**Министерство науки и высшего образования Российской Федерации** ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

**НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО**

**ITMO University**

**ОТЧЁТ ПО УЧЕБНОЙ, ОЗНАКОМИТЕЛЬНОЙ ПРАКТИКЕ**

**Обучающийся / Student** Никандров Сергей Андреевич

**Факультет/институт/кластер/ Faculty/Institute/Cluster** факультет систем управления и робототехники

**Группа/Group** R34352

**Направление подготовки/ Subject area** 15.03.06 Мехатроника и робототехника **Образовательная программа / Educational program** Робототехника и искусственный интеллект 2021

**Язык реализации ОП / Language of the educational program** Русский

**Квалификация/ Degree level** Бакалавр

**Тема ВКР/ Thesis topic** Исследование методов и алгоритмов искусственного интеллекта для обнаружения и классификации поверхностных дефектов листового металлопроката

**Руководитель ВКР/ Thesis supervisor** Евстафьев Олег Александрович, кандидат технических наук, Университет ИТМО, факультет систем управления и робототехники, преподаватель (квалификационная категория "преподаватель")

**ВВЕДЕНИЕ**

В условиях современной экономики, где конкуренция на рынке металлопродукции достигла высокого уровня, обеспечение высокого

качества листового металлопроката является критически важным фактором для успешного функционирования предприятий. Поверхностные дефекты, такие как царапины, трещины, вмятины, загрязнения и ржавчина, не только снижают потребительскую ценность продукции, но и могут приводить к

значительным экономическим потерям, связанным с браком, затратами на исправление дефектов и потерей репутации. Традиционные методы контроля качества, включающие визуальный контроль, ультразвуковой,

рентгеновский, магнитопорошковый и вихретоковый методы, зачастую оказываются трудоемкими, подверженными влиянию человеческого фактора, и не всегда обеспечивают высокую скорость и точность обнаружения

дефектов, особенно при больших объемах производства.

В связи с этим, разработка автоматизированных систем обнаружения и классификации поверхностных дефектов металлопроката приобретает особую актуальность. Использование методов и алгоритмов искусственного интеллекта, в частности, глубокого обучения, открывает новые возможности для повышения эффективности контроля качества, обеспечивая высокую точность, скорость обработки данных и минимизацию влияния

человеческого фактора. Применение этих технологий позволяет не только выявлять дефекты, но и классифицировать их, что предоставляет ценную информацию для анализа причин возникновения дефектов и оптимизации производственного процесса.

Целью данной работы является исследование методов и алгоритмов обработки цифровых изображений, способных обеспечить высокую

достоверность контроля поверхностных дефектов листового металлопроката в разных условиях среды производства.

Для достижения поставленной цели необходимо решить следующие задачи:

* Аналитический обзор алгоритмов обработки изображения: Исследование существующих алгоритмов и методов обработки изображений для выявления поверхностных дефектов на металлических листах.
* Подготовка набора данных: подготовить набор данных включающий разнообразные цифровые изображения листового холоднокатаного металлопроката с различными типами поверхностных дефектов.
* Аналитический обзор алгоритмов машинного обучения: Исследование различных методов и алгоритмов машинного обучения для обработки изображений с целью выявления поверхностных дефектов.
* Выбор алгоритмов машинного обучения: рассмотреть классические алгоритмы машинного обучения и алгоритмы, основанные на искусственной нейронной сети.
* Программная реализация: Разработать программное обеспечения на языке программирования Python, для реализации и сравнения различных алгоритмов обработки изображений, включая методы морфологической обработки, алгоритмы машинного обучения и нейронные сети.
* Анализ результатов: произвести сравнительный анализ результатов выбранных алгоритмов.
* Апробация и выход.

**1 ТЕОРЕТИЧЕСКИЙ ОБЗОР**

* 1. **Актуальность**

Применение систем компьютерного зрения (Computer Vision, CV) для автоматизации контроля качества в производстве стремительно развивается. Алгоритмы CV позволяют обнаруживать дефекты продукции с высокой точностью и скоростью, значительно повышая эффективность производственных процессов.

