



PENINGKATAN RESOLUSI VIDEO MUSIK LAWAS INDONESIA MENGUNAKAN REAL-ESRGAN-X4PLUS

Abdul Ra'uf Alfansani^{1*}, Ema Utami², Dhani Ariatmanto³

^{1,2,3}Program Studi S2 PJJ Informatika, Universitas Amikom Yogyakarta, Indonesia

Email: rauf.alfansani@students.amikom.ac.id^{1*}

Abstrak: Video budaya Indonesia yang terekam dalam media analog seperti VCD sering mengalami degradasi visual khas, seperti *blur*, *jitter*, dan *noise* kompresi. Kondisi ini menyulitkan proses restorasi dan distribusi ulang dalam standar visual modern. Penelitian ini bertujuan untuk mengevaluasi efektivitas model *deep learning* Real-ESRGAN-X4plus dalam meningkatkan resolusi video budaya Indonesia beresolusi rendah. Lima video musik tradisional berdurasi ± 4 menit digunakan sebagai sampel. Setiap video diekstraksi menjadi total 5.760 *frame*, kemudian diproses menggunakan 13 kombinasi parameter *denoise_strength* dan *tile_size*. Kinerja model dibandingkan dengan dua metode *baseline*: *interpolasi bicubic* dan ESRGAN publik. Evaluasi objektif dilakukan menggunakan metrik PSNR dan SSIM, sementara kualitas perseptual dianalisis menggunakan pendekatan *objective* MOS berbasis rumus. Hasil menunjukkan bahwa konfigurasi *denoise_strength* = 0.0 dan *tile_size* = 256 memberikan hasil terbaik, dengan PSNR sebesar 35,10 dB, SSIM 0,956, dan skor MOS 5,01. Real-ESRGAN secara konsisten melampaui *baseline*, dan uji *paired t-test* mengonfirmasi perbedaan yang signifikan ($p < 0,01$). Kontribusi utama penelitian ini adalah evaluasi sistematis Real-ESRGAN untuk konteks restorasi video budaya Indonesia, yang sebelumnya belum banyak dikaji. Penelitian ini merekomendasikan eksplorasi lebih lanjut terhadap pelatihan ulang model pada data lokal dan penerapan model *spatio-temporal* untuk meningkatkan koherensi antar-*frame*. Pengembangan pipeline restorasi yang efisien dan adaptif akan menjadi kunci dalam pelestarian arsip audio-visual nasional ke depan.

Kata Kunci: Restorasi video, Real-ESRGAN, *super-resolution*, PSNR, MOS

PENDAHULUAN

Transisi dari media analog ke media digital berdampak signifikan terhadap kualitas dan aksesibilitas konten audiovisual, khususnya untuk video musik Indonesia yang diproduksi antara tahun 1980-an dan 2000-an. Banyak dari karya ini memiliki keterbatasan bawaan seperti resolusi rendah, *noise*, dan pemudaran warna karena asal-usulnya yang analog, sehingga mengurangi kelayakannya untuk *platform* tontonan modern. Kualitas persepsi media audiovisual memainkan peran penting; penelitian menunjukkan bahwa penurunan kualitas pada komponen audio dan video sangat memengaruhi pengalaman penonton secara keseluruhan [8]. Teknik canggih seperti pembelajaran mandiri dalam representasi data audiovisual dapat meningkatkan sinkronisasi dan kualitas artefak tersebut [15]. Lebih jauh lagi, model resolusi super video baru sedang dikembangkan untuk merekonstruksi bingkai berkualitas lebih tinggi dari sumber resolusi lebih rendah, yang menjanjikan untuk menghidupkan kembali konten lama [3]. Metodologi semacam itu dapat menjadi penting dalam melestarikan warisan budaya Indonesia dengan membuat video musik ini lebih cocok untuk penonton kontemporer.

