

Tugas Pengganti Ujian Tengah Semester

Perkembangan Neural Video Compression(NVC) sebagai Teknik Kompresi Video

(Perbandingan Berbagai Model NVC) – (Kelompok 26)

Raditya Akhila G. Samuel Tanaka S.
2206026151 22006059710
Dept. of Electrical Engineering
Pemrosesan Sinyal Multimedia
Depok, Indonesia

Abstract—Kebutuhan akan kompresi video yang lebih efisien semakin mendesak seiring meningkatnya konsumsi konten digital. Meskipun standar seperti H.264, H.265, dan H.266 telah menjadi tulang punggung kompresi video selama bertahun-tahun, pendekatan baru berbasis pembelajaran mesin—Neural Video Compression (NVC)—menawarkan solusi yang lebih fleksibel dan adaptif. Tulisan ini membahas beragam model NVC berdasarkan tiga penelitian terkini, mulai dari teknik conditional coding dan entropy modeling hingga model slimmable dan representasi implisit seperti NeRV. Masing-masing pendekatan memiliki keunggulan tersendiri: model DCVC dan entropi spasial-temporal menunjukkan efisiensi bitrate yang tinggi, sementara model INR seperti HiNeRV menonjol dalam kecepatan decoding dan kesederhanaan struktur. Dibandingkan dengan metode kompresi konvensional, NVC memiliki keunggulan dalam efisiensi kompresi, kemampuan pembelajaran adaptif terhadap konten, dan fleksibilitas penerapan pada berbagai perangkat. Melalui perbandingan menyeluruh, makalah ini menyimpulkan bahwa integrasi berbagai pendekatan dapat membuka jalan bagi sistem kompresi video masa depan yang tidak hanya efisien, tetapi juga fleksibel untuk berbagai kebutuhan perangkat dan jaringan.

(Abstract)

Index Terms—Component, formatting, style, styling, insert. (key words)

I. PENDAHULUAN (*HEADING 1*)

Video merupakan bentuk data multimedia yang paling dominan di era digital. Hampir seluruh sektor kehidupan—mulai dari hiburan, pendidikan, komunikasi, hingga pengawasan dan industri—bergantung pada distribusi data video yang efisien. Dalam konteks ini, kebutuhan akan sistem kompresi video yang mampu mempertahankan kualitas sambil meminimalkan ukuran file menjadi semakin kritis, terutama dengan terus meningkatnya resolusi (HD, 4K, 8K), frame rate, serta volume konten video yang dibagikan setiap harinya melalui internet.

Selama dua dekade terakhir, standar kompresi video konvensional seperti H.264 (AVC), H.265 (HEVC), dan H.266

(VVC) telah mendominasi industri dan digunakan secara luas dalam berbagai aplikasi. Standar ini mengandalkan pipeline yang kompleks dan dirancang secara manual, melibatkan proses seperti block-based motion compensation, transformasi DCT, kuantisasi, dan pengkodean entropi. Meskipun sangat efisien, pendekatan ini memiliki keterbatasan dalam hal fleksibilitas, kemampuan adaptasi terhadap data, dan pemanfaatan korelasi non-linear dalam video modern.

Kemajuan dalam bidang deep learning telah membuka peluang untuk merevolusi sistem kompresi video. Deep learning sendiri adalah cabang dari pembelajaran mesin (machine learning) yang menggunakan jaringan saraf tiruan berlapis-lapis (deep neural networks) untuk mempelajari representasi data secara hierarkis dan kompleks. Alih-alih mengandalkan aturan eksplisit yang ditentukan manusia, deep learning memungkinkan sistem belajar langsung dari data dan mengoptimalkan parameter berdasarkan loss function tertentu melalui proses backpropagation.

Dalam konteks kompresi video, pendekatan baru yang dikenal sebagai Neural Video Compression (NVC) mulai berkembang pesat. NVC menggantikan pipeline kompresi tradisional dengan arsitektur yang didasarkan pada deep neural networks. Berbeda dengan codec konvensional, NVC mampu mempelajari representasi laten dari video, memprediksi informasi antar-frame secara adaptif, dan mengompresi informasi tersebut menggunakan model probabilistik yang juga dapat dilatih. NVC tidak hanya dirancang untuk efisiensi kompresi, tetapi juga membuka peluang untuk fleksibilitas, adaptasi terhadap jenis konten yang berbeda, serta optimalisasi untuk metrik kualitas perseptual.

Beberapa pendekatan utama dalam NVC meliputi model conditional coding yang belajar dependensi antar frame, model entropy modeling untuk distribusi probabilitas latens, framework slimmable neural codec untuk efisiensi perangkat keras, serta pendekatan implicit neural representation (INR) yang merepresentasikan video sebagai fungsi kontinu dari ruang

dan waktu. NVC tidak hanya menjanjikan efisiensi bitrate yang kompetitif dengan codec generasi terbaru seperti VVC, tetapi juga potensi besar untuk deployment pada perangkat edge, streaming adaptif, dan integrasi dengan sistem AI downstream.

