

УДК 004.032.26; 687.1
DOI 10.47367/0021-3497_2023_5_236

**СРАВНЕНИЕ СВЕРТОЧНЫХ И ПОЛНОСВЯЗНЫХ НЕЙРОННЫХ СЕТЕЙ
ПРИМЕНИТЕЛЬНО К ЗАДАЧАМ РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ**
**COMPARISON OF CONVOLUTIONAL AND FULLY CONNECTED NEURAL
NETWORKS IN RELATION TO IMAGE RECOGNITION TASKS**

Е.Е. СМІРНОВ, В.В. КОСТЫЛЕВА, А.Р. МУРТАЗИНА, И.Б. РАЗИН

E.E. SMIRNOV, V.V. KOSTYLEVA, A.R. MURTAZINA, I.B. RAZIN

**(Российский государственный университет имени А.Н. Косыгина
(Технологии. Дизайн. Искусство))**

(Russian State University named after A.N. Kosygin (Technologies. Design. Art))

E-mail: evg7162@mail.ru

В статье рассматривается основополагающая проблема выбора архитектуры нейронной сети для решения задачи распознавания образов на плоском изображении. Показано, что баланс между скоростью обучения, временем проверки и количеством ошибок определяет затраты на обучение, которые существенно возрастают при повышении точности предсказаний. Представлены результаты прогнозов, полученные в ходе экспериментов с использованием архитектур сверточной нейронной сети, для тестовых наборов изображений изделий легкой промышленности.

The article deals with the fundamental problem of choosing a neural network architecture for solving the problem of pattern recognition on a flat image. It is shown that the balance between the learning rate, verification time, and the count of errors determines the training costs, which increase significantly with increasing prediction accuracy. The results of forecasts obtained in the course of experiments using convolutional neural network architectures for test sets of light industry products images are presented.

Ключевые слова: нейронная сеть, машинное обучение, обработка данных, распознавание изображений, классификация, изделия легкой промышленности, сверточные нейронные сети.

Keywords: neural network, machine learning, data processing, image recognition, classification, light industry products, convolutional neural networks.

Введение

В статьях [1-4] нами рассмотрено применение полносвязной нейронной сети (ПНС) для распознавания видов одежды на примере набора данных *mnist_fashion* [5]. В настоящей статье обсуждаются возможности альтернативного типа нейронных сетей – сверточных (СНС).

Идея сверточных сетей была основана на биологических исследованиях. В ходе изучения зрительной коры головного мозга кошки обнаружено, что существуют «простые клетки», которые особо активно реагируют на прямые линии под разными углами, и «сложные клетки», реагирующие на движение линий в одном направлении. Ян Ле Кун предложил использовать так называемые сверточные нейронные сети (Convolutional Neural Networks – CNN) [6].

Смысл заключается в чередовании сверточных (C-layers), субдискретизирующих (S-layers) и выходных полносвязных слоев (F-layers) (рис. 1 – пример архитектуры сверточной нейронной сети).

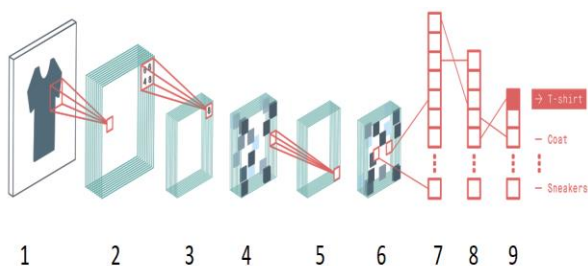


Рис. 1

Как видно на рис. 1, СНС имеет более сложную структуру, чем полносвязная сеть. В данном примере имеется 9 слоев: №1 – входной слой, на который подается изображение формата 28x28 с одним каналом; №2 и №5 – сверточные слои, которые позволяют выделять энные характерные признаки. Происходит это благодаря некоторым определенным маскам или сверткам [6, 7]. Чаще всего окно свертки принимают размером 3x3 со случайно инициализированными весами, после чего этой сверткой проходят по всему изображению с каким-то шагом, получая новую карту признаков.

На рис. 2 продемонстрирован пример с выделением признаков, подходящих под свертку. Очевидно, что матрица свертки с «1» в центральном столбце выделяет вертикальные линии, а с «1» в центральной строке – горизонтальные. Большое количество сверток позволяет вычлнить внушительный список значимых признаков. На первом этапе это, как правило, примитивные линии. На выходе сверточного слоя из одного изображения мы получаем некоторое количество новых.

