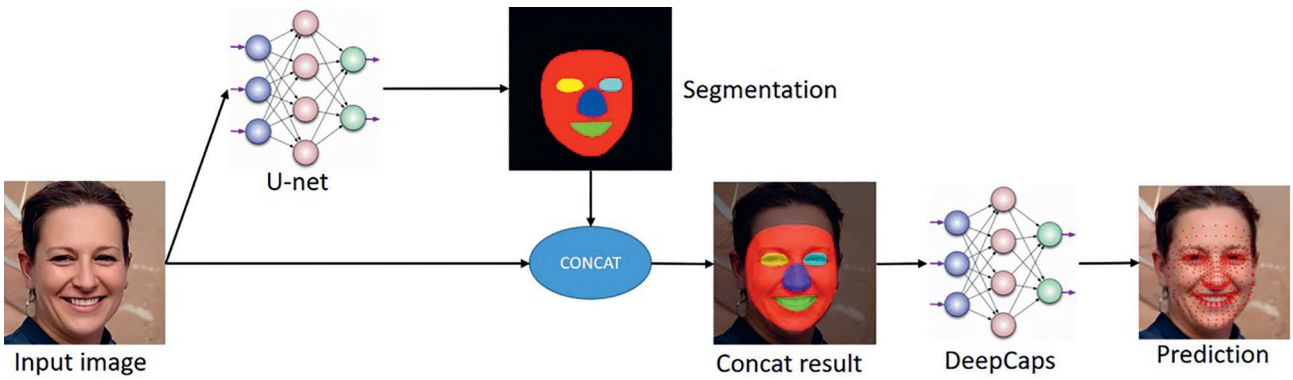



*Рис. 17*. Схема разработанной архитектуры с учетом сегментации и предсказания сетки лица глубокой капсульной сетью
*Fig. 17*. Developed deep capsule network architecture with additional segmentation for the facial keypoints detection

*Рис. 18*. Предсказание сетки лица глубокой капсульной сетью
*Fig. 18*. Deep capsule network's predictions

## Дискуссия

В связи с тем, что большинство современных решений по распознаванию ключевых точек лица используются в коммерческих проектах, исследователи, их создающие, не предоставляют подробной архитектуры своего решения и размеченных данных. В настоящей работе предложена архитектура, качественно сопоставимая с закрытыми решениями, однако являющаяся полностью открытой и доступной для других исследователей.

## Заключение

В результате работы сформирован набор данных для обучения и обучены две нейронные сети: капсульная сеть с динамической маршрутизацией и глубокая капсульная сеть. Обе сети показали высокие результаты в задаче детекции ключевых точек лица. Первая нейросеть распознавала 15 двумерных ключевых точек лица по набору данных, взятого из базы сайта Kaggle. Получена точность равная 0,86, что не является идеальным показателем, но при визуальном сравнении примеров, не входящих в обучающую выборку, результат оказался высоким. Вторая нейросеть распознавала 468 трехмерных ключевых точек лица. В открытом доступе отсутствуют наборы данных для решения такой задачи, поэтому был использован набор данных без разметки. Далее выполнена разметка с помощью решения корпорации Google — фреймворк MediaPipe, которая использована в обучении. В результате получена точность 0,90, что при визуальном сравнении дает крайне высокий результат, почти не отличимый от решений MediaPipe. Капсульные нейронные сети показали высокие результаты в решении задачи по детекции ключевых точек лица. На основе полученных результатов можно сделать вывод, что разработанное решение может быть использовано в технологиях по сопоставлению геометрии сетки лица реального человека с геометрией сетки лица трехмерной модели.

