

Научный журнал «Костюмология» / Journal of Clothing Science <https://kostumologiya.ru>

2023, Том 8, № 1 / 2023, Vol. 8, Iss. 1 <https://kostumologiya.ru/issue-1-2023.html>

URL статьи: <https://kostumologiya.ru/PDF/03TLKL123.pdf>

Дата поступления: 30.01.2023 / Дата публикации: 17.04.2023

Ссылка для цитирования этой статьи:

Смирнов, Е. Е. Распознавание вида одежды и обуви по изображению / Е. Е. Смирнов, В. В. Костылева, И. Б. Разин, А. Р. Муртазина // Костюмология. — 2023. — Т. 8. — № 1. — URL:

<https://kostumologiya.ru/PDF/03TLKL123.pdf>

For citation:

Smirnov E.E., Kostyleva V.V., Razin I.B., Murtazina A.R. Recognition of the type of clothing and shoes by image.

Journal of Clothing Science. 2023; 8(1): 03TLKL123. Available at: <https://kostumologiya.ru/PDF/03TLKL123.pdf>.

(In Russ., abstract in Eng.)

УДК 004.85; 687.1

ГРНТИ 28.23.37; 64.41.14

Смирнов Евгений Евгеньевич

ФГБОУ ВО «Российский государственный университет имени А.Н. Косыгина
(Технологии. Дизайн. Искусство)», Москва, Россия

Кандидат технических наук, доцент

E-mail: evg7162@mail.ru

РИНЦ: https://www.elibrary.ru/author_profile.asp?id=1097456

Костылева Валентина Владимировна

ФГБОУ ВО «Российский государственный университет имени А.Н. Косыгина
(Технологии. Дизайн. Искусство)», Москва, Россия

Заведующая кафедрой

Доктор технических наук, профессор

E-mail: kostyleva-vv@rguk.ru

РИНЦ: https://www.elibrary.ru/author_profile.asp?id=353612

Разин Игорь Борисович

ФГБОУ ВО «Российский государственный университет имени А.Н. Косыгина
(Технологии. Дизайн. Искусство)», Москва, Россия

Заведующий кафедрой

Кандидат технических наук, доцент

E-mail: razin-ib@rguk.ru

РИНЦ: https://www.elibrary.ru/author_profile.asp?id=850439

Муртазина Альфия Рустямовна

ФГБОУ ВО «Российский государственный университет имени А.Н. Косыгина
(Технологии. Дизайн. Искусство)», Москва, Россия

Кандидат технических наук, доцент

E-mail: aly1029@yandex.ru

РИНЦ: https://www.elibrary.ru/author_profile.asp?id=906389

Распознавание вида одежды и обуви по изображению

Аннотация. В статье предлагается решение проблемы поиска информации об изделии в условиях ограниченных данных о самом изделии и знаний о предметной области, в частности, распознавания и классификации видов одежды и обуви на изображении. Это позволит исключить следующие затруднения: (1) каким образом любому человеку не являющимся профессионалом в изделиях швейной и обувной промышленности быстро найти информацию

в сети интернет о конкретном изделии; (2) сбор статистики используемых одежды и обуви при помощи обработки фотографий в рамках промоакций с хештегами в социальных сетях и т. д. В рамках этого исследования проведены эксперименты с полносвязной архитектурой нейронной сети в задаче распознавания класса объекта на примере набора данных из 60 000 изображений с одним цветовым каналом в градациях серого.

Цель исследований состояла в оценке приемлемости архитектуры полносвязной нейронной сети для распознавания изделий легкой промышленности.

