

УДК 685.34

DOI 10.47367/0021-3497\_2025\_1\_285

**ОЦЕНКА ВОЗМОЖНОСТЕЙ НЕЙРОННОЙ СЕТИ В ПРЕДСКАЗАНИЯХ  
РАЗМЕРА ОБУВИ И ВЕСА ПАЦИЕНТА ПО ПОДОГРАММЕ**

**ASSESSMENT OF THE CAPABILITIES OF THE NEURAL NETWORK  
IN PREDICTING THE PATIENT'S SHOE SIZE AND WEIGHT BY PODOGRAM**

*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 им. 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 влияния различных параметров нейронных сетей на поиск оптимума и скорость сходимости в задачах распознавания плоских изображений. Показано, что в специфических задачах, таких как распознавание подометрических снимков, нужно гибко подходить к построению архитектуры нейронной сети. В частности, приходится отказываться от стандартных функций потерь и разрабатывать собственные, более точно описывающие предметную область. Представлены результаты прогнозов, полученные в ходе экспериментов с использованием различных архитектур сверточной нейронной сети для реальной базы данных пациентов врача ортопеда.*

*The article deals with the fundamental problem of the influence of various parameters of neural networks on the search for the optimum and the rate of convergence in flat image recognition problems. It is shown that in specific tasks, such as recognition of podometric images, it is necessary to take a flexible approach to the construction of a neural network architecture. In particular, we have to abandon the standard loss functions and develop our own ones that more accurately describe the subject area. The results of forecasts obtained during experiments using various convolutional neural network architectures for a real database of orthopedic patients are presented.*

**Ключевые слова:** нейронная сеть, машинное обучение, обработка данных, распознавание изображений, регрессия, подографический снимок, сверточные нейронные сети, глубокое обучение.

**Keywords:** neural network, machine learning, data processing, image recognition, regression, podogram, convolutional neural networks, deep learning.

### Введение

Для определения расположения отпечатков стоп на подограмме нами предложено использовать сверточную нейронную сеть (CNN) [1, 2, 3]. Результат применения такого типа сетей оказался удовлетворительным. В настоящих исследованиях поставлена задача определить размер обуви и вес пациента по компьютерному снимку плантарных поверхностей, полученных с помощью бароподометрической платформы. Платформа считывает величину давления стоп на опору и передает их в компьютер в виде изображений отделов стопы, испытывающих разные степени нагрузки. Эти степени отличаются цветовыми и цифровыми значениями. Снимки стопы выводятся на экран монитора. Первое, что необходимо сделать, это оценить динамику цве-

товой карты в зависимости от веса. Проблема может заключаться в том, что некоторые подометры показывают не абсолютную, а относительную карту. Последнее существенно усложняет задачу.

Для проверки возьмем двух пациентов со значительным разбросом показателей по весу и оценим цветовую шкалу. На рис. 1, а представлена цветовая шкала для пациента весом 39 кг и 35-м размером обуви, а на рис. 1, б – 81 кг и 38-м размером. Разница между 38-м и 35-м размерами составляет примерно 10% ( $\approx 24$  см и  $\approx 22$  см соответственно), а вес при этом отличается более чем в 2 раза. Однако цветовые шкалы давления одинаковые. Значит, мы имеем дело с абсолютными значениями.

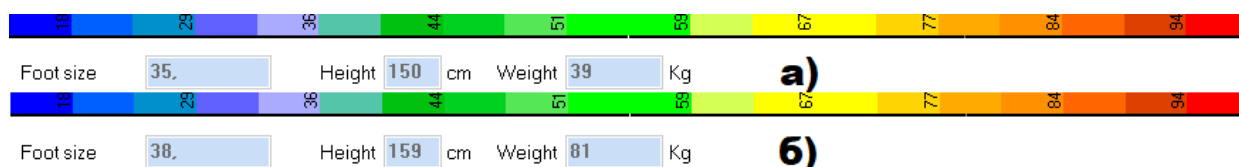


Рис. 1

### Обучающие данные

Набор данных возьмем из результатов предыдущих исследований [1]. Одним из выводов проведенного ранее эксперимента является предположение о недостатке исходных данных. Когда мы используем одно и то же изображение много раз, сеть переобучается и теряет свои обобщающие способности, вследствие чего не может

полноценно функционировать. В табл. 1 и на рис. 2 представлено распределение исходных данных. Количество рассчитывается по группировке размера и веса, округленного до десятка в большую сторону. Даже при объединении весов по десяткам большинство групп очень малочисленно. При этом в сеть данные будут поступать в исходном виде.

