

УДК 004.85:004.032.26

Использование генеративных сетей для создания синтетических обучающих выборок в задачах классификации изображений

Терехов Валерий Игоревич

terekchow@bmstu.ru

SPIN-код: 7370-1150

Колосов Максим Игоревич

mgtumax@yandex.ru

SPIN-код: 3474-5668

МГТУ им. Н.Э. Баумана, Москва, Россия

Аннотация. Рассмотрена проблема дефицита размеченных данных для обучения глубоких нейронных сетей. Представлен подход к созданию синтетических обучающих выборок с использованием генеративно-состязательных сетей. Показаны преимущества метода, включая возможность контроля над размещением объектов, вариативностью фонов и параметров освещения. Сделаны выводы, что предложенная методика позволяет эффективно расширять существующие наборы данных и может быть применена для задач, где сбор реальных данных затруднен.

Ключевые слова: генеративно-состязательные сети, синтетические данные, обучающие выборки, глубокое обучение, компьютерное зрение

Введение. Как следует из названия, синтетические данные — это данные, которые созданы искусственно, а не в результате реальных событий [1]. Синтетические данные способны с высокой точностью воспроизводить статистические свойства и скрытые закономерности реальных данных, сохраняя при этом контролируемые отличия [2]. Это позволяет эффективно применять их для дополнения и балансировки обучающих выборок. В случаях, когда сбор реальных данных невозможен или экономически нецелесообразен, синтетические данные, сгенерированные на основе экспертных знаний о предметной области или с использованием специализированных алгоритмов, становятся единственным практическим источником для обучения моделей машинного обучения.

Методы и материалы; результаты. Верификация практической применимости синтетических данных проводилась на задаче классификации изображений с использованием глубоких нейросетевых архитектур. Модель обучалась на трех различных наборах данных: исключительно на реальных данных, исключительно на синтетических данных, а также на комбинированной выборке. Наилучшие результаты продемонстрировала модель, обученная на комбинированном наборе данных, что свидетельствует о синергетическом эффекте при совместном использовании реальных и синтетических образцов. Важно отметить, что использование исключительно синтетических данных также позволило достичь удовлетворительных результатов, что подтверждает перспективность их применения в условиях дефицита реальных данных.

Заключение. Экспериментально подтверждена операционная эффективность генеративно-состязательных сетей для синтеза обучающих данных, позволяющих преодолевать ключевое ограничение современных систем ИИ — дефицита размеченных данных [3]. Установленный синергетический эффект от комбинации синтетических и реальных данных создает методологическую основу для применения технологии в критически важных областях: медицинской диагностике, автономных системах и промышленной автоматизации. Перспективы исследований заключаются в разработке формализованных метрик валидации и создании адаптивных систем генерации, учитывающих предметную специфику.

Список источников

- [1] Terekhov V., Zabelina V., Savchenko G., Chumachenko S. Classification of Tree Species by Trunk Image Using Conventional Neural Network and Augmentation of the Training Sample Using a Telegram-Bot. *Studies in Computational Intelligence*, 2022, vol. 1008 SCI, pp. 210–216.
- [2] Бессмертный И.А. *Системы искусственного интеллекта*. Москва, Изд-во Юрайт, 2025, 164 с.
- [3] Рабчевский А.Н. *Синтетические данные и развитие нейросетевых технологий*. Москва, Юрайт, 2025, 187 с.