## Литература


1. Волкова С.С., Матвеев Ю.Н. Применение сверточных нейронных сетей для решения задачи противодействия атаке спуфинга в системах лицевой биометрии // Научно-технический вестник информационных технологий, механики и оптики. 2017. Т. 17. № 4. С. 702–710. https://doi.org/10.17586/2226-1494-2017-17-4-702-710
2. Дикий Д.И., Артемьева В.Д. Исследование применимости искусственных нейронных сетей для верификации пользователей по динамике почерка // Научно-технический вестник информационных технологий, механики и оптики. 2017. Т. 17. № 4. С. 677–684. https://doi.org/10.17586/2226-1494-2017-17-4-677-684
3. Abiodun O.I., Kiru M.U., Jantan A., Omolara A.E., Dada K.V., Umar A.M., Linus O.U., Arshad H., Kazaure A.A., Gana U. Comprehensive review of artificial neural network applications to pattern recognition // IEEE Access. 2019. V. 7. P. 158820–158846. https://doi.org/10.1109/access.2019.2945545
4. Krizhevsky A., Sutskever I., Hinton G.E. ImageNet classification with deep convolutional neural networks // Communications of the ACM. 2017. V. 60. N 6. P. 84–90. https://doi.org/10.1145/3065386
5. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 770–778. https://doi.org/10.1109/cvpr.2016.90
6. Redmon J., Divvala S., Girshick R., Farhadi A. You only look once: Unified, real-time object detection // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 779–788. https://doi.org/10.1109/cvpr.2016.91


## References


1. Volkova S.S., Matveev Yu.N. Convolutional neural networks for face anti-spoofing. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2017, vol. 17, no. 4, pp. 702–710. (in Russian). https://doi.org/10.17586/2226-1494-2017-17-4-702-710
2. Dikiy D.I., Artemeva V.D. Research of artificial neural network applicability for user`s online handwritten signature verification. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2017, vol. 17, no. 4, pp. 677–684. (in Russian). https://doi.org/10.17586/2226-1494-2017-17-4-677-684
3. Abiodun O.I., Kiru M.U., Jantan A., Omolara A.E., Dada K.V., Umar A.M., Linus O.U., Arshad H., Kazaure A.A., Gana U. Comprehensive review of artificial neural network applications to pattern recognition. *IEEE Access*, 2019, vol. 7, pp. 158820–158846. https://doi.org/10.1109/access.2019.2945545
4. Krizhevsky A., Sutskever I., Hinton G.E. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 2017, vol. 60, no. 6, pp. 84–90. https://doi.org/10.1145/3065386
5. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. https://doi.org/10.1109/cvpr.2016.90
6. Redmon J., Divvala S., Girshick R., Farhadi A. You only look once: Unified, real-time object detection. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788. https://doi.org/10.1109/cvpr.2016.91