Т а б л и ц а 1

Размер	Вес	Количество	Размер	Вес	Количество	Размер	Вес	Количество
29	40	1	39	50	4	42	80	3
32	20	1	39	60	6	42	90	2
35	30	1	39	70	4	42	100	5
35	40	1	39	80	7	42	110	1
36	40	2	39	90	2	43	70	3
36	50	2	39	100	3	43	80	4
36	60	4	40	50	1	43	90	4
37	40	1	40	60	5	43	100	3
37	50	1	40	70	2	43	130	1
37	60	7	40	80	4	44	70	1
37	70	8	40	90	2	44	80	5
37	80	3	40	110	1	44	90	2
37	100	1	40	130	1	44	100	3
38	40	1	41	40	1	45	70	3
38	50	1	41	60	1	45	90	3

38	60	12	41	70	6	45	110	1
38	70	3	41	80	4	45	130	1
38	80	6	41	100	4	46	80	1
38	100	1	42	60	2	47	90	1
39	40	1	42	70	5	47	120	1

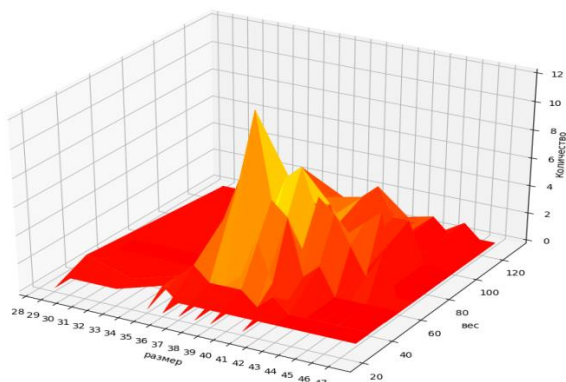


Рис. 2

Очевидно, что необходим дополнительный сбор информации, который не всегда можно быстро осуществить, а в некоторых ситуациях практически не реален. В таких случаях привлекателен метод генерации данных [4, 5]. Идея заключается в том, что исходное изображение меняется без потери смысловой нагрузки. Например, задача заключается в обучении сети распознавать на изображении лошадей. Для человека не имеет значения, в какую сторону смотрит лошадь: вправо или влево. Однако нейронная сеть будет воспринимать их как два разных изображения. Следовательно, для увеличения объема обучающих данных нужны простые операции, например, создание зеркальной копии, масштабирование, поворот на  $90^0$  или  $180^0$  и т. п. На этом этапе важно правильно оценить, какие деформации изображения допустимы в текущей задаче [6, 7].

В нашем случае использованы приемы «отражение по горизонтали» и «перенос расположения отпечатков». Изображение разбивали по вертикальной оси на две равные части, после чего обе части зеркально смещались по горизонтальной оси относительно вертикальной в диапазоне  $[-25px; +25px]$ . По вертикали оба фрагмента смещались синхронно в диапазоне  $[-50px; +50px]$ .

Ниже представлены примеры сгенерированных изображений. Как видно, смысловая нагрузка не нарушена. Мы просто имитировали различную постановку стоп на подограф. Для эксперимента было сгенерировано 2000 изображений из 200 исходных.

Epoch 1/50  
3/3 [=====] - 231s 73s/step

Далее необходимо хотя бы приблизительно оценить временные затраты на обучение [8, 9]. Для начала возьмем примерную архитектуру глубокой сверточной нейронной сети. Затем блок данных, состоящий из 50 исходных трехканальных (RGB) изображений размером  $500 \times 500$  px, отправим на обучение. Общее число тренировочных объектов, представленных в одном пакете, устанавливалось равным 16, поэтому в одной эпохе будет 3 шага. Как видно (рис. 3), одна эпоха обучения заняла 231 секунду.