Показано, что для потребителя при выборе изделия, поиске информации о нем, или покупке, предпочтительнее изображение с его описанием. Если работу по его фрагментации и выделению значимых признаков переложить на нейронную сеть, то это позволит сократить время на создание обучающей выборки, но увеличит время на обучение самой сети. Полносвязная нейронная сеть распознает образы простых изображений с точностью около 88 %. Тестирование различных функций активации показало, что в данной задаче выбор функции не является критичным. Результаты получились идентичными, однако, если возникнет необходимость построения более глубокой сети, то выбор функции активации скрытых слоев будет больше влиять на результат. Такие решения предполагается разрабатывать в рамках докторской диссертации Смирнова Е.Е. и использовать в учебном процессе кафедр «Художественное моделирование, конструирование и технология изделий из кожи», «Информационные технологии» РГУ имени А.Н.Косыгина (Технологии. Дизайн. Искусство) в виде учебных пособий.

Ключевые слова: нейронная сеть; машинное обучение; обработка данных; распознавание изображений; классификация образов; глубокое обучение; классификация; изделия легкой промышленности

Введение

В статьях [1; 2] были проведены простые эксперименты по классификации объектов при помощи нейронных сетей на основе некоторого набора параметров. Этот подход удобен в решении определенных задач, но имеет серьезное ограничение — эти параметры необходимо описать, что не всегда представляется возможным либо весьма затруднительным. В качестве примера возьмем описание различных видов обуви [3].

Таблица 1

Описание видов обуви

Вид обуви	Описание
Ремешково-сандальная обувь	Обувь, верх которой состоит из ремешков или деталей, близких по форме к ремешкам, а также обувь с сильно разрезанной передней частью, не полностью закрывающей тыльную часть стопы
Туфли	Обувь с верхом, не полностью закрывающим тыльную поверхность стопы, и берцами, расположенными не выше лодыжек
Полуботинки	Обувь с верхом, закрывающим всю тыльную поверхность стопы, берцами, расположенными не выше лодыжек; с приспособлениями для закрепления на стопе
Ботинки	Обувь с берцами, закрывающими лодыжки, с приспособлениями для закрепления на стопе
Сапоги	Обувь с высокими целыми голенищами, закрывающими стопу, голень или ее часть, а иногда и бедро

Как видно в таблице 1, каждому видов обуви соответствует большое количество параметров, которые в бытовой ситуации довольно проблематично определить с высокой точностью, а некоторые вообще не представляются возможными. С одеждой все обстоит ровно так же. В качестве такой «бытовой» ситуации в век современных цифровых технологий можно

представить следующую картину: человек в общественном месте увидел какое-то заинтересовавшее его изделие и захотел найти информацию или купить. В подходе, описанном выше, человек должен был бы перечислить параметры изделия, чтобы иметь возможность найти о нем информацию, а, учитывая наличие профессиональной терминологии и ограниченные данные об изделии (например, видно только часть изделия), успех данного процесса маловероятен. Также встречаются изделия, которые сложно отнести к правильной категории (рис. 1а, 1б). Так, исходя из описаний таблицы 1, например, ремешково-сандаляная обувь в интернет-магазине модели (рис. 1) представлены в разделе «туфли женские».



Рисунок 1. Пример ассортимента интернет-магазина в разделе «туфли женские»¹

Следовательно, в ключе клиентоориентированности гораздо логичнее запрашивать у пользователя изображение, а не текстовое описание, и, исходя из содержания данного файла давать пользователю описание содержимого. Таким образом мы можем сократить время на создание обучающей выборки, так как нужно просто соотнести изображение с видом изделия. При этом необходимо построить такую архитектуру сети, которая в процессе обучения и распознавания самостоятельно фрагментировать изображение и выделять значимые признаки.

Пример создания простой обучающей выборки представлен на рисунках 2 и 3. Рисунок 2 демонстрирует такой вариант создания обучающей выборки, в котором на изображении присутствует только одно изделие.

