

УДК 004.932.72

В. И. Чайко, А. В. Маркидонов

V. I. Chayko, A. V. Markidonov

Чайко Владимир Иванович, студент, КГПИ КемГУ, г. Новокузнецк, Россия.

Маркидонов Артем Владимирович, д. ф.-т. н., КГПИ КемГУ, г. Новокузнецк, Россия.

Chayko Vladimir Ivanovich, Student, Kuzbass Humanitarian Pedagogical Institute of Kemerovo State University, Novokuznetsk, Russia.

Markidonov Artem Vladimirovich, Doctor of Physics and Technology, Kuzbass Humanitarian Pedagogical Institute of Kemerovo State University, Novokuznetsk, Russia.

**ОПТИМИЗАЦИЯ АРХИТЕКТУРЫ СВЕРТОЧНОЙ
НЕЙРОННОЙ СЕТИ ДЛЯ КЛАССИФИКАЦИИ
ДЕФЕКТОВ ПОВЕРХНОСТИ МЕТАЛЛОВ**

**OPTIMIZING THE ARCHITECTURE OF A
CONVOLUTIONAL NEURAL NETWORK FOR
CLASSIFICATION OF METAL SURFACE DEFECTS**

Аннотация. Настоящая работа посвящена разработке и обоснованию архитектуры нейронной сети, специально адаптированной для задачи обнаружения дефектов на поверхности металла. В статье подробно рассматриваются ключевые аспекты проектирования архитектуры, обосновывается выбор конкретных слоёв и параметров, а также анализируются преимущества предложенного решения.

Annotation. This work is devoted to the development and justification of a neural network architecture specifically adapted for the task of detecting defects on the metal surface. The article discusses in detail the key aspects of architecture design, justifies the choice of specific layers and parameters, and analyzes the advantages of the proposed solution.

Ключевые слова: свёрточная нейронная сеть, классификация дефектов, поверхность металла, контроль качества.

Keywords: convolutional neural network, defect classification, metal surface, quality control.

Введение

Внедрение средств промышленной автоматизации представляет собой стратегическое направление развития производственных систем. Особое значение данный процесс приобретает в металлургической отрасли, где эффективность контроля качества продукции непосредственно определяет показатели безопасности и надежности конечных изделий. Визуальный контроль состояния металлических поверхностей остается критически важной частью технологического цикла.

Актуальность создания специализированных алгоритмов на основе нейронных сетей для мониторинга качества металлических поверхностей определяется растущими требованиями отраслевых стандартов к качеству продукции и необходимостью оптимизации производственных процессов.

Научная новизна данного исследования заключается в проектировании специализированной архитектуры сверточной нейронной сети, адаптированной к морфологическим особенностям дефектов металлических поверхностей и обеспечивающей повышенную точность классификации при соблюдении требований к вычислительной эффективности.

Объектом исследования является процесс детектирования дефектов на поверхности металлопроката.

Предметом исследования выступает архитектура сверточной нейронной сети для автоматизированной идентификации дефектов металлических поверхностей.

Цель работы заключается в разработке и верификации архитектуры нейронной сети, обеспечивающей эффективное распознавание дефектов на поверхности металлических изделий.

Для достижения поставленной цели требуется решение следующих задач:

1. Провести сравнительный анализ архитектур нейронных сетей, применяемых в задачах компьютерного зрения.
2. Исследовать методы оптимизации и повышения эффективности нейронных сетей.
3. Обосновать выбор и модифицировать структуру нейронной сети для решения специфической задачи классификации дефектов.
4. Провести экспериментальные исследования для оценки производительности предложенной архитектуры.

Практическая значимость работы определяется потенциалом внедрения разработанной модели в действующие технологические процессы для автоматизации операций контроля качества металлургической продукции.

1. Теоретическая часть

1.1. Актуальность задачи контроля качества металлопроката

Актуальность решения задач контроля качества металлопроката детерминирована комплексом факторов, включающим требования обеспечения эксплуатационной безопасности и надежности конструкций, экономические интересы субъектов производственно-технологической цепочки, а также интенсивное развитие технологических решений в области неразрушающего контроля.

Металлопрокат, выступая базовым конструкционным материалом в различных отраслях промышленности, определяет надежность характеристики и безопасность критически важных объектов: инфраструктурных сооружений (мосты, здания), транспортных средств и других ответственных конструкций.