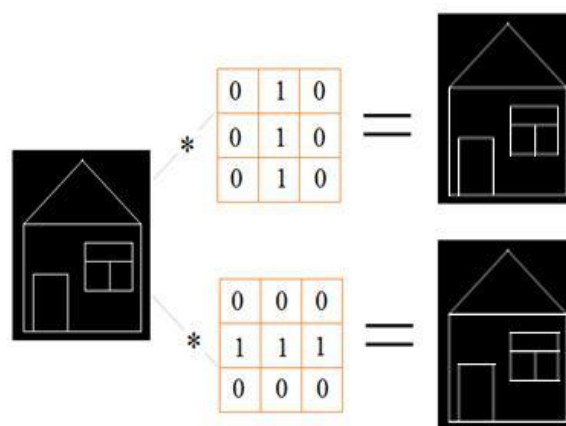


Рис. 2

В паре со сверточным часто идет слой субдискретизирующий, который служит снижению размера изображения с переносом необходимых признаков (на рис. 1 он отмечен под номером 3). Это позволяет получать более сложные значимые признаки, чем на предыдущем этапе. Субдискретизирующие слои делятся на максимизирующие, усредняющие, минимизирующие и выбираются исходя из поставленной задачи. Использование такого слоя при маске размером 2x2 уменьшает размер изображения в 4 раза.

Слой №№ 4, 6 называют слоями dropout [8, 9] или исключением. Они позволяют частично решать проблему переобучения и могут применяться как в СНС, так и ПНС. В рамках этой статьи мы на этом детально не останавливаемся. Слои №№ 7-9 – это классическая ПНС, описанная ранее.

Эксперимент состоял из трех этапов:

- 1) набор данных *mnist_fashion* [5];

2) набор данных из цветных изображений размера 132x102 пикселя с малым количеством посторонних объектов [10];

3) набор данных из цветных изображений размером 32x32 пикселя с большим количеством посторонних объектов.

Экспериментальные исследования

Этап 1

Архитектура сверточной нейронной сети представлена в виде табл. 1.

Таблица 1

Слой (функция активации)	Выходная размерность	Количество настраиваемых параметров
conv2d (ReLU)	(28, 28, 32)	320
субдискретизирующий	(14, 14, 32)	0
conv2d_1 (ReLU)	(14, 14, 64)	18496
max_pooling2d_1	(7, 7, 64)	0
flatten	(3136)	0
dense (ReLU)	(128)	401536
dropout	(128)	0
dense_1 (ReLU)	(64)	8256
dropout_1	(64)	0
dense_2 (softmax)	(10)	650

В качестве сверточной части сети взяты две пары сверточных (conv2d) и субдискретизирующих (max_pooling2d) сетей [11], после чего, вытягивая с последнего слоя сверточной части весь набор карт признаков, на первом слое (dense) полносвязной части нейронной сети получили $7*7*64 =$

3136 нейронов. Между скрытыми полносвязными слоями добавили слой dropout. В качестве функции активации везде взята ReLu. У выходного слоя по-прежнему SoftMax.

Количество настраиваемых параметров текущей архитектуры СНН (cnn_standart) составляет 429258 и отличается более чем в 4.2 раза в сравнении с ПНС (fashion_flatten_relu – 101770). Это говорит о больших временных затратах на обучение. В нашем случае они выросли приблизительно в 6.9 раза при увеличении количества параметров в 4.2. Это происходит из-за серьезной многократной обработки изображений [12]. Точность обучения данной архитектуры СНС составила приблизительно 91% (рис. 3, А), в то время как точность полносвязных – около 88% (рис. 3, Б).

А) cnn_standart точность = 91.13 %
 Б) fashion_flatten_relu точность = 88.38 %

Рис. 3

Таким образом, на этом простом наборе данных за счет сверточной структуры сократилось количество ошибок примерно на 25%. На рис. 4 приведен фрагмент предсказания, где рамками выделены различия в предсказаниях: красные – ошибка, зеленые – верное.