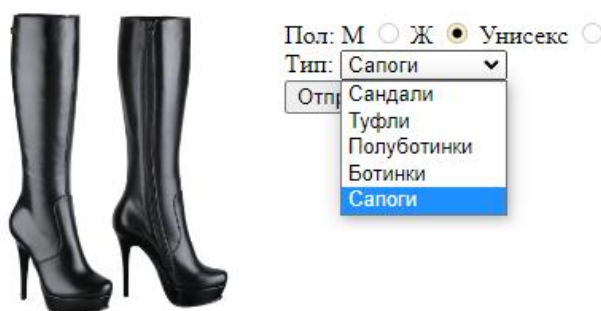


Рисунок 2. Создание относительно простой обучающей выборки (рисунок авторов)

На рисунке 3 показана разметка полноценной фотографии с разбивкой ее на фрагменты, соответствующие определенным типам изделий. На рисунке размечены сапоги женские черные, шуба женская белая, сумочка женская черная, перчатки женские черные. Такой подход к формированию обучающей выборки хотя и затратен, и менее точен, но позволяет реализовать решение поставленной задачи более полно [4].

¹ Женские туфли и лоферы // URL: <https://www.wildberries.ru/catalog/obuv/zhenskaya/tufli-lofery> (Дата обращения 12.01.23).

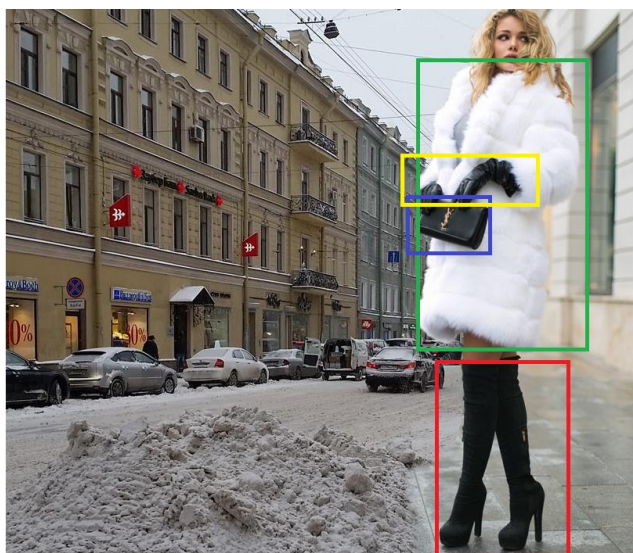


Рисунок 3. Разметка изображения для формирования обучающей выборки (рисунок авторов)

Описанная выше задача глобальна и тестировать приемлемые методы для ее решения весьма ресурсоемкая проблема. Поэтому обратимся к менее требовательным ресурсным данным, которые помогут оценить эффективность того или иного подхода.

Экспериментальные исследования

В качестве данных нами изучался набор `mnist_fashion`², который включал 60 000 изображений форматом 28x28 пикселей, с одним цветовым каналом в градациях серого. Все изображения отнесены к 10 различным классам (табл. 2).

Таблица 2

Классификация изображений в наборе данных `mnist_fashion`

Класс	Название класса на английском языке	Название класса на русском языке
0	T-shirt/top	Футболка/топ
1	Trouser	Брюки
2	Pullover	Свитер
3	Dress	Платье
4	Coat	Пальто
5	Sandal	Сандалии
6	Shirt	Рубашка
7	Sneaker	Кроссовки
8	Bag	Сумка
9	Ankle boot	Ботильоны

Кроме того, набор разделен на два фрагмента: обучающий — 50 000 изображений, тестовый — 10 000. Случайные 50 изображений из набора представлены на рисунке 4.

Ранее описывалась полносвязная нейронная сеть [2]. Проведем эксперимент с этим типом сети. Для этого сначала нужно спроектировать структуру. Самое простое определиться с выходным слоем: возьмем 10 нейронов, т. к. у нас есть 10 классов, а в качестве активационной функции — `softmax` [5].

² Fashion MNIST dataset, an alternative to MNIST // URL: https://keras.io/api/datasets/fashion_mnist/ (Дата обращения 15.01.23).