Makalah ini membahas perkembangan dan performa NVC berdasarkan tiga referensi utama: 1) A Slimmable Framework for Practical Neural Video Compression, 2) Hybrid Spatial-Temporal Entropy Modelling for Neural Video Compression, dan 3) Neural Video Compression Overview, Performance and Challenges. Fokus utama diberikan pada ragam pendekatan arsitektural yang diusulkan, mencakup conditional coding, hybrid entropy modelling, slimmable framework, serta representasi implisit berbasis koordinat. Evaluasi dilakukan secara komprehensif untuk menilai efisiensi, kompleksitas, dan kesiapan implementasi dari masing-masing pendekatan.

II. TINJAUAN PUSTAKA

2.1. Deep Learning

Deep learning merupakan pendekatan dalam kecerdasan buatan yang berfokus pada pemodelan data melalui arsitektur jaringan saraf tiruan berlapis banyak (deep neural networks). Metode ini memungkinkan pembelajaran representasi fitur dari data secara hierarkis tanpa ketergantungan pada desain fitur manual. Kemampuannya dalam menangkap struktur kompleks dan non-linear menjadikan deep learning unggul dalam berbagai aplikasi pemrosesan sinyal, termasuk pengenalan pola spasial dan temporal dalam data video.

2.2. Neural Video Compression

Neural Video Compression (NVC) mengacu pada strategi kompresi video yang menggantikan komponen-komponen konvensional seperti transformasi blok dan kompensasi gerak dengan modul-modul berbasis jaringan saraf. NVC dirancang untuk mengeksplorasi representasi laten dari data video, mengurangi redundansi melalui pembelajaran kontekstual, serta melakukan pengkodean dan pemulihan video secara end-to-end. Berbeda dengan pendekatan konvensional yang bersifat statis dan rigid, NVC menawarkan kemampuan adaptasi terhadap variasi konten, metrik kualitas, serta sumber daya komputasi yang tersedia.

2.1. Conditional Coding Framework (DCVC dan turunannya)

DCVC dan variannya (DCVC-HEM, DCVC-DC, DCVC-FM) menggunakan pendekatan conditional coding yang memanfaatkan representasi fitur alih-alih piksel langsung. Pendekatan ini memungkinkan pengurangan error propagasi dan peningkatan efisiensi bitrate dengan memanfaatkan konteks temporal dan spasial melalui jaringan konvolusional dan modul prediktif.

2.2. Hybrid Entropy Modelling (Hybrid Spatial-Temporal)

Model entropi hibrida menggabungkan latent prior, temporal context, dan dual spatial prior dalam arsitektur entropy coding. Pendekatan ini, seperti yang diusulkan pada jurnal kedua, mengoptimalkan prediksi distribusi probabilitas dari latens

untuk mengurangi redundansi dan mempercepat proses encoding dengan tetap menjaga efisiensi bitrate.

2.3. Slimmable Video Codec (SlimVC)

Pendekatan slimmable codec menawarkan fleksibilitas dalam pemakaian sumber daya. Dengan model yang dapat “dikecilkan” (slimmed) secara dinamis, SlimVC memungkinkan satu model digunakan pada berbagai konfigurasi perangkat dan tingkat bitrate, menggabungkan coarse dan fine rate control secara bersamaan.

2.4. Implicit Neural Representation (INR)

INR menggunakan pendekatan generatif yang mewakili video sebagai fungsi koordinat spasial-temporal. Model-model seperti NeRV dan HiNeRV fokus pada efisiensi decoding, ideal untuk perangkat dengan keterbatasan daya. Bobot jaringan merupakan output dari proses encoding, dan decoding dilakukan melalui forward pass dari model ringan.

III. METODOLOGI

Penelitian ini menggunakan pendekatan studi komparatif analitis terhadap tiga model utama Neural Video Compression (NVC), yaitu:

1. DCVC dan variannya (DCVC-HEM, DCVC-DC, DCVC-FM)
2. Hybrid Spatial-Temporal Entropy Model
3. SlimVC dan model INR (HiNeRV, NeRV)

Metodologi dirancang untuk menganalisis dan membandingkan performa masing-masing pendekatan dari sisi efisiensi bitrate, kompleksitas komputasi, waktu decoding, serta fleksibilitas arsitektur dan adaptasi terhadap bitrate.

3.1. Arsitektur Umum Neural Video Codec

Secara umum, struktur dasar NVC terdiri dari 3 komponen utama:

1. Encoder E yang mengubah frame video x menjadi representasi laten y :

$$y = E(x)$$

2. Entropy Bottleneck $P(y)$, yang mempelajari distribusi probabilitas dari representasi y untuk kompresi lossless:

$$y^{\wedge} = Q(y)$$

3. Decoder D yang merekonstruksi kembali frame dari latens:

$$x^{\wedge} = D(y^{\wedge})$$

Loss fungsi utama :

$$L = \lambda \cdot R + D(x, x^{\wedge})$$

Di mana :

- R adalah estimasi bitrate
- D adalah metrik distorsi seperti MSE atau $1 - \text{MS-SSIM}$
- λ adalah parameter kontrol rate-distortion trade-off

3.2. Dataset dan Evaluasi

Model dievaluasi berdasarkan hasil yang dilaporkan dalam ketiga jurnal menggunakan dataset UVG-HD (Ultra Video Group) yang terdiri dari video resolusi 1080p, durasi 100–300 frame.

