Моделирование и анализ данных 2025. Том 15. № 1. С. 7–18 ISSN: 2219-3758 (печатный) ISSN: 2311-9454 (online) © 2025 ΦΓБΟΥ ΒΟ ΜΓΠΠΥ

◇◇◇◇◇◇◇◇АНАЛИЗ ДАННЫХ◇◇◇◇◇◇◇

УЛК 004.8

# Обнаружение частей человеческого тела с помощью карт достоверности и карт ориентиров

### Потенко М.А.\*

Московский авиационный институт (национальный исследовательский университет) (МАИ) г. Москва, Российская Федерация ORCID: https://orcid.org/0009-0008-5222-2664 e-mail: potenkog@gmail.com

В данной статье рассматривается проблема поиска ключевых точек на объекте и определение его составных частей на примере человеческого тела, как одной из самых актуальных задач компьютерного зрения. Задача заключается в том, что для изображения с предполагаемым объектом (в данной публикации человеком) требуется найти положение всех важных составных частей/ точек, у человека таковыми являются суставы: плечи, локти, кисти, колени и т.д. Для решения данной проблемы предлагается использовать нейронные сети с применением «тепловых карт», которые представлены в виде карты достоверностей и карты ориентиров. Карта достоверности представляет собой матрицу, соответствующую исходному изображению, в каждой ячейке которой находится значение от 0 до 1 с вероятностью расположения искомого сустава (особой точки) в соответствующем пикселе. Карта ориентиров представляет собой матрицу, в каждой ячейке которой располагается двумерный вектор с направлением расположения следующего сустава. Главная особенность заключается в том, что карты помогают друг другу предсказывать правильные значения: благодаря карте достоверности, карта ориентиров знает примерное местоположение ориентиров и ей проще определить нужное направление и наоборот, зная примерное направление к следующему ориентиру, проще спрогнозировать местоположение ориентира. Данные карты вычисляются в несколько этапов, и на каждом новом этапе используется информация с предыдущего, таким образом, при вычислении каждой последующей карты точность полученных результатов повышается. В данной работе оптимальным количеством этапов было 6-7, но данное значение может меняться в зависимости от конечной архитектуры. Демонстрации работы происходит на наборе данных СОСО, включающая в себя 18 точек для каждого человеческого тела.

*Ключевые слова:* нейронные сети, свёрточные нейронные сети, Human Pose Estimation, тепловые карты, обработка изображений, ключевые точки, обнаружение объектов, сегментация объектов

#### Для цитаты:

Потенко М.А. Обнаружение частей человеческого тела с помощью карт достоверности и карт ориентиров // Моделирование и анализ данных. 2025. Том 15. № 1. С. 7–18. DOI: https://doi.org/10.17759/mda.2025150101

\*Потенко Максим Алексеевич, аспирант, Московский авиационный институт (национальный исследовательский университет) (МАИ), г. Москва, Российская Федерация, ORCID: https:// orcid.org/0009-0008-5222-2664, e-mail: potenkog@gmail.com

### 1. ВВЕДЕНИЕ

Обнаружение ключевых точек – одна из важнейших задач компьютерного зрения, на изображении располагается объект или группа объектов, у которых визуально можно выделить точки(пиксели), несущие определенную информацию, это могут быть границы автомобильного номера или дорожного знака, отдельные части автомобиля или какого-либо оборудования, суставы человеческого тела и другие. Эти точки могут применяться в огромном числе приложений, таких как распознавание объектов, отслеживание перемещений, построение 3D моделей и другие.

Данную задачу важно автоматизировать особенно в ситуациях, связанных с обработкой видео, содержащих большое число кадров, но это трудно из-за ряда проблем и ограничений:

- Изображения нередко бывают зашумлены
- Изменчивость среды. Фоны изображений, условия освещённости, особенности работы камеры могут значительно меняться
- Многие объекты имеют высокую степень свободы: могут вращаться, менять размер, положение в пространстве и точек относительно друг друга.
- Окклюзия. Части объекта могут перекрывать друг друга и тем самым мешать определять их положение.
- Работа в режиме реального времени. Для многих приложений необходимо, чтобы скорость работы была достаточно высокой, чтобы успевать обрабатывать отдельные кадры видео в режиме реального времени.