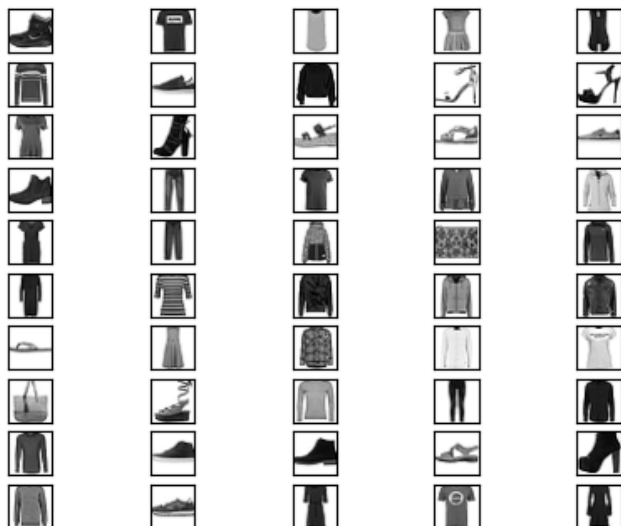


Рисунок 4. Фрагмент набора данных mnist_fashion (рисунок авторов)

Для исключения проблем на этом этапе необходимо провести преобразование номера класса в вектор размерностью 10 со значением 1 на индексе необходимого класса (табл. 3).

Таблица 3

Преобразование класса в вектор

Класс	Вектор
0	[1, 0, 0, 0, 0, 0, 0, 0, 0, 0]
1	[0, 1, 0, 0, 0, 0, 0, 0, 0, 0]
2	[0, 0, 1, 0, 0, 0, 0, 0, 0, 0]
3	[0, 0, 0, 1, 0, 0, 0, 0, 0, 0]
4	[0, 0, 0, 0, 1, 0, 0, 0, 0, 0]
5	[0, 0, 0, 0, 0, 1, 0, 0, 0, 0]
6	[0, 0, 0, 0, 0, 0, 1, 0, 0, 0]
7	[0, 0, 0, 0, 0, 0, 0, 1, 0, 0]
8	[0, 0, 0, 0, 0, 0, 0, 0, 1, 0]
9	[0, 0, 0, 0, 0, 0, 0, 0, 0, 1]

Чтобы использовать полносвязную сеть, на вход нужно подать вектор значений [6], т. е. изображение, преобразованное из двухмерного формата в одномерный (рис. 5).

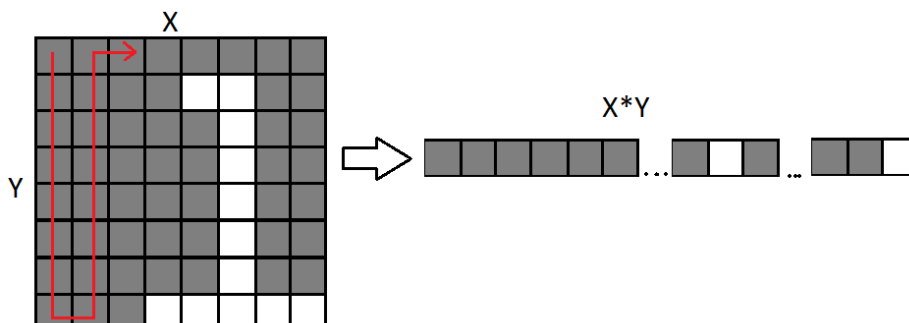


Рисунок 5. Вытягивание изображения в вектор (рисунок авторов)

Путем простых расчетов получаем $28 \times 28 = 784$ нейрона на входном слое. Так как значение пикселя может варьироваться от 0 до 255, то функции активации, вроде гиперболического тангенса, могут уходить в области насыщения. Это вызывает необходимость нормализации — деления значения каждого пикселя на 255, тогда у нас все пиксели будут иметь значения в диапазоне от 0 до 1 [7]. По результатам тестов скрытую часть сети следует

оставить в следующем виде: один скрытый слой на 128 нейронов [8]. Приведем результаты активации функциями ReLu и «гиперболический тангенс». Итоговая структура нейронной сети представлена на рисунке 6.