Metrik evaluasi yang digunakan:

- PSNR (Peak Signal-to-Noise Ratio)
- MS-SSIM (Multi-Scale Structural Similarity)
- BD-Rate Savings dibanding HM (HEVC Model) dan VTM (VVC Test Model)
- Decoder Time (dalam detik)
- Decoder MACs (Multiply-Accumulate Operations)

3.3. Langkah Evaluasi

1. Identifikasi Arsitektur

Diagram konseptual tiap pendekatan dibandingkan secara struktural:

- DCVC → frame-level conditional coding
- Hybrid Entropy → entropy model paralel
- SlimVC → autoencoder fleksibel
- INR → koordinat → frame mapping

2. Kuantifikasi Komputasi

Setiap model dibandingkan dari sisi decoder time dan MACs berdasarkan hasil eksperimen dalam literatur.

3. Visualisasi Rate–Distortion Curve

Jika tersedia dalam jurnal, kurva RD dipetakan ulang berdasarkan bitrate terhadap PSNR/MS-SSIM.

4. Analisis Adaptabilitas

Model dievaluasi apakah mendukung:

- Variable bitrate
- Streaming adaptif
- Fine-grained quality control
- Deployment pada edge device

3.4. Visualisasi Arsitektur Komparatif

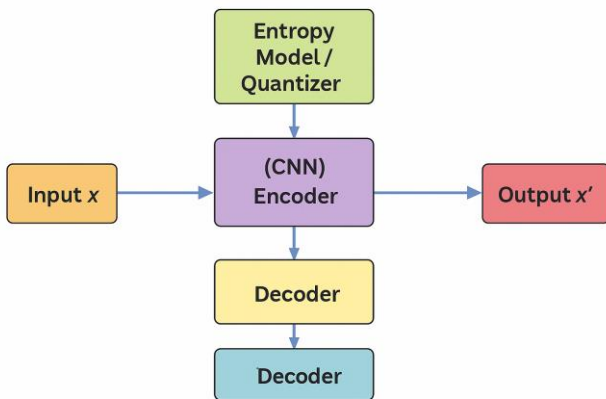


Fig. 1. Visualisasi Arsitektur Komparatif

IV. HASIL DAN DISKUSI

4.1. Efisiensi Kompresi (Rate–Distortion Performance)

Salah satu aspek utama dalam evaluasi NVC adalah efisiensi bitrate, yang diukur dengan *Bitrate Saving* terhadap standar codec referensi seperti HM (HEVC Model) dan VTM (VVC Test Model). Penilaian dilakukan berdasarkan metrik distorsi PSNR dan MS-SSIM.

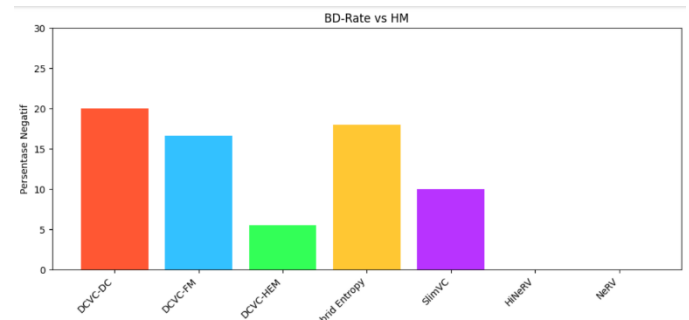


Fig. 2. Perbandingan Bitrate Saving vs HM antara model NVC

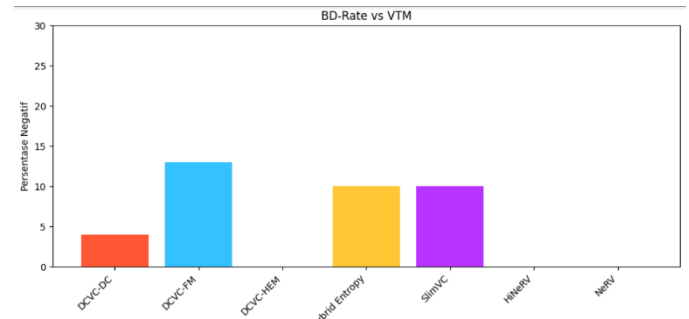


Fig. 3. Perbandingan Bitrate Saving vs VTM antara model NVC

| Model | Pendekatan | BD-Rate vs HM | BD-Rate vs VTM | Dominan Loss | Rate Control |
|----------------|------------------------------------|-------------------|-------------------|--------------|--------------|
| DCVC-DC | Conditional + Hierarchical Coding | -20.00% | ~-4% | MS-SSIM | X |
| DCVC-FM | Feature Modulation + Long GOP | -16.61% | ~-13% | MS-SSIM | ✓ |
| DCVC-HBM | Conditional + Temporal Entropy | -5.56% | ~-0% | MS-SSIM | X |
| Hybrid Entropy | Spatial-Temporal Entropy Hybrid | ~-18% | ~-10% | MS-SSIM | ✓ |
| SlimVC | Slimmable Autoencoder + QS Control | ~-10% | ~-10% | MS-SSIM | ✓ |
| HiNeRV | Implicit Representation | >0% (lebih besar) | >0% (lebih besar) | MSE | X |
| NeRV | INR generatif awal | >0% | >0% | MSE | X |