Анализ применяемых подходов сегодня походов проверки поверхностей металла позволяет выявить ряд системных ограничений:

- высокая зависимость от человеческого фактора;
- сложность выявления микродефектов;
- субъективность оценки некоторых параметров;
- значительные временные затраты на проведение полного контроля.

Внедрение современных методов контроля становится не просто желательным, а необходимым условием успешного функционирования металлургической отрасли в условиях жесткой конкуренции и повышенных требований к качеству продукции [3].

1.2. Проблема выбора архитектуры нейронных сетей для промышленного применения

Критерием оптимальности промышленной архитектуры является ее способность решать поставленную бизнес-задачу в рамках заданных ресурсных ограничений при обеспечении надежности и поддерживаемости в течение всего жизненного цикла системы. Фундаментальным условием для достижения этого является формализация требований и ограничений на начальной стадии проекта.

В контексте современного машинного обучения существует обширный набор архитектурных решений, обладающих различными характеристиками. Выбор конкретной архитектуры детерминируется рядом параметров, включая:

- модальность и структуру входных данных;
- сложность решаемой задачи;
- целевые метрики качества;
- доступные вычислительные ресурсы.

Ключевыми сравнительными параметрами архитектур являются точность, вычислительная сложность и совокупная стоимость владения.

Таким образом, выбор архитектуры для промышленного использования является итеративным процессом, требующим многокритериального анализа. Корректный выбор архитектуры обуславливает операционную эффективность, отказоустойчивость и масштабируемость системы [4].

1.3. Механизмы улучшения работы нейронной сети в данной архитектуре

В процессе обучения нейронных сетей выделяют два фундаментальных феномена предел обучения и переобучение. Для нивелирования указанных проблем применяются следующие методы:

- метод исключения (Dropout);
- субдискретизация (MaxPooling);

- адаптивное планирование скорости обучения (ReduceLROnPlateau);
- пакетная нормализация (Batch Normalization);
- преждевременная остановка (Early Stopping);
- сохранение чекпоинтов модели (Model Checkpointing) [1].

1. Практическая часть

2.1. Выбор типа нейронной сети

В области искусственного интеллекта существует множество архитектур нейронных сетей, специализированных для решения различных классов задач. Для обработки изображений доминирующим подходом являются сверточные нейронные сети (Convolutional Neural Networks, CNN). Принцип действия CNN основан на применении свертки (1) с использованием обучаемых фильтров (ядер), которые последовательно сканируют входное изображение, извлекая иерархические признаки – от простых градиентов яркости и контуров до составных объектов и сложных паттернов.

$$(I \cdot K)_{i,j} = \sum_m \sum_n I_{i+m,j+n} \cdot K_{m,n} \quad (1)$$

Ключевые преимущества сверточных архитектур для анализа изображений включают:

- высокую точность классификации и детекции, даже при наличии шумов и вариаций в данных;
- устойчивость к пространственным трансформациям, таким как смещение, масштабирование и незначительные изменения ракурса объекта [2].

2.2. Выбор структуры нейронной сети

Проектируемая сверточная нейронная сеть, как и большинство таких сетей, состоит из 3-х частей: сверточной, полносвязной и переходной [7].

Определение признаков на изображении лучше всего осуществлять в 3 этапа:

1. Определение простейших признаков – границы объектов, пятна, линий и т.д.
2. На основании простейших признаков, путем группировки, определить более сложные (составные) признаки – углы, простые формы.
3. На основании составных признаков четко определить признаки тех или иных типов повреждений.

Для экономии ОЗУ результат каждого из сверточных слоев сжимается при помощи MaxPooling2D [8].

Для увеличения точности и надежности результат работы 3 слоя целесообразно обработать при помощи еще одного сверточного слоя с целью увеличения точности определенных дефектов.

Для осуществления классификации повреждения на основании признаков полносвязной части достаточно иметь 3 слоя [6]. В качестве функций активаций, во всех слоях, где требуются функции активации, используется функция ReLU (2). Это позволяет избежать «сжатия» выхода в небольшой диапазон, избежать затухания градиента, создает эффект «разреженности», а ее вычисления требует гораздо меньше ресурсов от GPU [9]. Для предотвращения обучения используется Dropout [10].

$$ReLU(x) = \max(0, x) \quad (2)$$

Переходная часть состоит из простой функции «Flatten», которая преобразует двумерный набор данных в одномерный вектор [5].

