

В. С. КУВШИННИКОВ, канд. техн. наук

Е. Е. КОВШОВ, д-р техн. наук, проф.

А. В. СУДАРЕВ

АО «Научно-исследовательский и конструкторский институт монтажной технологии — Атомстрой» (Москва, Россия)

ГЕНЕРАЦИЯ СИНТЕТИЧЕСКИХ НАБОРОВ ДАННЫХ ДЛЯ ОБУЧЕНИЯ AI-МОДЕЛЕЙ В ЦИФРОВОЙ РАДИОГРАФИИ

Аннотация: В статье рассматривается проблема дефицита обучающих примеров для формирования наборов данных в задачах применения систем на основе искусственного интеллекта при анализе результатов радиационного вида неразрушающего контроля. Приведены предпосылки для разработки генератора примеров и рассмотрены альтернативные способы повышения качества обучения. Описан алгоритм применения математической модели цифрового двойника радиографического изображения при формировании размеченного набора данных. Предложена структура организации цифровых двойников при модификации условий экспонирования для повышения разнообразия примеров изображений. Представлены результаты автоматизированной генерации дефектов типа «продольная трещина».

Ключевые слова: цифровые и информационные технологии, искусственный интеллект, наборы данных, промышленная рентгенография, рентгеновское изображение

На сегодняшний день существует не одна архитектура искусственных нейронных сетей, практическое применение которых при решении прикладных производственных задач показывает обнадеживающие результаты. Особенную роль технологии искусственного интеллекта (ИИ) могут сыграть при замене человеческого труда в решении сложно формализуемых задач. К такому классу задач относятся задачи, связанные с распознаванием образов, например, при анализе результатов неразрушающего контроля.

Во многих отраслях производства, включая такие стратегически важные для государства, как машиностроение и энергетическая отрасль, незаменим радиационный вид неразрушающего контроля. По сравнению с визуальным и измерительным, а также ультразвуковым контролем, в радиационном виде используется ионизирующее излучение, крайне опасное для окружающих. Как действия рентгеновских и гамма-дефектоскопистов, так и отходы их производственной деятельности представляют более значимую угрозу для здоровья людей и окружающей среды, чем в других видах контроля.

В последние годы активно исследуются возможности применения сверточных нейронных сетей для решения задачи автоматической детекции дефектов на рентгенограммах, полученных при радиационном неразрушающем контроле сварных соединений металлических поверхностей деталей и конструкций. Многие из существующих решений используют однопроходные детекторы семейства YOLO (v5–v8), которые известны балансом точности и

скорости вывода. Баланс качества и стоимости вывода является важным показателем эффективности интеграции ИИ систем в производственные линии.

Уровень качества полученной ИИ системы во многом зависит от качества набора данных, использованных при обучении весов модели. Нередко используются разнородные датасеты, составленные из нескольких открытых источников. Такое решение вызвано дефицитом обучающих примеров, что, в свою очередь, объясняется сложностью и дороговизной организации процесса создания примеров размеченных изображений-результатов радиационного контроля. Анализ разнородных датасетов показал, что подобные данные нередко сопровождаются разметкой неудовлетворительного качества. Модели, обученные с применением разнородной выборки без учёта доменных смещений, обладают меньшей точностью. Исследователями предлагаются разные решения, позволяющие снизить негативный эффект от недостатков обучающих наборов. С одной стороны работа ведется в направлении развития архитектуры нейронных сетей, а с другой стороны совершенствуются средства обработки и аугментации изображений обучающих примеров. В рассматриваемой задаче применение многих стандартных аугментаций ограничено или исключено. Геометрические аугментации и аугментации цветового пространства, подходящие для фотоматериалов, некорректно модифицируют изображение швов и дефектов сварных соединений на рентгенограммах. Генеративный подход в создании обучающих примеров встречается реже. Среди способов генерации примеров наиболее перспективным направлением является симуляция изображений с применением математических моделей. Этот подход предусматривает применение приближенной к физически достоверной математической модели, формирующей нелинейную зависимость между уровнем яркости пикселей изображения и геометрически достоверными толщинами просвечиваемых материалов с учетом характеристик испускаемого источником проникающего излучения.

Одна из таких моделей была разработана при создании цифрового виртуального симулятора радиационного неразрушающего контроля. В его задачи входит снижение стоимости обучения специалистов по радиационному контролю, увеличение часов практических занятий с задействованием двигательной и зрительной памяти. Применение виртуальной среды производится при ознакомлении с оборудованием и в ходе развития навыков самостоятельной реализации этапов технологического процесса рентгеновского и гамма-контроля. Важной частью симулятора является комплексная система цифровых двойников объектов, позволяющая формировать симулированные радиографические изображения. При этом в интересах подготовки специалистов используемая модель позволяет отражать влияние параметров экспонирования на качество результатов. Моделируются изображения маркеров, индикаторов чувствительности, ограничительных меток и другого вспомогательного технологического оборудования. Соблюдается соответствие выполненным настройкам и различимости изображений объектов.

