

**Архипова Ангелина Александровна**

магистрант направления подготовки «бизнес-информатика в высокотехнологичных отраслях экономики»  
Национальный исследовательский ядерный университет «МИФИ»  
Россия, Москва  
arhipovaangelina1@gmail.com

**Angelina A. Arkhipova**

Master's student in the direction of training  
"Business Informatics in High-Tech Sectors  
of the Economy"  
National Research Nuclear University MPhI  
Russia, Moscow

**Зиннуров Ильнур Ильдарович**

магистрант направления подготовки «бизнес-информатика в высокотехнологичных отраслях экономики»  
Национальный исследовательский ядерный университет «МИФИ»  
Россия, Москва  
vip.zinnurov@mail.ru

**Ilnur I. Zinnurov**

Master's student in the direction of training  
"Business Informatics in High-Tech Sectors  
of the Economy"  
National Research Nuclear University MPhI  
Russia, Moscow

**ОБЗОР И АНАЛИЗ ПОДХОДОВ СОЗДАНИЯ  
НЕЙРОННЫХ СЕТЕЙ ДЛЯ РЕШЕНИЯ ЗАДАЧ  
ВИЗУАЛЬНОГО ПОИСКА**

**REVIEW AND ANALYSIS OF APPROACHES  
TO CREATING NN FOR SOLVING VISUAL  
SEARCH PROBLEMS**

**Аннотация**

Решение задачи визуального распознавания объектов и визуального поиска, как частного примера, становится всё более актуальным направлением в наши дни. С ростом вычислительных возможностей устройств как носимых, так и стационарных, открываются новые сферы применения для описываемой технологии и начинают применяться всё более совершенные методы работы с данными, которые подробно рассматриваются в работе.

**Ключевые слова:**

нейронные сети, визуальный поиск, распознавание образов

**Abstract**

Solving the problem of visual recognition of objects and visual search, as a particular example, is becoming more and more relevant today. With the growth of the computing capabilities of devices, both wearable and stationary, new areas of application for the described technology are opening up and more and more advanced methods of working with data are being applied, which are discussed in detail in the work.

**Keywords:**

neural networks, visual search, pattern recognition

Системы визуального распознавания и поиска изображений являются важным элементом современной жизни и постоянно совершенствуются, вслед за растущими потребностями пользователей. В 1992 году термин «поиск изображений на основе содержания» был использован впервые японским инженером Тошиказу Като для описания процесса извлечения изображений на основе их форм и цвета из базы данных. На сегодняшний день в мире хранятся десятки зеттабайт фото и видео информации, которые без обработки представляют неструктурированные и во многом малополезные данные. Используя нейронные сети, искусственный интеллект и машинное обучение, данные структурируются и используются для самых различных задач: профессиональный, повседневных, государственных, как пример, для

обеспечения государственной безопасности. Каждый человек, имеющий электронное устройство с выходом в интернет может отправить изображение и найти его составляющие или описание в интернете. В данной работе будет произведён анализ современных подходов, используемых для распознавания изображений. Основными задачами данной работы является:

- общая методика создания нейронной сети для распознавания изображений;
- анализ основных подходов в технологии распознавании изображений;
- коммерческое и некоммерческое использование технологии распознавания изображений.

### **Общая методика создания нейронной сети для решения задачи распознавания образов**

Для решения задачи распознавания образов применяется машинное обучение, изучающее методы построения алгоритмов, способных обучаться. Теория распознавания образов является приложением машинного обучения и описывает три типа методов распознавания графических образов: статистические, структурные и нейросетевые.

Статистические методы применяются, когда для распознавания достаточно простых численных и символических признаков, описывающих распознаваемый объект, представленный в виде вектора элементарных признаков. Примером применения статистического метода может быть распознавание шрифта, когда предварительно известны все признаки символов различных шрифтов.

Структурные методы применяются, когда важна информация, описывающая структуру объекта и необходимо не только детектировать признаки, позволяющие отнести объект к определённому классу, но и обнаружить признаки, исключающие отнесение объекта к другому классу. Структурные методы представляют объект, как иерархическую структуру признаков объекта. Применяются для распознавания изображений.