Generative Adversarial Networks (GAN) telah memajukan bidang *super-resolusi* (SR) secara signifikan dengan memungkinkan rekonstruksi detail halus, melampaui metode interpolasi tradisional. Model Real-ESRGAN yang diperkenalkan oleh [16] menggunakan proses pelatihan berdasarkan degradasi sintesis yang secara efektif meniru kondisi dunia nyata, yang menghasilkan kinerja yang unggul saat menangani gambar yang terdegradasi secara alami [18]. Model ini khususnya ditingkatkan dalam varian Real-ESRGAN-X4plus, yang memanfaatkan arsitektur *Residual-in-Residual Dense Block* (RRDB) yang diperkuat untuk mencapai faktor pembesaran 4x [18]. Lebih jauh lagi, integrasi teknik pembelajaran mendalam dalam rekonstruksi gambar menunjukkan kemampuan yang konsisten untuk meningkatkan resolusi spasial dan mengurangi *noise* secara bersamaan, yang menunjukkan dampak transformatif dari pendekatan ini dalam berbagai domain pencitraan [4]. Secara kolektif, kemajuan ini menyoroti kemanjuran metodologi pembelajaran mendalam dalam mengatasi tantangan dalam tugas *super-resolusi* dan rekonstruksi gambar [19].

Penelitian ini bertujuan: (1) menilai efektivitas Real-ESRGAN-X4plus dalam meningkatkan resolusi video musik lawas Indonesia; (2) membandingkan kualitas hasil dengan metode *baseline* (*bicubic* dan ESRGAN); serta (3) menyediakan pedoman teknis untuk proses restorasi arsip video lawas.

TINJAUAN PUSTAKA

Kajian Penelitian Terdahulu

Dalam lima tahun terakhir, pendekatan *Video Super-Resolution* (VSR) berbasis GAN telah mengalami kemajuan signifikan untuk meningkatkan kualitas visual video resolusi rendah. Beberapa model terkemuka telah dikembangkan, namun belum sepenuhnya menjawab kebutuhan khusus untuk restorasi video budaya Indonesia.

Penelitian [15] memperkenalkan Real-ESRGAN, yang dirancang untuk mengatasi artefak dunia nyata menggunakan strategi pelatihan berbasis degradasi sintesis. Meskipun efektif dalam meningkatkan kualitas gambar statis,



model ini belum mempertimbangkan koherensi temporal antar frame, sehingga masih kurang ideal untuk pemrosesan video secara menyeluruh. Sedangkan penelitian [13] mengembangkan ZS-RW-STSR (*Zero-Shot Real-World Spatio-Temporal Super-Resolution*) untuk meningkatkan resolusi video *spasio-temporal* tanpa pelabelan eksplisit. Pendekatan ini menjanjikan, namun dalam konteks video analog Indonesia dengan artefak khas seperti *jitter* dan *blur*, stabilitas performa masih menjadi tantangan. Di sisi lain penelitian [7] merancang model dengan *dual subnet* dan *multi-stage upsampling* untuk menangani pergerakan besar antar *frame*. Hasil yang dihasilkan cukup akurat, tetapi kebutuhan komputasi yang tinggi menjadi kendala dalam penerapannya pada digitalisasi arsip budaya yang umumnya dilakukan dalam infrastruktur terbatas. Selanjutnya penelitian [14] menawarkan pendekatan yang menggabungkan deteksi *region of interest* (ROI) dengan *super-resolusi* untuk efisiensi sistem video modern. Namun, model ini dirancang untuk konten digital kontemporer, dan tidak mempertimbangkan degradasi khas dari media analog seperti VCD. Sementara itu, penelitian [2] fokus pada peningkatan fidelitas video dengan skalabilitas tinggi menggunakan model AI modern. Meski efektif dalam konteks sistem berskala besar, pendekatan ini kurang spesifik untuk restorasi arsip budaya atau menangani tantangan *spasio-temporal* dari media historis.

Tabel 1. Perbandingan Penelitian Terdahulu

Peneliti	Tahun	Model	Kelebihan	Kelemahan
Wang et al.	2021	Real-ESRGAN	Menangani noise real-world, hasil tajam untuk gambar	Tidak mempertahankan kontinuitas antarframe
Shukla et al.	2024	ZS-RW-STSR	Integrasi <i>spasio-temporal</i> , tanpa data label	Tidak stabil pada domain degradasi analog khas Indonesia
Liu et al.	2021	Dual Subnet + Upsampling	Akurat untuk gerakan kompleks	Boros komputasi, tidak praktis untuk arsip budaya
Tang et al.	2024	ROI-VSR Unified Model	Efisien untuk video modern & streaming	Tidak mendeteksi artefak fisik khas video lawas
Das et al.	2024	AI Fidelity Enhancement	Scalable untuk sistem besar	Tidak spesifik untuk restorasi arsip budaya visual