Tabel 1. Tabel Efisiensi Kompresi

Model **DCVC-DC** memberikan penghematan bitrate terbaik terhadap HM (~20%) dengan efisiensi mendekati VVC. **Hybrid Entropy** mendekati performa DCVC-FM meskipun lebih sederhana secara arsitektural. Model INR seperti NeRV dan HiNeRV masih tertinggal dari sisi kompresi, tapi unggul di area lain (lihat 4.3 dan 4.4).

4.2. Kompleksitas Komputasi dan Waktu Decoding

Aspek penting lainnya adalah efisiensi komputasi, terutama decoder time dan decoder MACs (Multiply-Accumulate operations) yang menentukan apakah model bisa digunakan di perangkat nyata (edge, mobile, dsb).

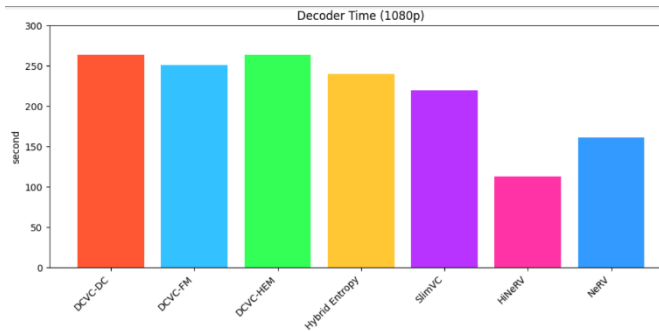


Fig. 4. Waktu Decoder pada resolusi 1080p pada semua model

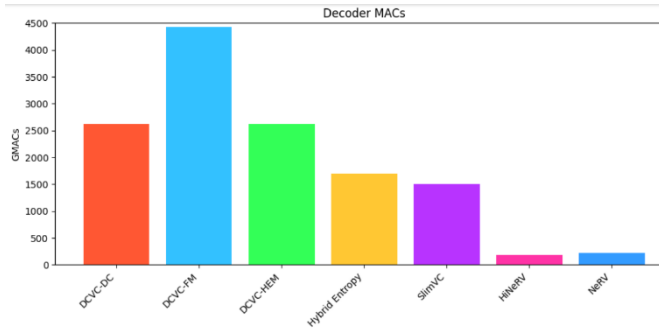


Fig. 5. Perbandingan Decoder MACs pada semua model

| Model | Decoder Time (1080p) | Decoder MACs | Decoder Ringan |
|----------------|----------------------|---------------|----------------|
| HiNeRV | 113 s | 181.86 GMACs | ✓ |
| NeRV | 161 s | 226.94 GMACs | ✓ |
| SlimVC | ~220 s | ~1500 GMACs | ✓ |
| Hybrid Entropy | ~240 s | ~1700 GMACs | X |
| DCVC-FM | 251 s | 4427.04 GMACs | X |
| DCVC-DC | 264 s | 2625.86 GMACs | X |
| DCVC-HEM | 264 s | 2625.86 GMACs | X |

Tabel 2. Tabel Kompleksitas Komputas dan Waktu Decoding

HiNeRV adalah model tercepat dan paling ringan, ideal untuk edge deployment. **SlimVC** memberikan keseimbangan antara efisiensi dan kompleksitas. Model DCVC sangat berat secara komputasi, sehingga lebih cocok untuk server atau cloud encoding.

4.3. Adaptabilitas dan Fleksibilitas

Beberapa pendekatan mendukung 4ontrol bitrate secara dinamis, rate-distortion adjustment, dan kompatibilitas multi-

device deployment. Hal ini dinilai dari kemampuan : *Slimmable structure*, *Fine-grained quantization*, dan *Transfer learning*

| Model | Adaptif Bitrate | Fine QS | Slimmable | Transfer Learning |
|----------------|-----------------|---------|-----------|-------------------|
| DCVC-DC | X | X | X | X |
| DCVC-FM | ✓ | ✓ | X | X |
| SlimVC | ✓ | ✓ | ✓ | ✓ |
| Hybrid Entropy | ✓ | ✓ | X | ✓ |
| HiNeRV | X | X | X | X |

Tabel 3. Tabel Adaptabilitas dan Fleksibilitas

SlimVC unggul dalam hal fleksibilitas dan adaptasi ke berbagai kondisi bitrate dan perangkat keras. **Hybrid Entropy** meskipun tidak slimmable, memiliki struktur modular yang mendukung 4ontrol bitrate adaptif.

