

PROPOSAL TESIS

**UPSCALING MUSIC VIDEO LAWAS INDONESIA BERBASIS FRAMES
VIDEO DENGAN MENGGUNAKAN ALGORITMA ESRGAN**



Disusun oleh:

Nama : ABDUL RA'UF ALFANSANI
NIM : 22.55.2307
Konsentrasi : Intelligence Animation

**PROGRAM STUDI S2 PJJ INFORMATIKA
PROGRAM PASCASARJANA
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

**UPSCALING MUSIC VIDEO LAWAS INDONESIA BERBASIS FRAMES VIDEO
DENGAN MENGGUNAKAN ALGORITMA ESRGAN**

**UPSCALING OLD INDONESIAN MUSIC VIDEO BASED ON FRAMES VIDEO USING
ESRGAN**

Dipersiapkan dan Disusun oleh

ABDUL RA'UF ALFANSANI

22.55.2307

Telah disetujui oleh Tim Dosen Pembimbing Tesis
pada tanggal 05 Februari 2025

Pembimbing Utama

Prof. Dr. Ema Utami, S.Si., M.Kom
NIK. 190302037

Pembimbing Pendamping

Dhani Ariatmanto, M.Kom., Ph.D.
NIK. 190302197

1. LATAR BELAKANG MASALAH

Dalam konteks pemrosesan citra dan video, teori super-resolusi telah berkembang menjadi komponen krusial. Super-resolusi, yang bertujuan meningkatkan resolusi citra atau video, memanfaatkan berbagai teknik komputasi untuk menghasilkan hasil yang lebih tajam dan jelas dari sumber beresolusi rendah (Wang et al, 2021). Evolusi dalam pembelajaran mendalam dan jaringan saraf tiruan telah membuka jalan bagi kemajuan besar dalam bidang ini. Metode ini mampu memperbaiki kualitas visual dengan cara yang tidak dapat dicapai oleh teknik tradisional, seperti interpolasi dan metode konvensional lainnya (Zhang et al., 2021). Perkembangan teknologi ini penting, mengingat pentingnya konten visual dalam berbagai aspek kehidupan, dari komunikasi hingga pemeliharaan warisan budaya (Ji et al., 2020). Khususnya, dalam konteks video musik lawas, teknologi ini menawarkan solusi untuk mengatasi keterbatasan teknologi perekaman masa lalu (Xie et al., 2021). Dengan demikian, penelitian dalam super-resolusi tidak hanya penting secara teknis tetapi juga memiliki dampak sosial dan budaya yang signifikan (Mou et al., 2022).

Penelitian ini secara spesifik akan memfokuskan pada penggunaan model RealESRGAN-X4plus, varian terbaru dari Real-ESRGAN, untuk upscaling video musik lawas Indonesia. RealESRGAN-X4plus merupakan pengembangan terkini dalam ranah super-resolusi, yang menggunakan teknik pembelajaran mendalam untuk meningkatkan kualitas citra dan video beresolusi rendah (Wang et al., 2021). Dengan memanfaatkan kecanggihan model pre-trained ini, penelitian bertujuan untuk mengatasi tantangan kualitas visual yang ada pada video musik lawas, seperti resolusi rendah, kejernihan yang buruk, dan warna yang pudar. RealESRGAN-X4plus, melalui pendekatannya yang inovatif dalam generative adversarial networks (GAN), memiliki potensi untuk menghasilkan citra yang lebih tajam dan detail dari konten beresolusi rendah yang ada pada video lawas, menjadikannya pilihan yang cocok untuk penelitian ini (Wang et al., 2021).

Implementasi RealESRGAN-X4plus pada video musik lawas akan melibatkan serangkaian proses evaluasi dan analisis. Proses ini akan dimulai dengan pemilihan video musik yang mewakili berbagai era dan gaya dalam sejarah

musik Indonesia, yang kemudian akan diolah melalui RealESRGAN-X4plus. Kualitas upscaling akan dievaluasi berdasarkan kriteria teknis dan estetika, termasuk perbaikan resolusi, kejernihan visual, dan pemeliharaan integritas artistik konten asli. Evaluasi ini tidak hanya akan memberikan wawasan tentang kinerja teknis RealESRGAN-X4plus tetapi juga akan menguji kemampuannya dalam mempertahankan karakteristik unik dari video musik lawas. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam upaya pelestarian warisan budaya Indonesia melalui teknologi pemrosesan citra canggih (Zhang et al., 2021; Ji et al., 2020).

Penelitian oleh Wang et al. tentang "Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data" memberikan dasar bagi penggunaan RealESRGAN dalam konteks dunia nyata (Wang et al., 2021). Penelitian ini menekankan pada pentingnya pelatihan model super-resolusi menggunakan data sintetis untuk menghasilkan hasil yang lebih aplikatif pada citra dunia nyata. Hal ini relevan untuk penelitian ini karena video musik lawas sering mengalami degradasi yang mirip dengan contoh-contoh dalam studi tersebut. Sementara itu, Liu dan Wang dalam "Random Degradation Model for Deep Blind Image Super-Resolution" mengeksplorasi model degradasi acak, yang memberikan pemahaman tentang bagaimana berbagai jenis degradasi dapat diatasi dalam super-resolusi (Liu & Wang; 2022). Ini penting karena video lawas mungkin mengalami berbagai bentuk degradasi yang tidak terduga.

Pada "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks," Wang et al. memperkenalkan ESRGAN, model yang meningkatkan kejelasan dan detail citra melalui jaringan adversarial generatif (Wang et al., 2021). Model ini menjadi landasan bagi RealESRGAN-X4plus dan studi ini akan memanfaatkan temuan mereka untuk memahami cara terbaik mengadaptasi model untuk video musik lawas. Selain itu, Zhang et al. dalam "Designing a Practical Degradation Model for Deep Blind Image Super-Resolution" memberikan wawasan tentang pentingnya mendesain model degradasi yang praktis untuk citra super-resolusi yang buta (Zhang et al., 2021). Ini memberikan panduan tentang cara mengatasi tantangan spesifik dalam upscaling video lawas.

Penelitian lebih lanjut, seperti "Blind Image Super-Resolution: A Survey and Beyond" oleh Liu et al., memberikan gambaran umum tentang perkembangan terkini di bidang super-resolusi yang buta dan aplikasinya pada berbagai jenis citra (Liu et al., 2021). Penggunaan teknik-teknik ini dalam konteks pelestarian warisan budaya, seperti video musik lawas, menunjukkan bagaimana teknologi dapat digunakan untuk memperbarui dan memelihara aset budaya penting. Ini menunjukkan bagaimana penelitian ini tidak hanya bertujuan untuk kemajuan teknis tetapi juga untuk pelestarian budaya.

Salah satu tantangan utama dalam penelitian ini adalah mengadaptasi teknologi super-resolusi, yang biasanya difokuskan pada citra, untuk upscaling video. Video musik lawas sering kali memiliki masalah kualitas yang lebih kompleks daripada citra, termasuk degradasi yang disebabkan oleh proses penyimpanan dan pengolahan sebelumnya (Wang et al., 2021). Selain itu, mempertahankan keaslian konten sambil meningkatkan resolusi adalah masalah yang rumit. Algoritma harus mampu meningkatkan kualitas visual tanpa mengubah konten asli secara substansial (Zhang et al., 2021; Liu & Wang, 2022). Masalah lain adalah variabilitas degradasi pada video lawas, yang memerlukan pendekatan super-resolusi yang fleksibel dan mampu menyesuaikan dengan berbagai jenis degradasi (Gu et al., 2019; Son et al., 2021). Akhirnya, penting untuk memahami bagaimana peningkatan resolusi dapat memengaruhi persepsi visual dan estetika dari video musik lawas, yang memiliki nilai historis dan budaya (Kim et al., 2020; Ko et al., 2021).

2. RUMUSAN MASALAH

1. Bagaimana batasan waktu optimal yang dapat dicapai dengan menggunakan algoritma RealESRGAN-X4plus untuk meningkatkan kualitas video musik lawas Indonesia?
2. Sejauh mana kualitas resolusi dan kejernihan visual yang dapat dicapai dengan algoritma RealESRGAN-X4plus pada video musik lawas?

3. Bagaimana mengukur kemiripan antara video asli dan hasil upscaling menggunakan *metrik Peak Signal-to-Noise Ratio (PSNR)* dan *Structural Similarity Index Measure (SSIM)*?

3. BATASAN MASALAH

1. Penelitian ini hanya menggunakan algoritma RealESRGAN-X4plus sebagai metode upscaling video.
2. Video yang diuji dalam penelitian ini terbatas pada video musik lawas Indonesia dengan format tertentu.
3. Evaluasi kualitas hasil upscaling dilakukan dengan metrik PSNR dan SSIM, tanpa mempertimbangkan aspek subjektif seperti preferensi visual pengguna.
4. Penelitian difokuskan pada efisiensi waktu pemrosesan dan peningkatan kualitas visual, tanpa mencakup aspek lain seperti perbaikan audio atau metadata.