Berdasarkan kajian penelitian terdahulu, dapat disimpulkan bahwa hingga saat ini belum terdapat pendekatan yang secara eksplisit diarahkan untuk restorasi video budaya Indonesia. Sebagian besar model yang dikembangkan masih berfokus pada peningkatan kualitas citra statis atau pengoptimalan kompresi video modern, tanpa mempertimbangkan karakteristik degradasi khas media analog seperti VCD, termasuk artefak visual berupa *jitter*, *blur*, dan *fading* warna. Selain itu, belum banyak studi yang secara sistematis menguji performa model berbasis GAN, khususnya dalam konteks video berdurasi panjang yang berasal dari sumber historis. Oleh karena itu, penelitian ini menjadi relevan dan penting untuk mengevaluasi performa Real-ESRGAN-X4plus dalam skenario nyata restorasi video budaya Indonesia, sekaligus menentukan konfigurasi parameter teknis yang optimal untuk mencapai keseimbangan antara kualitas visual yang tinggi dan efisiensi pemrosesan yang diperlukan dalam skala praktis.

Kajian Teori

Video Super-Resolution Berbasis GAN

Super-resolusi video (VSR) bertujuan meningkatkan kualitas spasial dari bingkai video resolusi rendah. Model SRGAN meletakkan dasar penggunaan GAN dalam tugas super-resolusi dengan menangkap detail frekuensi tinggi [20]. ESRGAN kemudian menyempurnakan pendekatan ini dengan jaringan *residual-in-residual* (RRDB) serta peningkatan *loss* fungsional.

Real-ESRGAN, sebagai pengembangan lebih lanjut, mengatasi tantangan *noise* dan artefak dunia nyata melalui pelatihan pada degradasi sintesis [15]. Namun, model ini tetap fokus pada pemrosesan gambar tunggal (*single-frame*). Upaya mengintegrasikan dimensi *spasio-temporal* dilakukan dalam model seperti ZS-RW-STSR [13], namun belum sepenuhnya dioptimalkan untuk domain video arsip dengan degradasi kompleks.

Secara teoritis, pendekatan GAN telah membuktikan efektivitasnya dalam super-resolusi spasial, namun belum maksimal dalam mengintegrasikan aspek temporal dan konteks budaya visual dalam satu kerangka pemrosesan yang utuh.

Pelestarian Warisan Video Budaya

Teori pelestarian digital menekankan pentingnya mempertahankan nilai historis, sosial, dan estetika dari konten visual saat dilakukan restorasi. Restorasi video musik, dokumenter, atau siaran lama tidak hanya bertujuan meningkatkan resolusi, tetapi juga memfasilitasi akses edukatif dan apresiasi budaya [1], [17].

Lembaga seperti Arsip *Rose Goldsen* menunjukkan bahwa restorasi harus kontekstual, memperhatikan nuansa visual asli dari karya budaya agar tidak hilang saat diproses secara digital. Oleh karena itu, pendekatan teknis dalam restorasi video seharusnya sejalan dengan prinsip-prinsip pelestarian budaya.

Metrik Evaluasi Kualitas Video

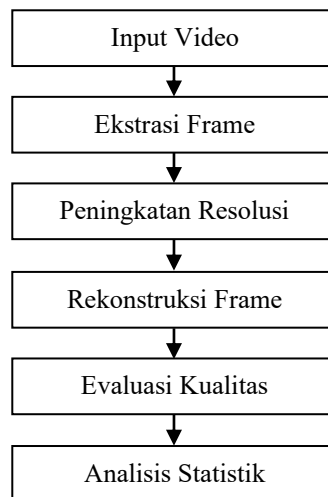


Penilaian hasil restorasi dilakukan dengan dua metrik utama, yaitu *Peak Signal-to-Noise Ratio* (PSNR) mengukur akurasi pixel, dengan nilai >30 dB mengindikasikan kualitas tinggi [5], [12] dan *Structural Similarity Index Measure* (SSIM) menilai kesamaan luminansi, kontras, dan struktur secara perseptual, dengan nilai mendekati 1 dianggap optimal [6]. Gabungan PSNR dan SSIM memberikan penilaian objektif dan perseptual yang menyeluruh.

METODE PENELITIAN

Diagram Alur Eksperimen

Gambar 1 memperlihatkan alur tahapan penelitian, dimulai dari input video VCD, ekstraksi frame, proses peningkatan resolusi menggunakan Real-ESRGAN-X4plus dan *baseline*, rekonstruksi video, evaluasi kualitas, hingga analisis statistik.



