

УДК 004.932.2
DOI: 10.18799/29495407/2024/1/47
Шифр специальности ВАК: 05.02.05

Подбор порога обнаружения дефектов трубы с целью устранения ошибок первого рода и уменьшения ошибок второго рода

Н.К. Половников✉

Московский государственный технический университет имени Н.Э. Баумана, Россия, г. Москва

✉pnk17m408@student.bmstu.ru

Аннотация. В данной статье рассмотрена проблема пропуска дефекта нейронной сетью при дефектоскопии внутренней поверхности трубы, а также проблема «ложного срабатывания», т. е. обнаружения дефекта там, где его нет. В результате исследования выбраны метрики, позволяющие устранить ошибку первого рода и уменьшить количество ошибок второго рода, экспериментально определено пороговое значение, которое устраняет вероятность пропуска дефекта, а также проведен анализ работы системы с подобранным пороговым значением.

Ключевые слова: нейронная сеть, дефектоскопия, ошибки первого и второго рода, пороговое значение, метрика полноты, метрика точности

Для цитирования: Половников Н.К. Подбор порога обнаружения дефектов трубы с целью устранения ошибок первого рода и уменьшения ошибок второго рода // Известия Томского политехнического университета. Промышленная кибернетика. – 2024. – Т. 2. – № 1. – С. 24–29. DOI: 10.18799/29495407/2024/1/47

UDC 004.932.2
DOI: 10.18799/29495407/2024/1/47

Selection of the threshold for detecting pipe defects to eliminate the errors of the first type and reduce the errors of the second type

N.K. Polovnikov✉

Bauman Moscow State Technical University, Moscow, Russian Federation

✉pnk17m408@student.bmstu.ru

Abstract. This article discusses the problem of missing a defect by a neural network during flaw detection of a pipe inner surface, as well as the problem of “false positives”, i. e. detecting a defect where there is none. As a result of the study, metrics were selected to eliminate the errors of the first type and reduce the number of errors of the second type. A threshold value was experimentally determined that eliminates the likelihood of missing a defect. Analysis of the system operation with a selected threshold value was carried out.

Keywords: neural network, defect detection, errors of the first and the second type, threshold value, recall metric, precision metric

For citation: Polovnikov N.K. Selection of the threshold for detecting pipe defects to eliminate the errors of the first type and reduce the errors of the second type. *Bulletin of the Tomsk Polytechnic University. Industrial Cybernetics*, 2024, vol. 2, no. 3, pp. 24–29. DOI: 10.18799/29495407/2024/1/47

Введение

На трубопрокатном производстве одной из важных задач на текущий момент является задача *дефектоскопии* [1]. Она заключается в обнаружении царапин, пробоин, потертостей, трещин и раковин на внутренней поверхности труб, предназначенных для строительства нефтяной или газовой магистрали. Если вовремя обнаружить дефект и определить его тип, можно решить сразу две задачи: не допустить отправки бракованной продукции клиентам и своевременно устранить нарушения и неисправности на производственной линии.

В исследовании [2] была представлена система технического зрения, которая находила дефекты на внутренней поверхности трубы с помощью нейронной сети U-Net. Изображение трубы разворачивалось в прямоугольный вид, после чего U-Net определяла предполагаемые области с дефектами и относила их к определенному классу [3].

Для обучения и тестирования модели U-Net использовался набор данных компании «Северсталь» с соревнования Kaggle [4]. Нейронная сеть определяла дефекты на представленном наборе с точностью 0,81. При этом высокая точность не гарантирует, что система сможет обнаружить все возможные дефекты, а также не исключает возможности «ложного срабатывания».

Примем в качестве *ошибки первого рода* такую ошибку, что нейронная сеть пропустила дефект

там, где он есть. В качестве *ошибки второго рода* – определение дефекта в том месте, где его нет.

Разрабатываемая система предназначена для работы совместно с человеком: при обнаружении дефекта изображение развернутой поверхности трубы отправляется на почту оператору, который осуществляет конечную проверку. Таким образом, ошибка второго рода может быть легко обнаружена, а ошибка первого рода, при которой изображение не отправляется оператору, *должна быть исключена*.