Human Pose Estimation область компьютерного зрения направленная на поиск и распознавание человеческого тела, его отдельных частей, определение их ориентации в пространстве. Есть различные подходы к решению данной задачи. В исследовании упор делается на подходе, связанным с поиском ключевых точек, таких как плечи, локти, кисти, колени, ступни, таз и т.д. Решение данной задачи актуально поскольку имеет широкие возможности для применения:

• Спорт. Автоматическое отслеживание спортсменов и анализ их движений. Можно проверять правильность выполнение упражнений, чтобы предотвратить травмы, давать рекомендации. Подсчёт количества подходов в упражнениях.



- VR/AR технологии. Можно взаимодействовать с реальными людьми в VR или AR пространстве.
- Мультипликация, анимация. Можно считывать движения людей, чтобы реалистично их представить в фильмах, сериалах, мультфильмах или компьютерных играх.
- Безопасность. Идентификация людей по движениям, отслеживание воровства.
- Умная обработка изображений. Удаление лишних людей с фото, замена фона, смена одежды и т.д.
- Медицина. Несмотря на то, что часть задач решается без применения ключевых точек [1], другая часть нуждается в их использовании, например анализ прогресса в восстановлении двигательной активности.
- Психология, реклама и т.д.
   К классическим алгоритмам решения задач поиска ключевых точек можно отнести:
- Harris Corner Detector[2]: Один из первых алгоритмов для поиска углов на изображении. Был предложен ещё в 1988 году и использует матрицу градиентов для определения угловых точек в изображении. Оценивает локальные изменения интенсивности в окрестностях пикселя и выдает точки, где эти изменения максимальны. Данный способ прост в реализации и довольно быстр, но крайне чувствителен к шуму и дает много ложных срабатываний, плохо устойчив к изменениям масштаба, поворотам и другим сложным сценам, а также не предоставляет дескрипторов точек, что усложняет их идентификацию и сопоставление.
- SIFT[3] (Scale-Invariant Feature Transform): находит ключевые точки и описывает их с помощью векторов признаков, инвариантных к масштабу и повороту. Алгоритм 1999 год, лучше справляется с инвариантностью масштабов и поворотов, чем предыдущий алгоритм, а также предоставляет дескрипторы точек. Проблемой алгоритма является его медленная работа.
- SURF[4] (Speeded-Up Robust Features): Улучшенная версия SIFT, более быстрая и эффективная. Был разработан в 2006 году. Несмотря на то, что алгоритм был доработан, он все еще слишком медленный в реальной работе.

Классические алгоритмы уступили место современным подходам, поскольку несмотря на все попытки их улучшить они все ещё плохо справлялись с инвариантностью условий и выдавали низкую точность с низкой скоростью работы. К тому же их применение сильно ограничено, поскольку эти алгоритмы не обладают достаточной универсальностью и многие точки в ряде задач не способны распознавать.

Все современные подходы основаны на нейронных сетях. В отличие от классических подходов, нейронные сети в теории можно сконфигурировать и обучить для решения любой задачи, связанной с поиском особых точек. Примеров таких решений могут являться

- Keypoint R-CNN[5] Модель, основанная на Faster R-CNN, которая добавляет ветвь для предсказания ключевых точек.
- AlpaPose[6]: Используется для оценки позы человека, выделяя ключевые точки на теле.
- OpenPose[7]: Схож с предыдущим вариантом.

Недостатком существующих систем главным образом является их низкая скорость работы и высокое потребление ресурсов, что сильно ограничивает область возможного применения. Алгоритмы с высокой скоростью обладают недостаточной точностью для применения в реальной работе.

В данной работе предлагается способ, который способен быстро решать задачу без потери точности. Обучение и тестирование выполнялось на наборе данных СОСО, но данный набор можно улучшить и расширит путём обработки изображений, например применить удаление теней с фотографии [8].

# 2. ТОПОЛОГИЯ МОДЕЛИ ЧЕЛОВЕЧЕСКОГО ТЕЛА

Человеческое тело можно представить в виде разного количества суставов или вообще без них в зависимости от предполагаемого применения.

