

# Критериальный анализ моделей обработки данных радиационного неразрушающего контроля

В.Д. Корчагин, В.С. Кувшинников, Е.Е. Ковшов

**Аннотация**—Проведено исследование нейросетевых моделей для задачи обработки данных радиационного неразрушающего контроля в контексте обнаружения производственных дефектов. Анализ основывается на результатах предшествующего авторского исследования актуальных SOTA-архитектур, применяемых для решения задач классификации изображений и детекции объектов. В ходе исследования рассмотрена эффективность следующих нейросетевых моделей: ResNet, EfficientNet, VGGNet, MobileNet и ViT. Представлено краткое описание структуры и ключевых инноваций рассматриваемых моделей.

Анализ был осуществлен на основе множественных измерений временных характеристик как отдельных экземпляров изображений, так и прохождения полного набора данных (датасета), а также скорости и точности обучения в зависимости от размера обучающей выборки и сложности базовой модели. В процессе обучения применялся метод обучения архитектуры без участия предварительно обученных весов. Из нескольких доступных источников сформирован датасет, включающий как размеченные, так и неразмеченные данные о дефектах в металле различных типов.

Резюмируется о недостаточной эффективности применения изображений в качестве входного тензора для достижения оптимальной точности результатов в рамках поставленной задачи. В связи с этим требуется дальнейшее исследование моделей, способных учитывать дополнительную метаинформацию. Отмечается, что полученные результаты обладают практической значимостью для проектирования архитектуры нейронной сети для решения задачи комплексирования алгоритмов поиска изображений по результатам радиационного контроля в промышленности.

**Ключевые слова**— критериальный анализ, неразрушающий контроль, компьютерное зрение, обучающая выборка, свёрточные нейронные сети.

Статья получена 15 марта 2024.

Корчагин Валерий Дмитриевич, аспирант РХТУ Д. И. Менделеева, (e-mail: valerak249@gmail.com).

Кувшинников Владимир Сергеевич, кандидат технических наук, старший научный сотрудник научно-инженерного центра Цифровых компьютерных систем и автоматизации, АО «НИКИМТ-Атомстрой», (e-mail: KuvshinnikovVS@atomrus.ru).

Ковшов Евгений Евгеньевич, доктор технических наук, профессор, начальник научно-инженерного центра Цифровых компьютерных систем и автоматизации, АО «НИКИМТ-Атомстрой», (e-mail: KovshovEE@atomrus.ru).

## I. ВВЕДЕНИЕ

Цифровизация является основным вектором развития современной экономики. На текущем этапе цифровые трансформации охватывают все сферы индустрии, при этом потребности сокращения временных затрат отражают актуальность и необходимость внедрения инновационных решений в производственные процессы. В данном контексте цифровая трансформация рассматривается как сложный процесс интеграции передовых достижений науки и техники, прежде всего, из области компьютерных технологий и информатизации в различных секторах промышленного производства.

Следует отметить, что данная тенденция тесно связана с глобальным трендом развития «новой экономики». Этот тренд определяется высокой степенью корреляции между повышением эффективности производственных и бизнес-процессов, прибылью предприятий и степенью внедрения компьютерных и телекоммуникационных технологий в отрасли промышленности.

Радиационный неразрушающий контроль (radiation non-destructive testing, RNDT) представляет из себя процесс регистрации и анализа ионизирующего излучения после взаимодействия с контролируемым объектом [1]. Существует три основных метода радиационного контроля: радиографический, радиоскопический и радиометрический. В зависимости от выбранного метода специалисту дефектоскопии необходимо произвести либо визуальную оценку изображения объекта контроля, либо оценку интенсивности излучения с применением специализированных детекторов. Радиационные методы позволяют обнаруживать различные дефекты, такие как нарушения однородности, прожоги, трещины, линии сварочного перехода и другие. Однако несмотря на их многочисленные преимущества, существуют определенные ограничения, требующие особого внимания со стороны специалистов по неразрушающему контролю. Получаемое изображение как результат лучевого воздействия существенно зависит от корректного расположения элементов, направления светового потока и пороговой чувствительности используемого дефектоскопического оборудования, что может приводить к появлению артефактов или к недостаточной выраженности дефектов на изображении. Подобные проблемы могут привести к неверному результату контроля. Таким образом,

этот процесс сопряжен с влиянием человеческого фактора на результаты контроля. Согласно ряду исследований [2], [3], [4] человеческий фактор оказывает существенное влияние на вероятность возникновения производственного дефекта, что может являться критичным показателем при дефектовке сложных конструкций. Применение искусственных нейронных сетей (ИНС) в рамках проведения неразрушающего контроля обеспечит повышение скорости и точности проведения контроля за счет автоматизации анализа входного набора данных для оценки состояния объекта контроля и локализации проблемного участка контролируемого объекта. Кроме того, результаты изучения исследований [5], [6], [7], [8] показывают, что применение программного обеспечения, разработанного с использованием модели ИНС для обнаружения производственных дефектов, будет иметь значительное влияние на область RNDT. Это позволит частично или полностью академическую сторону радиационного контроля путем автоматизации процессов подготовки и моделирования реальных условий с использованием виртуального окружения.

## II. ЦЕЛЬ РАБОТЫ

Цель данной работы заключается в исследовании эффективности применения моделей искусственных нейронных сетей для обнаружения производственных дефектов в металле. При этом необходимо выполнить следующие этапы исследования:

- осуществить отбор моделей ИНС с целью их применения в рамках исследовательской задачи;
- сформировать датасет, включающий в себя изображения различных типов структурных дефектов металла;
- произвести обучение каждой из отобранных моделей на сформированном датасете, варьируя размер тренировочной выборки;
- провести измерение временных характеристик для отдельного экземпляра изображения и полного датасета;
- проанализировать полученные результаты в зависимости от размера обучающей выборки.

Исходя из вышесказанного, можно заключить, что исследование эффективности моделей ИНС в контексте обнаружения производственных дефектов в металле, является одним путей повышения эффективности производственных процессов.