С точки зрения машинного обучения, одним из наиболее перспективных методов теории распознавания образов является использование нейронных сетей. Нейронная сеть – это математическая модель, а также её программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей (НС) – сетей нервных клеток живого организма [1]. Как и в случае с живым организмом, необходимо обучить нейронную сеть

распознавать объекты перед тем, как использовать её для такой задачи, для чего используются обучающие выборки данных, на которых разработчик предварительно размечает данные в дата сете [2]. В работе будут рассмотрены именно нейросетевые методы, так как они сильно превосходят по возможностям распознавания сложных образов прочие методы. Преимущества НС перед прочими методами теории распознавания образов:

- устойчивость к детектированию зашумлённых или искажённых изображений;
- решение задач с неизвестными закономерностями. В случае, когда неизвестны закономерности решения задачи, обучающаяся способность НС позволяет разрешить проблему. Эта особенность недоступна при применении прочих методов и экспертных систем;

- высокая скорость работы, которая обеспечивается поддержкой параллелизма обработки информации;

- высокая адаптируемость к изменениям. Возможность обучаться позволяет НС быть крайне адаптивными. Можно создать НС, которая будет дообучаться вслед за колебаниями параметров окружающей среды или быстро дообучить уже обученную НС вручную при изменении требований.

Наибольшей эффективностью для решения задачи распознавания сложных объектов на изображении обладают свёрточные НС, далее СНС. Структура СНС является однонаправленной и основана на чередовании сверточных и субдискретизирующих слоёв. Основной особенностью СНС при распознавании изображения является свёрточная особенность, позволяющая умножать каждый фрагмент изображения на ядро свёртки поэлементно, результат при этом суммируется и записывается в выходное изображение. Это позволяет СНС обеспечить инвариантность распознавания, всё больше расширяя область видимости сети и распознавая сначала детали, а в последствии более крупные объекты изображения.

Распознаваемые образы при обучении нейронной сети могут быть различны, в зависимости от целей, для которых она создаётся. Это может быть почерк, черты лица, объекты нашего мира и многое другое. Обучение происходит путём подачи в НС образца данных с описанием типов, к которым он относится. Данные представляют собой вектор значений признаков, которые должны позволить однозначно выявить класс. Общая структура подобной НС описана на рисунке 1.



Рисунок 1 – Структура создания НС для распознавания изображений

Существует несколько способов обучения нейронной сети. Каждый из них применяется для решения специализированных задач, стоящих перед разработчиком:

- обучение с учителем. Наиболее удобный подход быстрого обучения с низким показателем ошибки. НС обучается на предварительно размеченных данных;
- обучение без учителя. В данном случае перед НС ставятся задачи по поиску закономерностей в данных;
- обучение с частичным привлечением учителя. В данном случае размечена только часть данных. Это актуально, когда имеется большой объём данных, который сложно разметить вручную;
- обучение с подкреплением. При выполнении определённых действий НС получает «вознаграждение».

В свою очередь, НС для распознавания изображений решают следующие задачи:

- оптимизационные;
- извлечение ключевых признаков из заданных наборов (неявное извлечение признаков);
- классификация образов и характеристик, заранее из них извлечённых.

Методика создания НС для решения задач распознавания изображений представляет собой чётко структурированную последовательность, на определённых этапах которой разработчик реализует наиболее оптимальные для решения конкретной задачи подходы, от чего зависит эффективность и качество готового продукта.

## **Обзор основных подходов реализации НС для решения задачи распознавания изображений**

НС для распознавания изображений используются крайне эффективно, но при этом следует различать виды и подходы к построению таких сетей. Их существует несколько видов: однослойные, многослойные и свёрточные и их производные.

Многослойные нейронные сети (МНС) по архитектуре представляют собой последовательно соединённые слои. Слои соединяются таким образом, что каждый нейрон слоя своими входами связан со всеми нейронами предыдущего слоя, а выходами – последующего. К примеру, НС, имеющая два решающих слоя может эффективно аппроксимировать любую многомерную функцию. При уменьшении количества слоёв до одного, эффективность и круг решаемых такой сетью задач сильно снижается, так как она будет способна формировать лишь линейные разделяющие поверхности, что не позволит решать задачи типа «исключающее или» [3].

Эффективность решающей способности МНС, как и сложность её реализации и обучения, зависят от количества слоёв. НС, использующая два слоя и нелинейную функцию активации, способна создавать любые выпуклые области из пространства решений, а при добавлении ещё одного слоя – любые, в том числе и невыпуклые области. Важно учесть такую особенность, как обобщающая способность, у МНС она не теряется. Это связано с механизмом обучения. Обычно такие НС обучаются по принципу обратного распространения ошибки (величины коррекции весов), являющегося методом градиентного спуска в пространстве весов с целью минимизации суммарной ошибки сети [4].