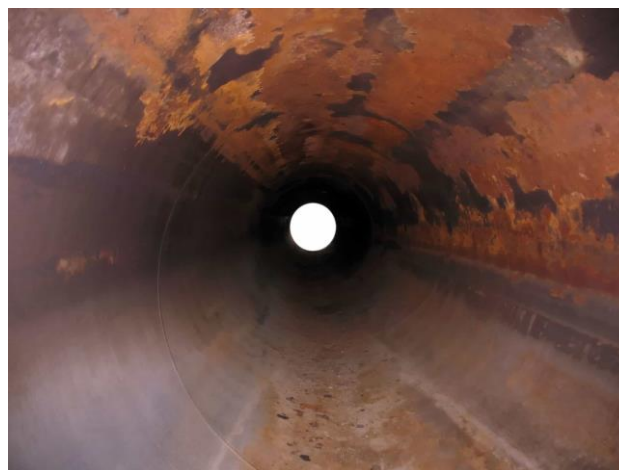


Рис. 1. Исходное изображение трубы
Fig. 1. Original image of the pipe

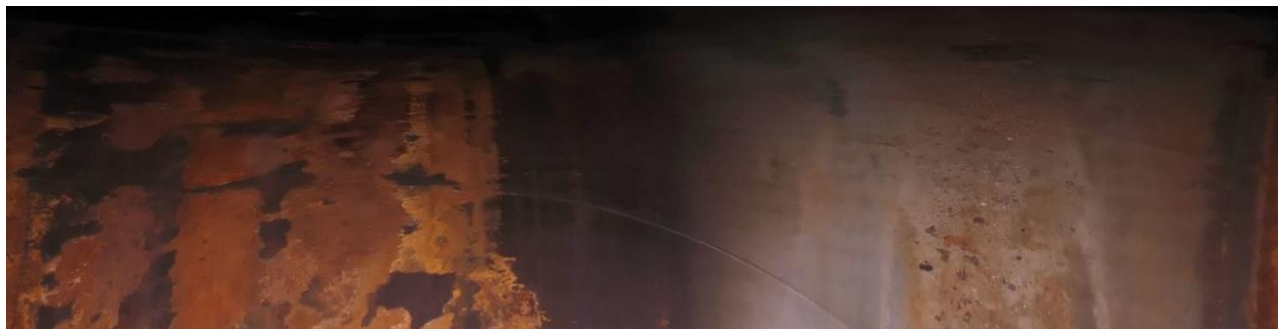


Рис. 2. Развернутое изображение трубы
Fig. 2. Expanded image of the pipe



Рис. 3. Выделенные моделью U-Net области с дефектами
Fig. 3. Areas with defects identified by the U-Net model

Метрики полноты и точности для оценки ошибок

Нейронная сеть U-Net для каждого входного изображения выделяет области предполагаемых дефектов. Для каждой области определяются вероятности нахождения в ней одного из четырех классов дефектов. Введем переменную *prob_defect*, которая показывает *вероятность нахождения хотя бы одного дефекта на всем изображении*: для каждой обнаруженной области определим максимальное значение из всех классов. Если значение *prob_defect* хотя бы одной области на изображении превышает *пороговое значение* (по умолчанию равное 0,5), то на изображении присутствует дефект, иначе – отсутствует. Таким образом, задача сведена к *бинарной классификации*: определению есть дефект на изображении или нет. Примем, что 1 – *наличие дефекта на изображении*, а 0 – *отсутствие дефекта*.

Классификация дефект/не дефект осуществляется с помощью *порогового значения (threshold)*. Данное граничное условие является настраиваемым параметром модели. За счет изменения граничного значения можно влиять на предсказания модели, а также на метрики, характеризующие качество работы модели: чем меньше пороговое значение, тем большее количество изображений модель отнесет к дефектам.

Метрика в задаче машинного обучения – это внешний, объективный критерий качества, зависящий не от параметров модели, а только от ее предсказаний [5]. Например, часто встречаемой метрикой является число верно угаданных моделью меток (*accuracy*). Метрика вычисляется на основе размеченных данных (*target*) и предсказаний модели (*preds*).

