УДК 004.89

Разработка методов генерации видео для анализа однородных данных в промышленности

Золотарев Д.В. (Университет ИТМО), **Румянцева М.Ю.** (Университет ИТМО)

Научный руководитель – Ефимова В. А.

(Университет ИТМО)

Введение. Флотация — это процесс, используемый для разделения полезных минералов от пустой породы посредством разницы в смачиваемости частиц, что приводит к образованию пузырей и их подъёму к поверхности. Анализ характеристик движения пузырей – скорости, направления и изменения формы – является важным этапом контроля и оптимизации технологического процесса обогащения руды.

При решении задач отслеживания множества однородных объектов и их анализа основная проблема — это ограниченность размеченных данных. Сбор и ручная разметка видеоданных требуют значительных затрат, особенно для динамичного процесса флотации. Один из подходов — разметка отдельных кадров для детекции (20-50 кадров), с последующим использованием алгоритмов трекинга. Однако этот вариант не обеспечивает высокой точности, так как не позволяет полностью отследить все пространственно-временные зависимости.

Поэтому в этой работе предлагается использовать синтетическую генерацию видеоданных для получения достоверного Ground Truth. В рамках работы обучены модели интерполяции FILM, RIFE, EMA-VFI и авторегрессивные модели предсказания следующего кадра IAM4VP DMSVFN в этом домене.

Основная часть. Для генерации синтетических видеоданных используются два подхода:

1. **Интерполяция кадров**. Создаются промежуточные кадры между ключевыми и с применением оптического потока и алгоритмов FILM [1], RIFE [2] и EMA-VFI [3]. При этом EMA-VFI является моделью, основанной на XVFI [4] с более высокими метриками на открытых датасетах. Преимуществами этих методов являются высокое качество интерполируемых кадров и реалистичность движения. Однако качество результата во многом зависит от способности модели корректно оценивать динамику сцены и передавать текстурные особенности однородных объектов.
2. **Предсказание следующих кадров**. Модель прогнозирует будущее состояние сцены по последовательности кадров с использованием сверточных сетей и трансформеров. Были рассмотрены модели: SimVP [5], IAM4VP [6] и DMSVFN [7]. Также в работе был использован фреймворк OpenSTL [8], в котором собраны различные архитектурные решения для задачи предсказания следующих кадров.

Синтетическая генерация позволяет получать значительные объёмы данных с точной разметкой, что упрощает обучение сегментационных и мультитрекинговых моделей, однако могут возникать артефакты при отсутствии учёта специфики домена.

Выводы. Использование синтетической генерации видеоданных позволяет существенно ускорить и упростить процесс подготовки обучающих выборок для задач анализа однородных объектов. В рамках данной работы были успешно обучены модели FILM, RIFE, EMA-VFI и IAM4VP DMSVFN, что подтвердило эффективность предложенных методов. Методы интерполяции кадров и предсказания следующего кадра являются перспективными инструментами, обеспечивающими высокое качество генерируемых видеопоследовательностей. В дальнейшем планируется добавить маски мультитрекинга к генерации, что позволит создавать готовые датасеты мультитрекинга с масками GT. Также планируется расширить домен с флотации до всех однородных данных.

Список использованных источников:

[1] Fitsum Reda и др. "FILM: Frame Interpolation for Large Motion." arXiv:2202.04901, 2022. URL: https://arxiv.org/abs/2202.04901.

[2] Zhewei Huang и др. "Real-Time Intermediate Flow Estimation for Video Frame Interpolation." arXiv:2011.06294, 2020. URL: https://arxiv.org/abs/2011.06294.

[3] Guozhen Zhang и др. "Extracting Motion and Appearance via Inter-Frame Attention for Efficient Video Frame Interpolation." arXiv:2303.00440, 2023. URL: https://arxiv.org/abs/2303.00440.

[4] Hyeonjun Sim и др. "XVFI: eXtreme Video Frame Interpolation." arXiv:2103.16206, 2021. URL: https://arxiv.org/abs/2103.16206.

[5] Zhangyang Gao, Cheng Tan, Lirong Wu, Stan Z. Li. SimVP: Simpler yet Better Video Prediction. arXiv preprint arXiv:2206.05099, 2022. URL: https://arxiv.org/abs/2206.05099.

[6] Minseok Seo, Hakjin Lee, Doyi Kim, Junghoon Seo. Implicit Stacked Autoregressive Model for Video Prediction. arXiv preprint arXiv:2303.07849, 2023. URL: https://arxiv.org/abs/2303.07849.

[7] Xiaotao Hu, Zhewei Huang, Ailin Huang, Jun Xu, Shuchang Zhou. A Dynamic Multi-Scale Voxel Flow Network for Video Prediction. arXiv preprint arXiv:2303.09875, 2023. URL: https://arxiv.org/abs/2303.09875.

[8] Cheng Tan. OpenSTL: Open Framework for Spatiotemporal Learning. URL: https://github.com/chengtan9907/OpenSTL, 2023.