4. TUJUAN PENELITIAN

1. Menentukan batasan waktu optimal dalam proses upscaling video musik lawas Indonesia menggunakan algoritma RealESRGAN-X4plus untuk mencapai efisiensi tanpa mengorbankan kualitas hasil.
2. Mengevaluasi sejauh mana algoritma RealESRGAN-X4plus dapat meningkatkan resolusi dan kejernihan visual video musik lawas Indonesia.
3. Mengukur kemiripan antara video asli dan hasil upscaling menggunakan metrik kuantitatif PSNR dan SSIM guna memberikan penilaian objektif terhadap efektivitas teknik upscaling.

5. MANFAAT PENELITIAN

1. Memberikan wawasan bagi peneliti dan pengembang teknologi upscaling mengenai batasan waktu dan kualitas optimal dari algoritma RealESRGAN-X4plus.

2. Membantu industri kreatif dalam proses restorasi dan peningkatan kualitas video musik lawas Indonesia agar dapat digunakan kembali dalam format digital yang lebih baik.
3. Menyediakan panduan evaluasi kuantitatif yang dapat diterapkan dalam penelitian dan pengembangan metode upscaling video lainnya.
4. Mendorong penggunaan teknologi modern untuk melestarikan dan meningkatkan aset budaya berupa video musik lawas Indonesia.

6. TINJAUAN PUSTAKA

Super-resolution adalah bidang penting dalam pemrosesan citra dan video, yang bertujuan untuk meningkatkan resolusi dan kualitas visual konten berbasis sumber low-resolution. Penelitian oleh Wang et al. (2021) memperkenalkan Real-ESRGAN, sebuah model berbasis generative adversarial networks (GAN) yang dilatih menggunakan data sintetis untuk menangani degradasi pada citra dunia nyata (Wang et al., 2021). Zhang et al. (2021) juga mengembangkan model degradasi praktis untuk mendukung deep blind image super-resolution, memberikan kemampuan adaptif terhadap degradasi kompleks pada citra (Zhang et al., 2021).

Model ESRGAN pertama kali diperkenalkan oleh Wang et al. sebagai pengembangan dari SRGAN, dengan fokus pada peningkatan detail tekstur dan kejelasan visual (Wang et al., 2021). Teknologi ini kemudian disempurnakan menjadi Real-ESRGAN yang menawarkan performa lebih baik dengan pelatihan berbasis degradasi sintetis, memungkinkan aplikasi yang lebih luas pada data dunia nyata (Wang et al., 2021). Model terbaru, RealESRGAN-X4plus, menunjukkan peningkatan yang signifikan dalam menangani data visual kompleks, membuatnya potensial untuk diadaptasi dalam penelitian ini.

Penggunaan teknologi super-resolusi dalam pelestarian budaya telah menjadi topik penting, meskipun hingga saat ini masih berfokus pada citra digital atau video umum. Penelitian oleh Ji et al. (2020) menyoroti pentingnya injeksi noise dan estimasi kernel dalam memperbaiki kualitas citra dunia nyata tanpa mengorbankan keaslian visual (Ji et al., 2020). Dalam konteks degradasi kompleks, Xie et al. (2021) mengembangkan filter diskriminatif untuk

mengatasi degradasi spesifik, yang relevan dengan kondisi video musik lawas (Xie et al., 2021).

Penanganan degradasi yang kompleks pada video lawas membutuhkan pendekatan yang fleksibel dan adaptif. Liu dan Wang (2022) memperkenalkan random degradation model, yang memberikan solusi untuk variasi degradasi yang sering dijumpai pada data dunia nyata (Liu et al., 2022). Son et al. (2021) juga mengembangkan adaptive downsampling untuk meningkatkan efisiensi proses super-resolusi pada data dengan degradasi heterogen (Son et al., 2021).

Hingga saat ini, penelitian-penelitian sebelumnya belum mengaplikasikan teknologi super-resolusi pada video musik lawas, terutama dalam konteks pelestarian konten visual historis. Sebagian besar penelitian masih berfokus pada pengolahan citra atau video umum tanpa mempertimbangkan nilai artistik, estetika, dan keaslian konten budaya. Penelitian ini menawarkan pendekatan baru dengan mengadaptasi RealESRGAN-X4plus untuk upscaling video musik lawas Indonesia. Fokus pada peningkatan resolusi, kejernihan visual, dan pelestarian karakteristik artistik asli dari video musik lawas adalah kontribusi unik yang diusulkan dalam tesis ini.

Evaluasi kualitas *upscaling* akan menggunakan metrik teknis seperti PSNR dan SSIM untuk menilai kejelasan visual, serta analisis subjektif untuk mengukur persepsi estetika. Penelitian oleh Baghel et al. (2023) menyoroti pentingnya penggunaan arsitektur Transformer dalam jaringan adversarial generatif untuk meningkatkan kualitas gambar hasil *super-resolution*. Pendekatan ini memungkinkan model untuk menghasilkan gambar dengan detail yang lebih baik dan distorsi minimal, yang diukur melalui peningkatan signifikan pada metrik PSNR dan SSIM. Kim et al. (2020) juga menunjukkan bagaimana dual back-projection dapat digunakan untuk meningkatkan efisiensi proses pembelajaran internal dalam super-resolusi buta (Kim et al., 2020).

Dengan mengaplikasikan teknologi super-resolusi pada video musik lawas, penelitian ini tidak hanya berkontribusi pada pengembangan teknologi tetapi juga mendukung pelestarian warisan budaya Indonesia. Peningkatan kualitas visual video musik lawas memungkinkan apresiasi yang lebih luas terhadap

karya seni dari masa lalu, sekaligus menjadikannya relevan dalam konteks budaya modern.

7. LANDASAN TEORI

7.1. Super-Resolusi (Super-Resolution)

Super-resolusi merupakan salah satu teknik dalam bidang pemrosesan citra dan video yang bertujuan untuk meningkatkan resolusi spasial dari konten beresolusi rendah menjadi resolusi tinggi. Teknik ini memungkinkan citra atau video menjadi lebih tajam dan kaya detail, yang mendekati kualitas konten dengan resolusi asli yang lebih tinggi. Dalam prosesnya, super-resolusi tidak hanya memperbesar ukuran gambar atau video secara fisik, tetapi juga memperkaya detail visual yang mungkin hilang dalam proses degradasi. Dengan berkembangnya teknologi pembelajaran mesin, khususnya *deep learning*, metode super-resolusi telah mengalami transformasi besar, menggantikan pendekatan tradisional dengan solusi yang lebih presisi dan realistis (Wang et al., 2021).

Super-resolusi dapat diterapkan pada berbagai jenis data visual, termasuk gambar tunggal dan video. Untuk gambar tunggal, proses ini dikenal sebagai *Single Image Super-Resolution (SISR)*, di mana peningkatan resolusi hanya dilakukan pada satu frame atau citra. Teknik ini telah banyak digunakan dalam berbagai aplikasi seperti pengenalan wajah, pemrosesan medis, dan pengawasan keamanan. Di sisi lain, penerapan super-resolusi pada video, yang dikenal sebagai *Video Super-Resolution (VSR)*, menghadapi tantangan tambahan, yaitu menjaga konsistensi temporal antar frame. Selain meningkatkan kualitas tiap frame secara individual, metode VSR harus mampu mengurangi artefak visual yang dapat mengganggu aliran video secara keseluruhan (Liu dan Wang, 2022).

Sebelum teknologi berbasis jaringan saraf tiruan berkembang, super-resolusi dilakukan dengan metode konvensional seperti interpolasi dan rekonstruksi berbasis model matematis. Metode interpolasi, termasuk interpolasi bilinear dan bicubic, memanfaatkan nilai rata-rata dari piksel

tetangga untuk memperkirakan nilai piksel baru dalam proses pembesaran citra. Meski metode ini cepat dan sederhana, hasilnya sering kali kurang tajam dan tidak mampu merekonstruksi detail halus yang hilang. Di sisi lain, pendekatan berbasis rekonstruksi menggunakan model degradasi tertentu untuk memprediksi citra resolusi tinggi dari citra resolusi rendah. Meskipun lebih canggih, pendekatan ini sering kali terbatas pada asumsi degradasi tertentu yang tidak selalu sesuai dengan data dunia nyata (Ji et al., 2020).

Kemajuan dalam super-resolusi, terutama dengan menggunakan metode berbasis pembelajaran mendalam seperti ESRGAN (*Enhanced Super-Resolution Generative Adversarial Networks*), telah memungkinkan peningkatan kualitas citra yang jauh lebih signifikan dibandingkan metode tradisional. ESRGAN dan varian lanjutannya, RealESRGAN, menggunakan pendekatan *generative adversarial networks* (GAN) untuk menghasilkan citra atau video dengan detail tinggi yang realistis, sehingga lebih relevan dalam aplikasi dunia nyata (Wang et al., 2021; Zhang et al., 2021).



Gambar 1. Hasil super-resolusi dengan metode Bicubic dan ESRGAN. Sumber: Wang et al., 2021.