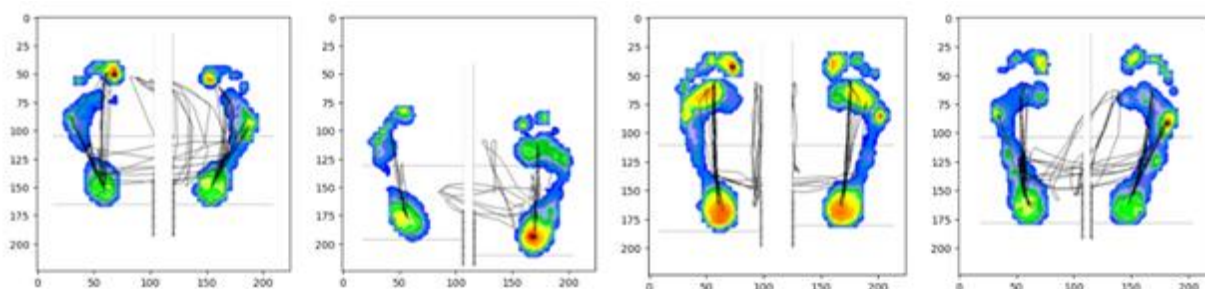
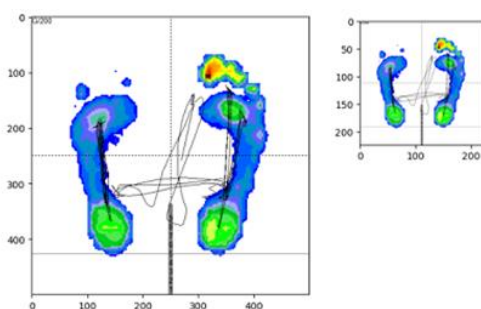


Рис. 3

Проецируя на набор данных в 2000 изображений, можно предположить, что общее время выполнения одной эпохи равно примерно 9240 секунд или  $\approx 2,5$  часа. Это слишком большие показатели: проведение одного эксперимента может продлиться несколько десятков дней. Конечно, время не будет расти пропорционально, так как в каждую эпоху включена также валидация, сохранение весов модели и много других операций, что при росте входных данных будет увеличивать пропорцию на само обучение, а значит, уменьшать абсолютные показатели. Точное время будет понятно при полноценном обучении, но это необходимо учитывать уже на начальных этапах построения архитектуры сети.

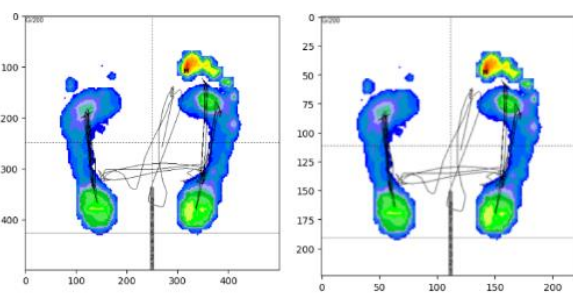
Для решения данной проблемы можно:

- уменьшить количество входных данных;
- увеличить вычислительные мощности;
- уменьшить нейронную сеть;



**а) б)**

Рис. 4



**а) б)**

Рис. 5

На основе визуальной оценки такой подход можно рекомендовать для подготовки данных. Новые изображения получаются размерностью  $224 \times 224 \times 3$  и состоят из 150528 чисел, что почти в 5 раз меньше исходного изображения. Возьмем такой же блок из 50 изображений и отправим в ту же сеть. Получаем время эпохи 46 секунд. Выигрыш во времени в 5,02 раза.

Epoch 1/50  
3/3 [=====] - 46s 14s/step

- уменьшить размерность входных данных.