## III. НАБОР ДАННЫХ

Для исследования был использован сформированный датасет, включающий в себя как цветные, так и монохромные изображения, содержащий информацию о местоположении дефекта металла. Датасет для исследования был сформирован из наборов данных, доступных на ведущей онлайн-платформе для проведения соревнований и исследований в области Data Science – Kaggle, а также из открытых репозиториях на портале Github таких, как:

- KolektorSDD;
- Steel surface defects (NEU-CLS);

- Metal surface defects (NEU);
- Severstal Steel Defect Detection;
- BSDData;
- Defect class location;
- DAGM 2007.

На рис. 1-3 приведены примеры изображений, хранящихся в сформированном датасете.

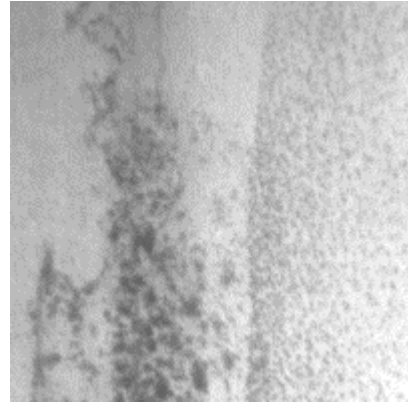


Рис. 1. Случайное изображение из датасета Metal surface defects

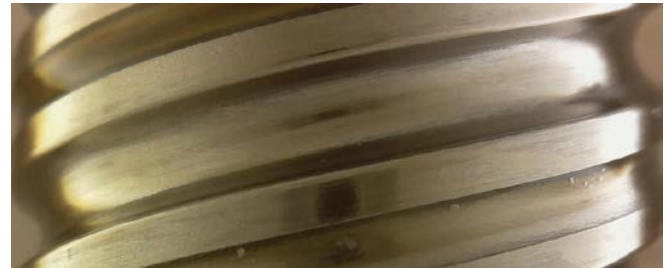


Рис. 2. Случайное изображение из датасета BSDData

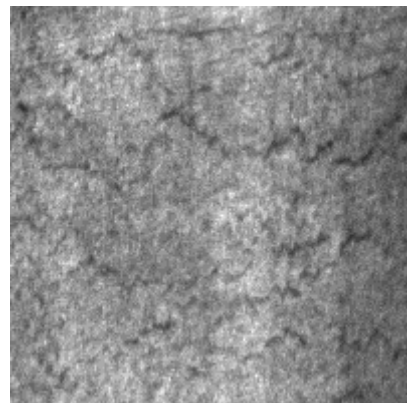


Рис. 3. Случайное изображение из датасета Steel surface defects

Большинство анализируемых датасетов включают предварительно сформированное размещение данных в каталогах, в соответствии с наличием дефекта на изображении. Кроме того, значительная часть изображений сопровождается специальным файлом формата XML, содержащим информацию о координатах обнаруженного дефекта. Пример структуры такого XML-файла приведен на рис. 4.

```

<annotation>
  <folder>3</folder>
  <filename>img_01_424799600_00001.jpg</filename>
  <path>E:\002_Data\suFace detect\2-fuxinzhe\msos1\3\img_01_424799600_00001.jpg</path>
  <source>
    <database>Unknown</database>
  </source>
  <size>
    <width>2048</width>
    <height>1000</height>
    <depth>1</depth>
  </size>
  <segmented>0</segmented>
  <object>
    <name>3_yueyawan</name>
    <pose>Unspecified</pose>
    <truncated>1</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>1688</xmin>
      <ymin>1</ymin>
      <xmax>1904</xmax>
      <ymax>340</ymax>
    </bndbox>
  </object>
</annotation>

```

Рис. 4. Содержание XML-файла из датасета Defect class location

Для обучения моделей ИНС применялись следующие ключевые параметры:

- width/height – размеры изображения;
- xmin/xmax/ymin/ymax – координаты ограничивающей рамки по оси x и y.

По результатам изучения сформированного датасета следует заметить, что значительная часть изображений демонстрирует наличие однородных дефектов, что позволяет провести их классификацию по 10 основным категориям, а именно:

- Crease (складка);
- Crescent gap (серповидный зазор);
- Inclusion (включение);
- Oil spot (масляное пятно);
- Punching hole (дефект штамповки);
- Rolled pit (отпечаток);
- Silk spot (волосовина);
- Waist folding (клиновидный дефект);
- Water spot (водяное пятно);
- Welding line (линия сварного соединения).

При обнаружении учтенного дефекта в отдельном датасете было осуществлено их агрегирование в единый каталог данных, структурированный в соответствии с классификацией хранимых дефектов. В случае отсутствия информации о местоположении дефекта, данные использовались в качестве проверочной выборки изображений после проведения обучения и скоринга модели. Такая организация данных позволяет контролировать процесс обучения и моделировать применение ИНС в реальных промышленных условиях. Общий объем сформированного датасета составляет около 28 тысяч изображений.

#### IV. КОНФИГУРАЦИЯ ОБОРУДОВАНИЯ

При обучении исследуемых ИНС использовались различные средства вычислительной техники, не объединённые в локальную вычислительную сеть. Обучение производилось независимо на каждом вычислительном устройстве в 50 эпох для каждой модели. Это позволяет выявить требования к вычислительным ресурсам и имитировать работу в реальных сценариях. Для проведения исследования использовались следующие конфигурации средств вычислительной техники:

##### 1) Конфигурация №1:

- CPU: Ryzen 7 3700X 8C/16T 4 ГГц;

- GPU: NVIDIA GeForce RTX 3080 10 ГБ;
- RAM: 32 GB DDR4 3800 МГц CL16;
- SSD: Samsung 970 Evo.

##### 2) Конфигурация №2:

- CPU: Intel Core I7-8750H 6C/12T 3.5 ГГц;
- GPU: NVIDIA GeForce GTX 1060 6 ГБ;
- RAM: 16 GB DDR4 2666 МГц CL19;
- SSD: Phison SATA OEM.