Перед переходом к самим метрикам необходимо ввести важную концепцию для описания этих метрик в терминах ошибок классификации – матрицу ошибок (*confusion matrix*) [6]. Матрица ошибок – это таблица с четырьмя различными комбинациями прогнозируемых и фактических значений, с помощью которой можно сопоставить предсказания модели и реальные значения. Прогнозируемые значения описываются как положительные и отрицательные, а фактические – как истинные и ложные (табл. 1).

Здесь $\hat{y}$ – это предсказание алгоритма на данном изображении, а y – истинное значение класса изображения, полученное из разметки.

Набор данных компании «Северсталь» состоит из 801 jpeg изображений с разметкой. Введем массив *target* из 0 (дефекта на изображении нет) и 1 (дефект на изображении есть), соответствующий *размеченным данным*, а также массив *preds*, состоящий из *предсказанных вероятностей prob_defect*

для каждого изображения. Переведем значения *prob_defect* в бинарный вид: если *prob_defect* меньше *порогового значения (threshold)*, то заменяем его на 0 (дефект не обнаружен), иначе – на 1 (дефект присутствует). Всего в каждом массиве будет 801 элемент, равный 0 или 1 ($\hat{y}=\text{preds}=[1, 0, 1, 1, \dots]$; $y=\text{target}=[0, 1, 0, 1, 1, \dots]$). Таким образом, для уже обученной модели U-Net (веса слоев определены и зафиксированы) мы можем получать различные значения массива *preds*, изменяя параметр *порогового значения (threshold)*.

Таблица 1. Оценка истинных и предсказанных значений (матрица ошибок)

Table 1. Estimation of true and predicted values (error matrix)

Реальное значение Real value	Положительное предсказание Positive prediction: $\hat{y}=1$	Отрицательное предсказание Negative prediction: $\hat{y}=0$
Позитив Positive: $y=1$	Верный позитив (TP) True Positive (TP)	Ошибочный негатив (FN) – ошибка первого рода False Negative (FN) – Type I Error
Негатив Negative: $y=0$	Ошибочный позитив (FP) – ошибка второго рода False Positive (FP) – Type II Error	Верный негатив (TN) True Negative (TN)

Сравнивая между собой элементы массивов $\hat{y}=\text{preds}$ и $y=\text{target}$, можно получить матрицу ошибок для каждого параметра *порогового значения (threshold)*.

Для оценки влияния ошибок первого и второго рода воспользуемся метриками *полноты (recall)* и *точности (precision)*.

Метрика полноты (*recall*) – это отношение верных предсказаний дефекта (TP) к общему числу истинных дефектов (как обнаруженных, так и нет) [5]. С помощью данной метрики можно оценить количество ошибок первого рода, которое допускает система:

$$\text{Полнота}(\text{recall}) = \frac{TP}{TP + FP}.$$

Метрика точности (*precision*) – это отношение верных предсказаний дефекта (TP) к общему числу предсказанных дефектов (как верных, так и нет) [5]. С помощью данной метрики можно оценить количество ошибок второго рода, которое допускает система:

$$\text{Точность}(\text{precision}) = \frac{TP}{TP + FN}.$$

Для полного исключения ошибок первого рода необходимо, чтобы метрика полноты (*recall*) была

максимальна, т. е. равна 1,0. Благодаря этому нейронная сеть не будет пропускать дефекты там, где они есть. Однако при увеличении метрики полноты возможен рост ошибок второго рода, которые данная метрика не учитывает. Для борьбы с «ложными срабатываниями» примем, что метрика точности (precision) не должна упасть ниже 0,5.