Первые три варианта имеют явные недостатки или сложности в реализации, а вот вариант №4 можно рассмотреть подробнее. Одно цветное изображение состоит из 750000 чисел. Вместе с глубокой сетью это представляет огромную вычислительную задачу. Можно попробовать оценить потери информации при сжатии изображения. Поскольку CNN в некоторой степени имитирует зрение [10, 11], то для начала их можно оценить визуально. За основу возьмем известные CNN vgg16 и vgg19. У этих сетей на вход подаются изображения размерностью  $224 \times 224 \times 3$ . На рис. 4 показаны исходное (а) и пропорционально сжатое (б) изображения. На рис. 5 показаны исходное (а) и растянутое до размеров исходного уменьшенное (б) изображения. Как видно, сжатое изображение (рис. 5, б) почти не теряет в качестве по сравнению с исходным (рис. 5, а).

### Обучение

В процессе подбора приемлемой архитектуры зачастую встречаются некачественные модели. В начале эксперимента при использовании относительно простых сверточных архитектур [1] почти все модели имели общую черту – для всех входных данных предсказывали примерно одно и то же значение. Результат их работы очень похож на классическую линейную аппроксимацию (рис. 6, б) или вообще оторван от реальности (рис. 6, а).

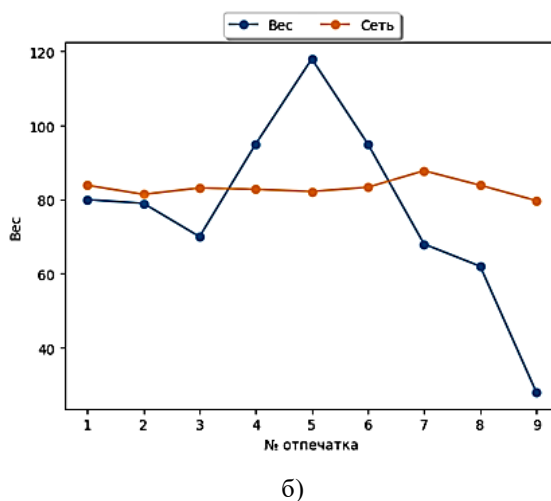
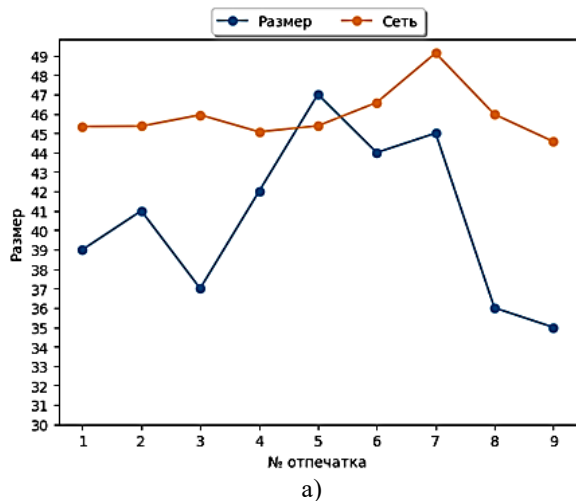


Рис. 6

Исходя из постановки задачи сеть имеет 2 выхода. Основные функции потерь для задач регрессии – это mse и mae. Разброс размеров в этом наборе составляет  $47,0-29,0=16$ , при этом разброс веса равен  $131,0 \text{ кг}-25,0 \text{ кг}=106 \text{ кг}$ . В абсолютных единицах диапазон изменчивости веса превышает диапазон размеров в 6,6 раза, следовательно, ошибка в предсказании веса будет влиять на значение этих функций сильнее, чем в предсказании размера. При этом ошибка в предсказании веса в 3 кг в данной задаче не несет никакой смысловой

нагрузки, в то время как ошибка на 3 размера катастрофична. Отчасти этим объясняется картина предсказаний (рис. 6), что подтверждается и данными табл. 2, содержащей информацию по тестовой выборке и результаты функций MSE и MAE по каждому показателю. При этом mae по весу в  $\approx 3,4$  раза больше, чем по размеру.