В данной статье будет использоваться топология СОСО [9], которая является практически стандартом для представления человеческого тела. В данной топологии человеческое тело представлено в виде 18 точек. Схема данной топологии представлена на рис. 1.



Рис. 1. Модель человеческого тела



Рис. 2. Запись точек

Каждому изображению ставится в сопоставление список, включающий в себя информацию по каждому человеку на изображении, элементом списка является другой список с информацией по каждой отдельной точке. Точка состоит из двух значений: координаты х и у с её положением на изображении. Индекс каждой точки обязательно фиксирован, чтобы можно было идентифицировать её с определённой частью тела. Шаблон записи точек представлен на рис. 2.

# 3. КАРТЫ ДОСТОВЕРНОСТИ

Карта достоверностей представляет собой матрицу с размером как у исходного изображения (после некоторых преобразований). Идея была вдохновлена работой Alejandro Newell [10]. В каждой ячейке этой матрицы находится число от 0 до 1, обозначающее вероятность присутствия в соответствующем пикселе исходного изображения определённого сустава. Для каждого сустава строится своя карта достоверностей, поэтому общее число карт для набора СОСО равняется 18, еще один слой желательно выделить под фон. Упрощенный пример матрицы представлен в табл. 1. На ней демонстрируется матрица достоверности 10 на 10, с координатой максимума в точке (4, 3).

Таблица 1

0	0	0	0	0	0	0	0	0	0
0	0	0	0.4	0.5	0	0	0	0	0
0	0	0	0.6	0.71	0.58	0.34	0	0	0
0	0	0.45	0.72	0.9	0.67	0.65	0	0	0
0	0	0	0.65	0.76	0.59	0	0	0	0
0	0	0	0	0.5	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

#### Карта достоверности

Пример наложения одной из таких матриц можно увидеть на Рис. 3. В этом примере карта достоверности левого плеча наложена на исходное изображение. Найдя максимальное значение карты, мы получаем предполагаемую точку расположения сустава. По мере удаления от этой точки вероятности нахождения там точки уменьшаются.



Рис. 3. Матрица достоверности

Математическое представление карты достоверности:

$$C = (C1, C2, C3, ..., C18)$$

 $C_i \in \mathbb{R}^{w \times h}$ , где w – ширина изображения, h – высота изображения (1)  $i \in \{1, 2, ..., 18\}, 18$  – общее число суставов(особых точек)

### 4. КАРТЫ ОРИЕНТИРОВ

По аналогии с картой достоверности карта ориентиров является матрицей такого же размера, что и исходное изображение (после некоторых преобразований), но каждым элементом такой матрицы является двумерный вектор с направлением к связанной точке. Всего таких связей 19, как и соответствующих матриц:

- $0 \rightarrow 14$
- $0 \rightarrow 15$
- $1 \rightarrow 0$
- $1 \rightarrow 2$
- $1 \rightarrow 5$
- $1 \rightarrow 8$
- $1 \rightarrow 11$
- $2 \rightarrow 3$
- $2 \rightarrow 16$
- $3 \rightarrow 4$
- $5 \rightarrow 6$
- $5 \rightarrow 17$
- $6 \rightarrow 7$
- $8 \rightarrow 9$
- 9  $\rightarrow 10$
- $11 \rightarrow 12$
- $12 \rightarrow 13$
- $14 \rightarrow 16$
- $15 \rightarrow 17$

Все индексы точек соответствуют схеме на Рис. 1. Графическая демонстрация карт ориентиров представлена на Рис. 4, на нём упрощённая карта ориентиров для точек (5, 6) – от левого плеча до левого локтя.



Рис. 4. Матрица ориентиров



У матрицы ориентиров учитывается не один пиксель с предполагаемым суставом, а весь путь от начальной точки до конечной, включая вероятные области вокруг.

Математическое представление карт ориентиров:

$$0 = (01, 2, 03, \dots, 018)$$

 $O_i \in \mathbb{R}^{w \times h \times 2}$ , где w – ширина изображения, h – высота изображения (2)  $i \in \{1, 2, ..., 18\}, 18$  – общее число суставов(особых точек)

## 5. СХЕМА СЕТИ

Для вычисления карт достоверности и карт ориентиров, организуем параллельную поэтапную схему сети. Каждый этап использует информацию с предыдущего этапа, таким образом повышая точность вычислений. Экспериментальным путём было выбрано 7 этапов.