Одним из интересных направлений применения однослойной НС в области распознавания изображений является восстановление зашумлённых или искажённых изображений. При этом возможно оценить качество восстановления, сравнивая качество исходного и итогового изображений. Это также именуется автоассоциативной памятью [5].

Процесс распознавания МНС изображений, а в приведённом примере – лиц, изображён наглядно на рисунке 2. В данном случае нейронная сеть определила отношение входного объекта к первому классу, так как первый нейрон выходного слоя был наиболее активен.

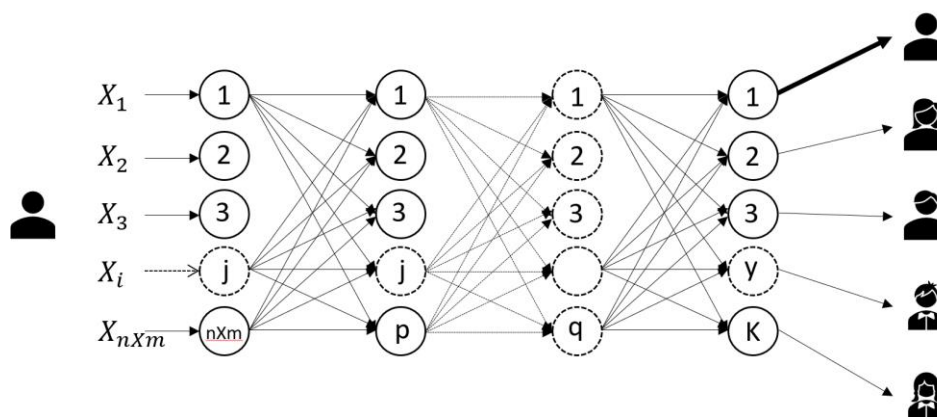


Рисунок 2 – Распознавание МНС лиц

Если в процессе распознавания не один нейрон не набрал достаточной пороговой активности, то считается, что образ не относится ни к одному из изученных классов. В процессе работы НС устанавливается соответствие образа классу, что и называется «обучением с учителем». Этот подход требует большой вычислительной и передающей мощности и отлично подходит для небольших групп образов. Если необходимо распознавание большого количества различных классов, то используется передача не целого образа, а лишь его наиболее важных признаков, которых достаточно для необходимого уровня распознавания.

К примеру, в работе «Multi-valued and Universal Binary Neurons: Learning Algorithms, Applications to Image Processing and Recognition» [6] описывается подход с применением частотного подхода для распознавания изображений. Эксперимент производился в университете MIT и по результатам НС показала 100% распознавание объектов. При разработке НС применялась однослойная сеть, которая была основана на многозначных нейронах (они основаны на многозначной пороговой логике и определены на множестве комплексных чисел).

Один из наиболее популярных подходов распознавания лиц сегодня, – это распознавание по специфичным признакам лица. Такой подход был подробно описан в работе «Hybrid approaches to frontal view face recognition using the Hidden Markov Model and Neural Network» [7]. При разработке НС использовались гибридные подходы, а на вход сети подавали расстояния между ртом, носом, глазами. Это значительно увеличило скорость и уменьшило необходимые вычислительные мощности для эффективной работы. Сама же НС строилась таким образом, что в первом варианте на вход подавались результаты обработки скрытой марковской моделью [8], а во втором

– результат НС подавался на вход скрытой марковской модели. Отмечено, что во втором случае не наблюдалось никаких преимуществ, что означало то, что эффективность НС была достаточной.

Не менее интересным подходом является использование НС, когда на её вход подаются результаты декомпозиции исходного изображения по методу главных компонент [9]. Такая работа была проведена и подробно описана в работе «Face recognition using transform features and neural networks» [10]. При обработке двумерных изображений есть особенность, что в классической МНС изображение раскладывается на одномерный вектор, хотя оно двумерно. Эту проблему авторы работы решили использованием свёрточной структуры НС (СНС) [11], которая позволила анализировать изображение как двумерное. При разработке были применены [12]:

- локальное рецепторное поле. Оно обеспечило двумерную кольную связь между нейронами;
- общий вес. Это позволило распознавать некоторые черты в произвольных местах изображения;
- иерархическая организация, дополненная пространственными подвыборками.