$$\begin{aligned} \text{MSE} &= \frac{1}{n} \sum_{t=1}^n e_t^2 \\ \text{MAE} &= \frac{1}{n} \sum_{t=1}^n |e_t| \end{aligned}$$

Т а б л и ц а 2

Выборка									
Размер					Вес				
Показатели			Ошибки предсказаний		Показатели			Ошибки предсказаний	
ср. зн.	мин	макс	mse	mae	ср. зн.	мин	макс	mse	mae
40.67	35.0	47.0	40.65	5.62	77.22	28.0	118.0	589.69	19.17

#### Решение проблем разброса параметров и ошибки предсказаний

Первое, что можно предложить, – снизить влияние ошибки веса на значение результата. В пакете Keras [5] нет возможности устанавливать произвольные коэффициенты в стандартных функциях. Для этого необходимо написать собственную пользовательскую функцию потерь:

$$MY\_MAE = \frac{1}{n} \sum_{t=1}^n k_r * |e_{tr}| + k_w * |e_{tw}|$$

где  $k_r$  – коэффициент важности размера;  $k_w$  – коэффициент важности веса;  $e_{tr}$  – ошибка по размеру;  $e_{tw}$  – ошибка по весу.

Примем  $k_r$  равным 1, а  $k_w$  – 0,05. Графики функций потерь представлены на рис. 7. Как можно заметить, график стандартной функции менее плавный, общее значение ошибки существенно меньше, а значит, и вклад в поиск минимума функции параметра размера выше.

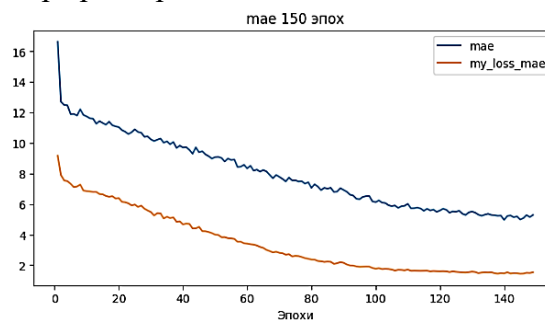


Рис. 7

Функция потерь – это только один фактор, влияющий на скорость и точность обучения. Кроме него, нужно применять и другие подходы.

*Нормализация по мини-батчам и отключение нейронов*

В некоторых случаях, когда обучение затягивается или не приносит желаемого результата, можно попробовать применить нормализацию по мини-батчам (batchnormalization – bn) [12, 13] и/или отключение части нейронов (dropout – do) [14].

BN позволяет приводить данные к единичной дисперсии и нулевым математиче-

ским ожиданиям и избавиться от внутреннего ковариационного сдвига:

$$a) m_v = M\{V\} = \frac{1}{m} \sum_{i=1}^m v_k^i$$

$$б) \sigma_v^2 = M\{(V - m_v)^2\} = \frac{1}{m-1} \sum_{i=1}^m (v_k^i - m_v)^2 \approx \frac{1}{m} \sum_{i=1}^m (v_k^i - m_v)^2$$

$$в) z_k^i = \frac{v_k^i - m_v}{\sqrt{\sigma_v^2 + \varepsilon}}$$

На рис. 8 представлен пример обучения некоторой сети с применением BN, DO и без них на наборе данных cifar10. Метод BN существенно ускоряет процессы обучения и сходимости модели.

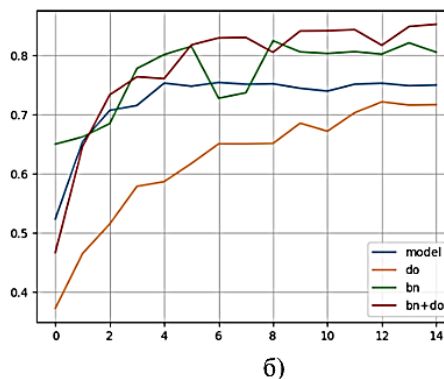
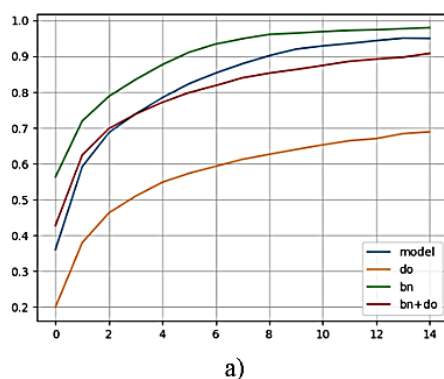