## V. ИСПОЛЬЗУЕМЫЕ МОДЕЛИ

### A. EfficientNet

Общее название семейства нейронных сетей, основанных на архитектуре MnasNet [9], представленные в 2019 году с целью достижения высокой эффективности и точности обучения при ограниченности вычислительных ресурсов. Основная идея EfficientNet (сокращенно EffNet) [10] заключается в автоматическом масштабировании собственной архитектуры, определяющей на основании коэффициента масштабирования, регулирующего размер и разрешение ИНС. Данный факт обеспечивает EffNet возможность проявлять высокую эффективность как в узкоспециализированных задачах, так и при работе с большими данными. Согласно исследованию [11], EffNet на сегодняшний день является одной из наиболее успешно применяемых в задаче классификации на 1000 классов ImageNet свёрточных архитектур. На рис.5 представлена схема первой версии архитектуры – EffNetB0.

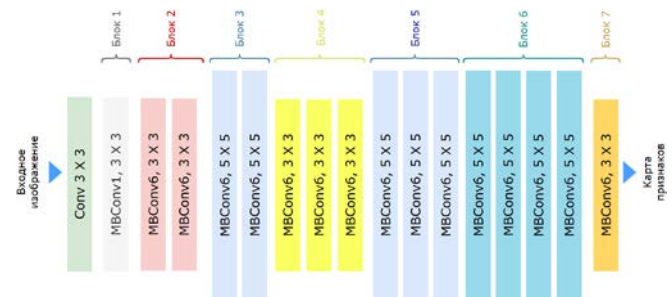


Рис. 5. Архитектура EffNetB0 [12]. Здесь: Conv – сверточная операция; MBConv – инвертированный остаточный блок сжатия и возбуждения.

### B. ResNet

ResNet представляет собой глубокую нейронную сеть, представленную в 2015 исследователями из Microsoft Research [13].

Основной концепцией ResNet является внедрение блоков с остаточным соединением (residual connections), что обеспечивает эффективное обучение глубоких нейронных сетей и снижает проблему затухания градиента. В блоках с остаточным соединением градиент проходит через скрытые слои, после чего складывается с величиной, полученной перед прохождением скрытых слоев. Этот механизм позволяет эффективно сохранять информацию о градиенте даже при значительной глубине сети.

Типичный ResNet блок состоит из двух свёрточных слоев с ядром одинакового размера, разделенных функцией активации, за которыми следует агрегация полу-

ченных величин с исходными результатами и повторно применяется функция активации на выходе. Пример типовой архитектуры ResNet отражён на рис.6.



Рис. 6. Архитектура ResNet50 [14]. Здесь Batch norm – нормализация пакета данных; ReLu – функция активации; Max pool – пулинг по максимальным значениям; ID Block – ResNet блок.

C. MobileNet

MobileNet – архитектура, представленная в 2017 году [15]. Была разработана специалистами отделения Google, специализирующихся на исследованиях искусственного интеллекта (Google Brain) для применения в мобильных и встраиваемых устройствах с ограниченными вычислительными ресурсами.

Основным принципом MobileNet является использование глубоких свёрточных слоев с операцией свертки с ядром размером 3x3, за исключением начальных и конечных слоев. Вместо стандартных свёрточных операций, которые требуют большого количества вычислений, MobileNet использует свертки в глубину (depthwise separable convolutions), которые разделяют операцию свертки на две отдельные части: операцию свертки по каналам (depthwise convolution) и операцию свертки по точкам (pointwise convolution). Это позволяет значительно снизить количество вычислений и параметров сети, что делает ее идеальной для мобильных и встраиваемых устройств.

Еще одним важным элементом архитектуры MobileNet является использование параметра «ширина сети» (width multiplier), который контролирует число каналов в свёрточных слоях. Этот параметр позволяет настраивать архитектуру сети под конкретные требования к скорости работы и точности на различных вычислительных устройствах. Таким образом, MobileNet обеспечивает баланс между эффективностью использования ресурсов и точностью модели, что делает ее популярным выбором для создания различных мобильных приложений.

MobileNetV3, используемая в качестве одного из объектов анализа представляет собой третью итерацию исходной архитектуры, которая была улучшена с использованием адаптивных технологий. Эти изменения позволяют сети автоматически подстраиваться под различные характеристики входных данных в соответствии с доступными вычислительными ресурсами. На рис. 7 представлена главная часть блочной архитектуры,

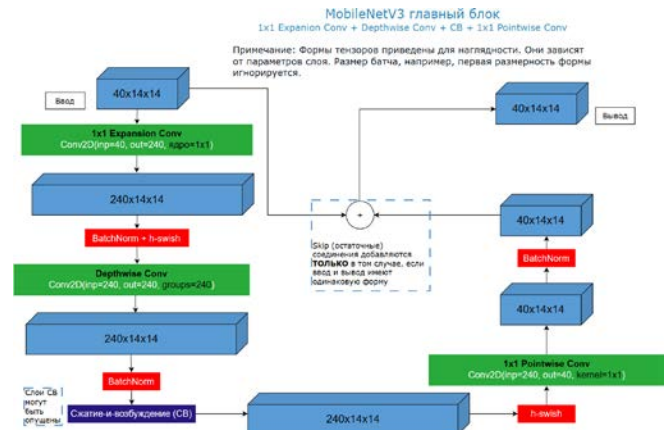


Рис. 7. Главная часть блочной архитектуры MobileNetV3 [16]. Здесь: h-swish – жесткий нелинейный сдвиг; Expansion Conv – блок расширения (выделения) пространства признаков.

D. VGGNet

VGGNet [17] представляет собой глубокую ИНС классической структуры, характеризующуюся значительным количеством свёрточных слоев различной глубины. Это обеспечило возможность достижения значительных результатов в области классификации, сегментации и детекции изображений на момент её представления в 2014 году. На рис. 8 показана архитектурная конфигурация сети, содержащей 16 свёрточных слоев.