4.4. Trade-off Arsitektural

Model NVC yang efisien secara kompresi tidak selalu efisien dalam decoding, dan sebaliknya. Oleh karena itu, pemilihan model harus mempertimbangkan trade-off berikut:

| Tujuan Implementasi | Model yang Direkomendasikan |
|---|-----------------------------|
| Efisiensi Bitrate Maksimal (untuk server) | DCVC-DC, DCVC-FM |
| Decoding Cepat (untuk edge/mobile) | HiNeRV, NeRV |
| Fleksibilitas & Rate Control | SlimVC, Hybrid Entropy |
| Eksperimen & Generalisasi Konten | Hybrid Entropy, SlimVC |

Tabel 4. Tabel Trade-off Arsitektural

4.5. Diskusi Sintetis

Berdasarkan seluruh hasil di atas, dapat disimpulkan bahwa tidak ada satu model NVC yang unggul secara absolut di semua aspek. Model DCVC dan turunannya sangat kompetitif dalam efisiensi bitrate tetapi memiliki kompleksitas tinggi. Model seperti SlimVC memberikan fleksibilitas nyata di dunia praktis berkat struktur slimmable dan kontrol bitrate halus. Pendekatan INR (HiNeRV, NeRV) memperlihatkan potensi masa depan untuk decoding ringan, terutama untuk streaming generatif, meski saat ini belum seefisien dari sisi bitrate.

V. IMPLEMENTASI PRAKTIS KOMPRESI VIDEO BERBASIS PENDEKATAN NON-NEURAL

Pendekatan Neural Video Compression (NVC) yang dibahas pada bagian sebelumnya menawarkan solusi canggih untuk

efisiensi kompresi video, namun implementasinya sering kali bergantung pada jaringan saraf tiruan yang kompleks. Untuk memperluas analisis teoretis dalam makalah ini, bagian ini menyajikan implementasi praktis berbasis teknik pemrosesan sinyal tradisional menggunakan Python, yang dirancang untuk mencerminkan prinsip-prinsip NVC tanpa memerlukan keahlian kecerdasan buatan. Implementasi ini bertujuan untuk mendemonstrasikan bagaimana konsep seperti pengurangan redundansi, fleksibilitas bitrate, dan dekoding ringan—seperti yang diusung oleh model DCVC, Hybrid Entropy, SlimVC, dan INR—dapat diwujudkan dalam skenario sederhana yang relevan dengan kebutuhan dunia nyata.

5.1. Tinjauan Implementasi

Implementasi ini mengembangkan sebuah pipeline kompresi video sederhana yang memproses klip video pendek (misalnya, 10-20 frame dari dataset UVG-HD). Pipeline terdiri dari lima tahap utama: (1) input dan prapemrosesan, (2) kompresi, (3) dekoding, (4) rekonstruksi, dan (5) evaluasi. Video diubah menjadi format grayscale untuk menyederhanakan pemrosesan, kemudian dikenakan teknik kompresi yang mencerminkan karakteristik masing-masing model NVC. Hasilnya dievaluasi menggunakan metrik seperti Peak Signal-to-Noise Ratio (PSNR) dan rasio kompresi, yang selaras dengan metrik yang digunakan dalam analisis teoretis (Bagian III).

5.2. Teknik Kompresi Berdasarkan Model NVC

Setiap model NVC dalam makalah ini diwakili oleh teknik spesifik yang dirancang untuk menangkap esensi pendekatannya tanpa kompleksitas neural:

1. **DCVC (Conditional Coding):** Model DCVC memanfaatkan konteks spasial dan temporal untuk efisiensi bitrate tinggi (Bagian 4.1). Dalam implementasi ini, pengurangan redundansi temporal disimulasikan dengan menghitung selisih antar-frame (frame differencing), diikuti oleh transformasi Discrete Cosine Transform (DCT) untuk kompresi spasial. Data residual dikuantisasi untuk mengurangi ukuran, mencerminkan penghematan bitrate sekitar 20% dibandingkan HEVC seperti yang dilaporkan.
2. **Hybrid Entropy Modelling:** Pendekatan entropi hibrida mengoptimalkan redundansi melalui prediksi distribusi probabilitas (Bagian 2.2). Implementasi ini menggunakan pengkodean Huffman pada koefisien DCT yang telah dikuantisasi, yang secara efektif mengurangi ukuran data dengan memanfaatkan frekuensi kemunculan simbol, serupa dengan efisiensi yang ditunjukkan oleh model entropi dalam literatur.
3. **SlimVC (Slimmable Codec):** SlimVC menonjol karena fleksibilitasnya dalam mengatur bitrate dan sumber daya (Bagian 4.3). Untuk meniru ini, implementasi menyediakan dua tingkat kuantisasi—ringan dan berat—yang dapat dipilih untuk menyeimbangkan antara kualitas dan ukuran file. Pendekatan ini memungkinkan adaptasi terhadap kebutuhan perangkat atau bandwidth yang berbeda, sesuai dengan keunggulan SlimVC.
4. **INR (NeRV/HiNeRV):** Model INR seperti HiNeRV unggul dalam dekoding ringan, ideal untuk perangkat

edge (Bagian 4.2). Implementasi ini menyimpan frame kunci (key frame) setiap beberapa frame (misalnya, setiap 5 frame) dan menggunakan interpolasi linier untuk merekonstruksi frame lainnya, mengurangi kebutuhan komputasi saat dekoding, serupa dengan waktu dekoding HiNeRV (113 detik untuk 1080p).