Всего нейронная сеть имеет 15 слоев. Список всех слоев и их характеристики представлены на таблице 1.

2.3. Выбор датасета

Для обучения и валидации модели был использован датасет «NEU metal surface defects data», доступный на платформе Kaggle. Выбор обусловлен следующими критериями:

1. Репрезентативность данных. Датасет состоит из изображений, полученных в ходе технического контроля на промышленных производствах, что обеспечивает соответствие данных реальным условиям эксплуатации.
2. Аутентичность. Изображения содержат фотографии фактических дефектов поверхности металла, а не синтетические или смоделированные повреждения.
3. Структурированность. Датасет обладает предопределенной раздельной выборкой для обучения, валидации и тестирования, что стандартизирует процесс оценки модели.
4. Сбалансированность классов. Равное количество примеров для каждого типа дефектов минимизирует риск смещения модели (bias) в процессе обучения.
5. Достаточный объем. Объем датасета, составляющий 1800 изображений, является репрезентативным для решения задачи классификации дефектов поверхности металла [11].

Таблица 1. – Список слоев и их характеристики

| № слоя | Характеристики слоев | | | |
|--------|----------------------|-------------------|------------------------------|--|
| | Тип слоя | Функция активации | Количество нейронов/фильтров | |
| 1 | Conv2D | ReLU | 32 фильтра | Поиск простейших признаков |
| 2 | MaxPooling2D | - | 0 | Уменьшает карту признаков |
| 3 | Conv2D | ReLU | 64 фильтра | Комбинирует простейшие признаки |
| 4 | MaxPooling2D | - | 0 | Уменьшает карту признаков |
| 5 | Conv2D | ReLU | 128 фильтров | Ключевой слой для обнаружения конкретных типов дефектов |
| 6 | MaxPooling2D | - | 0 | Уменьшает карту признаков |
| 7 | Conv2D | ReLU | 128 фильтров | Работает с результатом 6 и 7 дефекта |
| 8 | MaxPooling2D | - | 0 | Уменьшает карту признаков |
| 9 | Flatten | - | 18432 элементов в векторе | Преобразует карту признаков в полносвязную часть нейронной сети |
| 10 | Dropout | - | 18432 элементов в векторе | Слой, отключающий нейроны для предотвращения переобучения. |
| 11 | Dense | ReLU | 512 нейронов | Анализирует всю совокупность признаков для классификации |
| 12 | Dropout | - | 512 нейронов | Слой, отключающий нейроны для предотвращения переобучения. |
| 13 | Dense | ReLU | 256 нейронов | Абстрагирование от уникальных признаков для более точной классификации |
| 14 | Dropout | - | 256 нейронов | Слой, отключающий нейроны для предотвращения переобучения. |
| 15 | Dense | SoftMax | 6 нейронов | Определяет вероятности различных классов |

2.4. Результаты обучения

Точность результатов нейронной сети определяется по формуле (3):

$$A = \frac{m}{n} \quad (3)$$

Результаты обучения представлены на рисунках 1 и 2 в виде графиков.