На сегодняшний день можно с уверенностью сказать, что автоматизированная визуальная дефектоскопия, основанная на компьютерном зрении, способна в значительной степени сократить непосредственное участие работников в процессе проверки качества на всех видах производственных линий, отведя человеку роль руководителя процесса. Современные системы контроля качества способны провести подсчет объектов, снять измерения, проверить цвет, комплектность, наличие маркировок и штрихкодов, выявить дефекты и, при необходимости, сопоставить изделие с эталоном.

* 1. **Обзор существующих подходов**

Качество листового металлопроката является критически важным параметром, определяющим его пригодность для широкого спектра промышленных применений. Поверхностные дефекты, возникающие в процессе производства, могут значительно снижать механические свойства, коррозионную стойкость и эстетическую привлекательность материала, приводя к браку и увеличению затрат. В связи с этим, автоматизация процесса обнаружения и классификации таких дефектов представляет собой актуальную и востребованную задачу.

Традиционные методы контроля качества, основанные на визуальном осмотре, являются трудоемкими, субъективными и не всегда позволяют выявлять дефекты на ранних стадиях. В последние годы, благодаря развитию технологий искусственного интеллекта (ИИ), появились новые возможности для автоматизации этой задачи.

В данном разделе представлен обзор существующих подходов к обнаружению и классификации поверхностных дефектов листового металлопроката с использованием методов и алгоритмов ИИ. Особое внимание уделяется анализу преимуществ и недостатков различных подходов, а также определению перспективных направлений исследований в этой области. Рассматриваются как классические методы машинного обучения, основанные на ручном извлечении признаков, так и современные методы глубокого обучения, позволяющие автоматически извлекать сложные иерархические представления из изображений поверхности. Анализируются гибридные подходы, объединяющие преимущества различных методов, а также подходы, основанные на анализе текстуры и спектральных характеристик поверхности. Данный обзор послужит основой для выбора наиболее эффективных методов и алгоритмов для решения задачи обнаружения и классификации дефектов в рамках настоящей выпускной квалификационной работы.

**1.2.1 Обзор методов машинного обучения**

На данный момент существует две основные группы методов машинного зрения: традиционные методы и методы глубоко обучения.

Методы глубокого обучения (МГО), значительно превосходят традиционные методы машинного обучения во многих задачах, особенно в тех, где данные обладают сложной структурой. Вот основные преимущества глубокого обучения перед традиционными подходами:

Во-первых, МГО автоматически извлекает признаки из данных. Нейронные сети учатся выявлять наиболее релевантные признаки для решения задачи. Это особенно важно для сложных данных, таких как изображения, текст и звук.

Во-вторых, разработаны специальные архитектуры для работы с неструктурированными данными. Например, рассматриваемые нами сверточные нейронные сети (СНС) для изображений.

В-третьих, Производительность обычно улучшается при увеличении объема данных. Глубокие нейронные сети способны извлекать сложные зависимости из больших объемов данных, что позволяет достигать высокой точности.

Одними из самых популярных СНС на данный момент являются: EfficientNet, ResNet, VGG. Их мы и будем использовать в нашей работе.

**1.2.2 Трансферное обучение**

Это метод машинного обучения, при котором знания, полученные при решении одной задачи, используются для решения другой, похожей задачи. Вместо того чтобы обучать модель с нуля для каждой новой задачи, трансферное обучение позволяет переносить уже изученные признаки и паттерны, что значительно ускоряет процесс обучения и улучшает производительность модели на новой задаче.

Так как мы не обладаем большим количеством данных для обучения с нуля нашей СНС, то будем использовать уже предобученные, чтобы улучшить наши результаты.

Трансферное обучение работает, потому что многие реальные задачи имеют общие черты и закономерности. Например, модель, обученная распознавать объекты на изображениях, может быть использована для распознавания объектов на новых изображениях, даже если эти изображения взяты из другой сферы (например, медицинские изображения вместо обычных фотографий). Модель уже научилась выделять базовые признаки, такие как края, углы и текстуры, которые полезны для многих задач распознавания изображений и тем самым, она будет лучше справляться с новой задачей.

* 1. **Обзор набора данных**

Для обнаружения и классификации поверхностных дефектов листового металлопроката будем использовать набор данных NEU-surface-defect-database.