СНС имеет массу положительных особенностей при работе с изображениями. Среди них устойчивость к небольшим изменениям освещения, растяжению, сжатию или размытию изображения. Архитектура такой НС сильно напоминает работу нейронов в зрительной коре человека, её нейроны только частично связаны с нейронами предыдущего слоя из небольшой области. В рамках одного свёрточного слоя(плоскости) веса нейронов являются одинаковыми, а последующий слой уменьшает размерность предыдущего путём локального усреднения, что создаёт иерархию слоёв, где более поздний слой извлекает всё более общие признаки. Такие признаки менее требовательны к искажениям исходного изображения.

Сравнивая СНС и МНС подходы к построению НС в рамках технологии распознавания изображений, важно отметить значительное преимущество СНС как по надёжности классификации, так и по скорости работы алгоритма. Это было доказано в работе «Face recognition: a convolutional neural-network approach» [13]. Авторы доказали, что при замене СНС на МНС (многослойный персептрон) сильно упало качество распознавания, что наглядно показано на таблице 1. Сравнивались СНС и МНС, использующие метод главных компонент и самоорганизующуюся карту

Кохонена, результат указывался, как средняя из трёх компиляций. Это было ожидаемо, так как многослойный персептрон не имеет встроенной инвариантности к незначительной трансляции и локальной деформации, которая реализуется в СНС с использованием локальных рецептивных полей, общих весов и пространственной субдискретизации.

Таблица 1 – Сравнение показателей ошибки для СНС и МНС

|     | Linear PCA | SOM   |
|-----|------------|-------|
| МНС | 41,2%      | 39,6% |
| СНС | 5,33%      | 3,83% |

### **Коммерческое и некоммерческое использование технологии распознавания изображений**

Технологии распознавания изображений применяются повсеместно в самых различных отраслях от развлечений до государственной безопасности. Подробное распределение использования технологии на 2019 год по данным исследования Grand View Research [14] наглядно показано на рисунке 3.

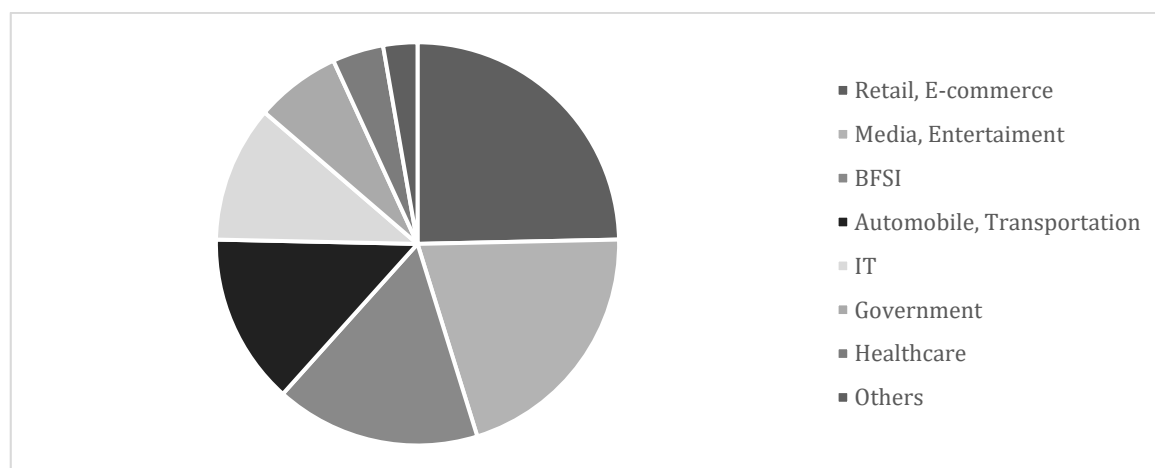


Рисунок 3 – Доля мирового рынка распознавания изображений.

Не менее интересны и феноменальные прогнозы по росту рынка данного сегмента с разбивкой по странам, которые можно увидеть на рисунке 4. Ожидается, что к 2027 году азиатско-тихоокеанский рынок распознавания изображений составит 27.396,83 миллиона долларов США.