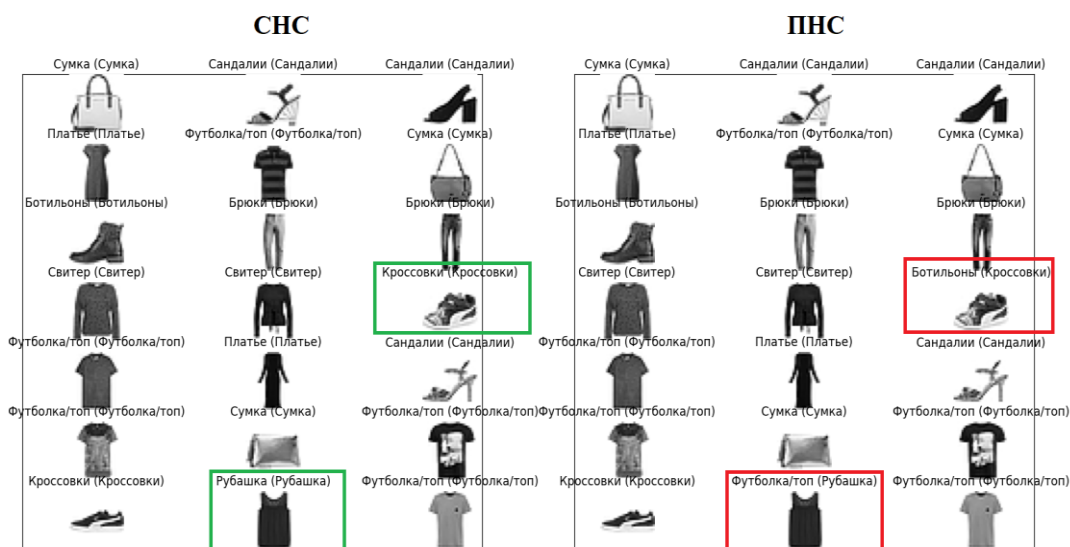


Рис. 4

Снижение количества ошибок довольно значительно, но и временные затраты на обучение в 7 раз выше. Может сложиться впечатление, что использование СНС не имеет особого смысла. Чтобы проверить, так это или нет, проведем еще один эксперимент.

Этап 2

На этом этапе воспользуемся большим набором данных об обуви, состоящим из 50 000 изображений каталога, собранных с zappos.com [13, 14]. Изображения разделены на 4 основные категории: туфли, сандалии, тапочки и сапоги (shoes, sandals, slippers, boots). Для удобства анализа [15] обувь центрирована на белом фоне и изображена в той же ориентации. Изображения представлены в формате 132x102 пикселя с тремя цветовыми каналами. Фрагмент набора показан на рис. 5.

Структуры сетей взяты из предыдущего этапа с изменением только входного и выходного слоев. Их необходимо скорректировать под характеристики выборки. В ней присутствует четыре класса, значит, на выходе будет четыре нейрона. Входной слой

для полносвязной сети скорректируется на $132 \times 102 \times 3 = 40\,392$ нейрона.



Рис. 5

Скорость одной эпохи обучения в этом случае существенно изменилась – в 8,92 раза. Относительно предыдущего этапа время на обучение сверточной сети увеличилось в 12,5 раза, а полносвязной – в 9,5 раза. Зависимость времени на обучение от размерности входных данных близка к линейной.

На рис. 6 представлены результаты прогона тестового набора через нейронные сети. Рамками выделены расхождения в предсказаниях: красная – ошибка; зеленая – верное предсказание. Из рисунка видно, что полносвязная сеть ошибается чаще.



Рис. 6

Полные данные о точности предсказания сетей по всей тестовой выборке составили: для полносвязной – 87%, сверточной – 94%.

В этом случае количество ошибок при использовании сверточной сети сокращается

на 49%, что явно просматривается на графике точности обучающей и валидационной выборок обеих сетей (рис. 7 – графики обучения нейронных сетей)