Перед формированием тепловых карт требуется извлечь признаки из изображения, для этого берётся предварительно обученная свёрточная нейронная сеть. В данной работе использовалась MobileNetV3 [11], но подходят и другие модели, например VGG-19 [12] или ResNet [13]. Выбор MobileNetV3 был сделан из-за того, что его скорость извлечения признаков в 5–10 раз быстрее, чем у VGG-19, что позволят использовать его и на маломощных устройствах.

На первом этапе вычисляются начальные приближения карты достоверностей и карты ориентиров, на каждом последующем результаты с предыдущего этапа объединяются и используются для вычисления новых карт. Пример уточнения результатов при смене этапов приведён на рис. 5. На этом примере видно, как по мере прохождения этапов ошибочные предположения отметались и осталось довольно точное.



Рис. 5. Поэтапное предсказание

Обозначим множество признаков символом Р. Зададим функции Fc<sup>1</sup> и Fo<sup>1</sup>, которые будут принимать на вход признаки изображения и выдавать карты достоверности и карты ориентиров соответственно для первого этапа:

$$C^{1} = Fc^{1}(P)$$

$$O^{1} = Fo^{1}(P)$$
(3)

На каждом этапе после первого для вычисления каждой из карт будем использовать помимо признаков карты с предыдущего этапа с помощью функций Fc<sup>i</sup> и Fo<sup>i</sup>.

$$C^{i} = Fc^{i} \left( P, C^{i-1}, O^{i-1} \right), \quad \forall i > 1$$
  

$$O^{i} = Fo^{i} \left( P, C^{i-1}, O^{i-1} \right), \quad \forall i > 1$$
(4)

где і-номер этапа

Общая схема работы представлена на Рис. 6.



Рис. 6. Схема работы

В ходе экспериментов над свёрточными слоями было определено, что вместо стандартного подхода с использованием крупных сверток можно использовать последовательность небольших свёрток: вместо свертки 7х7 три свертки 3х3, а вместо 5х5 две свёртки 3х3, это снижает количество операций и увеличивает производительность без потери в точности вычислений. Также оказалось полезным использование шагов сокращения со сверткой 1х1, двух таких слоев достаточно.

Несмотря на то, что в итоге получаются карты ориентаций и достоверностей с размером как у исходного изображения, в процессе работы нейронной сети будут



вычисляться карты 64х64, так как при размере 256х256 они занимают крайне много памяти GPU. При таком сокращении карт немного снижается точность, но не так существенно, как выигрыш в памяти.

Вычисление 7 этапов для создания тепловых карт весьма затратная по времени операция. Но если отследить поведение карт, то видно, что по мере обучения сети, требуется меньше этапов для получения хорошего результата, таким образом становится понятно, что такое количество вычислений в конечной системе избыточно, поэтому перед непосредственным запуском программы лишние этапы будут отключены. Большое количество этапов хорошо способствует обучению, но при использовании в реальной работе можно от них отказаться.

## 6. ФУНКЦИЯ ПОТЕРЬ

Для корректного обучения сети необходимо определить функцию потерь (Loss функция), для этого используется стандартная функция L2 (Наименьшая квадратичная ошибка).

Определим некоторые символы:

- t номер этапа
- х точная позиция (пиксель)
- С карта достоверности, одномерный вектор
- О карта ориентиров, двумерный вектор
- І количество суставов и ориентиров (обоих по 19)
- W-веса.
- \* эталонная (верная) запись. Математическая запись:

$$fc^{t} = \sum_{i=1}^{I} \sum_{x} W(x) (C_{i}^{t}(x) - C_{i}^{*}(x))$$
  
$$fo^{t} = \sum_{i=1}^{I} \sum_{x} W(x) (O_{i}^{t}(x) - O_{i}^{*}(x))$$
  
(5)

### 7. ЗАКЛЮЧЕНИЕ

Поиск ключевых точек объекта крайне важный компонент, дающий возможность применять его в огромном числе задач по автоматизации и анализу. И человек не является исключением: возможность для программ визуально отслеживать и интерпретировать движения людей является актуальной областью для исследований и разработок.