Набор данных содержит 1800 изображений в оттенках серого: по 300 образцов с шестью различными типичными дефектами поверхности на каждом из них.

Он содержит следующие 6 видов поверхностных дефектов: Прокатная окалина (RS), заплатка (Pa), деформация (Cr), поверхность с ямками (PS), металловключения (In) и царапины (Sc).

* 1. **Вывод по теоретическому обзору**

Проведенный анализ текущего состояния в области контроля качества листового металлопроката подчеркивает актуальность разработки автоматизированных систем обнаружения и классификации поверхностных дефектов.

Традиционные методы, основанные на визуальном осмотре и других неавтоматизированных подходах, обладают существенными ограничениями в скорости, точности и объективности. В связи с этим, в работе исследуется возможность применения методов искусственного интеллекта, а именно глубокого обучения, для решения поставленной задачи.

Определено, что для реализации практической части исследования будет использован набор данных NEU-surface-defect-database, содержащий изображения с шестью различными типами дефектов поверхности листового металлопроката.

Результаты теоретического анализа и выбор набора данных послужат основой для последующей разработки и сравнения алгоритмов автоматического обнаружения и классификации дефектов, что позволит оценить эффективность применения методов глубокого обучения в данной предметной области.

1. **Анализ существующих архитектур**
   1. **Архитектура СНС**

Свёрточная нейронная сеть — это класс глубоких нейронных сетей, которые особенно хорошо подходят для обработки данных, имеющих структуру, похожую на сетку, таких как изображения, видео, аудио и текст. СНС стали доминирующим подходом во многих задачах компьютерного зрения, а также успешно применяются в других областях.

Основные компоненты СНС:

1. **Свёрточный слой:**

* **Операция свёртки:** Ядром СНС является операция свёртки. Свёрточный слой состоит из набора обучаемых фильтров (или ядер). Каждый фильтр скользит по входным данным (например, по изображению), выполняя поэлементное умножение между фильтром и соответствующей частью входных данных, а затем суммирует результаты. Этот процесс порождает карту признаков (feature map), которая показывает, где в входных данных находятся признаки, на которые реагирует данный фильтр.
* **Локальная связность:** В отличие от полносвязных слоев, свёрточные слои имеют локальную связность. Это означает, что каждый нейрон в свёрточном слое связан только с небольшой областью во входных данных, называемой рецептивным полем (receptive field). Это позволяет сети эффективно обнаруживать локальные признаки и снижает количество обучаемых параметров.
* **Разделение весов:** Все нейроны в карте признаков используют один и тот же фильтр. Это называется разделением весов (weight sharing) и еще больше снижает количество параметров, а также делает сеть устойчивой к сдвигам входных данных.
* **Несколько фильтров:** Свёрточный слой обычно содержит несколько фильтров, каждый из которых извлекает разные признаки из входных данных.

1. **Функция активации:**

* **ReLU:** После каждого свёрточного слоя обычно применяется функция активации. ReLU — это простая, но эффективная функция, которая возвращает входное значение, если оно положительное, и ноль в противном случае. ReLU помогает сети быстрее обучаться и избежать проблемы затухающего градиента.

1. **Слои пулинга:**

* **Уменьшение размерности:** Слои пулинга уменьшают пространственное разрешение карт признаков, уменьшая количество параметров и вычислительных операций.
* **Инвариантность к сдвигам:** Пулинг также делает сеть более устойчивой к небольшим сдвигам и искажениям входных данных.
* **Max-pooling:** Наиболее распространенный тип пулинга. Выбирает максимальное значение в каждом пуле (небольшой области карты признаков).

1. **Полносвязные слои:**

* **Классификация:** В конце СНС обычно находятся один или несколько полносвязных слоев, которые выполняют классификацию признаков, извлеченных свёрточными слоями.
* **Каждый нейрон связан со всеми нейронами предыдущего слоя:** В полносвязных слоях каждый нейрон связан со всеми нейронами предыдущего слоя.