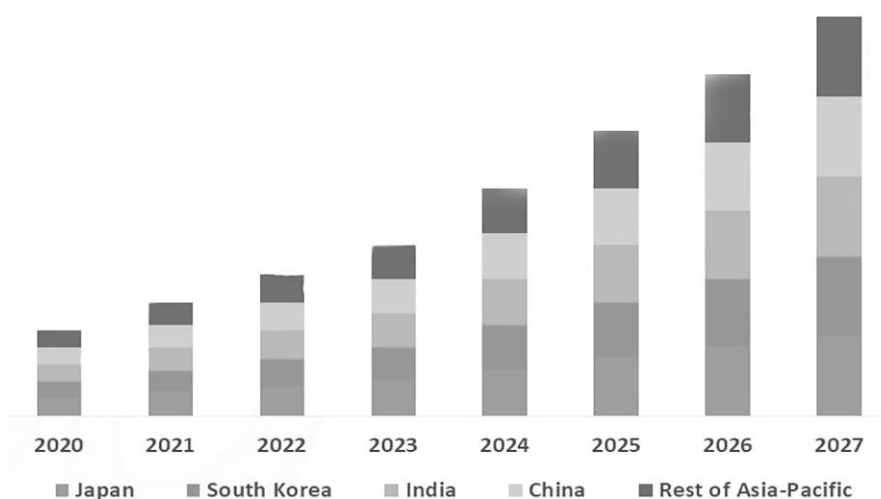


Рисунок 4 – Прогноз рынка распознавания изображений 2020-2027

В России наиболее крупными разработчиками являются Яндекс, Сколково и крупные университеты. К примеру, компания Центр2М применяет данную технологию для распознавания объектов на предприятиях. При помощи этой технологии, на предприятии контролируется соблюдение норм безопасности сотрудниками и соответствие работ установленному регламенту.

Не менее активное развитие распознавание изображений получило и в сфере медицины. Осенью 2020 года Spline.ai совместно с разработчиком компьютерных технологий Xilinx, продемонстрировали алгоритм с открытым кодом, способный качественно анализировать рентгеновские снимки. Данная разработка получила особенно быстрое развитие, так как была специализирована для обнаружения пневмонии и COVID-19, что позволило значительно разгрузить врачей узких специальностей для их более эффективного использования. Наиболее интересным проектом является государственная система распознавания лиц. Система представляет собой один из компонентов умного города, предоставляющий в государственную базу информацию о гражданине и рассчитывающую «рейтинг общественной благонадёжности». Для реализации подобной системы страна создаёт крупнейшую сеть видеонаблюдения в мире, – это 570 млн камер [15]. Алгоритм распознавания является практически совершенным и показывает точность распознавания лиц в маске 95% [16]. По задумке властей, система всего за три секунды должна распознавать человека. Рейтинг гражданина будет влиять на его статус в обществе, государственные и коммерческие привилегии. Аналогичные системы распознавания используются в крупных городах мира.

Одним из применений визуального поиска может быть внедрение его на коммерческие сайты. Данная функция представляет подсистему, которая находит товар из каталога, наиболее соответствующий товару на изображении, отправленному клиентом. Согласно исследованиям Gartner, использование визуального и голосового поиска брендом позволит ему повысить результативность онлайн торговли на 30% [17]. Frisbuy также провели исследование и пришли к выводу, что внедрение системы визуального поиска на сайт позволит обеспечить следующие показатели [18]:

- рост глубины просмотров сайта на 48%;
- увеличение показателя возвратов пользователя на 75%;
- рост среднего чека при заказе онлайн на 11%.

### **Заключение**

В работе рассмотрен полный цикл создания нейронной сети для реализации технологии распознавания изображений, описаны все ключевые этапы и нюансы, необходимые для создания качественного продукта. Проанализированы различные подходы написания НС и особенности их использования от однослойных НС до СНС.

Также, приведено сравнение СНС и МНС на примере многослойного персептрона, где ключевым показателем эффективности был уровень ошибки распознавания. Результаты показали крайне высокую эффективность СНС на сложных, деформирующихся объектах. Хотя НС и обладают высоким уровнем распознавания изображений, в случае распознавания 3D объектов на 2D изображениях в реальных условиях возникает немало сложностей, связанных с освещением, углам поворота, различным состоянием самого объекта (как выражение эмоций у человека).

В заключении, проведён обзор коммерческих и некоммерческих проектов, использующих технологию распознавания объектов в реальном мире. Широта возможностей использования технологии поражает воображение, а рынок растёт

### **Список использованных источников**

1. Нейронная сеть. – Текст : электронный // Wikipedia : [сайт]. – URL: [https://ru.wikipedia.org/wiki/Нейронная\\_сеть](https://ru.wikipedia.org/wiki/Нейронная_сеть) (дата обращения: 17.12.2020).
2. Data set. – Текст : электронный // Wikipedia : [сайт]. – URL: [https://ru.wikipedia.org/wiki/Data\\_set\\_\(IBM\)](https://ru.wikipedia.org/wiki/Data_set_(IBM)) (дата обращения: 17.12.2020).