Рис. 8

Метод DO позволяет избежать переобучения посредством случайного отключения нейронов. Таким образом, происходит повышение специализации каждого нейрона в отдельности и имитация обучения сразу нескольких нейронных сетей, объединяемых в процессе эксплуатации в одну сеть. Именно этим определяется график сходимости на обучающей выборке. В то время как BN и обычная модель обучают одну сеть, DO обучает несколько более обобщающих, поэтому на выборке валидации точность предсказаний выше, чем на тренировочной, т. к. на этом этапе все версии сети сводятся в одну. Наиболее хороший результат показала комбинация BN+DO. На рис. 8, б самая быстрая сходимость и наиболее высокая точность.

Таким образом, обоснована необходимость проверки этих методов в текущей задаче.

### Обобщенная модель нейронной сети

Выбранная нами архитектура нейронной сети (рис. 9) включает в себя 5 сверточных блоков (Conv Block), 2 полносвязные независимые сети и выходной слой из двух нейронов (FCNN). Каждый сверточный блок содержит от 2 до 4 сверточных слоев, слой BN до функции активации, слой с функцией активации ReLu и DO с вероятностью выброса нейронов от 10% до 25%. Полносвязные сети одинаковы и имеют по 2 слоя с применением как BN, так и DO и по одному нейрону на выходных слоях с линейной функцией активации, которые объединяются в общий выходной слой.

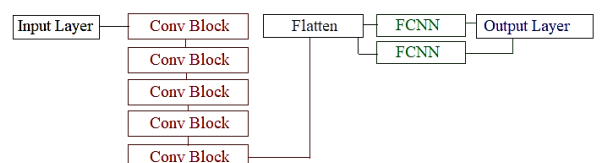


Рис. 9

За время обучения прошло 150 эпох. Срезы по обучению создавались каждые 50 эпох. На рис. 10 и в табл. 3 представле-

ны исходные данные и предсказания по эпохам.

Таблица 3

№	Размер	50 эпох	100 эпох	150 эпох	Вес	50 эпох	100 эпох	150 эпох
1	39	37,10	38,67	40,01	80	75,29	73,28	85,86
2	41	39,33	44,41	45,24	79	81,20	80,72	82,30
3	37	35,88	37,78	39,11	70	67,92	64,99	76,97
4	42	37,88	42,56	43,64	95	76,45	95,87	95,96
5	47	39,52	44,27	44,28	118	76,48	93,32	97,53
6	44	37,74	42,35	43,41	95	76,84	77,96	87,49
7	45	39,54	42,62	42,93	68	79,64	75,40	74,82
8	36	35,29	35,67	38,70	62	61,58	57,91	78,70
9	35	34,17	33,61	38,19	28	57,48	45,18	63,25