7.2. Generative Adversarial Networks (GAN)

Generative Adversarial Networks (GAN) merupakan salah satu inovasi utama dalam pembelajaran mesin yang diperkenalkan oleh Ian Goodfellow

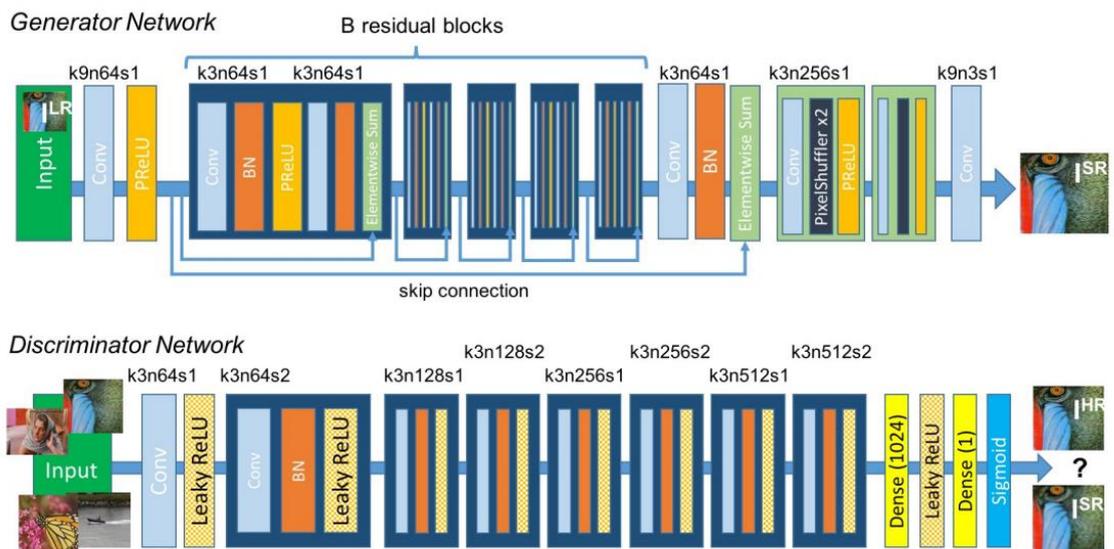
pada tahun 2014. GAN dirancang sebagai framework pembelajaran mesin yang melibatkan dua jaringan saraf, yaitu *Generator* dan *Discriminator*, yang dilatih secara bersamaan dalam suatu sistem adversarial. Tujuan utama GAN adalah menghasilkan data yang terlihat realistis dengan cara mensimulasikan distribusi data yang mendekati data asli. GAN terdiri dari dua komponen utama:

1. Generator

Generator bertugas untuk menghasilkan data baru yang mirip dengan data asli. Dalam konteks gambar, Generator akan menghasilkan gambar-gambar baru yang secara visual menyerupai data pelatihan. Jaringan ini mencoba untuk "menipu" Discriminator dengan menciptakan data yang realistis.

2. Discriminator

Discriminator berfungsi sebagai pengklasifikasi yang membedakan antara data nyata (data asli) dan data palsu (yang dihasilkan oleh Generator). Tujuan Discriminator adalah untuk memberikan umpan balik kepada Generator tentang seberapa realistis data yang dihasilkannya.



Gambar 2. *Arsitektur Generative Adversarial Network (GAN)* yang digunakan dalam SRGAN (Super-Resolution Generative Adversarial Network). Diagram ini menunjukkan hubungan antara *Generator* dan *Discriminator*, di mana *Generator* berfungsi untuk menghasilkan citra

super-resolusi (*SR*), sementara Discriminator bertugas membedakan antara citra asli (*HR*) dan citra yang dihasilkan (*SR*). Sumber: Ledig et al., 2017.

Proses pelatihan Generative Adversarial Networks (GAN) bersifat adversarial, di mana dua jaringan saraf tiruan—Generator dan Discriminator—saling berkompetisi. Generator berusaha menghasilkan data yang menyerupai data asli untuk menipu Discriminator, sementara Discriminator dilatih untuk membedakan antara data asli dan data yang dihasilkan oleh Generator. Selama pelatihan, kedua jaringan ini saling meningkatkan kinerjanya melalui fungsi kehilangan (*loss function*) yang dirancang untuk menjaga keseimbangan antara keduanya. Pendekatan ini memungkinkan GAN untuk menghasilkan data yang semakin realistis seiring waktu (Chen et al., 2021).

Penerapan GAN dalam Super-Resolusi

GAN telah diterapkan secara luas dalam bidang super-resolusi, di mana model ini mampu memperbaiki detail visual gambar beresolusi rendah menjadi gambar beresolusi tinggi. Pendekatan ini lebih efektif dibandingkan metode tradisional karena mampu mempertahankan detail tekstur dan meningkatkan kualitas visual. (Wang et al., 2021).

1. SRGAN (Super-Resolution GAN)

SRGAN merupakan salah satu implementasi awal GAN dalam super-resolusi. Model ini memperkenalkan penggunaan fungsi kehilangan berbasis persepsi (*perceptual loss*), yang memanfaatkan fitur yang diekstrak dari jaringan pra-pelatihan, seperti VGGNet. SRGAN menghasilkan gambar super-resolusi yang tidak hanya tajam tetapi juga lebih realistis dibandingkan metode sebelumnya seperti interpolasi. (Wang et al., 2021).

2. ESRGAN (Enhanced Super-Resolution GAN)

ESRGAN adalah pengembangan lebih lanjut dari SRGAN. Model ini memperbaiki kelemahan SRGAN dengan menambahkan fungsi kehilangan Relativistic Discriminator, yang membantu Generator menghasilkan gambar yang lebih realistis. ESRGAN juga berfokus pada peningkatan

detail visual dan tekstur, sehingga lebih cocok untuk tugas yang memerlukan kualitas visual tinggi. (Wang et al., 2018; Wang et al., 2021).

3. Real-ESRGAN

Real-ESRGAN adalah inovasi terbaru dalam super-resolusi berbasis GAN yang dirancang untuk mengatasi tantangan dunia nyata. Model ini memperkenalkan metode degradasi sintetis, yang memungkinkan pelatihan dengan berbagai jenis degradasi gambar, seperti noise, blur, dan kompresi. Pendekatan ini membuat Real-ESRGAN lebih tangguh dan aplikatif dalam berbagai skenario dunia nyata. (Wang et al., 2021; Zhang et al., 2021)

Salah satu varian penting dari Real-ESRGAN adalah **RealESRGAN-X4plus**, yang dirancang khusus untuk tugas-tugas yang memerlukan peningkatan resolusi gambar atau video dengan tingkat detail tinggi. Model ini menggunakan arsitektur Generator dan Discriminator yang ditingkatkan untuk menghasilkan hasil yang lebih tajam dan realistis(Wang et al., 2021). Dalam penelitian ini, RealESRGAN-X4plus digunakan untuk upscaling video musik lawas, mengatasi tantangan seperti resolusi rendah, warna yang pudar, dan artefak visual lainnya.

4. Perbedaan Utama Antara SRGAN, ESRGAN, dan Real-ESRGAN

SRGAN adalah implementasi awal dari GAN untuk super-resolusi. Pendekatan ini memperkenalkan fungsi perceptual loss, yang memungkinkan model menghasilkan gambar super-resolusi yang lebih tajam dan realistis dibandingkan metode interpolasi tradisional. Namun, SRGAN masih memiliki keterbatasan dalam menangkap detail halus, terutama pada gambar dengan tekstur kompleks.

ESRGAN merupakan pengembangan dari SRGAN yang lebih fokus pada detail visual. Dengan Residual-in-Residual Dense Block (RRDB) dan Relativistic Discriminator, ESRGAN mampu menghasilkan gambar dengan tekstur yang lebih baik dan detail visual yang realistis. ESRGAN lebih stabil dalam pelatihan dan menghasilkan gambar yang lebih tajam dibandingkan SRGAN.

Real-ESRGAN melangkah lebih jauh dengan mengatasi degradasi dunia nyata, seperti noise, blur, dan kompresi. Model ini dilatih dengan degradasi sintesis, membuatnya lebih tangguh dan aplikatif dalam kondisi dunia nyata. Real-ESRGAN juga memiliki varian seperti RealESRGAN-X4plus, yang dirancang khusus untuk menghasilkan gambar atau video dengan resolusi sangat tinggi dan detail yang lebih tajam, menjadikannya pilihan yang optimal untuk tugas upscaling video dengan berbagai tantangan visual.

7.3. Pengolahan Video Digital

Pengolahan video digital merupakan salah satu bidang yang terus berkembang di dunia teknologi informasi. Bidang ini berfokus pada manipulasi, analisis, dan pemrosesan video digital untuk berbagai tujuan, seperti kompresi, peningkatan kualitas, transmisi, hingga analisis data visual. Video digital, pada dasarnya, terdiri atas serangkaian gambar diam yang disebut *frame* yang ditampilkan secara berurutan dalam interval tertentu. Dengan memanfaatkan prinsip dasar dan teknologi mutakhir, pengolahan video digital memungkinkan peningkatan kualitas visual, efisiensi penyimpanan, dan pemrosesan yang lebih cepat. (Moeslund, 2021).

Prinsip Dasar Video Digital

Video digital dikarakterisasi oleh berbagai parameter utama yang memengaruhi kualitas visual, efisiensi, dan pengalaman pengguna. Parameter-parameter ini meliputi:

1. Frame Rate:

Frame rate menunjukkan jumlah gambar atau *frame* yang ditampilkan per detik dalam sebuah video, yang biasa dinyatakan dalam satuan frame per second (fps). Misalnya, video dengan *frame rate* 24 fps berarti ada 24 gambar yang ditampilkan setiap detik. Semakin tinggi *frame rate*, semakin halus gerakan yang terlihat dalam video. Standar umum seperti 24 fps digunakan dalam film, sedangkan video gim dan konten realitas

virtual sering kali menggunakan *frame rate* 60 fps atau lebih untuk pengalaman yang lebih responsif dan realistis. (Moeslund, 2021).

