

УДК 004.932.2:004.8

Духній В.—ст. гр. СП-41

Тернопільський національний технічний університет імені Івана Пулюя

РОЗРОБКА ПРОГРАМНОЇ СИСТЕМИ ДЛЯ АПСКЕЙЛІНГУ ВІДЕОПОТОКУ З ВИКОРИСТАННЯМ МЕТОДІВ ГЛИБОКОГО НАВЧАННЯ ТА ЧАСОВОЇ УЗГОДЖЕНОСТІ КАДРІВ

Науковий керівник: к.т.н., доц. каф. ПІ Стоянов Ю. М.

Dukhnii V.

Ternopil Ivan Puluj National Technical University

DEVELOPMENT OF A SOFTWARE SYSTEM FOR VIDEO UPSCALING USING DEEP LEARNING AND TEMPORAL FRAME COHERENCE METHODS

Supervisor: PhD, Assoc. Prof. Stoyanov Yu. M.

Ключові слова: апскейлінг відео, глибоке навчання, Real-ESRGAN, часова узгодженість, обробка відеопотоку.

Keywords: video upscaling, deep learning, Real-ESRGAN, temporal coherence, video stream processing.

Вступ. Відео низької роздільної здатності є поширеною проблемою в архівній обробці, системах відеоспостереження та стримінгових платформах. На відміну від обробки окремих зображень, апскейлінг відеопотоку має специфічну складність: незалежна покадрова обробка породжує часові артефакти — мерехтіння, нестабільність текстур та непослідовність деталей між суміжними кадрами. Ці ефекти суттєво знижують суб'єктивну якість відтворення навіть при формально високій роздільній здатності кожного кадру [1].

Мета роботи — розробка програмної системи для підвищення роздільної здатності відеопотоку на основі методів глибокого навчання з урахуванням часової узгодженості між кадрами.

Основна частина. Розроблена система реалізує конвеєрну архітектуру обробки відеопотоку, яка складається з трьох основних компонентів: модуля зчитування відео, ядра попиксельної обробки кадрів та модуля кодування вихідного відео.

Модуль зчитування відео (`video_reader`) реалізований як генератор кадрів — він послідовно витягує кадри у форматі PIL-зображень та передає метадані потоку: частоту кадрів, загальну кількість кадрів, тривалість і роздільну здатність. Генераторний підхід дозволяє уникнути повного розпакування відео на диск, що критично для тривалих кліпів і обмежених ресурсів пам'яті.

Ядром системи підвищення роздільної здатності є модель Real-ESRGAN [2] на базі RRDB-генератора (Residual-in-Residual Dense Block) з підтримкою масштабування 2×, 4× та 8×. Для відео з людьми до конвеєру підключається модуль GFPGAN [3], який відновлює деталі обличчя з використанням генеративних пріорів StyleGAN2. Фреймворк PyTorch забезпечує інференс на трьох апаратних бекендах: CUDA, MPS та

СРУ. Для відео великої роздільної здатності застосовується тайлінг — покадрова обробка по фрагментах з перекриттям із подальшим безшовним зшиванням.

Часова узгодженість між кадрами забезпечується на рівні постобробки: суміжні кадри аналізуються на наявність різких змін яскравості та структури текстур, після чого застосовується адаптивне змішування результатів. Це пригнічує мерехтіння, характерне для незалежної покадрової обробки, і забезпечує плавніший відеоряд без необхідності використання окремих мереж оптичного потоку.

Модуль запису відео (video_writer) кодує оброблені кадри у формат MP4 з кодеком H.264 засобами бібліотеки imageio-ffmpeg. На етапі фіналізації оригінальний аудіопотік мультіплексується у вихідний файл без перекодування, що зберігає якість звуку незалежно від коефіцієнта масштабування відеоряду. Передбачена коректна обробка тимчасових файлів при скасуванні або аварійному завершенні обробки.

Взаємодія з системою реалізована через графічний інтерфейс на базі Streamlit. Користувач може задати діапазон кадрів, крок пропуску кадрів, коефіцієнт масштабування, політику частоти кадрів на виході та перемикач збереження аудіо. Інтерфейс також відображає прогрес обробки з індикацією поточного кадру і загальної кількості, а при відсутності встановленого FFmpeg виводить діагностичне повідомлення з інструкцією зі встановлення.

Результати. Система забезпечує повний цикл обробки відеофайлів: від зчитування MP4-потіку до отримання MP4-виходу з підвищеною роздільною здатністю та збереженням аудіотреком. Генераторна потокова обробка дозволяє опрацьовувати відеокліпи довільної тривалості без зростання пікового споживання оперативної пам'яті. Застосування часової узгодженості між кадрами усуває мерехтіння, характерне для незалежної покадрової обробки.

Висновки. Розроблено програмну систему для апскейлінгу відеопотоку на основі глибокого навчання. Тонкошарова архітектура відео-адаптерів дозволила повністю перевикористати існуючий конвеєр обробки зображень без переписування моделей та фільтрів. Реалізована підтримка часової узгодженості кадрів підвищує суб'єктивну якість вихідного відео порівняно з незалежною покадровою обробкою. Подальший розвиток системи передбачає підтримку форматів MOV та MKV, а також додавання вихідного формату WebM.

Список використаних джерел:

1. Caballero J. et al. Real-Time Video Super-Resolution with Spatio-Temporal Networks and Motion Compensation // Proceedings of the IEEE CVPR. – 2017.
2. Wang X. et al. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data // Proceedings of the IEEE/CVF ICCV Workshops. – 2021.
3. Wang X. et al. Towards Real-World Blind Face Restoration with Generative Facial Prior // Proceedings of the IEEE/CVF CVPR. – 2021.