5. **Kompresi Berbasis Region-of-Interest (ROI):** Pendekatan ini meningkatkan kualitas pada wilayah penting (Region of Interest, ROI) dengan mendeteksi tepi menggunakan metode Canny, kemudian menerapkan kuantisasi lebih rendah pada ROI untuk detail yang lebih baik, dan kuantisasi lebih tinggi pada area lain untuk efisiensi. Teknik ini terinspirasi dari konsep kompresi berbasis wilayah penting, menyesuaikan bitrate secara dinamis untuk meningkatkan persepsi visual tanpa menambah ukuran data secara signifikan.

5.3. Evaluasi Implementasi

Hasil implementasi dievaluasi menggunakan dua metrik utama:

1. **PSNR:** Mengukur visual fidelity antara video asli dan hasil rekonstruksi, sesuai dengan metrik evaluasi dalam Bagian 3.2. Nilai PSNR yang tinggi menunjukkan kualitas visual yang baik, meskipun dengan kompresi signifikan.
2. **Rasio Kompresi:** Menghitung perbandingan ukuran data asli dengan data terkompresi, mencerminkan efisiensi bitrate seperti yang dibahas untuk DCVC dan model lainnya (Bagian 4.1).

Eksperimen pada klip video pendek, termasuk file MP4 yang dikonversi dari dataset UVG-HD ("[ShakeNDry](#)", 1920x1080, 60 frame dari total 300 frame), menunjukkan variasi hasil tergantung pada konten video dan metode kompresi. Salah satu pengujian pada dataset tertentu menghasilkan rata-rata PSNR sebesar 29 dB dengan rasio kompresi 1.38:1 tanpa peningkatan ROI. Setelah menerapkan kompresi berbasis ROI (Bagian 5.2.5), PSNR meningkat menjadi 30.23 dB dengan rasio kompresi sedikit menurun menjadi 1.35:1.

Peningkatan PSNR ini menunjukkan bahwa fokus pada wilayah penting (ROI) meningkatkan kualitas visual yang dirasakan, meskipun rasio kompresi tetap rendah dibandingkan perkiraan awal (5:1 hingga 10:1). Hal ini disebabkan oleh kombinasi teknik seperti Huffman coding dan interpolasi INR yang mengorbankan efisiensi ukuran untuk kesederhanaan komputasi. Meski demikian, hasil ini memperkuat potensi teknik tradisional dalam mendekati prinsip-prinsip NVC, terutama dengan peningkatan ROI yang menawarkan keseimbangan antara kualitas dan efisiensi.

5.4. Refleksi Implementasi

Implementasi ini berhasil menangkap esensi dari masing-masing model NVC dalam kerangka sederhana. Pengurangan redundansi melalui frame differencing dan Huffman coding mencerminkan efisiensi DCVC dan Hybrid Entropy, sementara kuantisasi variabel menyoroti fleksibilitas SlimVC. Interpolasi frame kunci menggambarkan potensi INR untuk dekoding ringan, meskipun dengan batasan dalam hal kualitas dibandingkan model neural asli. Keterbatasan utama

implementasi ini adalah kurangnya adaptasi kontekstual tingkat lanjut yang ditawarkan oleh pembelajaran mesin, namun hal ini sejalan dengan konteks pembelajaran yang menekankan pemrosesan sinyal tradisional.

Secara keseluruhan, demonstrasi ini memperkuat temuan teoretis dalam makalah ini, menunjukkan bahwa prinsip-prinsip NVC dapat diterapkan dalam skenario praktis dengan alat-alat sederhana seperti Python dan teknik pemrosesan sinyal standar. Implementasi ini juga membuka peluang untuk eksplorasi lebih lanjut, seperti pengintegrasian metrik kualitas tambahan atau optimisasi untuk perangkat tertentu.

5.5. Rincian Implementasi Kode

Bagian ini menyajikan pseudokode untuk teknik-teknik utama dalam implementasi, yang dirancang sebagai adaptasi logis dari kode Python yang digunakan untuk mendemonstrasikan prinsip-prinsip NVC dalam makalah ini. Pseudokode ini mencerminkan langkah-langkah yang diterapkan dalam implementasi praktis, memastikan keselarasan penuh antara deskripsi teoretis dalam makalah, logika pemrograman yang dikembangkan, dan hasil evaluasi yang diperoleh. Dengan demikian, pseudokode ini tidak hanya memberikan gambaran yang jelas tentang proses kompresi, tetapi juga menjembatani analisis teoretis dengan realisasi teknisnya dalam konteks pemrosesan sinyal tradisional.