Gambar 1. Diagram Alur Penelitian

Desain Eksperimen

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan desain *pre-test-post-test* untuk mengukur pengaruh algoritma peningkatan resolusi berbasis pembelajaran mendalam terhadap kualitas visual video budaya Indonesia. Lima video musik tradisional Indonesia yang berasal dari media VCD tahun 1985–2000 digunakan sebagai sampel, masing-masing berdurasi sekitar empat menit dan beresolusi asli 480p.

Pemilihan lima video dilakukan secara purposif berdasarkan kriteria: (1) representasi periode distribusi media VCD, (2) keberagaman konten visual budaya (tari, busana, arsitektur), dan (3) variasi degradasi visual khas seperti *noise*, *blur*, *jitter*. Lima video dianggap mencukupi untuk pengujian multivariabel parameter karena mampu merepresentasikan pola umum degradasi media analog, dengan pertimbangan efisiensi sumber daya eksperimen.

Prosedur Eksperimen

Ekstraksi Frame

Setiap video dikonversi menjadi *frame* individu menggunakan FFmpeg versi 4.4 pada laju 24 fps. Dengan durasi rata-rata 240 detik per video, total frame per video mencapai sekitar 1.150-1.200 frame, menghasilkan total sekitar 5.760 frame yang dianalisis dalam evaluasi.

Proses Peningkatan Resolusi

Model Real-ESRGAN-X4plus dijalankan menggunakan framework PyTorch 2.2 pada perangkat GPU NVIDIA RTX 4060 (8 GB). Parameter yang diuji meliputi *denoise_strength* : 0.0 hingga 1.0 (inkremental 0.1) dan *tile_size* : 256 (utama), 128, dan 64 sebagai variasi.

Proses inferensi dilakukan per frame dalam mode *batch*, menggunakan segmentasi *tile* otomatis untuk menjaga efisiensi memori. Total 13 konfigurasi diuji.

Metode *Baseline* Pembanding

Dua metode pembanding digunakan adalah (1) *Interpolasi Bicubic* : Metode konvensional peningkatan resolusi yang menggunakan perhitungan matematis piksel tetangga. Hasilnya halus, namun tidak dapat merekonstruksi detail yang hilang. Diimplementasikan melalui OpenCV; (2) ESRGAN Publik: Model *deep learning* berbasis GAN yang menghasilkan detail visual realistis, namun versi default-nya tidak dilatih ulang untuk data video analog, sehingga tidak optimal untuk domain budaya Indonesia. Digunakan sebagai *baseline* AI tanpa *fine-tuning*.

Rekonstruksi Video

Frame hasil *upscaling* disatukan kembali menjadi video format MP4 (H.264, 24 fps) dengan FFmpeg. Audio asli dari VCD disisipkan ulang untuk mempertahankan kesesuaian *temporal* dan pengalaman menonton utuh. Total output video uji adalah 65 (13 konfigurasi Real-ESRGAN $\times$ 5 video + 10 video *baseline*).



Evaluasi Kualitas

Evaluasi Objektif

Nilai PSNR dan SSIM dihitung per *frame* menggunakan pustaka *scikit-image* dan OpenCV. Nilai rata-rata dihitung untuk masing-masing video dan konfigurasi parameter.

Evaluasi Subjektif (*Objective MOS*)

Karena tidak dilakukan penilaian persepsi visual oleh panelis manusia, penelitian ini menggunakan pendekatan *objective MOS (Mean Opinion Score)* berbasis kombinasi PSNR dan SSIM. Pendekatan ini memungkinkan estimasi persepsi visual manusia secara kuantitatif dan efisien. Perhitungan dilakukan dengan rumus:

$$\text{MOS} = 0.1 \times \text{PSNR} + 5 \times \text{SSIM} - 3.5 \quad (1)$$

Rumus ini digunakan karena telah terbukti memiliki korelasi tinggi dengan penilaian subjektif, dan memungkinkan komparasi antar konfigurasi parameter Real-ESRGAN secara sistematis.

Mitigasi Bias Evaluasi

Seluruh evaluasi kualitas dilakukan secara otomatis berbasis skrip, tanpa intervensi visual manual selama interpretasi hasil. Hal ini memastikan pendekatan yang objektif dan menghindari potensi bias visual, sesuai prinsip blind evaluation.