1. **Выходной слой:**

* **Softmax:** Для задач многоклассовой классификации используется функция softmax, которая преобразует выходные значения в вероятности для каждого класса.
* **Sigmoid:** Для задач бинарной классификации используется функция sigmoid, которая выдает вероятность принадлежности к одному из двух классов.

1. **Как работает СНС:**
2. **Вход:** СНС принимает на вход данные, имеющие структуру сетки (например, изображение).
3. **Свёртка:** Свёрточные слои извлекают локальные признаки из входных данных с помощью фильтров.
4. **Активация:** Функция активации ReLU применяет нелинейность к картам признаков.
5. **Пулинг:** Слои пулинга уменьшают размерность карт признаков и делают сеть более устойчивой к сдвигам.
6. **Полносвязные слои:** Полносвязные слои выполняют классификацию признаков, извлеченных на предыдущих слоях.
7. **Выход:** Выходной слой выдает вероятность принадлежности к каждому классу (для классификации) или прогноз (для регрессии).
   1. **Архитектура EfficientNet**

**EfficientNet** — семейство сверточных нейронных сетей, которое позволяет достичь высокой точности при меньшем количестве вычислительных ресурсов по сравнению с предыдущими архитектурами. Главная особенность EfficientNet заключается в эффективном масштабировании базовой сети, получившей название **EfficientNet-B0**, для достижения высокой точности и эффективности (меньшего количества параметров и вычислений) по сравнению с другими CNN.

Традиционно, CNN масштабируют, увеличивая глубину (количество слоев), ширину (количество каналов в слоях) или разрешение входного изображения. Однако, такое масштабирование часто выполняется эмпирически (путем проб и ошибок) и может привести к:

* **Снижению точности:** Несбалансированное масштабирование может привести к переобучению.
* **Увеличению вычислительных затрат:** Простое увеличение всех параметров может привести к чрезмерному увеличению количества операций и, как следствие, к замедлению работы.

EfficientNet предлагает новый подход к масштабированию, названный **Compound Scaling**. Идея заключается в том, чтобы масштабировать все три измерения (глубину, ширину и разрешение) согласованно, используя набор коэффициентов, определенный с помощью архитектурного поиска (Neural Architecture Search - NAS).

Формально:

Глубина:

Ширина:

Разрешение:

Где – константы, определённые с помощью NAS, задающие пропорции масштабирования каждого измерения. – коэффициент масштабирования, определяемый пользователем, который контролирует общее количество ресурсов, выделяемых на масштабирование.

Основные компоненты EfficientNet-B0:

1. **MBConv:**

* Использует inverted residual blocks, как в MobileNetV2, для уменьшения количества параметров и вычислений.
* Использует depthwise separable convolutions для дальнейшего снижения вычислительной сложности.

1. **Squeeze-and-Excitation blocks:**

* Используются для адаптивной перекалибровки каналов в feature maps, что позволяет сети уделять больше внимания наиболее важным каналам.

1. **Compound Scaling:** Как описано выше, согласованное масштабирование глубины, ширины и разрешения.

**Преимущества EfficientNet:**

* **Высокая точность:** Достигает высокой точности на различных задачах компьютерного зрения (классификация изображений, обнаружение объектов).
* **Высокая эффективность:** Имеет значительно меньшее количество параметров и вычислений, чем другие СНС, что делает ее более быстрой и энергоэффективной.
* **Универсальность:** Может быть использована для различных задач компьютерного зрения.
* **Простота масштабирования:** Согласованное масштабирование позволяет легко создавать модели разной вычислительной сложности.
  1. **Архитектура ResNet**

**ResNet –** семейство СНС, которая совершила прорыв в области глубокого обучения, решив проблему затухающего градиента и позволяя успешно обучать очень глубокие сети (до 152 слоев и более).

Одна из основных проблем при обучении глубоких нейронных сетей – это затухающий градиент. При распространении градиента ошибки обратно по сети для обновления весов, градиент может экспоненциально уменьшаться с каждым слоем, особенно в очень глубоких сетях. Это приводит к тому, что веса в ранних слоях не обновляются эффективно, и сеть не обучается.