2. Resolusi:

Resolusi mengacu pada jumlah piksel yang membentuk gambar dalam setiap frame video. Resolusi rendah seperti 480p (720×480 piksel) sering digunakan dalam konten lama atau siaran televisi tradisional, sedangkan resolusi modern seperti 1080p (Full HD) atau 4K menawarkan kualitas visual yang jauh lebih tajam dan detail. Resolusi lebih tinggi memungkinkan video menampilkan detail yang lebih halus, tetapi menuntut kapasitas penyimpanan yang lebih besar dan bandwidth lebih tinggi untuk transmisi. (Blokdyk, 2020).

3. Codec:

Codec (compressor-decompressor) adalah teknologi inti dalam pengolahan video digital, memungkinkan video dikompresi untuk menghemat ruang penyimpanan dan memudahkan transmisi. Codec seperti H.264, H.265 (HEVC), dan VP9 banyak digunakan dalam aplikasi modern karena efisiensinya. Codec bertugas memastikan bahwa video tetap dapat diakses dengan kualitas tinggi tanpa memakan ruang penyimpanan yang berlebihan. (Moeslund, 2021).

4. Format Video:

Format video seperti MP4, AVI, dan MKV menentukan cara data video dan audio dikemas untuk penyimpanan dan distribusi. Setiap format memiliki kelebihan masing-masing, seperti kompatibilitas, kualitas, dan fleksibilitas. (Moeslund, 2021).

Struktur Frame dalam Video

Video digital tidak hanya terdiri atas serangkaian gambar statis, tetapi memiliki struktur yang memungkinkan pengolahan yang efisien. Struktur frame dirancang untuk mendukung kompresi data tanpa terlalu mengorbankan kualitas visual. Frame dalam video dibagi menjadi tiga jenis utama:

- I-frame (Intra-coded frame): Merupakan frame mandiri yang menyimpan informasi lengkap tentang gambar. I-frame biasanya digunakan sebagai titik referensi utama untuk P-frame dan B-frame.
- P-frame (Predicted frame): Berisi informasi perubahan dari frame sebelumnya, yang biasanya berupa I-frame atau P-frame. P-frame lebih hemat dalam penggunaan data karena hanya menyimpan perbedaan visual antar frame.
- B-frame (Bidirectional frame): Memanfaatkan informasi dari frame sebelum dan sesudahnya untuk menghasilkan gambar yang lebih akurat dan detail.

Struktur ini memungkinkan video digital dikompresi secara signifikan, menghasilkan ukuran file yang lebih kecil tanpa mengorbankan pengalaman visual pengguna. (Moeslund, 2021).

Proses Upscaling Video

Proses upscaling video adalah upaya untuk meningkatkan resolusi video asli ke resolusi yang lebih tinggi. Upscaling menjadi penting dalam berbagai aplikasi, termasuk restorasi video lama, produksi film, dan adaptasi konten untuk layar dengan resolusi lebih tinggi. Proses ini melibatkan peningkatan jumlah piksel dalam setiap frame video, biasanya dengan memprediksi nilai piksel tambahan berdasarkan data yang sudah ada. (Blokdyk, 2020).

Pendekatan tradisional untuk upscaling video melibatkan metode interpolasi, seperti bilinear dan bicubic. Metode ini menggunakan algoritma matematis untuk memperkirakan nilai piksel baru berdasarkan nilai piksel tetangga. Namun, metode ini memiliki keterbatasan dalam menghasilkan detail tekstur dan sering kali menghasilkan gambar yang tampak buram atau kurang tajam. (Szeliski, 2022).

Dalam beberapa tahun terakhir, metode berbasis *deep learning* telah mengubah pendekatan upscaling video secara signifikan. Teknologi seperti ESRGAN (Enhanced Super-Resolution Generative Adversarial Network) dan Real-ESRGAN menawarkan hasil yang jauh lebih baik dengan memanfaatkan

kemampuan jaringan saraf untuk mempelajari pola-pola kompleks dalam data visual. ESRGAN, misalnya, mampu menghasilkan detail tekstur yang realistis dan konsisten, bahkan pada resolusi tinggi. (Wang et al., 2021).

Namun, tantangan utama dalam upscaling video bukan hanya meningkatkan kualitas setiap frame individu, tetapi juga menjaga *temporal consistency* atau konsistensi temporal antar frame. Ketidakkonsistenan antar frame, seperti flickering atau perubahan tekstur yang tidak konsisten, dapat merusak pengalaman menonton video. Oleh karena itu, metode modern sering kali memproses beberapa frame secara bersamaan, mempertimbangkan hubungan temporal untuk menghasilkan hasil yang lebih mulus dan konsisten. (Moeslund, 2021; Blokdyk, 2020).

Dengan pemahaman mendalam tentang prinsip dasar, struktur frame, dan proses upscaling, pengolahan video digital menjadi elemen penting dalam pengembangan teknologi multimedia. Aplikasi teknologi ini tidak hanya terbatas pada dunia hiburan, tetapi juga mencakup bidang medis, keamanan, hingga analisis data berbasis video.

7.4. Model Degradasi pada Citra dan Video

Model degradasi pada citra dan video merujuk pada berbagai jenis gangguan yang dapat menurunkan kualitas visual suatu media. Gangguan ini dapat berupa noise, blur, artefak kompresi, atau penurunan kualitas warna, dan sangat memengaruhi pengalaman visual serta interpretasi informasi (Blokdyk, 2020; Moeslund, 2021). Dalam konteks video musik lawas, degradasi sering kali lebih kompleks karena melibatkan kombinasi dari berbagai jenis gangguan akibat teknologi perekaman dan penyimpanan yang sudah usang. Memahami jenis degradasi ini penting untuk pengembangan metode super-resolusi yang mampu memulihkan kualitas visual secara efektif.

Jenis-Jenis Degradasi

Degradasi pada citra dan video dapat dikelompokkan menjadi beberapa jenis utama:

1. Noise

Noise adalah gangguan acak yang sering kali muncul dalam bentuk piksel yang tidak sesuai dengan pola atau warna sekitarnya (Wang & Bovik, 2009). Noise dapat dihasilkan oleh berbagai faktor seperti sensor kamera, kompresi video, atau transmisi sinyal.

- Pada citra, noise sering muncul dalam bentuk *Gaussian noise*, *salt-and-pepper noise*, atau *Poisson noise*.
- Pada video musik lawas, noise sering kali terjadi akibat konversi format dari analog ke digital, atau karena degradasi fisik pada pita rekaman.

2. Blur

Blur merujuk pada hilangnya detail akibat penyebaran intensitas cahaya pada piksel. Blur dapat disebabkan oleh gerakan kamera (motion blur), ketidakfokusan lensa (out-of-focus blur), atau proses pengolahan seperti kompresi (Szeliski, 2022). Blur pada video musik lawas sering kali diperburuk oleh ketidaksempurnaan perangkat perekam yang digunakan pada zamannya (Moeslund, 2021).

3. Artefak Kompresi

Artefak kompresi adalah gangguan visual yang muncul akibat proses kompresi data video atau citra untuk mengurangi ukuran file. Kompresi berbasis blok, seperti yang digunakan dalam format JPEG dan MPEG, sering kali menghasilkan gangguan seperti blok-blok yang terlihat jelas atau garis kasar di sekitar objek (Blokdyk, 2020).

4. Warna yang Pudar

Warna yang pudar merupakan masalah umum pada video musik lawas. Hal ini terjadi karena degradasi kimiawi pada pita magnetik atau hilangnya informasi warna selama proses konversi analog ke digital (Moeslund, 2021).

5. Degradasi Kompleks pada Video Musik Lawas

Video musik lawas sering kali menunjukkan degradasi yang bersifat kompleks dan gabungan dari berbagai jenis gangguan. Misalnya, noise

akibat usia media, blur karena pergerakan kamera manual, serta artefak yang muncul selama proses remastering (Szeliski, 2022). Kompleksitas ini menjadi tantangan utama dalam pengembangan model super-resolusi yang efektif.

Pendekatan Model Degradasi dalam Super-Resolusi

Untuk mengatasi degradasi yang terjadi, model super-resolusi sering kali menggunakan pendekatan yang dirancang untuk mensimulasikan dan memahami pola degradasi. Pendekatan ini melibatkan dua metode utama:

1. Random Degradation Model

Random degradation model adalah pendekatan yang mensimulasikan degradasi acak untuk melatih model super-resolusi. Dengan menciptakan variasi degradasi pada data pelatihan, model dapat mempelajari bagaimana mengatasi berbagai jenis degradasi secara lebih umum.

- Contoh penerapan model ini dapat ditemukan dalam penelitian seperti *Real-ESRGAN* (Wang et al., 2021), yang menggunakan degradasi acak untuk meningkatkan performa model di berbagai kondisi dunia nyata.