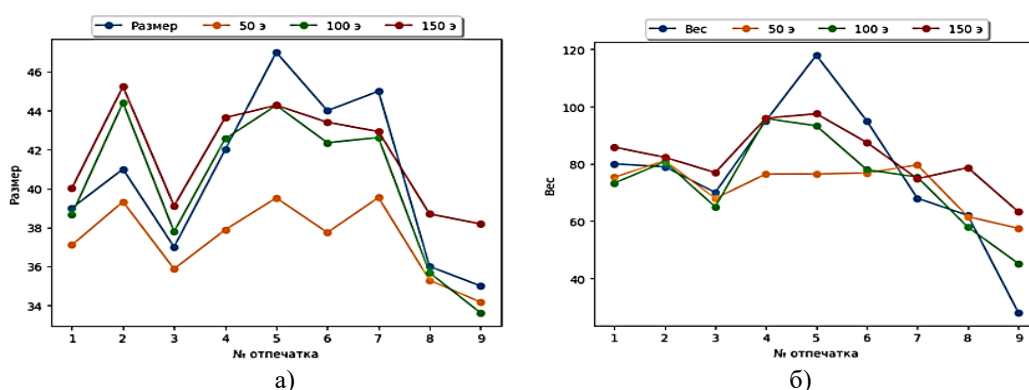


Рис. 10

Таблица 4

/ Эпохи	MSE		MAE	
	Размер	Вес	Размер	Вес
50	16.7465	381.5423	3.283388	14.306202
100	3.389278	148.89667	1.5050193	9.41308
150	6.1762815	237.55756	2.2520332	11.537127

Представленные в табл. 3 и 4 данные показывают, что сеть пришла в оптимальное состояние примерно на 100-й эпохе. Среднее абсолютное отклонение по размеру составило 1.5, что довольно много. Однако необходимо учесть очень ограничен-

ный набор данных, в котором редкие события серьезно искажают статистику. По наиболее распространенным размерам 36, 37 и 39 мы получили отклонения 0.33, 0.78 и 0.33 соответственно, что выглядит уже вполне приемлемо.

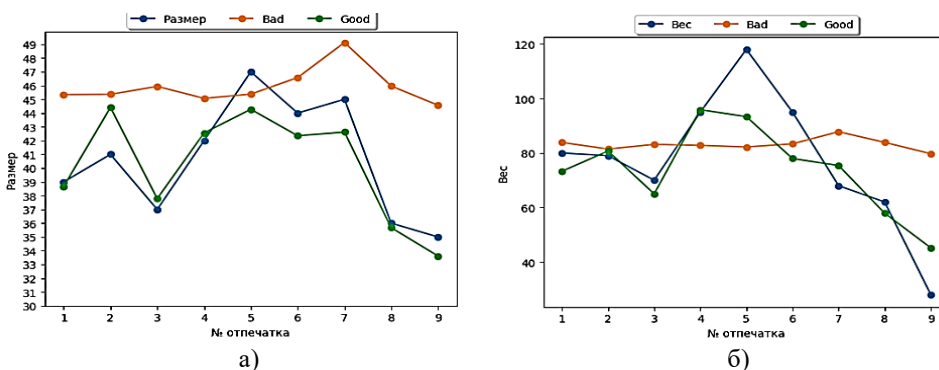


Рис. 11

В заключение приведем сравнение предложенной сети с описанной в [1]. На рис. 11 видно, насколько полученная сеть стала точнее обрабатывать новые данные.

Однако недостаток исходных данных крайне усложняет текущую задачу. Для улучшения результатов можно предложить:

- использовать более сложную функцию потерь, учитывающую, например, больший штраф за предсказание меньшего размера, добавить процентное отношение и т. д.;
- оптимизировать архитектуру сети;
- усложнить генерацию изображений;
- увеличить количество и разнообразие исходных данных.