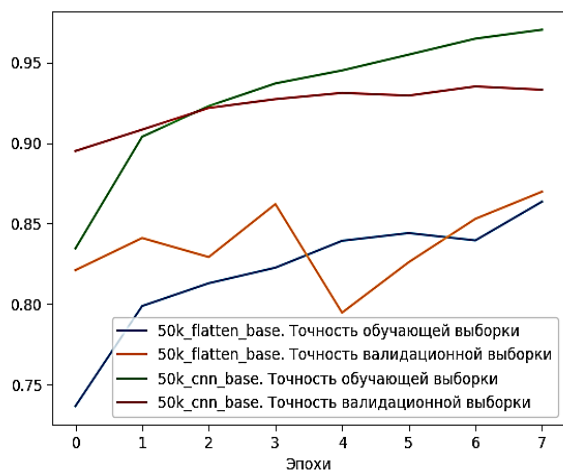


Рис. 7

Этап 3

На этом этапе протестировано выделение целевых объектов среди посторонних. Возьмем похожий по параметрам набор cifar10. Его отличия от mnist_fashion: три цветовых канала вместо одного; на изображении кроме основного объекта присутствуют второстепенные. В этот набор также включены 60 000 изображений формата 32x32 с тремя цветовыми каналами, разбитых в пропорциях 50 000/10 000 на 10 классов. Тогда нам необходимо всего лишь скорректировать входные слои под новую размерность изображений, а все остальные настройки сетей можно оставить прежними. Таким образом эксперимент будет максимально точно показывать разницу между обработками цветных изображений с дополнительными деталями и черно-белых без таковых.

В результате точность полносвязной сети составила всего 43.59%, а сверточной – 74.18% (рис. 8).

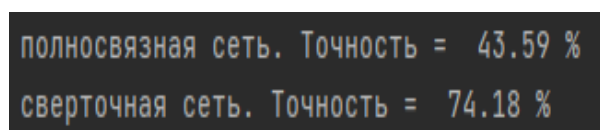


Рис. 8

Этот разрыв отчетливо виден на графиках точности (рис. 9).

На этом примере разница между рассмотренными типами сетей стала заметнее (количество ошибок ниже почти в 2.2 раза).

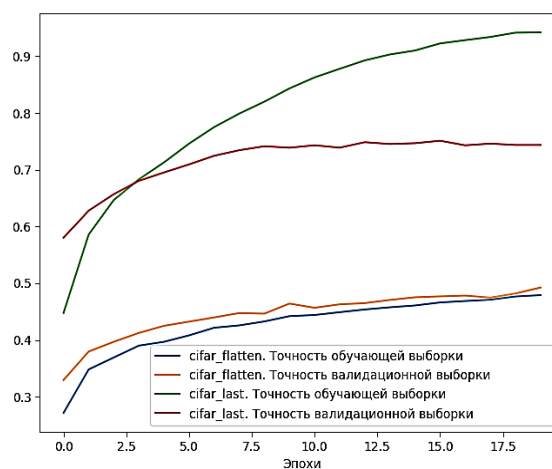


Рис. 9

Это обусловлено как раз способностями сверточных сетей обобщать фрагменты изображений, чего в свою очередь лишены полносвязные: при вытягивании изображения в линейную цепочку байт мы теряем топологию изображения [16], и, чем большего размера само изображение и количество деталей на нем, тем заметнее этот эффект.

В аналогичных работах точность распознавания сверточной сети составляет от 61% до 82% (в зависимости от области применения: медицина, промышленность, карты местности и др.) [17-19]. Подобные результаты обусловлены сложностью распознавания объектов, качеством изображений, объемом выборки и архитектурой сети.

ВЫВОДЫ

Результаты экспериментов показывают, что применение сверточной архитектуры нейронной сети в задачах распознавания изображений позволяет получить существенное сокращение ошибочных предсказаний в сравнении с полносвязной. При этом, чем больше размер собственно изображения и количество деталей на нем, тем значительнее будет проявляться этот эф-

факт. Но за более высокую точность приходится платить существенным ростом потребности в вычислительных мощностях и времени. Вопросам поиска приемлемых компромиссных решений будут посвящены последующие исследования. Полученные результаты служат развитию, например, научной теории автоматизированного проектирования вкладных ортопедических приспособлений обуви. С использованием нейронной сети сверточной архитектуры для определения состояния стоп, их размеров, особенностей и зон наибольшего давления нами исследуются подометрические снимки стоп. Это имеет существенное практическое и социальное значение, так как здоровьесберегающий эффект конструкции стельки в устранении перегрузок нижних конечностей, обеспечении их правильного анатомического положения и пресечении прогрессирования деформаций тем более действенен, чем точнее ее ложемент соответствует индивидуальному рельефу плантарной поверхности стопы.