7. Sabour S., Frosst N., Hinton G.E. Dynamic routing between capsules // Advances in Neural Information Processing Systems. 2017. V. 30. P. 3856–3866.
8. Nguyen H.H., Yamagishi J., Echizen I. Capsule-forensics: Using capsule networks to detect forged images and videos // Proc. of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2019. P. 2307–2311. https://doi.org/10.1109/icassp.2019.8682602
9. Wang N., Gao X., Tao D., Yang H., Li X. Facial feature point detection: A comprehensive survey // Neurocomputing. 2018. V. 275. P. 50–65. https://doi.org/10.1016/j.neucom.2017.05.013
10. Beham M.P., Roomi S.M.M. A review of face recognition methods // International Journal of Pattern Recognition and Artificial Intelligence. 2013. V. 27. N 4. P. 1356005. https://doi.org/10.1142/S0218001413560053
11. Li H., Lin Z.L., Brandt J., Shen X., Hua G. Efficient boosted exemplar-based face detection // Proc. of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. 2014. P. 1843–1850. https://doi.org/10.1109/cvpr.2014.238
12. Sun Y., Wang X., Tang X. Deep convolutional network cascade for facial point detection // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition. 2013. P. 3476–3483. https://doi.org/10.1109/cvpr.2013.446
13. Zhang Z., Luo P., Loy C.C., Tang X. Facial landmark detection by deep multi-task learning // Lecture Notes in Computer Science. 2014. V. 8694. P. 94–108. https://doi.org/10.1007/978-3-319-10599-4_7
14. Trigeorgis G., Snape P., Nicolaou M.A., Antonakos E., Zafeiriou S. Mnemonic descent method: A recurrent process applied for end-to-end face alignment // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 4177–4187. https://doi.org/10.1109/cvpr.2016.453
15. Zhu X., Lei Z., Liu X., Shi H., Li S.Z. Face alignment across large poses: A 3D solution // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 146–155. https://doi.org/10.1109/cvpr.2016.23
16. Jourabloo A., Liu X. Large-pose face alignment via CNN-based dense 3D model fitting // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 4188–4196. https://doi.org/10.1109/cvpr.2016.454
17. Kartynnik Y., Ablavatski A., Grishchenko I., Grundmann M. Real-time facial surface geometry from monocular video on mobile GPUs // arXiv. 2019. arXiv:1907.06724. https://doi.org/10.48550/arXiv.1907.06724
18. Wood E., Baltrušaitis T., Hewitt Ch., Johnson M., Shen J., Milosavljević N., Wilde D., Garbin S., Sharp T., Stojiljković I., Cashman T., Valentin J. 3D face reconstruction with dense landmarks // Lecture Notes in Computer Science. 2022. V. 13673. P. 160–177. https://doi.org/10.1007/978-3-031-19778-9_10
19. Pechyonkin M. Understanding Hinton's Capsule Networks. Part I: Intuition. Medium. 2018, December 18 [Электронный ресурс]. URL: https://medium.com/ai%C2%B3-theory-practice-business/understanding-hintons-capsule-networks-part-i-intuition-b4b559d1159b (дата обращения: 12.12.2022).
20. Yu D., Wang H., Chen P., Wei Z. Mixed pooling for convolutional neural networks // Lecture Notes in Computer Science. 2014. V. 8818. P. 364–375. https://doi.org/10.1007/978-3-319-11740-9_34
21. Ding Y., Chen F., Zhao Y., Wu Z., Zhang C., Wu D. A stacked multi-connection simple reducing net for brain tumor segmentation // IEEE Access. 2019. V. 7. P. 104011–104024. https://doi.org/10.1109/access.2019.2926448