#### ЛИТЕРАТУРА

1. Смирнов Е.Е., Костылева В.В., Муртазина А.Р. и др. Определение позиции стоп на изображении с применением сверточной нейронной сети // Костумология. 2023. Т 8. №3. – <https://kostumologiya.ru/PDF/29TLKL323.pdf> (дата обращения: 20.12.2023).
2. Смирнов Е.Е., Костылева В.В., Разин И.Б. и др. Распознавание вида одежды и обуви по изображению // Костумология. 2023. Т 8. №1. – <https://kostumologiya.ru/PDF/03TLKL123.pdf> (дата обращения: 20.12.2023).
3. Белгородский В.С., Гусева М.А., Андреева Е.Г., Рогожина Ю.В. Искусственный интеллект в оценке качества готовой швейной продукции // Известия вузов. Технология текстильной промышленности. 2022. № 2 (398). С. 168...177. – DOI 10.47367/0021-3497\_2022\_2\_168
4. Srinnivas S. A Taxonomy of Deep Convolutional Neural Nets for Computer Vision // arXiv, 2016. – <http://arxiv.org/abs/1601.06615> (дата обращения: 20.12.2023).
5. Коул А., Ганджу С., Казам М. Искусственный интеллект и компьютерное зрение. Реальные проекты на Python, Keras и TensorFlow. СПб.: Питер, 2023. 624 с.
6. LeCun Y., Denker J. S., Solla S. etc. Optimal Brain Damage, in Touretzky, David (Eds), Advances in Neural Information Processing Systems 2 (NIPS\*89), Morgan Kaufman, Denver, CO, 1990.
7. LeCun Y., 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. LeCun Y., Bottou L., Orr G., Muller K. Efficient BackProp, in Orr, G. and Muller K. (Eds), Neural Networks: Tricks of the trade, Springer, 1998.
9. Marc'Aurelio R., Poultney C., Chopra S., LeCun Y. 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. Samarasinghe S. Neural Networks for Applied Sciences and Engineering. From Fundamentals to Complex Pattern Recognition // Auerbach Publications, 2007. 570 с.
11. Simon H. Neural Networks: A Comprehensive Foundation Second Edition. М.: Вильямс, 2006. 1104 с.

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

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

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

#### REFERENCES

1. Smirnov E. E., Kostyleva V. V., Murtazina A. R. Determination of stop positions in an image using a vertical neural network // Costumology. 2023. Т 8. No. 3. – <https://kostumologiya.ru/PDF/29TLKL323.pdf> (access date: 12/20/2023).
2. Smirnov E.E., Kostyleva V.V., Razin I.B. Recognition of the type of clothing and shoes from an image // Costumology. 2023. Т 8. No. 1. – <https://kostumologiya.ru/PDF/03TLKL123.pdf> (access date: 12/20/2023).
3. Belgorodsky V.S., Guseva M.A., Andreeva E.G., Rogozhina Yu.V. Artificial intelligence in quality assessment of finished sewing products // Izvestiya Vysshikh Uchebnikh Zavedenii, Seriya Tekhnologiya Tekstil'noi Promyshlennosti. 2022. № 2 (398). P. 168...177. – DOI 10.47367/0021-3497\_2022\_2\_168
4. Srinnivas S. A Taxonomy of Deep Convolutional Neural Nets for Computer Vision // arXiv, 2016. – <http://arxiv.org/abs/1601.06615> (date of access: 20.12.2023)
5. Anirudh K., Siddha G., Meher K. Practical Deep Learning for Cloud, Mobile, and Edge. Real-World AI and Computer-Vision Projects Using Python, Keras, and TensorFlow // SPb.: Piter, 2023. 624 s.
6. LeCun Y., Denker J. S., Solla S. etc. Optimal Brain Damage, in Touretzky, David (Eds), Advances in Neural Information Processing Systems 2 (NIPS\*89), Morgan Kaufman, Denver, CO, 1990.
7. LeCun Y., 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. LeCun Y., Bottou L., Orr G., Muller K. Efficient BackProp, in Orr, G. and Muller K. (Eds), Neural Networks: Tricks of the trade, Springer, 1998.
9. Marc'Aurelio R., Poultney C., Chopra S., LeCun Y. 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. Samarasinghe S. Neural Networks for Applied Sciences and Engineering. From Fundamentals to Complex Pattern Recognition // Auerbach Publications, 2007. 570 с.
11. Simon H. Neural Networks: A Comprehensive Foundation Second Edition // М.: Вильямс, 2006. 1104 с.
12. Yu A., Grauman K. Fine-Grained Visual Comparisons with Local Learning. In CVPR, 2014.
13. Yu A., Grauman K. Semantic Jitter: Dense Supervision for Visual Comparisons via Synthetic Images. In ICCV, 2017.
14. Brink H., Richards J.W., Fetherolf M. Real-World Machine Learning // Shelter Island: Manning, 2017.

Рекомендована кафедрами художественного моделирования, конструирования и технологии изделий из кожи, информационных технологий РГУ имени А.Н. Косыгина. Поступила 08.02.24.