ЛИТЕРАТУРА

1. Смирнов Е.Е., Костылева В.В., Разин И.Б., Муртазина А.Р. Распознавание вида одежды и обуви по изображению // Костумология. 2023. Т. 8. № 1. – <https://kostumologiya.ru/PDF/03TLKL123.pdf>.
2. Смирнов Е.Е., Костылева В.В., Разин И.Б., Миронов В.П. Искусственные нейронные сети в решении задач классификации обуви // Костумология. 2021. Т. 6. № 4. – <https://kostumologiya.ru/PDF/15TLKL421.pdf>
3. Белгородский В.С., Гусева М.А., Андреева Е.Г., Рогожина Ю.В. Искусственный интеллект в оценке качества готовой швейной продукции // Изв. вузов. Технология текстильной промышленности. 2022. № 2 (398). С. 168...177. – DOI 10.47367/0021-3497_2022_2_168
4. Романовский Р.С., Петросова И.А., Андреева Е.Г., Шитлова Е.А. Разработка новых моделей мужской одежды с учетом рекомендаций искусственного интеллекта // Изв. вузов. Технология текстильной промышленности. 2021. No 4 (394). С. 145...152. – DOI 10.47367/0021-3497_2021_4_145
5. Fashion MNIST dataset, an alternative to MNIST. – https://keras.io/api/datasets/fashion_mnist/ (Дата обращения 15.01.23).
6. Cun Y.L., Denker J. S., Solla S., Howard R.E., Jackel L.D. Optimal Brain Damage, in Touretzky, David (Eds), Advances in Neural Information Processing Systems 2 (NIPS*89), Morgan Kaufman, Denver, CO, 1990.

7. Cun Y.L., Bengio Y. Convolutional Networks for Images, Speech, and Time-Series, in Arbib, M. A. (Eds), The Handbook of Brain Theory and Neural Networks, MIT Press, 1995.

8. Cun Y.L., Bottou L., Orr G., Muller K. Efficient Back Prop, in Orr, G. and Muller K. (Eds), Neural Networks: Tricks of the trade, Springer, 1998.

9. Aurelio R.M., Poulitney Ch., Chopra S., Cun Y.L. 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.

10. UTZappos50K – <https://vision.cs.utexas.edu/projects/finegrained/utzap50k/> (Дата обращения: 20.02.2023).

11. Samarasinghe S. Neural Networks for Applied Sciences and Engineering. From Fundamentals to Complex Pattern Recognition // Auerbach Publications, 2007. 570 p.

12. Simon H. Neural Networks: A Comprehensive Foundation Second Edition. М.: Вильямс, 2006. 1104 с.

13. Shoes, Sneakers, Boots, & Clothing + FREE SHIPPING. – <https://www.zappos.com/> (дата обращения 20.02.2023).

14. Yu A., Grauman K. Fine-Grained Visual Comparisons with Local Learning. In CVPR, 2014.

15. Yu A., Grauman K. Semantic Jitter: Dense Supervision for Visual Comparisons via Synthetic Images. In ICCV, 2017

16. Brink H., Richards J.W., Fetherolf M. Real-World Machine Learning // Shelter Island: Manning, 2017.

17. Van Nguyen S., Tran H.M. & Le T.S. Application of Geometric Modeling in Visualizing the Medical Image Dataset. SN COMPUT. SCI. 2020, 1, 254. - <https://doi.org/10.1007/s42979-020-00266-0>

18. Ruiz R.B., Velásquez J.D. Artificial Intelligence for the Future of Medicine. In: Lim, C.P., Vaidya, A., Chen, YW., Jain, V., Jain, L.C. (eds) Artificial Intelligence and Machine Learning for Healthcare. Intelligent Systems Reference Library, 2023. vol. 229. Springer, Cham. https://doi.org/10.1007/978-3-031-11170-9_1

19. Guang Jia, Xunan Huang, Sen Tao etc. Artificial intelligence-based medical image segmentation for 3D printing and naked eye 3D visualization, Intelligent Medicine, Volume 2, Issue 1, 2022, Pages 48-53. – <https://doi.org/10.1016/j.imed.2021.04.001>.

REFERENCES

1. Smirnov E.E., Kostyleva V.V., Razin I.B., Murtaзина A.R. Recognition of the type of clothing and shoes by image. Journal of Clothing Science. 2023, 8(1) – <https://kostumologiya.ru/PDF/03TLKL123.pdf>.