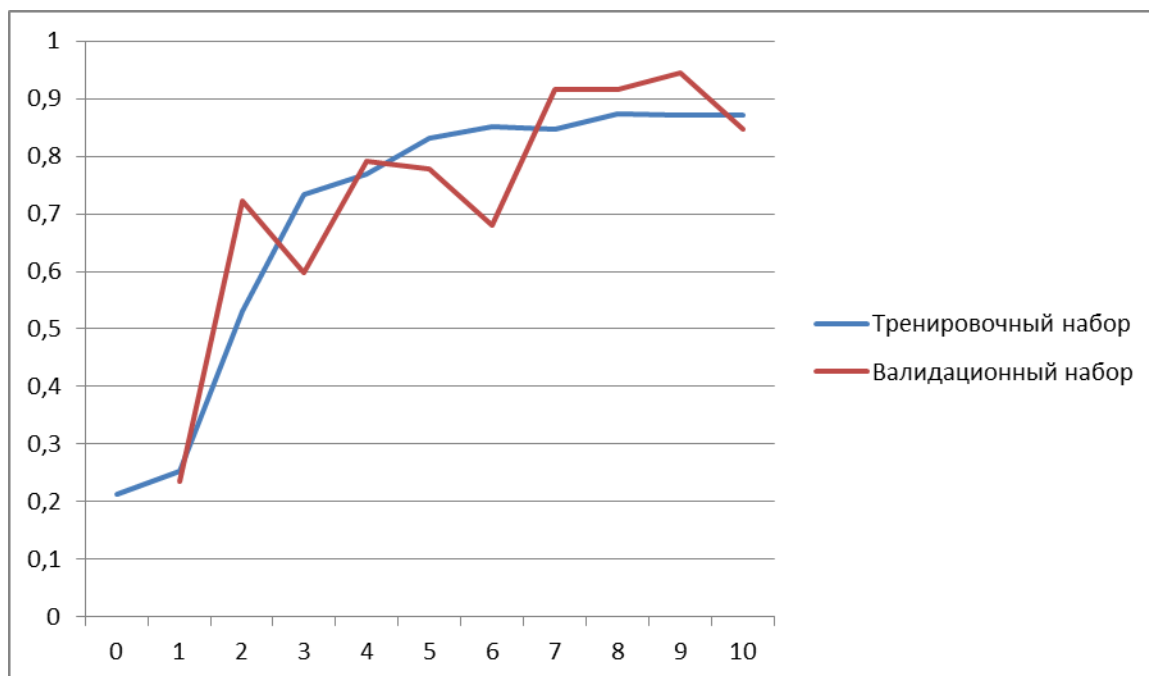


Рисунок 1. Точность спроектированной нейронной сети

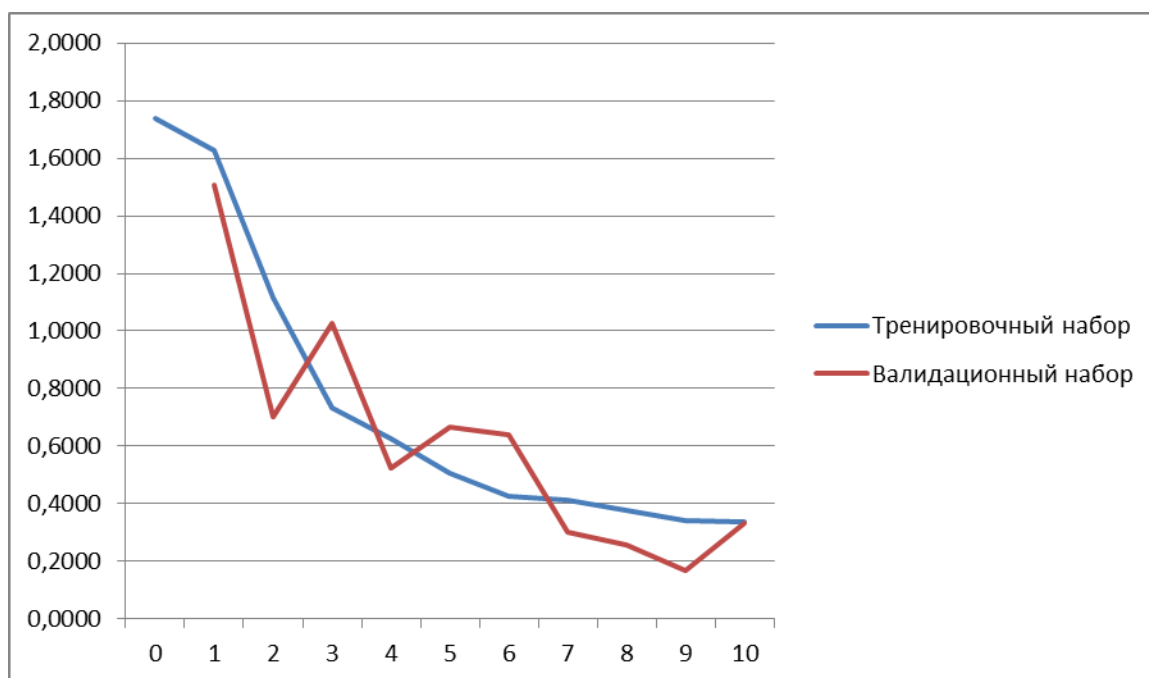


Рисунок 2. График функции потерь

Выводы

В результате проведенного исследования:

1. Разработана и обоснована эффективная архитектура CNN. Предложенная 15-слойная структура, сочетающая последовательность сверточных блоков (Conv2D + MaxPooling2D) и полносвязный классификатор с слоями

регуляризации (Dropout), продемонстрировала свою способность к эффективному определению повреждений поверхности металла.