2. Kernel Estimation

Kernel estimation adalah metode untuk memperkirakan kernel degradasi yang mewakili pola gangguan pada citra atau video. Kernel ini kemudian digunakan untuk membalikkan proses degradasi dan memulihkan kualitas visual.

- Teknik ini sangat efektif untuk menangani blur, di mana kernel blur dapat diperkirakan berdasarkan analisis frekuensi atau spasial. Penelitian seperti *Blind Image Super-Resolution* oleh Gu et al. (2019) menggunakan pendekatan ini untuk meningkatkan ketajaman gambar hasil super-resolusi.

Model Degradasi pada Video Musik Lawas

Video musik lawas memerlukan pendekatan khusus dalam model degradasi karena sifat gangguan yang kompleks. Model degradasi konvensional sering kali tidak cukup untuk menangani kombinasi gangguan seperti noise, blur, dan artefak warna (Moeslund, 2021). Oleh karena itu, penelitian terbaru cenderung mengembangkan pendekatan yang lebih canggih, seperti menggabungkan random degradation model dengan kernel estimation untuk menciptakan model yang lebih adaptif.

Penelitian seperti *Real-ESRGAN* juga menyoroti pentingnya menciptakan dataset degradasi sintetis yang mencerminkan karakteristik degradasi dunia nyata (Wang et al., 2021). Hal ini memungkinkan model untuk mempelajari pola degradasi spesifik pada video musik lawas, sehingga mampu memberikan hasil yang lebih realistis dan memuaskan.

7.5. Evaluasi Kualitas Super-Resolusi

Super-resolusi, baik pada citra maupun video, tidak hanya bertujuan untuk meningkatkan resolusi spasial tetapi juga menjaga kualitas visual yang sesuai dengan ekspektasi pengguna. Oleh karena itu, evaluasi hasil super-resolusi menjadi langkah penting dalam mengukur keberhasilan algoritma atau model yang digunakan. Evaluasi kualitas ini dapat dikelompokkan menjadi dua kategori utama: metrik kuantitatif dan evaluasi subjektif (Wang & Bovik, 2009; Szeliski, 2022).

1. Metrik Kuantitatif

Metrik kuantitatif digunakan untuk menilai perbedaan antara hasil super-resolusi dan citra atau video dengan resolusi tinggi sebagai referensi (*ground truth*). Beberapa metrik yang sering digunakan meliputi:

- **Peak Signal-to-Noise Ratio (PSNR)**

PSNR merupakan metrik standar yang digunakan untuk menilai seberapa baik hasil super-resolusi dibandingkan dengan *ground truth*. PSNR diukur dalam desibel (dB) dan dihitung berdasarkan logaritma rasio antara nilai maksimum sinyal dengan tingkat distorsi atau noise

yang terdapat pada citra atau video (Blokdyk, 2020).
Rumus PSNR adalah:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right)$$

Di mana MAX_I adalah nilai maksimum piksel (biasanya 255 untuk citra 8-bit), dan MSE adalah rata-rata kesalahan kuadrat antara hasil super-resolusi dan *ground truth*. Nilai PSNR yang lebih tinggi menunjukkan kualitas yang lebih baik (Wang et al., 2004).

- **Structural Similarity Index Measure (SSIM)**

SSIM adalah metrik yang dirancang untuk menilai kesamaan struktural antara dua citra. Tidak seperti PSNR, SSIM tidak hanya memperhatikan intensitas piksel tetapi juga mempertimbangkan faktor-faktor seperti luminansi, kontras, dan struktur lokal (Wang et al., 2004).

Rumus SSIM adalah:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

Di mana μ_x dan μ_y adalah rata-rata dari dua citra, σ_x^2 dan σ_y^2 adalah variansi, dan σ_{xy} adalah kovarians antara citra x dan y. SSIM menghasilkan nilai antara 0 hingga 1, di mana nilai 1 menunjukkan kesamaan sempurna (Wang et al., 2004).

Meskipun metrik-metrik ini banyak digunakan, mereka memiliki keterbatasan, terutama dalam menangkap aspek visual atau estetika yang relevan dengan persepsi manusia (Szeliski, 2022).

2. Evaluasi Subjektif

Selain metrik kuantitatif, evaluasi subjektif menjadi penting, terutama dalam konteks video musik lawas yang memiliki elemen estetika dan keaslian artistik. Penilaian ini melibatkan persepsi manusia terhadap hasil super-resolusi yang dihasilkan, yang meliputi:

- **Penilaian Estetika dan Keaslian Artistik**

Video musik lawas memiliki karakteristik unik yang berkaitan dengan gaya visual dan estetika pada zamannya. Oleh karena itu, evaluasi subjektif mencakup penilaian apakah hasil super-resolusi berhasil mempertahankan karakteristik tersebut tanpa mengorbankan kualitas visual. Hal ini penting karena perubahan yang terlalu drastis dapat menghilangkan nilai historis dari video tersebut (Moeslund, 2021).

- **Preferensi Visual dari Penonton**

Dalam evaluasi subjektif, panel penonton diminta untuk memilih versi hasil super-resolusi yang lebih disukai. Preferensi ini mencakup aspek seperti kejernihan gambar, ketajaman detail, dan kesan alami (*naturalness*) (Szeliski, 2022). Pendekatan ini sering kali digunakan dalam studi yang membandingkan model super-resolusi untuk aplikasi tertentu, termasuk pada video musik lawas.

Evaluasi subjektif sering kali melibatkan metode standar seperti *Mean Opinion Score* (MOS), di mana penonton memberikan skor pada skala tertentu (misalnya 1-5) berdasarkan pengalaman visual mereka (Wang & Bovik, 2009).

Meskipun metrik kuantitatif seperti PSNR dan SSIM sangat berguna, mereka tidak selalu mencerminkan persepsi visual manusia. Sebaliknya, evaluasi subjektif memberikan wawasan yang lebih mendalam tetapi memerlukan banyak waktu dan sumber daya. Kombinasi kedua pendekatan ini sering kali digunakan untuk memberikan penilaian yang lebih komprehensif terhadap hasil super-resolusi.

7.6. Warisan Budaya Digital

Digitalisasi telah menjadi bagian penting dari upaya pelestarian warisan budaya. Dengan berkembangnya teknologi, berbagai upaya untuk menyimpan dan melestarikan artefak budaya telah beralih ke format digital, termasuk dokumen, foto, musik, film, dan video. Dalam konteks ini, teknologi super-resolusi menjadi relevan karena memungkinkan peningkatan kualitas konten budaya digital, termasuk video musik lawas, yang sering kali mengalami

degradasi kualitas akibat usia dan metode penyimpanan (Moeslund, 2021; Blokdyk, 2020).

Pelestarian konten budaya digital menjadi semakin penting, mengingat banyak artefak budaya yang berisiko hilang karena degradasi fisik, perubahan teknologi, atau hilangnya media penyimpanan asli (Szeliski, 2022). Video musik lawas adalah salah satu bentuk artefak budaya yang memiliki nilai historis tinggi tetapi sering kali ditemukan dalam kondisi teknis yang tidak ideal, seperti resolusi rendah, artefak kompresi, dan warna yang pudar. Teknologi modern, termasuk super-resolusi, berperan penting dalam mengatasi tantangan ini (Wang et al., 2021).

Restorasi konten budaya digital tidak hanya memperbaiki kualitas teknis tetapi juga memastikan aksesibilitasnya bagi generasi mendatang. Menurut Gonzalez dan Woods (2018), pemrosesan citra digital memungkinkan rekonstruksi data visual yang hilang atau rusak, termasuk melalui teknik peningkatan resolusi. Dalam konteks video musik lawas, super-resolusi tidak hanya meningkatkan kualitas visual tetapi juga membantu menghidupkan kembali nilai-nilai estetika yang menjadi ciri khas era tertentu.

Super-resolusi memainkan peran kunci dalam proses ini karena kemampuannya untuk mengembalikan detail yang hilang dan meningkatkan kejernihan visual. Misalnya, metode seperti Real-ESRGAN memungkinkan restorasi detail halus dan tekstur yang sering hilang dalam video musik lawas, menjadikannya lebih mendekati aslinya sambil tetap relevan untuk standar visual modern (Wang et al., 2021).

Video musik lawas Indonesia merupakan bagian integral dari warisan budaya digital yang tidak hanya mencerminkan perkembangan seni dan musik, tetapi juga dinamika sosial, budaya, dan sejarah lokal. Video-video ini sering kali merekam elemen budaya, seperti mode, gaya arsitektur, dan adat istiadat yang mencerminkan *zeitgeist* (semangat zaman) dari era tertentu.

Menurut sumber sejarah, video musik lawas memiliki dampak signifikan terhadap penyebaran seni dan budaya di Indonesia. Pada era 1980-an hingga 2000-an, video musik menjadi salah satu medium utama untuk menyampaikan

pesan-pesan budaya kepada masyarakat luas. Namun, karena keterbatasan teknologi pada masa itu, banyak video yang direkam dalam format dengan resolusi rendah dan kualitas kompresi yang buruk. Kondisi ini menjadi tantangan dalam proses restorasi karena diperlukan pendekatan teknologi yang mampu mengatasi degradasi kualitas tanpa mengorbankan keaslian konten (Moeslund, 2021).