Подбор порогового значения

Подберем *пороговое значение* для максимизации метрики полноты (recall). Построим графики зависимости метрик полноты и точности от различных пороговых значений:

1. Введем массив *trs*, в котором переберем значения порогового значения *tr* от 0 до 1 с шагом 0,1.
2. Для каждого элемента *tr* из массива *trs* сформируем массив *preds*:
 - Пропустим через обученную модель U-Net тестовый набор из 20 изображений, получив значение *prob_defect* для каждого изображения.
 - Каждый *prob_defect* сравним с текущим значением *tr*. Если *prob_defect* больше *tr*, то текущее изображение содержит дефект (добавляем в массив *preds* 1). Иначе изображение не содержит дефекта (добавляем 0).
3. С помощью библиотеки *scikit-learn* [7] языка Python [8] рассчитаем значение метрики полноты (recall) для каждого значения *tr*: сравним соответствующий массив предсказаний *preds* с размеченными данными из массива *target* (содержит истинные значения). В итоге получим массив со значениями метрики полноты *arr_recall*.
4. По полученным значениям *arr_recall* построим график зависимости метрики полноты (recall) от граничных значений (threshold). На рис. 4 изображен полученный график.

Из графика видно, что при граничном значении, равном 0,3, метрика полноты (recall) начинает убывать, т. е. в предсказаниях модели начинают появляться *ошибки первого рода*. Таким образом, все граничные значения не более 0,3 удовлетворяют поставленной задаче.

Проверим показатели точности (precision) при граничных значениях не более 0,3. Создадим массив со значениями метрики полноты *arr_precision* и построим соответствующий график, представленный на рис. 5.

На основании полученных графиков можно сделать вывод, что при граничном значении, равном 0,3, нейронная сеть *не допускает ошибок первого рода*. При этом значение метрики точности равно 0,6: *ошибка второго рода* будет допущена в 40 % случаев. Это удовлетворяет поставленному требованию, что значение точности не должно быть ниже 0,5.

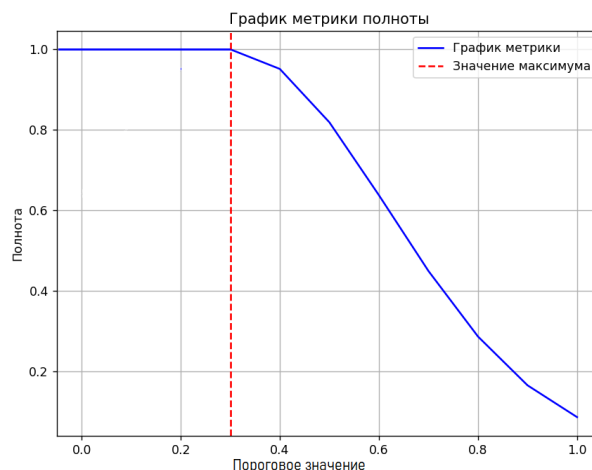


Рис. 4. Зависимость метрики полноты от порогового значения

Fig. 4. Dependence of the completeness metric on the threshold value

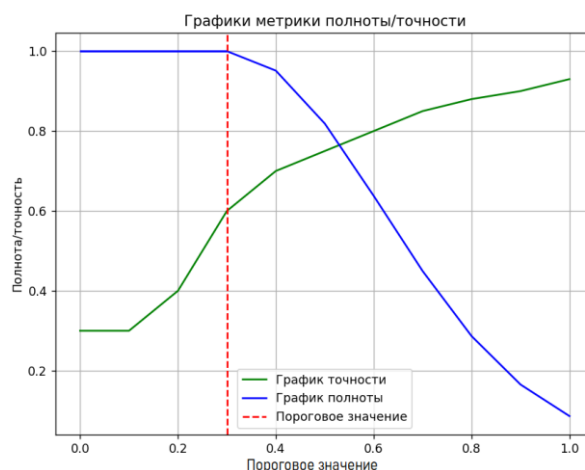


Рис. 5. Зависимость метрик полноты и точности от порогового значения

Fig. 5. Dependence of recall and accuracy metrics on the threshold value

Анализ работы системы

Сравним состояние системы до и после изменения порогового значения *tr*. По умолчанию параметр *tr* был равен 0,5, а после подбора значение *tr* стал равным 0,3.

Создадим сбалансированную выборку из 10 изображений набора данных «Северсталь», пять из которых имеют дефекты, а другие пять – нет. У каждого изображения есть свой порядковый номер за счет чего можно сопоставить предсказание сети и изображение на входе.