2. Smirnov E.E., Kostyleva V.V., Razin I.B., Mironov V.P. Artificial neural networks in solving problems of shoe classification. Journal of Clothing Science, 4(6). – <https://kostumologiya.ru/PDF/15TLKL421.pdf>

3. Fashion MNIST dataset, an alternative to MNIST. – https://keras.io/api/datasets/fashion_mnist/ (date of the application 15.01.23).

4. *Belgorodsky V.S., Guseva M.A., Andreeva E.G., Rogozhina Yu.V.* Artificial intelligence in quality assessment of finished sewing products // *Izvestiya Vysshikh Uchebnykh Zavedenii, Seriya Tekhnologiya Tekstil'noi Promyshlennosti*. 2022. № 2 (398). P. 168...177. – DOI 10.47367/0021-3497_2022_2_168
5. *Romanovsky R.S., Petrosova I.A., Andreeva E.G., Shipilova E.A.* Development of new models men's clothing taking into account the recommendations of artificial intelligence // *Izvestiya Vysshikh Uchebnykh Zavedenii, Seriya Tekhnologiya Tekstil'noi Promyshlennosti*. 2021, No 4 (394). P. 145...152. – DOI 10.47367/0021-3497_2021_4_145
6. *Cun Y.L., Denker J. S., Solla S., Howard R.E., Jackel L.D.* Optimal Brain Damage, in Touretzky, David (Eds), *Advances in Neural Information Processing Systems 2 (NIPS*89)*, Morgan Kaufman, Denver, CO, 1990
7. *Cun Y.L., Bengio Y.* Convolutional Networks for Images, Speech, and Time-Series, in Arbib, M. A. (Eds), *The Handbook of Brain Theory and Neural Networks*, MIT Press, 1995.
8. *Cun Y.L., Bottou L., Orr G., Muller K.* Muller: Efficient BackProp, in Orr, G. and Muller K. (Eds), *Neural Networks: Tricks of the trade*, Springer, 1998.
9. *Aurelio R.M., Poultney Ch., Chopra S., Cun Y.L.* 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.
10. UTZappos50K – <https://vision.cs.utexas.edu/projects/finegrained/utzap50k/> (Дата обращения: 20.02.2023).
11. *Samarasinghe S.* Neural Networks for Applied Sciences and Engineering. From Fundamentals to Complex Pattern Recognition // Auerbach Publications, 2007. 570 с.
12. *Simon H.* Neural Networks: A Comprehensive Foundation Second Edition. M.: Williams, 2006. 1104 p.
13. Shoes, Sneakers, Boots, & Clothing + FREE SHIPPING. – <https://www.zappos.com/> (date of the application 20.02.2023).
14. *Yu A., Grauman K.* Fine-Grained Visual Comparisons with Local Learning. In CVPR, 2014.
15. *Yu A., Grauman K.* Semantic Jitter: Dense Supervision for Visual Comparisons via Synthetic Images. In ICCV, 2017.
16. *Brink H., Richards J.W., Fetherolf M.* Real-World Machine Learning // Shelter Island: Manning, 2017.
17. *Van Nguyen S., Tran H.M. & Le T.S.* Application of Geometric Modeling in Visualizing the Medical Image Dataset. SN COMPUT. SCI. 1, 254 (2020). <https://doi.org/10.1007/s42979-020-00266-0>
18. *Ruiz R.B., Velásquez J.D.* Artificial Intelligence for the Future of Medicine. In: Lim, C.P., Vaidya, A., Chen, YW., Jain, V., Jain, L.C. (eds) *Artificial Intelligence and Machine Learning for Healthcare*. Intelligent Systems Reference Library, 2023. vol 229. Springer, Cham. https://doi.org/10.1007/978-3-031-11170-9_1
19. *Guang Jia, Xunan Huang, Sen Tao etc.* Artificial intelligence-based medical image segmentation for 3D printing and naked eye 3D visualization, *Intelligent Medicine*, Volume 2, Issue 1, 2022, Pages 48-53. – <https://doi.org/10.1016/j.imed.2021.04.001>.

Рекомендована кафедрой художественного моделирования, конструирования и технологии изделий из кожи РГУ имени А.Н. Косыгина (Технологии. Дизайн. Искусство). Поступила 06.06.23.