Restorasi video musik lawas tidak hanya berkontribusi pada pelestarian seni dan budaya tetapi juga memberikan kesempatan kepada generasi baru untuk memahami nilai-nilai historis melalui medium yang dapat diterima secara visual. Misalnya, video musik lawas Indonesia dari era 1980-an yang direstorasi dapat menjadi media edukasi bagi masyarakat modern untuk mengenal lebih dekat konteks budaya pada masa itu, termasuk pengaruh globalisasi terhadap musik dan seni local (Szeliski, 2022).

Teknologi super-resolusi memberikan solusi inovatif dalam mengatasi tantangan restorasi video musik lawas. Dibandingkan dengan metode konvensional, super-resolusi berbasis *deep learning* seperti ESRGAN atau Real-ESRGAN memungkinkan peningkatan kualitas visual yang lebih baik dengan mempertahankan detail asli. Metode ini juga dapat mengatasi degradasi kompleks, seperti noise, blur, dan artefak kompresi, yang sering ditemukan pada video musik lawas. Salah satu contoh penerapannya adalah penggunaan super-resolusi untuk meningkatkan resolusi video dari 480p menjadi 1080p atau lebih, sambil memastikan bahwa tekstur dan detail visual tetap konsisten (Wang et al., 2021).

7.7. Teknologi Pendukung

Perkembangan teknologi pendukung dalam bidang pengolahan video dan super-resolusi telah membuka banyak peluang untuk implementasi metode yang lebih canggih. Dalam penelitian yang memanfaatkan teknik super-resolusi berbasis *deep learning*, penggunaan framework, library, dan perangkat keras yang tepat menjadi elemen penting untuk mencapai hasil yang optimal. Poin utama dari teknologi pendukung ini meliputi pemanfaatan framework seperti PyTorch atau

TensorFlow, alat pengolahan video seperti FFmpeg, serta akselerasi komputasi menggunakan GPU. Sebagai contoh, penggunaan framework deep learning seperti PyTorch dan TensorFlow telah terbukti efektif dalam mengimplementasikan model berbasis jaringan saraf tiruan untuk tugas-tugas seperti super-resolusi (Stevens et al, 2020; Hartman et al, 2021).

Framework deep learning modern seperti PyTorch dan TensorFlow telah menjadi pilihan utama dalam implementasi model berbasis jaringan saraf tiruan, termasuk ESRGAN (Enhanced Super-Resolution GAN). PyTorch, yang dikembangkan oleh Meta AI (sebelumnya Facebook AI), menawarkan fleksibilitas tinggi untuk eksperimen penelitian karena sifatnya yang dynamic computation graph. Hal ini memudahkan para peneliti untuk melakukan modifikasi pada arsitektur model dan eksperimen lainnya. Sebagai alternatif, TensorFlow, yang dikembangkan oleh Google Brain, menawarkan efisiensi yang baik untuk produksi dengan dukungan ekosistem seperti TensorBoard untuk visualisasi pelatihan model. Penggunaan PyTorch dan TensorFlow dalam penelitian super-resolusi telah terbukti memberikan hasil yang sangat baik dalam meningkatkan kualitas gambar dan video (Hartman et al, 2021; Stevens et al, 2020).

Dalam konteks super-resolusi, framework ini mendukung berbagai fungsi seperti pelatihan model, inferensi, dan evaluasi hasil dengan metrik seperti PSNR dan SSIM. PyTorch memiliki pustaka pendukung seperti torchvision untuk manipulasi dataset gambar dan video, yang sangat relevan untuk proyek berbasis super-resolusi. Sebaliknya, TensorFlow menawarkan alat seperti Keras untuk penyederhanaan implementasi model dengan abstraksi tingkat tinggi. Peneliti menggunakan kedua framework ini untuk meningkatkan efektivitas pelatihan dan evaluasi model dalam tugas super-resolusi (Stevens et al, 2020; Hartman et al, 2021).

Selain framework deep learning, pengolahan video juga memerlukan alat bantu tambahan. FFmpeg, sebuah perangkat lunak berbasis CLI (Command Line Interface), memainkan peran penting dalam ekstraksi, pemrosesan, dan manipulasi video. FFmpeg mendukung berbagai format file video dan codec, memungkinkan peneliti untuk mengonversi video menjadi rangkaian frame individu yang dapat

diolah lebih lanjut menggunakan model super-resolusi. Sebagai contoh, video lawas dapat dipecah menjadi ribuan frame, di mana masing-masing frame dapat ditingkatkan kualitasnya secara individu sebelum digabung kembali menjadi video resolusi tinggi. Penggunaan FFmpeg untuk pengolahan video dalam penelitian super-resolusi telah banyak dilaporkan dalam berbagai studi.

Pemanfaatan GPU (Graphics Processing Unit) telah menjadi komponen krusial dalam implementasi super-resolusi berbasis deep learning. GPU dirancang untuk menangani komputasi paralel dalam jumlah besar, yang sangat cocok untuk tugas-tugas intensif seperti pelatihan dan inferensi model jaringan saraf tiruan. Dibandingkan dengan CPU (Central Processing Unit), GPU dapat mempercepat proses pelatihan model hingga puluhan kali lipat, memungkinkan eksperimen yang lebih cepat dan efisien. Dalam penelitian yang melibatkan super-resolusi berbasis deep learning, GPU sangat penting untuk mempercepat pelatihan dan inferensi model, khususnya dengan platform seperti NVIDIA CUDA dan cuDNN (Kirk & Hwu, 2022).

Dalam pelatihan model seperti ESRGAN, GPU digunakan untuk mempercepat kalkulasi pada lapisan-lapisan jaringan saraf, termasuk convolutions, backpropagation, dan optimasi parameter. Selain itu, GPU juga sangat efisien untuk menangani batch data yang besar, yang diperlukan dalam pelatihan model dengan dataset resolusi tinggi. Platform seperti NVIDIA CUDA dan cuDNN menyediakan API dan pustaka yang memungkinkan framework seperti PyTorch dan TensorFlow memanfaatkan penuh kemampuan GPU. Penggunaan GPU dalam pelatihan dan inferensi telah meningkatkan efisiensi dalam penelitian super-resolusi (Kirk & Hwu, 2022).

Pada tahap inferensi, GPU memungkinkan pemrosesan gambar atau video secara real-time, yang penting untuk aplikasi praktis seperti restorasi video musik lawas. GPU seperti NVIDIA RTX 3090 atau A100 memiliki VRAM (Video Random Access Memory) yang cukup besar untuk menangani video resolusi tinggi dan model deep learning kompleks secara bersamaan. Penelitian sebelumnya menunjukkan bahwa GPU memainkan peran penting dalam aplikasi super-resolusi dengan efisiensi tinggi (Liu et al., 2020).

Dalam proyek berbasis video, GPU juga membantu mempercepat proses pengolahan frame menggunakan FFmpeg, terutama ketika digunakan bersama filter-filter khusus untuk decoding dan encoding video. FFmpeg dan GPU bekerja bersama-sama untuk meningkatkan efisiensi pengolahan video dalam aplikasi super-resolusi berbasis deep learning (Liu et al., 2020).

Dalam penelitian ini, PyTorch digunakan untuk mengimplementasikan model Real-ESRGAN, dengan dukungan perangkat keras berupa GPU NVIDIA untuk pelatihan dan inferensi. FFmpeg digunakan untuk mengekstraksi frame dari video musik lawas sebelum frame tersebut diproses menggunakan model super-resolusi. Setelah semua frame selesai diproses, FFmpeg kembali digunakan untuk merangkai frame menjadi video resolusi tinggi. Integrasi teknologi ini memungkinkan proses restorasi video menjadi efisien dan menghasilkan output berkualitas tinggi.

8. KEASLIAN PENELITIAN

Table 1. Matriks literature review dan posisi penelitian

Upscaling Music Video Lawas Indonesia Berbasis Frames Video Dengan Menggunakan Algoritma ESRGAN

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data	Wang et al., arXiv, 2021	Mengembangkan metode super-resolusi menggunakan degradasi sintetis untuk meningkatkan performa pada citra dunia nyata.	Real-ESRGAN memberikan hasil visual lebih baik dibandingkan metode tradisional, terutama pada citra dengan degradasi kompleks.	Kelemahan utama adalah membutuhkan banyak data sintetik untuk pelatihan, yang memakan waktu dan sumber daya tinggi.	Penelitian ini menggunakan Real-ESRGAN, tetapi fokus pada domain video musik lawas, memberikan konteks budaya yang spesifik.
2	ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks	Wang et al., ECCV Proceedings, 2018	Meningkatkan performa SRGAN dengan menambahkan residual-in-residual dense blocks untuk memperbaiki detail visual.	ESRGAN menghasilkan detail lebih realistis dibandingkan SRGAN tetapi kurang efektif untuk degradasi kompleks.	Tidak dirancang untuk menangani degradasi acak dan situasi dunia nyata dengan variasi signifikan.	Penelitian ini menggunakan ESRGAN sebagai referensi teknis tetapi menerapkan teknologi terbaru (Real-ESRGAN) untuk tantangan unik pada video lawas.