**Ключевая идея ResNet -** Остаточные связи**:** ResNet решает проблему затухающего градиента с помощью остаточных связей (residual connections), также известных как skip connections или shortcut connections. Вместо того, чтобы каждый слой обучался отображению входов  в новые признаки  ResNet обучается остаточному отображению . Затем выход блока ResNet получается путем добавления входа  к остаточному отображению:

**Преимущества остаточных связей:**

* **Решение проблемы затухающего градиента:** Остаточные связи позволяют градиенту ошибки распространяться непосредственно через “shortcut” connections, минуя сверточные слои. Это помогает сохранить градиент и обеспечивает эффективное обучение даже очень глубоких сетей.
* **Упрощение обучения:** Остаточные связи облегчают обучение, позволяя сети “выучить” тождественное отображение (identity mapping), если это необходимо. Если сверточные слои в блоке ResNet не нужны для улучшения производительности, сеть может просто установить их веса в ноль, и остаточная связь будет передавать вход без изменений.
* **Возможность обучения очень глубоких сетей:** Благодаря остаточным связям, ResNet позволяет обучать сети с сотнями и даже тысячами слоев.

**Архитектура ResNet:**

ResNet состоит из нескольких этапов, каждый из которых содержит несколько **residual blocks**.

**Основные компоненты ResNet**:

* **Stem:** Начальный слой, состоящий из свертки 7x7 с шагом 2, за которой следует batch normalization и ReLU.
* **Residual Blocks:** Основные строительные блоки ResNet, содержащие две или три сверточные слоя с остаточной связью.
* **Average Pooling:** Глобальное среднее объединение в конце сети.
* **Fully Connected Layer:** Полносвязный слой для классификации.

**Типы residual blocks:** ResNet использует два основных типа residual blocks:

* **Basic Block:** Используется для ResNet-18 и ResNet-34. Состоит из двух сверточных слоев 3x3 с batch normalization и ReLU.
* **Bottleneck Block:** Используется для более глубоких сетей (ResNet-50, ResNet-101, ResNet-152). Состоит из трех сверточных слоев: 1x1, 3x3 и 1x1. Свертки 1x1 используются для уменьшения и последующего увеличения количества каналов, что снижает вычислительную сложность.

**Преимущества ResNet:**

* **Решает проблему затухающего градиента:** Остаточные связи обеспечивают эффективное обучение даже очень глубоких сетей.
* **Простота обучения:** Остаточные связи облегчают обучение, позволяя сети “выучить” тождественное отображение.
* **Высокая точность:** ResNet достигает высокой точности на различных задачах компьютерного зрения.
* **Широкое распространение:** Существует множество предварительно обученных моделей ResNet, доступных для использования.
  1. **Архитектура VGGNet:**

VGGNet использует последовательные слои сверточных фильтров для изучения особенностей изображений. Каждый сверточный слой применяет несколько фильтров, за которыми следует нелинейная функция активации. Фильтры захватывают разные аспекты изображения, такие как края и текстуры. VGGNet стала важным этапом в развитии СНС, продемонстрировав, что увеличение глубины сети с использованием небольших фильтров свертки может значительно улучшить производительность.

До VGGNet часто использовались большие фильтры свертки (например, 11x11, 7x7) в ранних слоях СНС. VGGNet ставила целью исследовать, как глубина сети влияет на точность, используя только небольшие фильтры свертки размером 3x3. Авторы предполагали, что более глубокие сети с меньшими фильтрами могут быть более эффективными в извлечении признаков.

**Ключевые особенности VGGNet:**

* **Глубина:** VGGNet характеризуется большой глубиной – сети VGG16 и VGG19 имеют 16 и 19 сверточных слоев соответственно.
* **Однородность:** В основном используются только сверточные слои с фильтрами 3x3 и слои max-pooling 2x2 с шагом 2. Такая однородная структура упрощает проектирование и понимание сети.
* **Маленькие фильтры свертки:** Использование небольших фильтров свертки 3x3.

**Архитектура VGGNet:**

VGGNet состоит из нескольких последовательных блоков, каждый из которых содержит несколько сверточных слоев 3x3, за которыми следует слой max-pooling 2x2. В конце сети расположены три полносвязных слоя.