В этой статье изучается подход с использованием тепловых карт: карт достоверности и ориентации. Разрабатывается каждая из этих карт и включается в общую параллельную и поэтапную сеть. Распараллеливание вычислений для обоих карт между собой и их внутренних компонентов позволяет быстро и эффективно справляться с задач, а поэтапное построение с использованием информации с предыдущего этапа, позволяет сильно повысить точность обучения нейронной сети.

В работе использовалась топология тела от СОСО и также использовался их набор данных для обучения и тестирования сетей, но при этом разработанный подход легко адаптируется и под другие топологии, основанные на скелетном (фиксированного числа ключевых точек) представлении человека.

Вычисление карт достоверности и ориентаций крайне трудоемкий и затратная как по времени, так и по ресурсам задача, поэтому крайне важно свести к минимуму лишние операции. Для этого в статье предлагается способ с откидыванием лишних вычислений карт, после завершения этапа обучения, что позволяет повысить скорость работы в несколько раз. Также слабым звеном может являться выделение признаков изображения, поэтому в данной работе используется модель свёрточной сети MobileNetV3, которая достаточно быстра и оптимальна по ресурсам, что даёт возможность портировать её даже на мобильные устройства.

#### Литература

- 1. *Ахмед С.Х., Скородумов С.В.* Использование нейросетевых подходов в диагностировании заболеваний // Моделирование и анализ данных. 2020. Том 10. № 2. С. 49–61. DOI:10.17759/ mda.2020100204
- 2. *Крисс Харрис, Майк Стивенс.* «A Combined Corner and Edge Detector». // Конференция Alveo Vision. 1988. Том. 15.
- David G. Lowe. Object recognition from local scale-invariant features // Proceedings of the International Conference on Computer Vision. 1999. C. 1150–1157 // DOI: 10.1109/ ICCV.1999.790410
- 4. *Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool.* «Speeded Up Robust Features», ETH Zurich, Katholieke Universiteit Leuven, 2006.
- 5. *Alexander G. Schwing, Abhinav Gupta.* «Keypoint R-CNN» // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
- 6. *Yash Goyal, Abhinav Gupta, et al.* «AlphaPose: Real-Time and Accurate Multi-Person Pose Estimation» // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
- 7. *Zhe Cao, Tomas Simon.* «OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields» // IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2019.
- 8. Алексейчук А.С., Мукин Ю.Д. Применение свёрточных нейронных сетей в задаче устранения теней с фотографий // Моделирование и анализ данных. 2024. Том 14. № 1. С. 41–51. DOI: 10.17759/mda.2024140103
- Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollar, C Lawrence' Zitnick. Microsoft coco: Common objects in context. // European conference on computer vision, P. 740–755. Springer, 2014.
- Tompson, J.J., Jain, A., LeCun, Y., Bregler, C. «Joint training of a convolutional network and a graphical model for human pose estimation. // Advances in Neural Information Processing Systems.», 2014, 1799–1807.
- 11. Andrew G. Howard, Mark Sandler, Huiyu Wang, et al. «Searching for MobileNetV3», 2019.
- 12. *Karen Simonyan, Andrew Zisserman.* «Very Deep Convolutional Networks for Large-Scale Image Recognition» // International Conference on Learning Representations (ICLR), 2015.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. «Deep Residual Learning for Image Recognition» // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.