Analisis Statistik

Analisis statistik menggunakan uji t-berpasangan (*paired t-test*) dengan $\alpha = 0,05$ untuk mengukur signifikansi perbedaan antara metode Real-ESRGAN dan *baseline*, berdasarkan nilai PSNR, SSIM, dan MOS. Pengujian dilakukan per video dan per konfigurasi parameter, untuk mengidentifikasi konfigurasi optimal secara kuantitatif dan perseptual. Dengan rancangan eksperimen ini, penelitian diharapkan dapat memberikan evaluasi menyeluruh terhadap efektivitas Real-ESRGAN-X4plus dalam restorasi video budaya Indonesia serta menghasilkan rekomendasi parameter teknis terbaik untuk kualitas visual optimal dengan efisiensi pemrosesan yang layak.

HASIL DAN PEMBAHASAN

Rekonstruksi Video

Seluruh eksperimen dilakukan pada lima video musik budaya Indonesia berformat VCD berdurasi ± 4 menit dengan resolusi asli 852×480 piksel. Setiap video diekstraksi menjadi *frame* (sekitar 1.150–1.200 *frame*/video, total 5.760 *frame*). Peningkatan resolusi dilakukan menggunakan Real-ESRGAN-X4plus (13 konfigurasi parameter) dan dua *baseline*. Hasil *frame* disatukan kembali menjadi video MP4 (H.264, 24 fps) dengan audio asli disisipkan ulang. Total video hasil rekonstruksi adalah 65 (13 $\times$ 5 Real-ESRGAN dan 10 *baseline*).

Evaluasi Ukuran Tile

Pengujian *tile_size* dilakukan untuk mengetahui pengaruh terhadap waktu dan kualitas. *Tile default* 256 dibandingkan dengan variasi 128 dan 64, menggunakan konfigurasi *denoise_strength* = 0.0.

Tabel 1. Pengaruh *tile_size* terhadap waktu dan kualitas

Tile Size	PSNR (dB)	SSIM	Time (s)
256	35.10	0.9560	411.42
128	35.10	0.9560	545.60
64	35.11	0.9561	1206.19

Ukuran tile tidak memengaruhi kualitas, namun berdampak besar pada waktu inferensi. Tile 256 terbukti paling efisien dan optimal. Hal ini mendukung prinsip *tile-aware inference* yang digunakan oleh [7] untuk efisiensi *spatio-temporal* pada VSR.

Pengaruh Parameter *Denoise Strength*

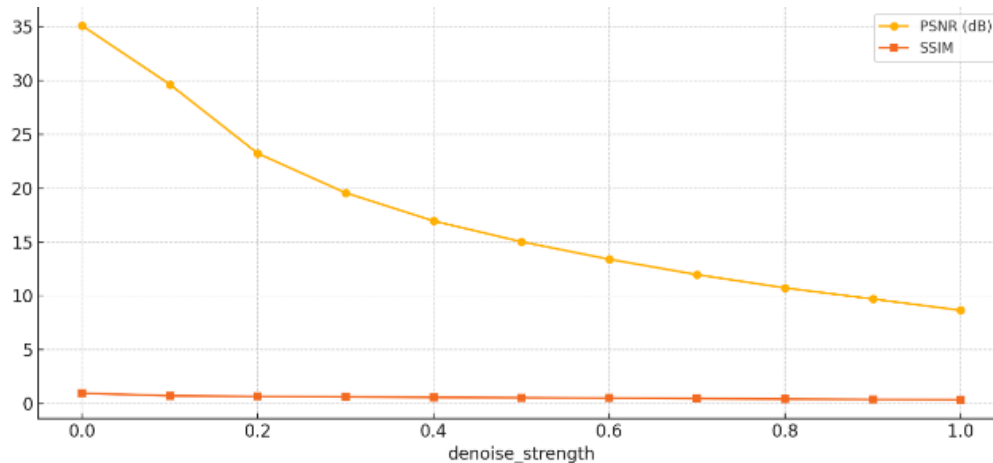
Pengujian *denoise_strength* dilakukan dari 0.0 hingga 1.0 (inkremen 0.1). Hasil menunjukkan bahwa semakin tinggi nilai parameter ini, semakin buruk kualitas hasil *upscaling*.