Рис. 8. Архитектура VGG16 [18]

E. Other Recommendations

SENet [19] является относительно новой архитектурой глубоких нейронных сетей, разработанной для повышения производительности и эффективности моделей в области компьютерного зрения. Предложенная в 2017 году специалистами из Университета Оксфорд и компании Momenta, основной концепцией SENet является интеграция модуля "сжатия и возбуждения" (squeeze-and-excitation, SE), который улучшает воздействие признаков, проходящих через сеть, учитывая их важность. На этапе сжатия выполняется операция глобального пулинга признаков, значительно снижающая размерность вектора параметров, а затем на этапе возбуждения применяется нелинейная операция для выделения наиболее значимых признаков и усиления их влияния. Иллюстрация структуры SE-блока представлена на рис. 9.

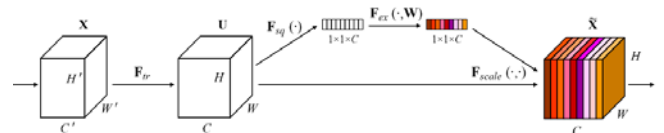


Рис. 9. Структура SE-блока [19]

F. ViT

ViT (Vision Transformer – визуальный трансформер) [20] представляет собой подход, представленный в 2020 году исследователями из компании Google.

Основная идея визуального трансформера заключает-

ся в адаптации архитектуры Transformer [21], успешно примененной в задачах обработки естественного языка (NLP), для работы с изображениями. В отличие от традиционных свёрточных архитектур, которые обрабатывают изображения посредством последовательного смещения фильтра с ядром свертки, ViT рассматривает изображение как последовательность наборов соседних пикселей (патчи). Изображение делится на небольшие патчи одинакового размера, которые затем преобразуются в векторное представление с помощью операции линейного развертывания и соединяются с результатами позиционного кодирования. Полученные данные подаются на вход классической архитектуре энкодера трансформера для дальнейшей обработки.

Одним из ключевых компонентов ViT является механизм внимания (self-attention), который позволяет модели выявлять и учитывать внутренние зависимости в данных, он позволяет модели фокусироваться на наиболее важных патчах изображения в процессе классификации.

Архитектура ViT демонстрирует лидирующие результаты в задачах классификации и сегментации изображений, превосходя ряд существующих моделей при обучении на датасетах размера JFT-300M. Более того, она обеспечивает возможность обучения моделей на изображениях различного разрешения без необходимости изменения архитектуры модели, что делает её универсальным инструментом в области компьютерного зрения. На рис. 10 представлена архитектура ViT/9.

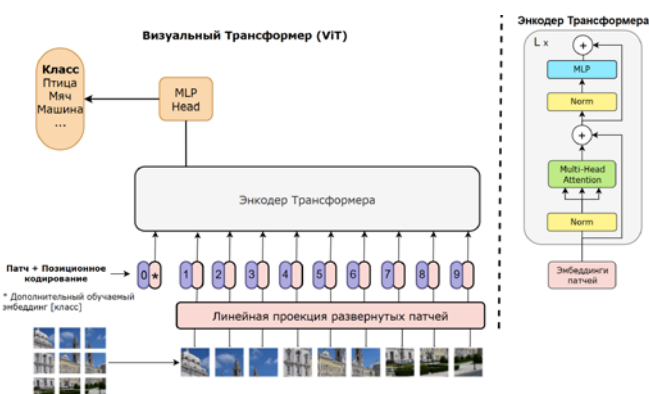


Рис. 10. Архитектура ViT/9 [20]. Здесь: MLP – многослойные перцептрон.

## VI. РЕЗУЛЬТАТЫ

Проведенный анализ моделей обработки данных основывается на следующих критериях:

- точность классификации;
- точность обнаружения;
- количество параметров модели;
- время прохождения одной эпохи;
- время достижения 90% точности;
- стабильность модели при уменьшении размерности выборки;
- требования к вычислительным ресурсам;
- комплексная оценка объединенных нормированных показателей.

В ходе обучения фиксировались следующие метрики:

- Ассигасу (точность);

- Categorical crossentropy (категориальная кросс-энтропия, далее - энтропия);
- OneHotMeanIoU (пересечение через объединение).

Применение указанных метрик позволяет оценить эффективность модели при решении задачи классификации с любым количеством классов, а также её точность при обнаружении дефектов. Применение метрики времени достижения средней точности в 90% обусловлено тем, что данный показатель является средним значением точности по всем рассмотренным моделям.

На рис. 11-12 изображены графики метрик точности классификации (ассигасу) и IoU (пересечение над объединением) для задачи обнаружения дефектов в течении 50 эпох обучения. Обучение проводилось без применения предварительно обученных весов, поскольку в датасетах ImageNet, CIFAR и других отсутствуют изображения дефектов металла. Вместо этого для обучения применялся сформированный набор собственных тренировочных и тестовых данных.