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
3	Random Degradation Model for Deep Blind Image Super-Resolution	Liu et al., CVPR, 2022	Memperkenalkan model degradasi acak untuk memperbaiki performa super-resolusi pada citra dengan berbagai degradasi dunia nyata.	Model degradasi acak efektif dalam meningkatkan kualitas pada berbagai jenis degradasi kompleks.	Memiliki keterbatasan dalam menerapkan model degradasi pada video karena kurang mempertimbangkan konsistensi antar frame.	Penelitian ini mengambil inspirasi dari model degradasi acak ini tetapi menerapkannya pada video musik lawas, yang menekankan tantangan konsistensi temporal.
4	Dual Back-Projection-Based Internal Learning for Blind Super-Resolution	Kim et al., IEEE, 2020	Memanfaatkan pembelajaran internal berbasis back-projection untuk mengatasi degradasi kompleks pada citra.	Teknik ini efektif untuk mengurangi noise pada citra dengan degradasi berat.	Kurang optimal untuk pemrosesan video karena hanya dirancang untuk citra diam (still images).	Penelitian ini memperluas ide ini dengan fokus pada degradasi kompleks dan konsistensi temporal pada video lawas.
5	Deep Degradation Prior for Real-	Ko et al., CVPR, 2021	Memperkenalkan prior degradasi untuk super-resolusi dunia nyata	Hasil visual sangat baik untuk citra dengan artefak	Tidak mempertimbangkan efek temporal pada video,	Penelitian ini memperbaiki kelemahan ini

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
	World Super-Resolution		dengan fokus pada citra terdegradasi secara signifikan.	kompresi berat, tetapi tidak cukup baik untuk video dengan variasi tinggi.	sehingga hasilnya kurang konsisten antar frame.	dengan menyesuaikan teknologi super-resolusi untuk video musik lawas, dengan konsistensi antar frame sebagai fokus.
6	Unsupervised Image Super-Resolution with an Indirect Supervised Path	Han et al., NeurIPS, 2019	Menerapkan jalur supervisi tidak langsung untuk meningkatkan kualitas super-resolusi tanpa menggunakan data anotasi penuh.	Metode ini efektif dalam situasi dengan data terbatas, tetapi kurang akurat dibandingkan metode supervisi penuh.	Sulit diterapkan pada domain video karena keterbatasan data pelatihan tanpa anotasi video yang memadai.	Penelitian ini memanfaatkan pendekatan supervisi penuh dengan data video musik lawas untuk hasil super-resolusi yang lebih baik.

9. METODE PENELITIAN

9.1. Jenis, Sifat dan Pendekatan Penelitian

Penelitian ini termasuk dalam jenis penelitian eksperimental. Penelitian eksperimental bertujuan untuk menguji pengaruh suatu variabel (dalam hal ini, algoritma ESRGAN) terhadap objek penelitian (video musik lawas Indonesia). Dalam penelitian ini, eksperimen dilakukan untuk mengetahui apakah penggunaan algoritma ESRGAN dapat meningkatkan kualitas video musik lawas Indonesia dalam hal resolusi dan kejernihan gambar. Penelitian ini bersifat kuantitatif karena mengukur perubahan kualitas gambar menggunakan metrik yang dapat dihitung, seperti PSNR, SSIM, dan MSE.

Sifat penelitian ini adalah aplikatif dan deskriptif. Aplikatif karena penelitian ini menerapkan algoritma ESRGAN untuk tujuan praktis, yaitu untuk meningkatkan kualitas video musik lawas Indonesia. Deskriptif karena penelitian ini menggambarkan dan menjelaskan hasil eksperimen yang dilakukan, terutama dalam hal pengukuran kualitas video sebelum dan sesudah di-upscale. Penelitian ini bertujuan untuk memberikan penjelasan mengenai seberapa efektif algoritma ESRGAN dalam meningkatkan kualitas video melalui pengukuran yang objektif.

Pendekatan penelitian yang digunakan adalah kualitatif, meskipun evaluasi objektif menggunakan metrik teknis. Pendekatan ini fokus pada pemahaman mendalam tentang pengaruh algoritma ESRGAN terhadap kualitas visual video. Pendekatan kualitatif dalam penelitian ini digunakan untuk mengeksplorasi perubahan yang terjadi pada video setelah proses upscaling, dengan penekanan pada analisis kualitas gambar secara objektif. Meskipun menggunakan data numerik dari metrik seperti PSNR, SSIM, dan MSE, pendekatan ini tetap bertujuan untuk memberikan pemahaman yang lebih luas dan mendalam tentang efek teknis dari penerapan algoritma ESRGAN pada video musik lawas Indonesia.

9.2. Metode Pengumpulan Data

Metode pengumpulan data dalam penelitian ini dilakukan melalui beberapa tahap yang terstruktur untuk memperoleh data yang relevan dan valid mengenai

video musik lawas Indonesia yang akan diproses dengan algoritma ESRGAN. Berikut adalah langkah-langkah dalam pengumpulan data:

Langkah pertama adalah mengumpulkan satu video musik lawas Indonesia yang akan dijadikan objek penelitian. Video yang dipilih harus memiliki kualitas asli yang mencerminkan variasi, seperti resolusi rendah atau kualitas gambar yang buram, untuk menunjukkan potensi perbaikan melalui proses upscaling. Video ini akan digunakan untuk menguji berbagai variabel, sehingga hasil penelitian dapat menggambarkan efektivitas ESRGAN dalam berbagai kondisi pengolahan. Pemilihan video mempertimbangkan faktor teknis seperti format video, durasi, dan kesesuaian dengan algoritma ESRGAN untuk diproses.

Setelah video tersebut dipilih, langkah selanjutnya adalah mengonversi video menjadi frame-frame gambar individu. Video akan dipecah menjadi serangkaian gambar atau frame dengan menggunakan perangkat lunak pengolah video. Pemecahan ini bertujuan untuk memudahkan pengolahan gambar pada tahap selanjutnya, karena ESRGAN akan diterapkan pada setiap frame dengan menggunakan berbagai variabel pengujian secara terpisah.

Setelah video dipisah menjadi frame-frame, setiap frame akan diproses menggunakan algoritma ESRGAN untuk meningkatkan kualitas gambar. Proses ini dilakukan dengan menggunakan perangkat lunak yang telah dilengkapi dengan implementasi algoritma ESRGAN untuk upscaling gambar. Hasil dari proses ini adalah frame-frame yang telah memiliki resolusi lebih tinggi dan kualitas yang lebih baik dibandingkan dengan gambar asli.

Setelah semua frame di-upscale, frame-frame yang telah diproses akan digabungkan kembali menjadi video dengan resolusi yang lebih tinggi. Video ini akan dibandingkan dengan versi asli untuk mengukur sejauh mana kualitas visual meningkat setelah proses upscaling.

Sebagai bagian dari pengumpulan data, kualitas gambar yang dihasilkan dari proses upscaling akan diukur menggunakan metrik objektif seperti PSNR (Peak Signal-to-Noise Ratio), SSIM (Structural Similarity Index), dan MSE (Mean Squared Error). Metrik ini akan digunakan untuk mengukur perbedaan antara video asli dan video yang telah di-upscale. Data yang diperoleh dari pengukuran ini akan

digunakan untuk menganalisis apakah algoritma ESRGAN dapat meningkatkan kualitas video musik lawas Indonesia secara signifikan.

9.3. Metode Analisa Data

1. Analisis Deskriptif Data

Langkah pertama dalam analisis adalah melakukan analisis deskriptif terhadap hasil pengukuran kualitas gambar yang diperoleh dari video asli dan video yang telah di-upscale menggunakan ESRGAN. Hasil pengukuran metrik PSNR, SSIM, dan MSE akan disajikan dalam bentuk tabel atau grafik untuk memberikan gambaran umum mengenai perbandingan antara video asli dan video yang telah di-upscale. Dalam hal ini:

- PSNR (Peak Signal-to-Noise Ratio) akan dianalisis untuk mengetahui tingkat perbedaan antara video asli dan video hasil upscaling. Nilai PSNR yang lebih tinggi menunjukkan kualitas gambar yang lebih baik.
- SSIM (Structural Similarity Index) akan dianalisis untuk melihat kesamaan struktural antara gambar asli dan hasil upscaling. Nilai SSIM yang lebih tinggi menunjukkan kesamaan yang lebih baik.

2. Analisis Perbandingan

Selanjutnya, analisis perbandingan dilakukan untuk membandingkan hasil kualitas video yang dihasilkan oleh ESRGAN dengan video asli. Perbandingan dilakukan dengan melihat perbedaan nilai metrik PSNR, SSIM antara kedua versi video (asli dan upscaled).

- Perbandingan PSNR: Jika nilai PSNR video yang di-upscale lebih tinggi dibandingkan dengan video asli, itu menunjukkan bahwa video tersebut memiliki kualitas gambar yang lebih baik setelah upscaling.
- Perbandingan SSIM: Jika nilai SSIM video yang di-upscale lebih tinggi dibandingkan dengan video asli, itu menunjukkan bahwa struktur gambar (luminance, kontras, dan struktur) lebih terjaga dengan baik setelah upscaling.