Вышеперечисленные качества позволяют применять модель в качестве базы при генерации дополнительных примеров изображений. Реализация законченной системы генерации требует решения ряда дополнительных задач, среди которых можно выделить: разработка циклического алгоритма создания примеров, разработка алгоритма формирования разнообразных условий экспонирования, автоматизация генерации разнообразных объектов контроля с дефектами разных типов (ГОСТ Р ИСО 6520-1-2012), автоматизация процесса разметки сгенерированных изображений. Схема алгоритма автоматизированного формирования размеченных данных приведена на рис. 1.

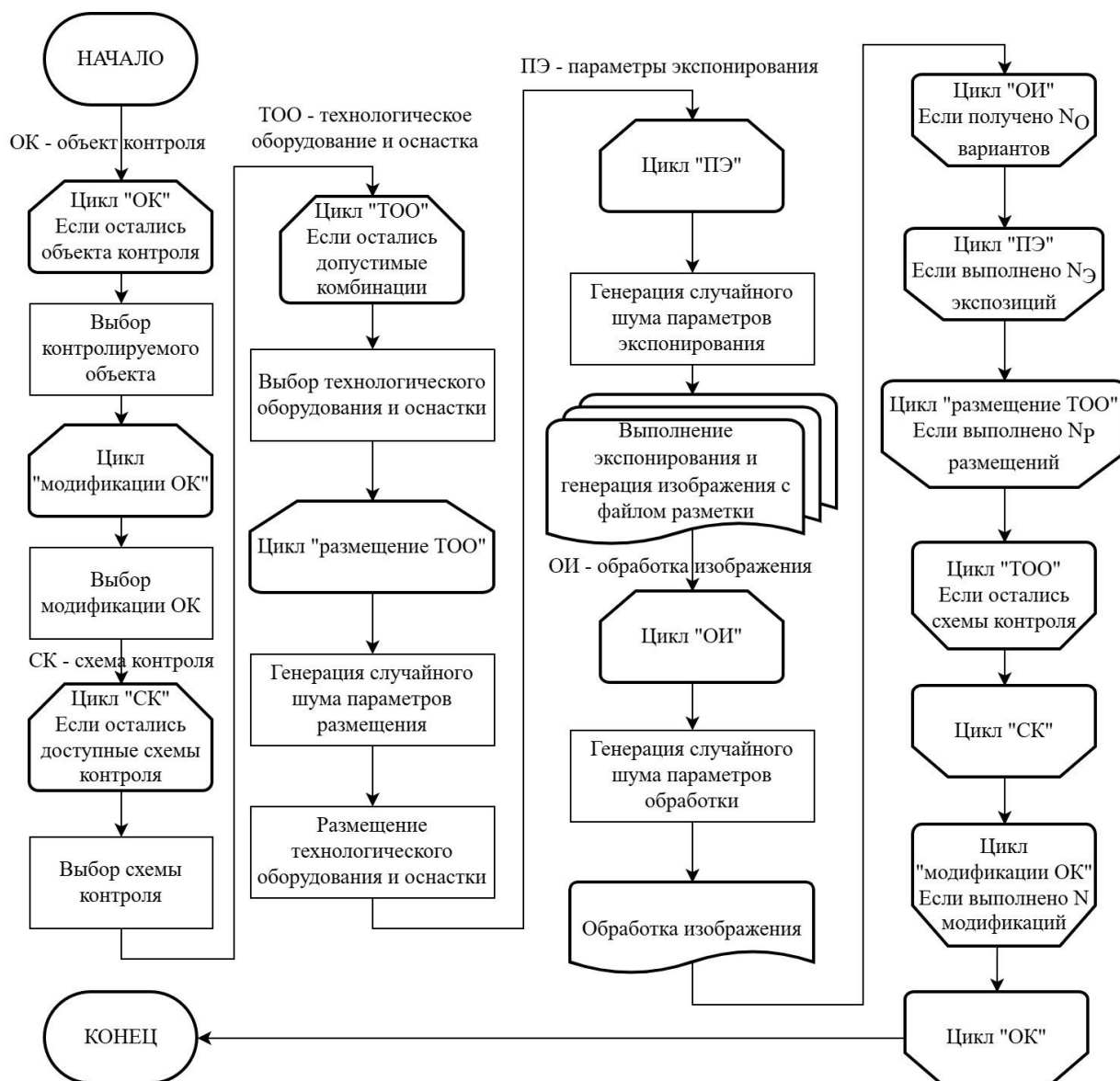


Рис. 1. Алгоритм автоматизированного формирования размеченных данных

При решении задачи реализации алгоритма формирования условий экспонирования предложен способ организации цифровых двойников по группам, позволяющий в удобной форме модифицировать примеры, получая высокую репрезентативность изображений, представленная на рис. 2. Для