Общая структура VGGNet:

1. **Сверточные слои:** Последовательность сверточных слоев 3x3 с ReLU. Количество сверточных слоев в каждом блоке варьируется.
2. **Max-Pooling:** Слой max-pooling 2x2 с шагом 2 для уменьшения размера feature maps.
3. **Полносвязные слои:** Три полносвязных слоя: два слоя по 4096 нейронов и выходной слой с количеством нейронов, равным количеству классов (например, 1000 для ImageNet).
4. **Softmax:** Softmax функция для получения вероятностей классов.

**Преимущества VGGNet:**

* **Простота:** Однородная и легко понимаемая архитектура.
* **Увеличение глубины:** Демонстрация эффективности увеличения глубины сети.
* **Маленькие фильтры:** Использование маленьких фильтров свертки 3x3.
* **Хорошая производительность:** Достижение высокой точности на ImageNet.

1. **РАЗРАБОТКА**
   1. **Метрики для оценки модели**

Выбор правильной метрики для оценки модели машинного обучения - критически важен для понимания её реальной производительности и сравнения с другими моделями. Разные метрики подходят для разных задач и типов данных. Так как у нас исследование посвящено классификации, то рассмотрим метрики для задач классификации:

**Основные термины:**

* **TP (True Positive):** Правильно предсказанный положительный объект.
* **TN (True Negative):** Правильно предсказанный отрицательный объект.
* **FP (False Positive):** Ложно предсказанный положительный объект.
* **FN (False Negative):** Ложно предсказанный отрицательный объект.

**Метрики для задач классификации:**

1. **Accuracy (Точность) -** доля правильно классифицированных объектов:
   * Формула**:**
   * Применение**:** Подходит для сбалансированных классов (когда количество объектов каждого класса примерно одинаково).
   * Ограничения**:** Плохо работает при несбалансированных классах (модель может давать высокую точность, просто предсказывая наиболее распространенный класс).
2. **Precision (Точность) -** доля правильно предсказанных положительных объектов среди всех объектов, предсказанных как положительные:

* Формула**:**
* Применение**:** Важна, когда нужно минимизировать количество ложноположительных срабатываний (FP). Например, в медицинской диагностике, когда важно не поставить ложный диагноз.
* Ограничения**:** Не учитывает ложноотрицательные срабатывания (FN).

1. **Recall (Полнота) -** доля правильно предсказанных положительных объектов среди всех реальных положительных объектов:

* Формула**:**
* Применение**:** Важна, когда нужно минимизировать количество ложноотрицательных срабатываний (FN). Например, в системе обнаружения мошенничества, когда важно не пропустить ни одного мошеннического действия.
* Ограничения**:** Не учитывает ложноположительные срабатывания (FP).

1. **F1-Score (F1-мера) –** среднее гармоническое между точностью и полнотой:

* Формула**:**
* Применение**:** Хорошо подходит для несбалансированных классов, так как учитывает и точность, и полноту.
* Ограничения**:** Может быть сложнее интерпретировать, чем точность или полноту.

1. **Area Under the ROC Curve (AUC-ROC) -** площадь под кривой ROC (Receiver Operating Characteristic). ROC показывает зависимость между True Positive Rate (TPR, он же Recall) и False Positive Rate (FPR) при различных порогах классификации.

* Применение**:** Отличная метрика для оценки бинарной классификации, особенно при несбалансированных классах. Позволяет сравнить модели, не выбирая конкретный порог классификации.
* Интерпретация**:** AUC-ROC = 0.5 означает, что модель не лучше случайного угадывания. AUC-ROC = 1 означает, что модель идеально классифицирует объекты.

1. **Log Loss (Binary Cross-Entropy, Cross-Entropy) -** функция потерь, которая используется для оценки вероятностных моделей классификации.

* Применение**:** Часто используется в задачах бинарной и мультиклассовой классификации.
* Интерпретация**:** Чем меньше значение Log Loss, тем лучше модель предсказывает вероятности классов.

1. **Confusion Matrix (Матрица ошибок) -** таблица, которая показывает количество объектов, правильно и неправильно классифицированных для каждого класса.

* Применение**:** Позволяет получить подробную информацию об ошибках модели и выявить, какие классы она путает.