После подачи изображений на вход нейронной сети и визуального анализа результата была составлена табл. 2, показывающая работу системы до и после изменения порогового значения.

Таблица 2. Результаты работы нейронной сети на проверочной выборке

Table 2. Results of the neural network operation on the test set

Пороговое значение tr Threshold value tr	Верно классифицированные изображения, шт. Correctly classified images, pcs.	Допущена ошибка, шт. An error was made, pcs.	
		первого рода first type	второго рода second type
0,5	7	2	1
0,3	8	0	2

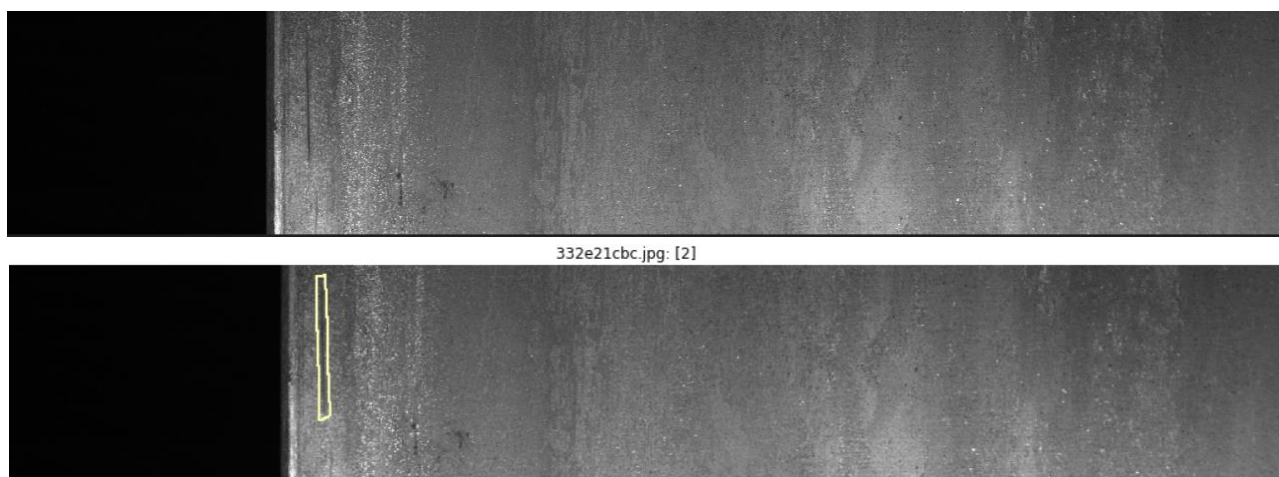


Рис. 6. Сравнение предсказаний нейронной сети при $tr=0,5$ (сверху) и $tr=0,3$ (снизу)

Fig. 6. Comparison of neural network predictions with $tr=0.5$ (top) and $tr=0.3$ (bottom)

На рис. 6 приведены предсказания нейронной сети до и после изменения параметра граничного значения. Видно, что на верхнем рисунке система допускает ошибку первого рода, тогда как при измененном значении tr определен дефект типа 2 (трещина). Классификация типа дефекта в задаче дефектоскопии необходима, так как на основании определенного класса и его выраженности определяются дальнейшие работы по его устранению.

Анализ показывает, что изменение порогового значения привело к устранению ошибки первого рода.

Заключение

В ходе данного исследования было определено пороговое значение, равное 0,3, а также проведен анализ работы системы с подобранным пороговым значением.

За счет изменения порогового значения нейронная сеть U-Net исключает пропуск дефекта на наборе данных компании «Северсталь». Использование технологий технического зрения и машинного обучения позволяет существенно повысить темпы производства на трубопрокатных заводах и автоматизировать контроль выпускаемой продукции.