3. Нейронные сети. – Текст : электронный // Большая российская энциклопедия : [сайт]. – URL: [https://bigenc.ru/technology\\_and\\_technique/text/4114009](https://bigenc.ru/technology_and_technique/text/4114009) (дата обращения: 18.12.2020).

4. Романов Дмитрий Евгеньевич Нейронные сети обратного распространения ошибки // ИВД. 2009. №3. URL: <https://cyberleninka.ru/article/n/neyronnye-seti-obratnogo-rasprostraneniya-oshibki> (дата обращения: 18.12.2020).

5. Valentin, D. Connectionist models of face processing: a survey. / D. Valentin, H. Abdi, A.J., and, Cottrell, G. O'Toole. – Текст : непосредственный // Pattern Recognition. – 1994. – № 27. – С. 1209-1230.

6. Aizenberg, I. N. Multi-valued and Universal Binary Neurons: Learning Algorithms, Applications to Image Processing and Recognition. / I. N. Aizenberg, N. N. Aizenberg, G. A. Krivosheev. – Текст : непосредственный // Lecture Notes in Artificial Intelligence - Machine Learning and Data Mining in Pattern Recognition. – 1999. – № . – С. 21-35.

7. Yoon, K. S. Hybrid approaches to frontal view face recognition using the Hidden Markov Model and Neural Network. / K.S. Yoon, Y.K. Ham, R. Park. – Текст : непосредственный // Pattern Recognition. – 1998. – № 31. – С. 283-293.

8. Скрытая марковская модель. – Текст : электронный // Wikipedia : [сайт]. – URL: [https://ru.wikipedia.org/wiki/Скрытая\\_марковская\\_модель](https://ru.wikipedia.org/wiki/Скрытая_марковская_модель) (дата обращения: 18.12.2020).

9. Метод главных компонент. – Текст : электронный // Machinelearning : [сайт]. – URL: [http://www.machinelearning.ru/wiki/index.php?title=Метод\\_главных\\_компонент](http://www.machinelearning.ru/wiki/index.php?title=Метод_главных_компонент) (дата обращения: 18.12.2020).

10. Ranganath, S. Face recognition using transform features and neural networks. / S. Ranganath, K. Arun. – Текст : непосредственный // Pattern Recognition . – 1997. – № 30. – С. 1615-1622.

11. Lawrence, S. Face Recognition: A Convolutional Neural Network Approach. / S. Lawrence, C. L. Giles, A. C. Tsoi, A. D. Back. – Текст : непосредственный // IEEE Transactions on Neural Networks, Special Issue on Neural Networks and Pattern Recognition. – С. 1-24.

12. Levina, A. Subsampling scaling / A. Levina, V. Priesemann. – Текст : непосредственный // Nature Communications. – 2017. – № 8.

13. Lawrence, S. Face Recognition: A Convolutional Neural Network Approach. / S. Lawrence, C. L. Giles, A. C. Tsoi, A. D. Back. – Текст : непосредственный // IEEE

Transactions on Neural Networks, Special Issue on Neural Networks and Pattern Recognition. – 1997. – № 8. – С. 1-24.

14. Image Recognition Market Size, Share & Trends Analysis Report By Technique, By Application, By Component (Hardware, Software), By Deployment Mode, By Vertical, By Region, And Segment Forecasts, 2020-2027. – Текст : электронный // Grand View Research : [сайт]. – URL: <https://www.grandviewresearch.com/industry-analysis/image-recognition-market> (дата обращения: 20.12.2020).

15. Распознавание лиц: почему в Китае не скрыться даже в 60-тысячной толпе. – Текст : электронный // BBC : [сайт]. – URL: <https://www.bbc.com/russian/news-43751391> (дата обращения: 20.12.2020).

16. В Китае научились распознавать 95% лиц в масках. – Текст : электронный // Forbes : [сайт]. – URL: <https://www.forbes.ru/newsroom/biznes/395425-v-kitae-nauchilis-raspoznavat-95-lic-v-maskah> (дата обращения: 20.12.2020).

17. Gartner Top Strategic Predictions for 2018 and Beyond. – Текст : электронный // Gartner : [сайт]. – URL: <https://www.gartner.com/smarterwithgartner/gartner-top-strategic-predictions-for-2018-and-beyond/> (дата обращения: 17.12.2020).

18. Визуальный поиск для e-commerce. – Текст : электронный // Frisbuy : [сайт]. – URL: <https://frisbuy.ru/s/visualsearchforecom/> (дата обращения: 17.12.2020).