# **Detecting Human Body Parts Using Confidence Maps and Landmark Maps**

### Maxim A. Potenko\*

Moscow Aviation Institute (National Research University) (MAI) Moscow, Russian Federation, ORCID: https://orcid.org/0009-0008-5222-2664 e-mail: potenkog@gmail.com

This article discusses the problem of finding key points on an object and determining its components using the example of the human body as one of the most pressing computer vision problems. The problem is that for an image with a supposed object (in our case, a person), it is necessary to find the position of all important components/points, for a person these are joints: shoulders, elbows, hands, knees, etc. To solve this problem, it is proposed to use neural networks using «heat maps», which are presented in the form of a confidence map and a landmark map. The landmark map is a matrix corresponding to the original image, each cell of which contains a value from 0 to 1 with the probability of the location of the desired joint (special point) in the corresponding pixel. The landmark map is a matrix, each cell of which contains a two-dimensional vector with the direction of the next joint. The main feature is that the maps help each other to predict the correct values: thanks to the confidence map, the landmark map knows the approximate location of the landmarks and it is easier for it to determine the desired direction and vice versa, knowing the approximate direction to the next landmark, it is easier to predict the location of the landmark. These maps are calculated in several stages, and at each new stage, information from the previous one is used, so when calculating each subsequent map, the accuracy of the results obtained increases. In this work, the optimal number of stages was 6-7, but this value can change depending on the final architecture. Demonstrations of the work are carried out on the COCO dataset, which includes 18 points for each human body.

*Keywords:* neural networks, convolutional neural networks, Human Pose Estimation, heat maps, image processing, keypoints, object detection, object segmentation

#### For citation:

Potenko M.A. Detecting Human Body Parts Using Confidence Maps and Landmark Maps. *Modelirovanie i analiz dannykh = Modelling and Data Analysis*, 2025. Vol. 15, no. 1, pp. 7–18. DOI: https://doi.org/10.17759/mda.2025150101 (In Russ., abstr. in Engl.).

\**Maxim A. Potenko*, Graduate Student, Moscow Aviation Institute (national research university) (MAI), Moscow, Russian Federation, ORCID: https://orcid.org/0009-0008-5222-2664, e-mail: potenkog@gmail.com



Обнаружение частей человеческого тела с помощью карт достоверности и карт ориентиров Моделирование и анализ данных. 2025. Том 15. № 1.

#### References

- Ahmed S. Kh., Skorodumov S.V. The Use of Neural Network Approaches in the Diagnosis of Diseases. Modelirovanie i analiz dannykh = Modelling and Data Analysis, 2020. Vol. 10, no. 2, pp. 49–61. DOI:10.17759/mda.2020100204 (In Russ., abstr. in Engl.).
- Chris Harris, Mike Stevens «A Combined Corner and Edge Detector». // Alveo Vision Conference. 1988. Vol. 15
- David G. Lowe. Object recognition from local scale-invariant features // Proceedings of the International Conference on Computer Vision. 1999. C. 1150–1157 // DOI: 10.1109/ ICCV.1999.790410
- 4. Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool, «Speeded Up Robust Features», ETH Zurich, Katholieke Universiteit Leuven, 2006
- 5. Alexander G. Schwing, Abhinav Gupta, «Keypoint R-CNN» // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
- 6. Yash Goyal, Abhinav Gupta, et al. «AlphaPose: Real-Time and Accurate Multi-Person Pose Estimation» // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
- Zhe Cao, Tomas Simon, «OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields» // IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2019.
- Alekseychuk A.S., Mukin Yu. D. Application of Convolutional Neural Networks in the Problem of Removing Shadows from Photographs. Modelirovanie i analiz dannykh = Modelling and Data Analysis, 2024. Vol. 14, no. 1, pp. 41–51. DOI: https://doi.org/10.17759/ mda.2024140103 (In Russ., abstr. in Engl.).
- Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollar, C Lawrence ' Zitnick // Microsoft coco: Common objects in context. // European conference on computer vision, P. 740–755. Springer, 2014
- Tompson, J.J., Jain, A., LeCun, Y., Bregler, C.: «Joint training of a convolutional network and a graphical model for human pose estimation. // Advances in Neural Information Processing Systems.», 2014, 1799–1807
- 11. Andrew G. Howard, Mark Sandler, Huiyu Wang, et al. «Searching for MobileNetV3», 2019
- 12. Karen Simonyan, Andrew Zisserman, «Very Deep Convolutional Networks for Large-Scale Image Recognition» // International Conference on Learning Representations (ICLR), 2015.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. «Deep Residual Learning for Image Recognition» // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

Получена 05.02.2025 Принята в печать 19.02.2025 Received 05.02.2025 Accepted 19.02.2025