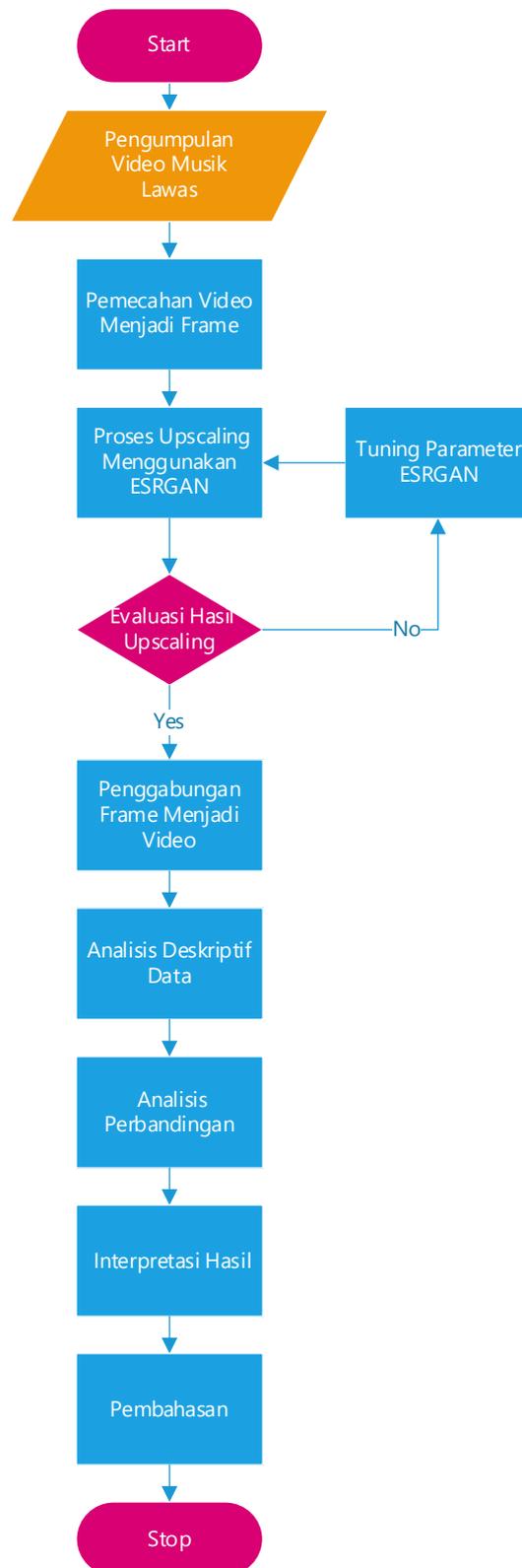
3. Interpretasi Hasil

Hasil analisis perbandingan akan diinterpretasikan untuk mengevaluasi seberapa efektif algoritma ESRGAN dalam meningkatkan kualitas video musik lawas Indonesia. Berdasarkan hasil PSNR, SSIM akan ditarik kesimpulan mengenai peningkatan kualitas gambar setelah menggunakan ESRGAN. Jika nilai PSNR dan SSIM menunjukkan peningkatan yang signifikan, maka dapat disimpulkan bahwa ESRGAN berhasil meningkatkan kualitas video musik lawas tersebut.

4. Pembahasan

Pembahasan akan mengkaji faktor-faktor yang memengaruhi hasil upscaling, seperti kualitas video asli, jenis video, dan karakteristik video musik lawas Indonesia yang digunakan dalam penelitian. Pembahasan ini juga dapat mencakup perbandingan hasil ESRGAN dengan metode upscaling lainnya jika relevan, untuk menilai keunggulan algoritma ESRGAN dalam konteks upscaling video musik lawas.

9.4. Alur Penelitian



Gambar 3. Alur Penelitian

Berdasarkan Gambar 3. Penelitian ini dimulai dengan pengumpulan data berupa video musik lawas Indonesia yang memiliki kualitas rendah. Data ini kemudian diolah dengan memecah video menjadi frame individu untuk mempermudah proses upscaling. Setelah frame-frame tersebut diperoleh, dilakukan proses upscaling menggunakan algoritma ESRGAN untuk meningkatkan resolusi setiap frame.

Hasil dari proses upscaling dievaluasi menggunakan metrik objektif seperti PSNR dan SSIM untuk menilai kualitas visual hasil. PSNR (*Peak Signal-to-Noise Ratio*) digunakan untuk mengukur seberapa baik kualitas frame hasil upscaling dibandingkan dengan frame asli. Nilai PSNR yang lebih tinggi menunjukkan bahwa perbedaan antara gambar asli dan gambar hasil upscaling lebih kecil, sehingga kualitas gambar lebih baik. Namun, PSNR lebih berfokus pada pengukuran kuantitatif dari segi kesalahan piksel secara keseluruhan dan tidak selalu mencerminkan persepsi visual manusia.

Sementara itu, SSIM (*Structural Similarity Index Measure*) digunakan untuk mengevaluasi sejauh mana struktur, pencahayaan, dan kontras antara gambar hasil upscaling menyerupai gambar asli. SSIM lebih relevan dalam menilai kualitas visual karena mempertimbangkan aspek-aspek yang lebih sesuai dengan persepsi manusia terhadap gambar. Nilai SSIM berkisar antara 0 hingga 1, di mana nilai mendekati 1 menunjukkan bahwa gambar hasil upscaling sangat mirip dengan gambar asli dalam hal struktur dan kualitas visual.

Jika hasil evaluasi dari PSNR dan SSIM tidak memenuhi standar yang telah ditetapkan, dilakukan proses tuning parameter ESRGAN untuk meningkatkan performanya. Proses upscaling kemudian diulang hingga hasilnya sesuai dengan standar yang diinginkan.

Setelah mendapatkan frame dengan kualitas yang lebih baik, frame-frame tersebut digabungkan kembali menjadi video. Tahap selanjutnya adalah melakukan analisis deskriptif terhadap data hasil upscaling dengan menyajikan nilai-nilai PSNR dan SSIM dalam bentuk tabel atau grafik. Hasil analisis ini dibandingkan

dengan data video asli untuk mengetahui sejauh mana algoritma ESRGAN berhasil meningkatkan kualitas video.

Interpretasi hasil dilakukan untuk mengevaluasi efektivitas algoritma ESRGAN, diikuti dengan pembahasan mengenai faktor-faktor yang memengaruhi hasil penelitian, seperti parameter ESRGAN, kualitas awal video, dan karakteristik video musik lawas Indonesia. Pembahasan ini menjadi dasar untuk memberikan kesimpulan serta rekomendasi untuk penelitian lanjutan.

10. RENCANA JADWAL PENELITIAN

No	Tahapan	Target Output	Bulan														
			Desember			Januari			Februari			Maret					
1	Penyusunan dan pengajuan judul	Judul Tesis	■	■													
2	Pengajuan Proposal	Bimbingan Proposal			■	■											
3	Persiapan dan pengumpulan data	Dataset video musik lawas dan tools					■	■									
4	Analisis Data	Dataset siap untuk diproses						■	■								
5	Data Wrangling/Processing	Dataset siap untuk upscaling							■	■							
6	Data Training	Menghasilkan model ESRGAN yang optimal								■	■	■					
7	Evaluating	Menentukan tingkat optimalisasi model													■		
8	Tuning	Memperbaiki hasil upscaling jika belum optimal														■	
9	Comparing	Membandingkan hasil video asli dan upscaled															■
10	Penyusunan Laporan	Bimbingan Tesis															■

Daftar Pustaka

- Wang, X., et al., 2021. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data.
- Zhang, K., et al., 2021. Designing a Practical Degradation Model for Deep Blind Image Super-Resolution.
- Liu, J., Wang, H., 2022. Random Degradation Model for Deep Blind Image Super-Resolution.
- Chen, X., et al., 2021. Adversarial Training Methods for Generative Neural Networks.
- Ji, X., et al., 2020. Real-World Super-Resolution via Kernel Estimation and Noise Injection.
- Liu, A., et al., 2021. Blind Image Super-Resolution: A Survey and Beyond.
- Baghel, A., et al., 2023. Leveraging Transformer-based Architectures in GANs for Enhanced Image Super-Resolution.
- Xie, L., et al., 2021. Finding Discriminative Filters for Specific Degradations in Blind Super-Resolution.
- Mou, C., et al., 2022. Metric Learning based Interactive Modulation for Real-World Super-Resolution.
- Luo, Z., et al., 2020. Unfolding the Alternating Optimization for Blind Super Resolution.
- Son, S., et al., 2021. Toward Real-World Super-Resolution via Adaptive Downsampling Models.
- Kim, J., et al., 2020. Dual Back-Projection-Based Internal Learning for Blind Super-Resolution.
- Ko, K., et al., 2021. Deep Degradation Prior for Real-World Super-Resolution.
- Blokdyk, G., 2020. *Digital Image Processing: A Complete Guide - 2020 Edition*. 5STARCOoks.
- Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Szeliski, R., 2022. *Computer Vision: Algorithms and Applications*. Springer, 2nd ed.

Hartman, J., Kopic, D., 2021. *Scaling TensorFlow to 300 Million Predictions per Second*.

Kirk, D. B., & Hwu, W. W., 2022. *Programming Massively Parallel Processors: A Hands-on Approach*. Morgan Kaufmann, 4th ed.

Stevens, E., Antiga, L., Viehmann, T., 2020. *Deep Learning with PyTorch*. Manning Publications.