Tabel 2. Pengaruh *denoise_strength* terhadap performa Real-ESRGAN-X4plus

<i>denoise_strength</i>	PSNR (dB)	SSIM	Waktu (s)
0.0	35.10	0.9560	411.42
0.1	29.64	0.7261	400.90
0.2	23.23	0.6583	396.31
0.3	19.57	0.6186	392.21
0.4	16.96	0.5827	381.67
0.5	15.03	0.5466	376.31
0.6	13.40	0.5080	382.17



0.7	11.97	0.4662	380.45
0.8	10.74	0.4222	372.25
0.9	9.71	0.3806	358.93
1.0	8.66	0.3484	345.32

Gambar 2. Kurva PSNR dan SSIM terhadap *denoise_strength*

PSNR turun drastis dari 35,10 (0.0) ke 8,66 dB (1.0), SSIM dari 0,956 ke 0,348. Hasil ini menunjukkan bahwa Real-ESRGAN terlalu agresif dalam reduksi noise pada nilai tinggi, sejalan dengan temuan [18] yang menyarankan batasan parameterisasi pada input dengan degradasi ringan. Dengan demikian, konfigurasi *denoise_strength* = 0.0 dipastikan paling optimal untuk kondisi video analog seperti VCD.

Evaluasi Objektif (PSNR & SSIM)

Evaluasi objektif dilakukan terhadap 13 konfigurasi Real-ESRGAN dan dua *baseline*. Nilai PSNR dan SSIM dihitung per *frame*, dirata-ratakan per video dan per konfigurasi. Hasil terbaik ditemukan pada konfigurasi *denoise_strength* = 0.0, mendukung validitas pelatihan Real-ESRGAN berbasis degradasi sintesis [15].

Evaluasi Subjektif (*Objective MOS*)

Nilai MOS pada Tabel 3 menunjukkan bahwa kualitas persepsi visual terbaik dicapai pada konfigurasi *denoise_strength* = 0.0, yang selaras dengan karakteristik persepsi manusia. Efektivitas MOS dalam merepresentasikan penilaian subjektif didukung oleh penelitian sebelumnya yang menunjukkan korelasi tinggi antara MOS dan metrik objektif seperti SSIM dan PSNR, terutama dalam menilai kejernihan dan struktur visual [9], [11], [10].

Tabel 3. Nilai MOS terhadap variasi *denoise_strength*

<i>denoise_strength</i>	PSNR (dB)	SSIM	MOS
0.00	35.10	0.9560	5.01
0.10	29.64	0.7261	3.74
0.20	23.23	0.6583	2.84
0.30	19.57	0.6186	2.24
0.40	16.96	0.5827	1.80
0.50	15.03	0.5466	1.45
0.60	13.40	0.5080	1.09
0.70	11.97	0.4662	0.69
0.80	10.74	0.4222	0.34
0.90	9.71	0.3806	0.01
1.00	8.66	0.3484	-0.29

Perbandingan dengan Metode Baseline

Real-ESRGAN dibandingkan dengan *interpolasi bicubic* dan ESRGAN publik.

Tabel 4. Perbandingan performa metode *upsampling*

Method	PSNR (dB)	SSIM
Original Video	21.37	0.643
Bicubic	26.58	0.746



ESRGAN	29.15	0.800
Real-ESRGAN-X4plus	35.10	0.956

Secara visual dan metrik, Real-ESRGAN menghasilkan peningkatan kualitas tertinggi. *Bicubic* sebagai metode resampling matematis gagal merekonstruksi detail tekstur karena tidak mempertimbangkan informasi frekuensi tinggi. Sementara ESRGAN publik lebih baik berkat penggunaan GAN, namun belum dioptimalkan untuk degradasi khas media analog seperti *jitter* atau *blur* khas VCD. Hal ini menjelaskan mengapa Real-ESRGAN unggul, sebagaimana dikonfirmasi [15], karena pelatihannya mencakup degradasi sintesis realistis dunia nyata.

Analisis Statistik

Uji *paired t-test* antara Real-ESRGAN dan kedua *baseline* menunjukkan $p < 0,01$, mengindikasikan peningkatan signifikan secara statistik pada semua metrik (PSNR, SSIM, MOS) di lima video uji.

Diskusi dan Implikasi Budaya

Real-ESRGAN telah menunjukkan efikasi yang signifikan dalam meningkatkan kualitas video melalui teknik resolusi super gambar. Real-ESRGAN terbukti mempertahankan fidelitas visual yang lebih tinggi sekaligus secara efektif mempertahankan detail yang rumit, seperti pola dan arsitektur tradisional, sehingga berkontribusi pada stabilitas antar-bingkai. Peningkatan ini menggarisbawahi kemampuan model dalam pemrosesan spasio-temporal, seperti yang disoroti oleh [2]. Kemampuan Real-ESRGAN untuk berfungsi secara efektif tanpa pelatihan ulang lokal memposisikannya sebagai alat yang berharga untuk pelestarian praktis konten digital budaya. Lebih lanjut, penelitian di masa mendatang sebaiknya mempertimbangkan pelatihan ulang Real-ESRGAN pada dataset lokal dan melibatkan evaluasi auditori bersama penilaian skor opini rata-rata (MOS) subjektif oleh panel independen untuk mencapai pemahaman yang komprehensif tentang kualitas persepsi.

KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa model Real-ESRGAN-X4plus merupakan solusi unggul dalam restorasi video budaya Indonesia berformat VCD yang memiliki degradasi khas seperti *blur*, *jitter*, dan *noise* kompresi. Dengan konfigurasi optimal *denoise_strength* = 0.0 dan *tile_size* = 256, model ini mampu meningkatkan kualitas visual secara signifikan baik secara objektif melalui nilai PSNR (35,10 dB) dan SSIM (0,956), maupun secara perseptual melalui skor MOS sebesar 5,01. Efektivitas Real-ESRGAN melampaui metode *interpolasi bicubic* dan ESRGAN publik yang tidak dirancang untuk data dunia nyata dengan degradasi khas media analog. Konfigurasi ini juga menunjukkan efisiensi inferensi terbaik tanpa mengorbankan kualitas, menjadikannya pilihan ideal untuk restorasi skala menengah dengan sumber daya komputasi terbatas.

Kontribusi orisinal dari penelitian ini terletak pada evaluasi sistematis Real-ESRGAN dalam konteks video budaya Indonesia, yang hingga kini belum banyak dieksplorasi. Selain itu, penelitian ini menyajikan analisis mendalam terhadap pengaruh parameter teknis (*denoise_strength*, *tile_size*) dan penerapan *objective* MOS sebagai pendekatan efisien untuk menilai persepsi kualitas secara kuantitatif. Hasil ini memperkuat temuan studi terdahulu yang menekankan pentingnya adaptasi model terhadap degradasi dunia nyata, serta membuka peluang baru untuk penerapan model pembelajaran mendalam dalam pelestarian arsip media budaya.

Penelitian ini masih memiliki beberapa keterbatasan. Pertama, evaluasi persepsi visual hanya dilakukan secara objektif melalui MOS berbasis rumus, tanpa melibatkan panelis manusia. Kedua, model Real-ESRGAN belum dilatih ulang (*fine-tuned*) dengan dataset lokal video budaya Indonesia, sehingga potensi peningkatan performa masih terbuka. Ketiga, fokus penelitian masih terbatas pada video musik, belum mencakup jenis video budaya lain seperti dokumenter, teater, atau siaran televisi lama. Oleh karena itu, penelitian selanjutnya disarankan untuk melibatkan penilaian subjektif melalui panelis pakar dan awam guna memvalidasi hasil persepsi, serta melakukan *fine-tuning* Real-ESRGAN terhadap dataset budaya lokal untuk meningkatkan adaptasi terhadap degradasi khas. Selain itu, pengujian model VSR berbasis *spatio-temporal* seperti BasicVSR++ atau VRT dapat memperbaiki stabilitas antar *frame* secara lebih optimal. Upaya optimalisasi proses inferensi melalui teknik *mixed precision* dan pemrosesan paralel juga penting untuk memastikan bahwa restorasi dapat dilakukan secara efisien dalam skala besar.

DAFTAR PUSTAKA

- [1] Casad, M., Rieger, O., and Alexander, D., 2015, 'Enduring access to rich media content: Understanding use and usability requirements', D-Lib Magazine, Vol. 21, No. 9/10. doi: 10.1045/september2015-casad.
- [2] Das, A., Paramaj, D., and BR, S., 2024, 'Scalable video fidelity enhancement: Leveraging the state-of-the-art AI models', Scalable Computing: Practice and Experience, Vol. 25, No. 3, hh. 1658–1666. doi: 10.12694/scpe.v25i3.2696.
- [3] Han, H., and Lv, J., 2022, 'Super-resolution-empowered adaptive medical video streaming in telemedicine systems', Electronics, Vol. 11, No. 18, Art. no. 2944. doi: 10.3390/electronics11182944.
- [4] Higaki, T., et al., 2020, 'Deep learning reconstruction at CT: Phantom study of the image characteristics', Academic Radiology, Vol. 27, No. 1, hh. 82–87. doi: 10.1016/j.acra.2019.09.008.
- [5] Hussain, M., Abbas, S., and Irshad, M., 2017, 'Quadratic trigonometric B-spline for image interpolation using GA', PLoS ONE, Vol. 12, No. 6, Art. no. e0179721. doi: 10.1371/journal.pone.0179721.