2. Достигнуты высокие практические показатели. Практический эксперимент на реальном промышленном датасете NEU metal surface defects data показал отличную сходимость модели. Всего за 10 эпох обучения была достигнута точность на валидационной выборке до 97%, а значение функции потерь снижено до 0.13.
3. Использование функции активации ReLU позволило избежать проблемы затухающего градиента и обеспечить высокую скорость обучения.
4. Применение слоев Dropout после каждого полносвязного слоя доказало свою необходимость для борьбы с переобучением, особенно в условиях большого количества параметров на стыке слоев Flatten и Dense.

Проведенное исследование подтвердило высокую эффективность сверточной нейронной сети для решения задачи автоматической классификации дефектов поверхности металла. В ходе работы была успешно достигнута поставленная цель – разработана и всесторонне обоснована структура нейронной сети, оптимально адаптированная под специфику промышленной задачи контроля качества.

Список литературы

1. Гончаренко, А. И. Высокопроизводительные нейронные сети глубокого обучения для устройств с низкими вычислительными ресурсами: дис. канд. техн. наук: 1.2.2. – Новосибирск, 2023. – 98 с. – Текст : непосредственный.
2. Как выбрать архитектуру нейронной сети под вашу задачу // DeepMe : [сайт]. – URL: <https://deepme.ru/art/552> (дата обращения: 29.09.2025). – Текст : электронный.
3. Контроль качества металлопроката на производстве: методы, этапы и современные технологии. Обеспечение безопасности и долговечности конструкций. // ЕК-ТОР :

- [сайт]. – URL: <https://ek-top.ru/articles/metalloprokat/metal-quality-control-importance/> (дата обращения: 24.09.2025). – Текст : электронный.
4. Нейронные сети в промышленном производстве // Upperator : [сайт]. – URL: <https://upperator.ru/industry#rec432138272> (дата обращения: 24.09.2025). – Текст : электронный.
 5. Борокин, А. Погружение в принципы изменения размерности массивов в машинном обучении / Андрей Борокин – Текст : электронный. // zentyx.ru : [сайт]. – URL: <https://zentyx.ru/posts/pogruzhenie-v-printsipyi-izmeneniya-razmernosti-massivov-v-mashinnom-obuchenii/> (дата обращения: 29.09.2025).
 6. Розенблатт, Ф. Принципы нейродинамики: перцептроны и теория механизмов мозга. / Ф. Розенблатт. – М.: Мир, 1965. – 480 с. – Текст : непосредственный.
 7. Сверточные нейронные сети (CNN): что это и как они работают // skypro : [сайт]. – URL: <https://sky.pro/wiki/python/svertochnye-nejronnye-seti-cnn-cto-eto-i-kak-oni-rabotayut/> (дата обращения: 29.09.2025). – Текст : электронный.
 8. Что такое MaxPooling2D. MaxPooling2D: Глубокое Погружение в Мир Сверточных Сетей // graph : [сайт]. – URL: <https://graph.org/CHto-takoe-MaxPooling2D-MaxPooling2D-Glubokoe-Pogruzhenie-v-Mir-Svertochnyh-Setej-01-04> (дата обращения: 29.09.2025). – Текст : электронный.
 9. Choosing the Right Activation Function for Your Neural Network // geeksforgeeks : [сайт]. – URL: <https://www.geeksforgeeks.org/deep-learning/choosing-the-right-activation-function-for-your-neural-network/> (дата обращения: 29.09.2025). – Текст: электронный.
 10. Китов, В. В. Машинное и глубокое обучение / В. В. Китов. – URL : <https://deepmachinelearning.ru/docs/Neural-networks/Regularization/DropOut> (дата обращения: 29.09.2025). – Текст : электронный.
 11. Fantacher, MD. Metal Surface Defects Dataset. NEU surface defect database with six kinds of typical surface defects / MD Fantacher – Текст : электронный. // kaggle: [сайт]. – URL: <https://www.kaggle.com/datasets/fantacher/>

neu-metal-surface-defects-data
30.09.2025).

(дата обращения:

© Чайко В. И., Маркидонов А. В., 2026