7. Sabour S., Frosst N., Hinton G.E. Dynamic routing between capsules. Advances in Neural Information Processing Systems, 2017, vol. 30, pp. 3856–3866.
8. Nguyen H.H., Yamagishi J., Echizen I. Capsule-forensics: Using capsule networks to detect forged images and videos. *Proc. of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 2307–2311. https://doi.org/10.1109/icassp.2019.8682602
9. Wang N., Gao X., Tao D., Yang H., Li X. Facial feature point detection: A comprehensive survey. *Neurocomputing*, 2018, vol. 275, pp. 50–65. https://doi.org/10.1016/j.neucom.2017.05.013
10. Beham M.P., Roomi S.M.M. A review of face recognition methods. *International Journal of Pattern Recognition and Artificial Intelligence*, 2013, vol. 27, no. 4, pp. 1356005. https://doi.org/10.1142/S0218001413560053
11. Li H., Lin Z.L., Brandt J., Shen X., Hua G. Efficient boosted exemplar-based face detection. *Proc. of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1843–1850. https://doi.org/10.1109/cvpr.2014.238
12. Sun Y., Wang X., Tang X. Deep convolutional network cascade for facial point detection. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3476–3483. https://doi.org/10.1109/cvpr.2013.446
13. Zhang Z., Luo P., Loy C.C., Tang X. Facial landmark detection by deep multi-task learning. *Lecture Notes in Computer Science*, 2014, vol. 8694, pp. 94–108. https://doi.org/10.1007/978-3-319-10599-4_7
14. Trigeorgis G., Snape P., Nicolaou M.A., Antonakos E., Zafeiriou S. Mnemonic descent method: A recurrent process applied for end-to-end face alignment. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 4177–4187. https://doi.org/10.1109/cvpr.2016.453
15. Zhu X., Lei Z., Liu X., Shi H., Li S.Z. Face alignment across large poses: A 3D solution. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 146–155. https://doi.org/10.1109/cvpr.2016.23
16. Jourabloo A., Liu X. Large-pose face alignment via CNN-based dense 3D model fitting. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 4188–4196. https://doi.org/10.1109/cvpr.2016.454
17. Kartynnik Y., Ablavatski A., Grishchenko I., Grundmann M. Real-time facial surface geometry from monocular video on mobile GPUs. *arXiv*, 2019, arXiv:1907.06724. https://doi.org/10.48550/arXiv.1907.06724
18. Wood E., Baltrušaitis T., Hewitt Ch., Johnson M., Shen J., Milosavljević N., Wilde D., Garbin S., Sharp T., Stojiljković I., Cashman T., Valentin J. 3D face reconstruction with dense landmarks. *Lecture Notes in Computer Science*, 2022, vol. 13673, pp. 160–177. https://doi.org/10.1007/978-3-031-19778-9_10
19. Pechyonkin M. *Understanding Hinton's Capsule Networks. Part I: Intuition. Medium*. 2018, December 18. Available at: https://medium.com/ai%C2%B3-theory-practice-business/understanding-hintons-capsule-networks-part-i-intuition-b4b559d1159b (accessed: 12.12.2022).
20. Yu D., Wang H., Chen P., Wei Z. Mixed pooling for convolutional neural networks. *Lecture Notes in Computer Science*, 2014, vol. 8818, pp. 364–375. https://doi.org/10.1007/978-3-319-11740-9_34
21. Ding Y., Chen F., Zhao Y., Wu Z., Zhang C., Wu D. A stacked multi-connection simple reducing net for brain tumor segmentation. *IEEE Access*, 2019, vol. 7, pp. 104011–104024. https://doi.org/10.1109/access.2019.2926448


## Авторы


**Бойцев Антон Александрович** — кандидат физико-математических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, SC 56401063400, https://orcid.org/0000-0002-3374-8256, boitsevanton@gmail.com

**Волчек Дмитрий Геннадьевич** — кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, SC 57197732532, https://orcid.org/0000-0002-0310-1654, dvolchek@itmo.ru

**Магазенков Егор Николаевич** — студент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0000-0002-7563-0846, egormaga04@mail.ru


## Authors


**Anton A. Boitsev** — PhD (Physics & Mathematics), Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, SC 56401063400, https://orcid.org/0000-0002-3374-8256, boitsevanton@gmail.com

**Dmitry G. Volchek** — PhD, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, SC 57197732532, https://orcid.org/0000-0002-0310-1654, dvolchek@itmo.ru

**Egor N. Magazenkov** — Student, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid.org/0000-0002-7563-0846, egormaga04@mail.ru

**Неваев Максим Кириллович** — системный проектировщик, ЗАО «Центр финансовых технологий», Санкт-Петербург, 191002, Российская Федерация, https://orcid.org/0000-0002-9000-7841, m.nevaev@alumni.nsu.ru
**Романов Алексей Андреевич** — кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 57194976341, https://orcid.org/0000-0002-6991-464X, romanov@itmo.ru

**Maxim K. Nevaev** — Systems Designer, ZAO "Center of Financial Technologies", Saint Petersburg, 191002, Russian Federation, https://orcid.org/0000-0002-9000-7841, m.nevaev@alumni.nsu.ru

**Aleksei A. Romanov** — PhD, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 57194976341, https://orcid.org/0000-0002-6991-464X, romanov@itmo.ru

*Статья поступила в редакцию 25.01.2023*
*Одобрена после рецензирования 22.02.2023*
*Принята к печати 16.05.2023*

*Received 25.01.2023*
*Approved after reviewing 22.02.2023*
*Accepted 16.05.2023*