СПИСОК ЛИТЕРАТУРЫ

1. Детектирование и распознавание дефектов внутренней поверхности металлических труб / А.И. Гайдар, П.Ю. Якимов, А.Е. Викторенков, А.В. Шустанов // Информационные технологии и нанотехнологии (ИТНТ-2020): Сборник трудов по материалам VI Международной конференции и молодежной школы. В 4-х т. Т. 4 / под ред. В.А. Фурсова. – Самара: Самарский национальный исследовательский университет им. академика С.П. Королева, 2020. – С. 741–749.
2. Ronneberger O., Fischer P., Brox T. U-Net: convolutional networks for biomedical image segmentation // Medical Image Computing and Computer-Assisted Intervention – MICCAI. – 2015. – P. 234–241. DOI: https://doi.org/10.1007/978-3-319-24574-4_28
3. Половников Н.К. Программный комплекс системы технического зрения для дефектоскопии внутренней поверхности трубы // Наукосфера. – 2024. – № 2 (2). – С. 210–215.
4. Severstal: Steel Defect Detection. Kaggle Competition. URL: <https://www.kaggle.com/competitions/severstal-steel-defect-detection/data> (дата обращения 21.02.2024).
5. Метрики классификации и регрессии в машинном обучении // Яндекс. URL: <https://education.yandex.ru/handbook/ml/article/metriki-klassifikatsii-i-regressii> (дата обращения 21.02.2024).
6. Оценка моделей ML/DL: матрица ошибок, Accuracy, Precision, Recall // Pythonru. URL: <https://pythonru.com/baza-znaniy/metriki-accuracy-precision-i-recall> (дата обращения 21.02.2024).

7. Библиотека машинного обучения Scikit-Learn // Scikit-learn.ru. URL: https://scikit-learn.ru/category/supervised_learning (дата обращения 21.02.2024).
8. Таршхоева Ж.Т. Язык программирования Python. Библиотеки Python // Молодой ученый. – 2021. – № 5 (347). – С. 20–21.

Информация об авторах

Николай Константинович Половников, магистрант, Московский государственный технический университет имени Н.Э. Баумана, Россия, 105005, г. Москва, 2-я Бауманская ул., 5, стр. 1. pnk17m408@student.bmstu.ru

Поступила: 25.02.2024

Принята: 16.03.2024

Опубликована: 21.03.2024

REFERENCES

1. Gaidar A.I., Yakimov P.U., Viktorenkov A.E., Shustanov A.V. Detection and recognition of defects in the inner surface of metal pipes. *Information technologies and nanotechnologies (ITNT-2020). Collection of works based on materials from the VI International Conference and Youth School*. In 4 vol. Vol. 4. Ed. by V.A. Fursov. Samara, Samara National Research University named after Academician S.P. Korolev, 2020. pp. 741–749. (In Russ.)
2. Ronneberger O., Fischer P., Brox T. U-Net: convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI*. 2015. pp. 234–241. DOI: https://doi.org/10.1007/978-3-319-24574-4_28
3. Polovnikov N.K. Software package for a technical vision system for flaw detection of the inner surface of a pipe. *Naukosfera*, 2024, no. 2 (2), pp. 210–215. (In Russ.)
4. *Severstal: Steel Defect Detection. Kaggle Competition*. Available at: <https://www.kaggle.com/competitions/severstal-steel-defect-detection/data> (accessed 21 February 2024).
5. Metrics of classification and regression in machine learning. *Yandex*. (In Russ.) Available at: <https://education.yandex.ru/handbook/ml/article/metriki-klassifikacii-i-regressii> (accessed 21 February 2024).
6. Evaluation of ML/DL models: error matrix, Accuracy, Precision, Recall. *Pythonru*. (In Russ.) Available at: <https://pythonru.com/baza-znaniy/metriki-accuracy-precision-i-recall> (accessed 21 February 2024).
7. Scikit-Learn machine learning library. Scikit-learn.ru. (In Russ.) Available at: https://scikit-learn.ru/category/supervised_learning (accessed 21 February 2024).
8. Tarshkoeva Zh.T. Python programming language. Python libraries. *Young scientist*, 2021, no. 5 (347), pp. 20–21. (In Russ.)

Information about the authors

Nikolai K. Polovnikov, Master's Student, Bauman Moscow State Technical University, 5, bld. 1, 2nd Bauman-skaya street, Moscow, 105005, Russian Federation. pnk17m408@student.bmstu.ru

Received: 25.02.2024

Revised: 16.03.2024

Accepted: 21.03.2024