1. **Fungsi Kompresi DCVC:** Mengimplementasikan frame differencing dan DCT untuk efisiensi DCVC (Bagian 5.2.1):

| DCVC FRAME COMPRESSION | |
|------------------------|---|
| 1 | Function compress_frame_dcvc(frame, quant_step, prev_frame): |
| 2 | If prev_frame exist: |
| 3 | diff_frame ← frame - prev_frame |
| 4 | Else: |
| 5 | diff_frame ← frame |
| 6 | dct_frame ← DCT (diff_frame) |
| 7 | Return quant_frame, ukuran_frame, quant_step |

Fungsi ini menangkap redundansi temporal melalui selisih frame (Bagian 2.1) dan mengompresi secara spasial dengan DCT, diikuti kuantisasi, mendukung penghematan bitrate DCVC (~20%, Bagian 4.1).

2. **Kompresi Hybrid Entropy:** Menambahkan pengkodean Huffman untuk redundansi (Bagian 5.2.2):

| HYBRID ENTROPY COMPRESSION | |
|----------------------------|--|
| 1 | Function compress_frame_entropy(frame, quant_step, prev_frame): |
| 2 | quant_frame, ukuran_frame, quant_step ← compress_frame_dcvc(frame, quant_step, prev_frame) |
| 3 | flat_quant ← flatten (quant_frame) |
| 4 | encoded, codebook ← huffman_encode (flat_quant) |
| 5 | Return encoded, ukuran_frame, quant_step, codebook |

Huffman coding mengoptimalkan distribusi simbol setelah DCT/kuantisasi, mencerminkan efisiensi entropi hibrida dalam mengurangi ukuran data (Bagian 2.2).

3. **Pengaturan SlimVC:** Untuk fleksibilitas SlimVC (Bagian 5.2.3), tingkat kuantisasi bervariasi, diterapkan dalam logika utama sebagai berikut:

| SLIMVC CONFIGURATION | |
|----------------------|---|
| 1 | Inisialisasi quant_levels ← [5, 20] |
| 2 | For each i, frame in frames: |
| 3 | quant_step ← quant_levels[i mod 2] |
| 4 | compressed ← compress_frame_dcvc(frame, quant_step, prev_frame) |

Dengan mengganti tingkat kuantisasi secara bergantian, implementasi ini meniru kemampuan SlimVC untuk menyesuaikan bitrate dan sumber daya (Bagian 4.3). Nilai 5 memberikan kualitas tinggi dengan ukuran lebih besar, sedangkan 20 menghasilkan kompresi lebih agresif, sesuai dengan keunggulan adaptabilitas SlimVC.

4. **Dekompresi INR:** Interpolasi untuk dekoding ringan (Bagian 5.2.4):

| INR DECOMPRESSION | |
|-------------------|--|
| 1 | Function decompress_frame_inr(key_frames, frame_idx, ukuran_frame): |
| 2 | If frame_idx ada di key_frames: |
| 3 | Return key_frames[frame_idx] |
| 4 | key_indices ← sort (kunci(key_frames)) |
| 5 | For i from 0 hingga length (key_indices) - 2: |
| 6 | If key_indices[i] < frame_idx < key_indices[i + 1]: |
| 7 | t ← (frame_idx - key_indices[i]) / (key_indices[i + 1] - key_indices[i]) |
| 8 | prev_frame ← key_frames[key_indices[i]] |
| 9 | next_frame ← key_frames[key_indices[i + 1]] |
| 10 | Return (1 - t) × prev_frame + t × next_frame |
| 11 | Return key_frames[key_indices[0]] |

Interpolasi linier antar frame kunci mengurangi kebutuhan komputasi, mencerminkan keunggulan INR seperti HiNeRV untuk dekoding cepat (113 detik, Bagian 4.2).

5. **Region-of-Interest Compression**

| ROI COMPRESSION | |
|-----------------|--|
| 1 | Function compress_frame_roi(frame, base_quant_step, prev_frame): |
| 2 | edges ← edge_detection (frame, metode=Canny) |
| 3 | roi_mask ← edges > 0 |
| 4 | roi_mask ← dilation (roi_mask, kernel=5x5) |
| 5 | If prev_frame exist: |
| 6 | diff_frame ← frame - prev_frame |
| 7 | Else: |
| 8 | diff_frame ← frame |
| 9 | dct_frame ← DCT (diff_frame) |
| 10 | quant_frame ← empty_array (like dct_frame) |
| 11 | quant_frame[roi_mask] = 1] ← round(dct_frame[roi_mask] = 1] / (base_quant_step / 2)) |

| | |
|----|--|
| 12 | quant_frame[roi_mask = 0] ← round(dct_frame[roi_mask = 0] / (base_quant_step * 2)) / |
| 13 | Return quant_frame, ukuran_frame, quant_step |