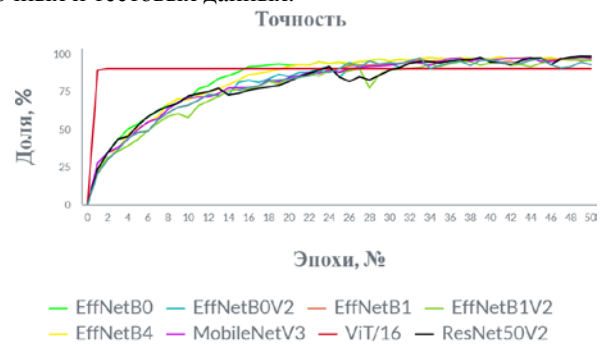


Рис. 11. График метрики Accuracy

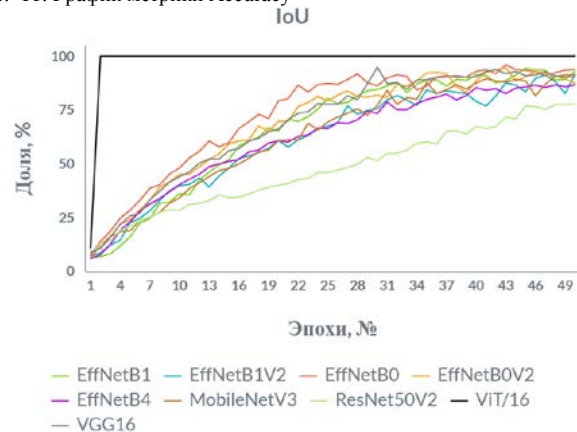


Рис. 12. График метрики IoU

На основании критериального анализа полученных данных следует отметить, что технология визуальных трансформеров доминирует в обнаружении производственных дефектов при неразрушающем контроле, достигая 100% точности локализации дефекта. Тем не менее, наблюдается относительно высокий уровень энтропии при классификации изображений, независимо от выбора базового механизма выделения признаков - свёрточной модели или модели внимания. Это наблюдение подтверждается результатами тестирования обученных моделей на проверочном наборе данных. На рис. 13 представлено несколько примеров полученных результатов.

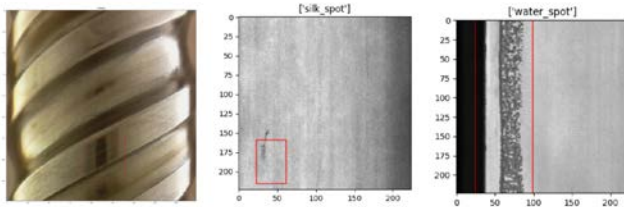


Рис. 13. Результаты тестирования обученных моделей

В ходе последующих вычислительных экспериментов, в которых применялась случайная выборка проверочных изображений, была выявлена общая закономерность, применимая независимо от рассматриваемой модели, и сделан вывод о том, что использование отдельного изображения в качестве основного источника обучения недостаточно для эффективного решения задач классификации и обнаружения дефектов при проведении неразрушающего контроля, поскольку наблюдается корреляция в снижении точности при решении задачи анализа схожих визуальных дефектов, вне зависимости от выбора архитектуры.

Кроме того, был проведен анализ эффективности моделей EfficientNetB7 и SEResNet50V2 в рамках решаемой задачи. Однако, использование EfficientNetB7 оказалось невозможным из-за высоких требований к видеопамяти для локального оборудования, независимо от размера пакета данных (батча). Для обучения данной модели была использована онлайн-платформа Google Colab, однако низкая скорость прохождения эпохи не позволила достичь конечного результата. Промежуточные результаты показали, что EfficientNetB7 неэффективна в рассматриваемой задаче из-за большого количества параметров, что подтверждается средней точностью на уровне 28% за 10 эпох и дальнейшим отсутствием положительной динамики. Для сравнения, любая из представленных на Рис. 11 моделей к этому моменту достигает точности не ниже 58% на том же наборе обучающих и тестовых данных и продолжает показывать положительную динамику.

Показатели SEResNet50V2 в рассматриваемой задаче демонстрируют отличия на уровне погрешности, в сравнении с базовой моделью ResNet50V2, в связи с чем данная модель в итоговой выборке не учитывается. Оценка погрешности осуществлялась путем визуального анализа графического представления результатов обучения. Проведенный анализ показал полное совпадение метрики Ассигасу и ухудшение метрики IoU на 1,2%.

Исследование устойчивости к переобучению моделей было достигнуто путем урезания обучающей выборки на 50% и 75%, соответственно, с последующим повторным обучением отобранных моделей с обнулением итоговых весов. Полученные результаты позволяют сделать вывод о повышении показателя случайного выброса энтропии в зависимости от увеличения глубины модели и размера обучающей выборки, что иллюстрируется на рис. 14-15. Наибольшую устойчивость к случайным выбросам демонстрирует EffNetB0 (рис. 16).

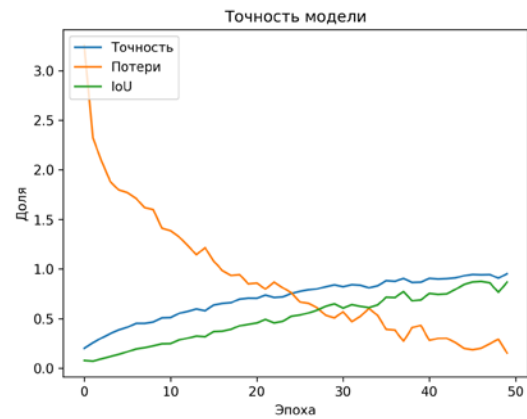


Рис. 14. Обучение с 50% датасета EffNetB1V2

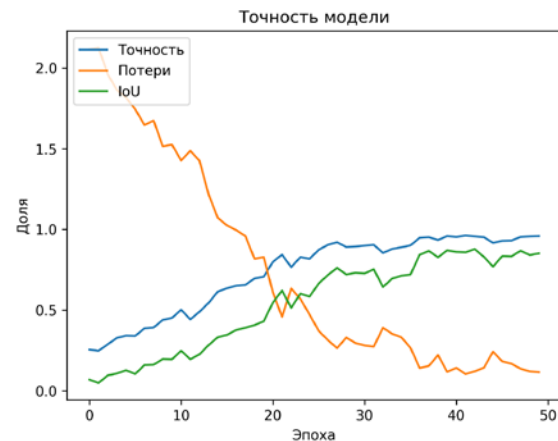


Рис. 15. Обучение с 25% датасета MobileNetV3

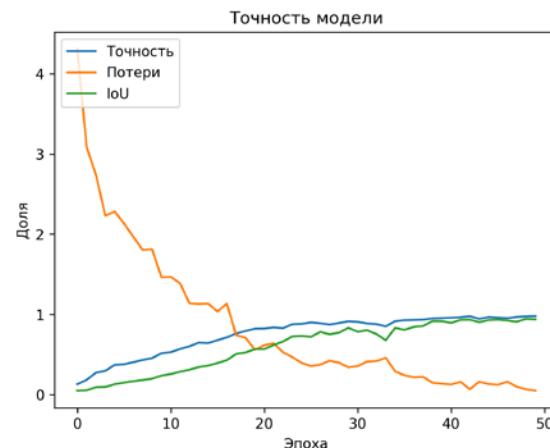


Рис. 16. Обучение с 25% датасета EffNetB0

Для определения наиболее эффективной из отобранных моделей была проведена комплексная оценка нормированных показателей с использованием сводного графика, представленного на Рис. 17.