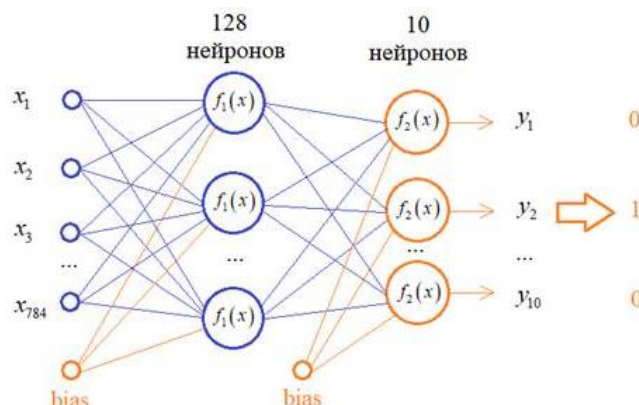


Рисунок 6. Структура нейронной сети (рисунок авторов)

Программа для проведения эксперимента написана на языке Python 3.9³ с использованием пакета Keras 2.10⁴.

Точность тестовой выборки при различных функциях активации практически одинакова (рис. 7).

```
fashion_flatten_relu точность = 88.38 %  
fashion_flatten_tanh точность = 87.96 %
```

Рисунок 7. Точность предсказания полносвязной нейронной сети (рисунок авторов)

Графики точности обучающей и валидационной выборок для каждого типа сети (рис. 8) показывают, что в какой-то момент точность предсказания на обучающей выборке существенно превышает точность валидационной. В целом точность на обучающей выборке должна быть выше, но когда это различие становится существенным, то наступает состояние «переобучения» сети. В данном примере десяти эпох обучения было бы достаточно.

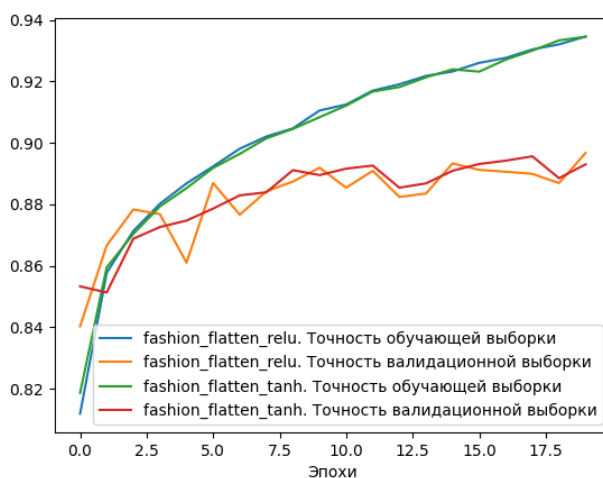


Рисунок 8. Графики обучения нейронной сети (рисунок авторов)

³ Welcome to Python.org // URL: <https://www.python.org/> (Дата обращения 14.01.23).

⁴ Keras: the Python deep learning API // URL: <https://keras.io> (Дата обращения 14.01.23).

На рисунке 9 представлен результат прогона 21 изображения через обе *нейронные сети*: вне скобок — предсказание сети, а внутри — класс, размеченный в наборе данных; без рамок верные предсказания у обеих сетей; красными рамками выделены ошибки; зелеными — верные предсказания в одной из сетей, если у другой результат ошибочен; синими рамками выделены спорные изображения.