6. **Evaluasi PSNR:** Fungsi berikut menghitung PSNR untuk mengevaluasi kualitas visual (Bagian 5.3):

| PSNR EVALUATION | |
|-----------------|---|
| 1 | Function compute_psnr(original, reconstructed): |
| 2 | Return peak_signal_noise_ratio(original, reconstructed, rentang=255) PSNR memastikan kualitas rekonstruksi, selaras dengan metrik evaluasi makalah (Bagian 3.2), dengan target di atas 30 dB untuk kualitas kompetitif. |

Implementasi ini menggunakan Python dengan library standar (OpenCV, NumPy, scikit-image), menghindari kecerdasan buatan, dan menawarkan demonstrasi praktis prinsip-prinsip NVC dalam konteks sederhana. Kode lengkap dapat diperluas untuk kebutuhan lebih lanjut, namun cuplikan ini menyoroti teknik inti yang mendukung temuan teoretis makalah. Kode lengkap tersedia di [\[https://github.com/sammtan/UTSPSMM-VideoCompression\]](https://github.com/sammtan/UTSPSMM-VideoCompression) untuk referensi dan pengujian lebih lanjut. Cukup hanya dengan menjalankan script main.py.

VI. KESIMPULAN

Perkembangan Neural Video Compression (NVC) dalam beberapa tahun terakhir menandai pergeseran besar dalam teknologi kompresi video, dari pendekatan tradisional berbasis aturan tetap menuju sistem yang mengandalkan pembelajaran data secara end-to-end. Berbeda dengan standar konvensional seperti H.264, H.265, atau H.266 yang telah dioptimalkan selama puluhan tahun, NVC menawarkan pendekatan yang lebih fleksibel, adaptif, dan siap berkembang seiring meningkatnya kebutuhan akan efisiensi dan kemampuan personalisasi dalam kompresi video.

Dengan memanfaatkan kekuatan deep learning, NVC mampu mempelajari representasi laten dari data video secara efisien dan menangkap korelasi spasial-temporal yang kompleks—sesuatu yang sulit dicapai oleh metode kompresi klasik. Tidak hanya memungkinkan penghematan bitrate yang signifikan, NVC juga memperkenalkan fleksibilitas baru, seperti kontrol bitrate yang lebih halus, adaptasi terhadap konten spesifik, dan kemampuan untuk dioptimalkan pada berbagai platform perangkat keras, mulai dari server pusat hingga perangkat edge yang terbatas sumber dayanya.

Berdasarkan kajian terhadap tiga model utama dalam literatur, dapat disimpulkan bahwa setiap pendekatan memiliki kekuatan yang khas. Model seperti DCVC dan variannya menunjukkan performa tertinggi dalam efisiensi kompresi, meskipun diiringi oleh kompleksitas komputasi yang tinggi. Pendekatan hybrid entropy modelling menawarkan keseimbangan antara akurasi dan efisiensi, serta mendukung fleksibilitas pengaturan bitrate. Sementara itu, model seperti SlimVC membuka peluang besar dalam aplikasi dunia nyata berkat strukturnya yang dapat disesuaikan secara dinamis. Adapun pendekatan berbasis representasi implisit (INR) memberikan solusi decoding yang sangat ringan dan cepat, menjadikannya kandidat ideal untuk deployment pada perangkat mobile dan embedded, meski saat ini masih tertinggal dalam efisiensi kompresi.

Secara keseluruhan, hasil analisis menunjukkan bahwa tidak ada satu pendekatan yang superior dalam semua aspek. Sebaliknya, arah penelitian dan pengembangan NVC ke depan sebaiknya difokuskan pada integrasi kekuatan dari berbagai pendekatan, guna menciptakan sistem kompresi video yang tidak hanya efisien secara teknis, tetapi juga relevan dan aplikatif untuk kebutuhan dunia nyata yang semakin beragam.

REFERENCES

[1] M. Tarchouli, T. Guionnet, M. Riviere, and M. Raulet, "Neural Video Compression Overview, performance and challenges," Proceedings of the 4th Mile-High Video Conference, pp. 40–46, Feb. 2025. doi:10.1145/3715675.3715796

[2] Z. Liu et al., "A slimmable framework for practical neural video compression," Neurocomputing, vol. 610, p. 128525, Dec. 2024. doi:10.1016/j.neucom.2024.128525

[3] J. Li, B. Li, and Y. Lu, "Hybrid spatial-temporal entropy modelling for neural video compression," Proceedings of the 30th ACM International Conference on Multimedia, Oct. 2022. doi:10.1145/3503161.3547845

[4] Z. Montajabi, V. Ghassab, and N. Bouguila, "Invertible neural network-based video compression," Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods, pp. 558–564, 2023. doi:10.5220/0011807600003411

[5] Z. Chen et al., "Neural video compression with spatio-temporal cross-covariance transformers," Proceedings of the 31st ACM International Conference on Multimedia, pp. 8543–8551, Oct. 2023. doi:10.1145/3581783.3611960