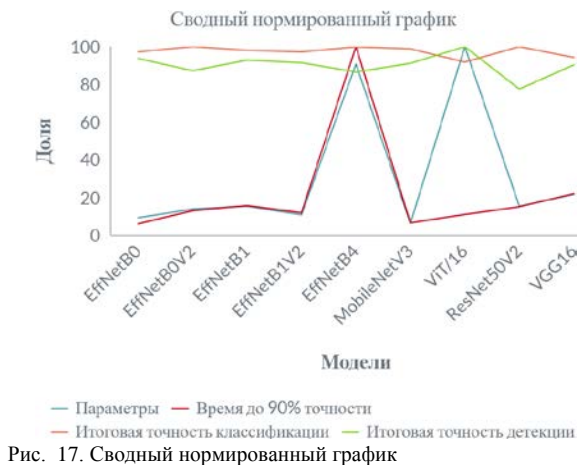


Рис. 17. Сводный нормированный график

На основе представленных данных можно сделать вывод о превосходстве адаптивных моделей EffNetB0 и MobileNetV3 с точки зрения общей эффективности в задаче анализа данных неразрушающего контроля. Принимая во внимание, что достижение оптимального значения метрики IoU не признается существенным в рамках более важной задачи правильной классификации дефектов, данный показатель оказывает минимальное воздействие на конечную оценку эффективности модели.

Для оценки вычислительной нагрузки был проведен анализ производительности рассматриваемых моделей как на графическом процессоре, так и на центральном процессоре каждой из экспериментальных конфигураций с использованием полного датасета. Поскольку вычислительные ресурсы не оказывают влияния на конечную точность модели, а являются временным аспектом обучения, оценка их использования на уровне в 50% и 25% от полного размера датасета не требуется.

Оценка вычислительных требований осуществлялась на основе измерений среднего времени прохождения одной эпохи при использовании полного размера тренировочного датасета. Результаты измерений в логарифмической форме при использовании конфигураций №1 и №2 представлены на рис. 18-19, соответственно.

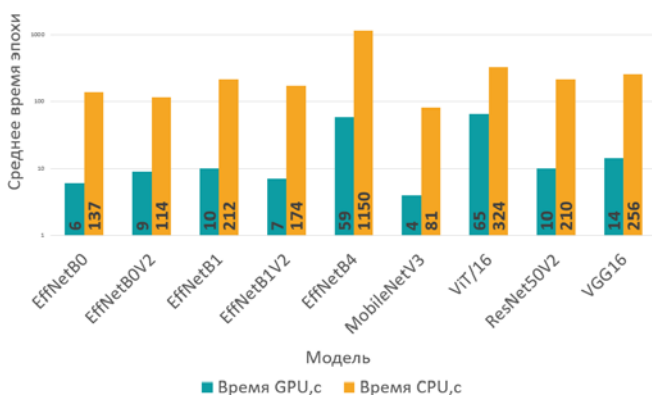


Рис. 18. Среднее время прохождения одной эпохи на конфигурации №1

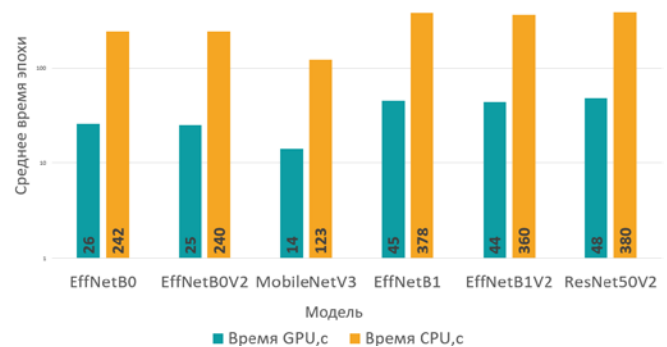


Рис. 19. Среднее время прохождения одной эпохи на конфигурации №2

Меньшее число рассмотренных моделей для конфигурации №2 обусловлено невозможностью осуществления процесса обучения даже при уменьшении размера батча до одного экземпляра из-за ограниченных ресурсов видеопамяти, необходимых для запуска отсутствующих моделей.

На основании представленной информации можно сделать вывод о более низких требованиях к вычислительным ресурсам при использовании модели MobileNetV3 в контексте классификации и обнаружения дефектов, что обеспечивает ее эффективное применение даже без дополнительного графического ускорителя.

## VII. ЗАКЛЮЧЕНИЕ

В ходе исследования были реализованы и обучены следующие модели ИНС: EffNetB0(V2), EffNetB1(V2), EffNetB4, EffNetB7, (SE)ResNet50V2, VGG16, MobileNetV3, ViT/16 при использовании собственного сформированного датасета. Программная реализация выполнена на языке Python3 в среде разработки Jupyter Notebook с использованием открытой библиотеки машинного обучения TensorFlow. Полученные результаты свидетельствуют о недостаточной эффективности рассмотренных моделей для решения поставленной задачи. Дальнейшие исследования целесообразно сфокусировать на изучении комбинированных моделей, объединяющих механизмы свертки и внимания, а также мультимодальных моделей для агрегирования текстовой метаинформации в качестве входного тензора. Примерами таких моделей могут быть сети CoAtNet [22] и CoCa [23].

Для последующих исследований рекомендуется использование стандартизированного набора данных в формате DICONDE с высоким контрастом в области дефектов и наличием метаинформации. Для обучения следует подготовить выборку, содержащую не менее 7 тысяч уникальных изображений.