- [6] Kim, S., Jang, H., Jang, J., Lee, Y., and Hwang, D., 2020, 'Deep-learned short tau inversion recovery imaging using multi-contrast MR images', *Magnetic Resonance in Medicine*, Vol. 84, No. 6, hh. 2994–3008. doi: 10.1002/mrm.28327.
- [7] Liu, H., Zhao, P., Ruan, Z., Shang, F., and Liu, Y., 2021, 'Large motion video super-resolution with dual subnet and multi-stage communicated upsampling', *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35, No. 3, hh. 2127–2135. doi: 10.1609/aaai.v35i3.16310.
- [8] Martínez, H., Hines, A., and Farias, M., 2021, 'Perceptual quality of audio-visual content with common video and audio degradations', *Applied Sciences*, Vol. 11, No. 13, Art. no. 5813. doi: 10.3390/app11135813.
- [9] Mittal, M., Kaushik, R., Verma, A., Kaur, I., Goyal, L., Roy, S., and Kim, T., 2020, 'Image watermarking in curvelet domain using edge surface blocks', *Symmetry*, Vol. 12, No. 5, Art. no. 822. doi: 10.3390/sym12050822.
- [10] Pan, A., Ma, Y., Hu, R., Cao, X., Wu, Y., Zhou, K., and Yu, A., 2023, 'Simultaneous real-time analysis of tear film optical quality dynamics and functional visual acuity in dry eye disease', *Eye and Vision*, Vol. 10, No. 1. doi: 10.1186/s40662-023-00333-6.
- [11] Ponomarenko, N., Марчук, B., and Egiazarian, K., 2022, 'Fivenet: joint image demosaicing, denoising, deblurring, super-resolution, and clarity enhancement', *Electronic Imaging*, Vol. 34, No. 14, hh. 218-1–218-6. doi: 10.2352/ei.2022.34.14.coimg-218.
- [12] Salah, Z., Al-Sit, W., Salah, Z., and Elsoud, E., 2024, 'Spatial domain noise removal filtering for low-resolution digital images', *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 34, No. 3, hh. 1627–1642. doi: 10.11591/ijeecs.v34.i3.pp1627-1642.
- [13] Shukla, A., et al., 2024, 'A novel zero-shot real world spatio-temporal super-resolution (ZS-RW-TSR) model for video super-resolution', *IEEE Access*, Vol. 12, hh. 123969–123984. doi: 10.1109/ACCESS.2024.3406476.
- [14] Tang, X., Ouyang, F., Xu, Y., Zhu, L., and Peng, B., 2024, 'A unified model fusing region of interest detection and super resolution for video compression', *Computational Materials and Continua*, Vol. 79, No. 3, hh. 3955–3975. doi: 10.32604/cmc.2024.049057.
- [15] Wang, S., Politis, A., Mesaros, A., and Virtanen, T., 2022, 'Self-supervised learning of audio representations from audio-visual data using spatial alignment', *IEEE Journal of Selected Topics in Signal Processing*, Vol. 16, No. 6, hh. 1467–1479. doi: 10.1109/JSTSP.2022.3180592.
- [16] Wang, X., Xie, L., Dong, C., and Shan, Y., 2021, 'Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data', *arXiv preprint, arXiv:2107.10833*. [Online]. Available: <https://arxiv.org/abs/2107.10833>
- [17] Winarni, F., 2018, 'Aspek hukum peran serta masyarakat dalam pelestarian cagar budaya', *Mimbar Hukum*, Vol. 30, No. 1, hh. 94–104. doi: 10.22146/jmh.29160.
- [18] Xu, Y., Guo, T., and Wang, C., 2024, 'A remote sensing image super-resolution reconstruction model combining multiple attention mechanisms', *Sensors*, Vol. 24, No. 14, Art. no. 4492. doi: 10.3390/s24144492.
- [19] Yao, H., et al., 2019, 'DR2-Net: Deep residual reconstruction network for image compressive sensing', *Neurocomputing*, Vol. 359, hh. 483–493. doi: 10.1016/j.neucom.2019.05.006.
- [20] Zerva, M., and Kondi, L., 2024, 'Video super-resolution using plug-and-play priors', *IEEE Access*, Vol. 12, hh. 11963–11971. doi: 10.1109/ACCESS.2024.3355195.