решения задач формирования разнообразных примеров размеченных изображений введен дополнительный вид цифрового двойника – цифровой двойник дефектоскописта.

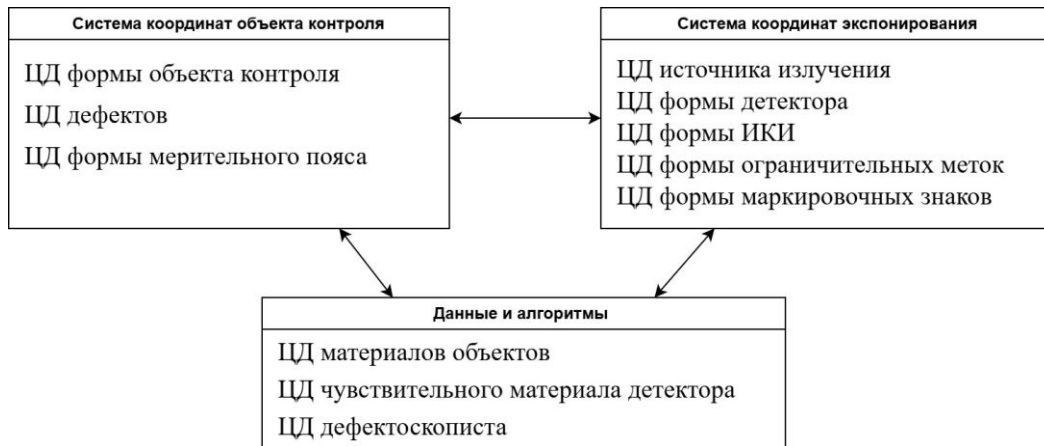


Рис. 2. Структура организации цифровых двойников при модификации условий экспонирования

Одной из важных составляющих задачи генерации объектов контроля является генерация разнообразных примеров дефектов заданного типа. Пример изображений автоматически сгенерированных продольных трещин представлен на рис. 3.

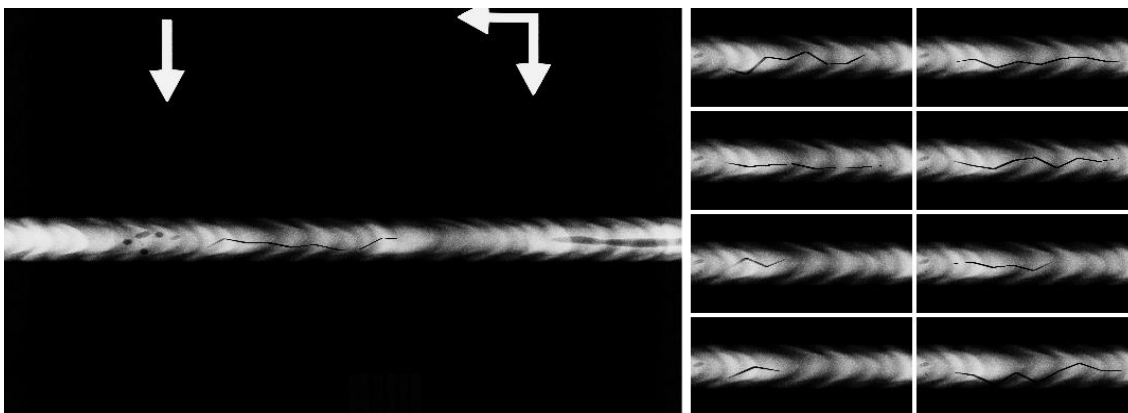


Рис. 3. Пример автоматически сгенерированных изображений трещин

Представленное решение позволяет создавать новые и балансировать имеющиеся наборы данных для решения узкоспециализированных задач. Достигнутых результатов достаточно для создания ранней версии ИИ системы поддержки принятия решений, что позволяет накапливать верифицированные результаты анализа реальных изображений для расширения обучающего набора.

Контакты:

KuvshinnikovVS@atomrus.ru (Кувшинников Владимир Сергеевич)

KovshovEE@atomrus.ru (Ковшов Евгений Евгеньевич)