В процессе исследования наибольшую репрезентативность и наличие размеченных данных продемонстрировали следующие наборы данных: Steel surface defects (NEU-CLS), Metal surface defects (NEU), Defect class location, Severstal Steel Defect Detection. Целесообразно использовать информацию из указанных наборов данных для расширения обучающей выборки путем проведения аугментации.

Количество цветных каналов не оказывает влияния

на окончательную точность модели в задаче обнаружения производственных дефектов. С целью сокращения количества подаваемых параметров на вход ИНС, рекомендуется использовать монохромный режим изображения на входном слое.

Для проведения обучения, в зависимости от доступных вычислительных ресурсов, рекомендуется использовать модель EffNetV0 или её модификации – для обучения с использованием графического ускорителя с объемом видеопамати не менее 10 ГБ. В случае использования центрального процессора (ЦП) для обучения рекомендуется применение архитектуры MobileNetV3 или её модификации, а также ЦП с 16 логическими ядрами и тактовой частотой не менее 4.2 ГГц при наличии в системе оперативной памяти, объёмом не менее 16 ГБ.

#### БИБЛИОГРАФИЯ

- [1] ГОСТ Р 55776-2013. Контроль неразрушающий радиационный. Термины и определения. М., 2019. 12 с.
- [2] Вишнеvский Д. А., Сахаров Б. А. Анализ влияния «человеческого фактора» на надёжность металлургического оборудования. // Сборник научных трудов ГОУ ВПО ЛНР «ДонГТУ». 2018. № 12(55).
- [3] Голдобин С. М. Влияние человеческого фактора на появление дефектов продукции // Методы менеджмента качества. 2017. № 8. С. 54 – 57.:
- [4] Вишнеvский Д. А. Влияние человеческого фактора на надёжность металлургического и машиностроительного оборудования. Сборник научных трудов ГОУ ВПО ЛНР «ДонГТУ». 2019. № 15(58).
- [5] Ковшов Е. Е., Кувшинников В. С., Казаков Д. Ф. Применение виртуальной реальности при разработке симулятора радиографии для обучения неразрушающему контролю // Контроль. Диагностика. 2021. Т. 24, № 7. С. 34 – 40. DOI: 10.14489/td.2021.07.pp.034-040.
- [6] Ковшов Е. Е., Кувшинников В. С., Казаков Д. Ф. Формирование рентгеновского изображения объекта неразрушающего контроля в среде виртуальной реальности // Контроль. Диагностика. 2021. Т. 24, № 8. С. 14 – 22. DOI: 10.14489/td.2021.08.pp.014-022.
- [7] Ковшов Е. Е., Кувшинников В. С. Симуляция физических свойств материала объекта контроля в VR-тренажере промышленной радиографии // Контроль. Диагностика. 2023. Т. 26, № 2. С. 4 – 12. DOI: 10.14489/td.2023.02.pp.004-012.
- [8] Ковшов Е. Е., Кувшинников В. С., Казаков Д. Ф. Применение моделей цифровых двойников при формировании радиографического изображения в среде виртуальной реальности // Контроль. Диагностика. 2023. Т. 26, № 9. С. 4 – 15. DOI: 10.14489/td.2023.09.pp.004-015
- [9] Tan M. Mnasnet: Platform-aware neural architecture search for mobile // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019. P. 2820 – 2828.
- [10] Tan M., Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks // International conference on machine learning. 2019. P. 6105-6114.
- [11] Корчагин, В. Д. Анализ современных SOTA-архитектур искусственных нейронных сетей для решения задач классификации изображений и детекции объектов // Программные системы и вычислительные методы. 2023. № 4. С. 73 – 87. DOI: 10.7256/2454-0714.2023.4.69306.
- [12] Ahmed T., Sabab N. H. N. Classification and understanding of cloud structures via satellite images with EfficientUNet // SN Computer Science. 2022. Т. 3, №. 1. P. 99.
- [13] He K. Deep residual learning for image recognition // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. P. 770 – 778.
- [14] Molčan S. Classification of Red Blood Cells Using Time-Distributed Convolutional Neural Networks from Simulated Videos // Applied Sciences. 2023. Т. 13, №. 13. P. 7967.
- [15] Howard A. et al. Searching for mobilenetv3 // Proceedings of the IEEE/CVF international conference on computer vision. 2019. P. 1314 – 1324.
- [16] A visual deep-dive into the building blocks of MobileNetV3. URL: <https://francescopochetti.com/a-visual-deep-dive-into-the-building-blocks-of-mobilenetv3/>
- [17] VGG16 From Scratch | Computer Vision With Keras. URL: <https://pysource.com/2022/10/04/vgg16-from-scratch-computer-vision-with-keras-p-7/>
- [18] Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition // Proceedings of the 3rd International Conference on Learning Representations. 2014.
- [19] Hu J., Shen L., Sun G. Squeeze-and-excitation networks // Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. P. 7132 – 7141.
- [20] Dosovitskiy A. An image is worth 16x16 words: Transformers for image recognition at scale // International Conference on Learning Representations. 2021.
- [21] Vaswani A. Attention is all you need // Advances in neural information processing systems. 2017. Т. 30.
- [22] Dai Z. Coatnet: Marrying convolution and attention for all data sizes // Advances in neural information processing systems. 2021. Т. 34. P. 3965 – 3977.
- [23] Yu J., Wang Z., et al.: Contrastive captioners are image-text foundation models // a rXiv preprint arXiv:2205.01917, 2022.



# Criteria analysis of radiation nondestructive testing data processing models

V.D. Korchagin, V.S. Kuvshinnikov, E.E. Kovshov

**Abstract** – A research of neural network models for the task of radiation nondestructive testing data processing in the context of production defect detection is done. The analysis is based on the results of the author's previous research of actual SOTA-architectures used for image classification and object detection tasks. The study considers the performance of the following neural network models: ResNet, EfficientNet, VGG-Net, MobileNet and ViT.