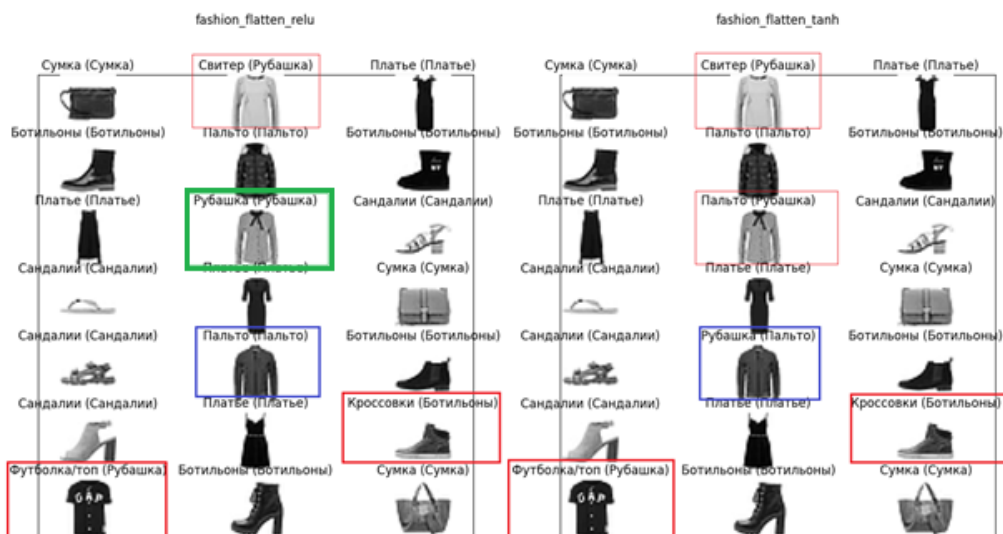


Рисунок 9. Тестирование нейронной сети (рисунок авторов)

На рисунке 10 более крупно изображены спорные результаты. То, что это изделие относится к «пальто» довольно странно: в таком виде оно больше похоже на «свитер» или «рубашку.» Но при этом нейронная сеть с ReLU предсказала правильно относительно разметки, а нейронная сеть с тангенсом больше совпала с «человеческим» восприятием.

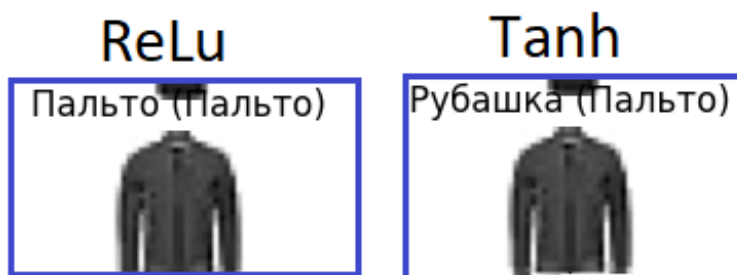


Рисунок 10. Спорное изображение (рисунок авторов)

Вывод

Использование полносвязной нейронной сети в распознавании образов на простых изображениях дает точность около 88 %. Вместе с тем, необходимо сравнительное тестирование с другими архитектурами нейронных сетей, например, со сверточными [9; 10], что и является предметом последующих исследований.

ЛИТЕРАТУРА

1. Смирнов Е.Е. Обучение многослойной нейронной сети определению размера изделий по антропологическим параметрам. / Смирнов Е.Е., Костылева В.В., Разин И.Б., Муртазина А.Р. // Костюмология. 2022. Т. 7. № 3. URL: <https://kostumologiya.ru/PDF/11TLKL322.pdf>.
2. Смирнов Е.Е. Искусственные нейронные сети в решении задач классификации обуви / Е.Е. Смирнов, В.В. Костылева, И.Б. Разин, В.П. Миронов // Костюмология. — 2021. — Т. 6. — № 4. — URL: <https://kostumologiya.ru/PDF/15TLKL421.pdf>.
3. Ключникова В.М., Костылева В.В., Конструирование изделий из кожи. Сер. Высшее образование: Бакалавриат. М.: Издательство: ООО "Научно-издательский центр Инфра-М", 2022. — 353 с.
4. S. Samarasinghe Neural Networks for Applied Sciences and Engineering. From Fundamentals to Complex Pattern Recognition // Auerbach Publications, 2007. 570 с.
5. Домингос П. Верховный алгоритм. Как машинное обучение изменит наш мир // М.: «МАНН, ИВАНОВ И ФЕРБЕР», 2016.
6. H. Simon Neural Networks: A Comprehensive Foundation Second Edition // М.: Вильямс, 2006. 1104 с.
7. H. Brink, J.W. Richards, M. Fetherolf — Real-World Machine Learning // Shelter Island: Manning, 2017.
8. A.Yu. and K. Grauman. "Semantic Jitter: Dense Supervision for Visual Comparisons via Synthetic Images". In ICCV, 2017.
9. Y. LeCun, L. Bottou, G. Orr and K. Muller: Efficient BackProp, in Orr, G. and Muller K. (Eds), Neural Networks: Tricks of the trade, Springer, 1998.
10. Ranzato Marc; Aurelio, Christopher Poultney, Sumit Chopra and Yann LeCun: Efficient Learning of Sparse Representations with an Energy-Based Model, in J. Platt et al. (Eds), Advances in Neural Information Processing Systems (NIPS 2006), MIT Press, 2006.

Smirnov Evgeniy Evgen'evich

Russian State University named A.N. Kosygin (Technology. Design. Art), Moscow, Russia
E-mail: evg7162@mail.ru
RSCI: https://www.elibrary.ru/author_profile.asp?id=1097456

Kostyleva Valentina Vladimirovna

Russian State University named A.N. Kosygin (Technology. Design. Art), Moscow, Russia
E-mail: kostyleva-vv@rguk.ru
RSCI: https://www.elibrary.ru/author_profile.asp?id=353612

Razin Igor Borisovich

Russian State University named A.N. Kosygin (Technology. Design. Art), Moscow, Russia
E-mail: razin-ib@rguk.ru
RSCI: https://www.elibrary.ru/author_profile.asp?id=850439

Murtazina Al'fiya Rustyamovna

Russian State University named A.N. Kosygin (Technology. Design. Art), Moscow, Russia
E-mail: aly1029@yandex.ru
RSCI: https://www.elibrary.ru/author_profile.asp?id=906389

Recognition of the type of clothing and shoes by image

Abstract. The article proposes a solution to the problem of searching for information about a product in conditions of limited data about the product itself and knowledge about the subject area, in particular, recognition and classification of types of clothing and footwear in the image. This will eliminate the following difficulties: (1) how any person who is not a professional in the clothing and footwear industry can quickly find information on the Internet about a particular product; (2) collecting statistics on used clothes and shoes by processing photos as part of promotions with hashtags on social networks, etc. As part of this study, experiments were carried out with a fully connected neural network architecture in the problem of object class recognition on the example of a dataset of 60,000 images with one color channel in grayscale. The purpose of the research was to assess the acceptability of the architecture of a fully connected neural network for recognizing light industry products. It is shown that for the consumer, when choosing a product, searching for information about it, or buying, an image with its description is preferable. If the work on its fragmentation and the selection of significant features is shifted to the neural network, then this will reduce the time for creating a training sample, but increase the time for training the network itself. A fully connected neural network recognizes patterns of simple images with an accuracy of about 88 %. Testing of various activation functions showed that in this problem the choice of function is not critical. The results were identical, however, if it becomes necessary to build a deeper network, then the choice of the activation function of the hidden layers will have a greater influence on the result. Such solutions are supposed to be developed within the framework of the doctoral dissertation of Smirnov E.E. and use in the educational process the departments "Artistic Modeling, Design and Technology of Leather Products", "Information Technology" of the Russian State University. A.N. Kosygin (Technology. Design. Art) in the form of textbooks.

Keywords: neural network; machine learning; data processing; image recognition; image classification; deep learning; classification; light industry products