The analysis was based on multiple measurements of the time characteristics of both individual image instances and passing the full dataset, as well as the speed and accuracy of training depending on the size of the training sample and the complexity of the base model. The training process utilized a learning architecture method without the participation of pre-trained weights. A dataset including both labeled and unlabeled data on defects in metal of various types was compiled from several public sources.

Results are summarized that the use of images as an input tensor is not effective enough to achieve optimal accuracy of the results for the task at hand. In this regard, further investigation of models capable of taking into account additional meta-information is required. The obtained results are of practical importance for designing the neural network architecture for solving the problem of completing the algorithms for image retrieval based on the results of radiation testing in industry.

**Keywords** – criteria analysis, nondestructive testing, computer vision, training sample, convolutional neural networks.

## References

- [1] Non-destructive radiation testing. Terms and definitions. Ru Standard No. GOST R 55776-2013. Moscow. 2019 [in Rus]
- [2] D.A. Vishnevsky. "Analysis of the influence of the human factor on the reliability of metallurgical equipment," *Sbornik nauchnykh trudov GOU VPO LNR «DonGTU»*, no. 12(55), 2018. [in Russian language]
- [3] S.M. Goldobin, "Influence of the human factor on the appearance of product defects," *Medogy menegjmenta kachestva*, no. 8, pp.54 – 57, 2017. [in Rus]
- [4] D.A. Vishnevsky, "Influence of human factor on reliability of metallurgical and machine-building equipment," *Sbornik nauchnykh trudov GOU VPO LNR «DonGTU»*, no. 15(58), 2019. [in Rus]
- [5] E.E. Kovshov, V.S. Kuvshinnikov, D.F. Kazakov, "Virtual reality usage in the radiography simulator development for non-destructive testing personnel training," *Kontrol'. Diagnostika*, vol. 24, no. 7, pp. 34–40, 2021. [in Rus]. DOI: 10.14489/td.2021.07.pp.034-04.
- [6] E.E. Kovshov, V.S. Kuvshinnikov, D.F. Kazakov, "Radiographic image of a non-destructive testing object generation in a virtual reality environment," *Kontrol'. Diagnostika*, vol. 24, no. 8, pp. 14–22, 2021 [in Rus]. DOI 10.14489/td.2021.08.pp.014-022.
- [7] E.E. Kovshov, V.S. Kuvshinnikov, "Testing object's material physical properties simulation in the industrial radiography VR environment," *Kontrol'. Diagnostika*, vol. 26, no. 2, pp. 4–12, 2023 [in Rus]. DOI: 10.14489/td.2023.02.pp.004-012.
- [8] E.E. Kovshov, V.S. Kuvshinnikov, D.F. Kazakov, "The use of digital twins models while a radiographic image formation in a virtual reality environment," *Kontrol'. Diagnostika*, vol. 26, no. 9, pp. 4–15, 2023. [in Rus]. DOI: 10.14489/td.2023.09.pp.004-015.
- [9] M. Tan et al., "Mnasnet: Platform-aware neural architecture search for mobile," In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019. pp. 2820 – 2828.
- [10] M. Tan, Q. Le, "Efficient net: Rethinking model scaling for convolutional neural networks," *International conference on machine learning*, 2019, pp. 6105–6114.
- [11] V.D. Korchagin, "Analysis of modern SOTA-architectures of artificial neural networks for solving problems of image classification and object detection," *Software systems and computational methods*, no. 4, pp. 73–87, 2023. [in Rus]. DOI: 10.7256/2454-0714.2023.4.69306.
- [12] T. Ahmed, N.H.N. Sabab, Classification and understanding of cloud structures via satellite images with EfficientUNet.SN, *Computer Science*, vol. 3, no. 1, pp. 99, 2022
- [13] K. He et al., Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770 – 778.
- [14] S. Molčan et al., "Classification of Red Blood Cells Using Time-Distributed Convolutional Neural Networks from Simulated Videos," *Applied Sciences*, vol. 13, no. 13, pp. 7967, 2023.
- [15] A. Howard et al., "Searching for mobilenetv3," In *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1314 – 1324.
- [16] A visual deep-dive into the building blocks of MobileNetV3. URL: <https://francescopochetti.com/a-visual-deep-dive-into-the-building-blocks-of-mobilenetv3/>
- [17] VGG16 From Scratch | Computer Vision With Keras. URL: <https://pysource.com/2022/10/04/vgg16-from-scratch-computer-vision-with-keras-p-7/>
- [18] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," In *Proceedings of the 3rd International Conference on Learning Representations*, 2014
- [19] J. Hu, L. Shen, G. Sun, "Squeeze-and-excitation networks," In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.
- [20] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," *International Conference on Learning Representations*, 2021
- [21] A. Vaswani et al., "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [22] Z. Dai et al., "Coatnet: Marrying convolution and attention for all data sizes," *Advances in Neural Information Processing Systems*, vol. 34, pp. 3965–3977, 2021
- [23] J. Yu et al. Coca: Contrastive captioners are image-text foundation models. arXiv preprint arXiv:2205.01917.

**Valeriy Korchagin**, Postgraduate student at Dmitry Mendeleev University of Chemical Technology of Russia (<https://www.muctr.ru/>), e-mail: valerak249@gmail.com, ORCID: <https://orcid.org/0009-0003-1773-0085>

**Vladimir Kuvshinnikov**, Can. of Sci. (Engineering), Senior Researcher of the Scientific and Engineering Center of Digital Computer Systems and Automation, JSC "NIKIMT-Atomstroy" (<https://ase-ec.ru/>), e-mail: KuvshinnikovVS@atomrus.ru, ORCID: <https://orcid.org/0000-0002-2565-2510>

**Evgeny Kovshov**, Dr. of Sci. (Engineering), Professor, Head of the Scientific and Engineering Center of Digital Computer Systems and Automation, JSC "NIKIMT-Atomstroy" (<https://ase-ec.ru/>), e-mail: KovshovEE@atomrus.ru, Researcher ID: B-1077-2016, ORCID: <https://orcid.org/0000-0003-4758-1730>