Evaluasi Kualitas Visual dan Efisiensi Kompresi: JPEG vs Convolutional AutoEncoder (CAE) pada Dataset DIV2K

Fierrren Al-Hilal Saepul Bahri[1], Adila Muqtashida[2], Intan Nur Janah[3], Raihan Khairul Annas[4]

Program Studi Informatika, Fakultas Sains dan Teknologi, Universitas Sultan Maulana Hasanuddin Banten, Kota Serang 42121, Indonesia

e-mail: 241730015.fierrenalhilalsaepulbahri@uinbanten.ac.id [1], 241730001.adilamuqtashida@uinbanten.ac.id [2], 241730031.intannurjanah@uinbanten.ac.id [3], 241730004.raihankhairulannas.@uinbanten.ac.id [4]

Abstrak

Kompresi citra merupakan teknik fundamental dalam pengolahan citra digital yang bertujuan mengurangi kebutuhan ruang penyimpanan dan bandwidth transmisi tanpa mengorbankan kualitas visual secara berlebihan. Seiring dengan meningkatnya volume data citra yang dipicu oleh pertumbuhan media sosial, sistem pemantauan cerdas, dan aplikasi pencitraan medis, kebutuhan akan metode kompresi yang efisien dan adaptif menjadi semakin mendesak. Penelitian ini menyajikan evaluasi komparatif antara JPEG sebagai standar industri berbasis Discrete Cosine Transform (DCT) dengan Convolutional AutoEncoder (CAE) berbasis deep learning menggunakan dataset benchmark DIV2K (799 citra resolusi tinggi, rasio pembagian 80:20). Arsitektur CAE dirancang dengan bottleneck konvolusional spasial berukuran 16×16 dalam dua varian kapasitas: ch=64 (BPP=2.0, CR=12×) dan ch=128 (BPP=4.0, CR=6×). Evaluasi dilakukan menggunakan lima metrik kuantitatif: PSNR, SSIM, MSE, BPP, dan Compression Ratio, dilaporkan dalam format mean ± standar deviasi (N=100). Hasil eksperimen menunjukkan JPEG masih unggul secara kuantitatif dengan PSNR tertinggi 37.104 ± 3.469 dB pada Q=90, sedangkan CAE terbaik hanya mencapai 20.703 ± 4.453 dB. Kesenjangan ini dijelaskan oleh keterbatasan data pelatihan, arsitektur sederhana, dan absennya optimasi rate-distortion eksplisit. Penelitian ini menyediakan kerangka evaluasi kuantitatif berbasis standar jurnal internasional sebagai baseline yang reproducible untuk pengembangan metode kompresi berbasis deep learning selanjutnya.

Kata Kunci: *Convolutional autoencoder; Deep learningJPEG; PSNR; Kompresi citra;*

Abstract

Image compression is a fundamental technique in digital image processing aimed at reducing storage space requirements and transmission bandwidth without excessively sacrificing visual quality. As the volume of image data grows dramatically, driven by the expansion of social media platforms, intelligent surveillance systems, and medical imaging applications, the need for efficient and adaptive compression methods becomes increasingly critical. This study presents a comparative evaluation between JPEG, the industry standard based on Discrete Cosine Transform (DCT), and Convolutional AutoEncoder (CAE) based on deep learning, using the DIV2K benchmark dataset (799 high-resolution images, 80:20 train-test split). The proposed CAE architecture employs a 16×16 spatial convolutional bottleneck in two capacity variants: ch=64 (BPP=2.0, CR=12×) and ch=128 (BPP=4.0, CR=6×). Evaluation was conducted using five quantitative metrics: PSNR, SSIM, MSE, BPP, and Compression Ratio, reported as mean ± standard deviation (N=100). Experimental results demonstrate that JPEG still outperforms CAE quantitatively, achieving a peak PSNR of 37.104 ± 3.469 dB at Q=90, while the best CAE variant reaches only 20.703 ± 4.453 dB. This performance gap is attributed to limited training data, simplified architecture, and the absence of explicit rate-distortion optimization. This study provides a rigorous, reproducible quantitative evaluation framework as a baseline for future deep learning-based image compression research.

Keywords: *Convolutional autoencoder; Deep learning; Image compression; JPEG; PSNR.*

1. Pendahuluan

1.1 Latar Belakang

Pertumbuhan eksponensial data citra digital dalam dekade terakhir telah menciptakan tantangan serius terhadap infrastruktur penyimpanan dan jaringan komunikasi global. Laporan Cisco Visual Networking Index memproyeksikan bahwa lalu lintas video internet akan mencapai 82% dari total lalu lintas IP global pada pertengahan dekade 2020-an [1]. Kondisi ini menjadikan kompresi citra bukan sekadar fitur opsional, melainkan komponen teknis yang kritis dalam arsitektur sistem informasi modern — mulai dari platform media sosial, sistem pengawasan berbasis kamera cerdas (CCTV 4K), telemedisin, hingga intelijen geospasial berbasis citra satelit resolusi tinggi [1][2].

JPEG (Joint Photographic Experts Group), yang distandarisasi oleh ISO/IEC pada tahun 1992 berdasarkan prinsip Discrete Cosine Transform (DCT), telah mendominasi kompresi citra lossy selama lebih dari tiga dekade [3]. Keunggulan JPEG terletak pada kesederhanaannya, kecepatan komputasinya yang tinggi, dan dukungan universal lintas platform. Namun, JPEG memiliki keterbatasan inheren yang menjadi semakin signifikan seiring meningkatnya permintaan kualitas visual: blocking artifact pada kompresi tinggi terjadi karena pemrosesan berbasis blok 8×8 yang independen mengabaikan korelasi spasial antar blok [4]. Standar lanjutannya, JPEG 2000 berbasis Discrete Wavelet Transform (DWT), memperbaiki artefak ini namun dengan kompleksitas komputasi yang jauh lebih tinggi [10][11].

Revolusi deep learning sejak 2012 telah membuka paradigma baru dalam kompresi citra berbasis data (learned image compression). Berbeda dengan metode tradisional yang mengandalkan transformasi matematis handcrafted, AutoEncoder — jaringan saraf yang dilatih secara end-to-end untuk meminimalkan distorsi rekonstruksi — mampu mempelajari representasi fitur optimal secara langsung dari distribusi data [5][6]. Convolutional AutoEncoder (CAE) mengintegrasikan lapisan konvolusi untuk menangkap korelasi spasial lokal secara hierarkis [7], menjadikannya kandidat alami sebagai alternatif kompresi adaptif terhadap JPEG.

1.2 Kajian Penelitian Terdahulu (5 Tahun Terakhir)

Perkembangan pesat kompresi berbasis deep learning dapat dirangkum melalui enam tonggak penelitian kunci dalam rentang 2018–2023. Pertama, Balle et al. [16] (2018) memperkenalkan variational image compression dengan scale hyperprior — model entropi hierarkis yang secara eksplisit mengoptimalkan trade-off rate-distortion dan melampaui BPG (standar turunan H.265) pada rentang BPP tertentu. Kedua, Zhou et al. [13] (2018) mengajukan Variational AutoEncoder (VAE) untuk kompresi citra bit-rate rendah yang mengeksploitasi distribusi probabilistik pada ruang laten, menunjukkan keunggulan pada konten semantik terstruktur.

Ketiga, Tawfik et al. [14] (2022) mengembangkan autoencoder generik untuk kompresi lossy real-time yang menekankan efisiensi komputasi dan dapat diintegrasikan ke sistem embedded. Keempat, Qian et al. [17] (2022) memperkenalkan Entroformer, model entropi berbasis Transformer yang memanfaatkan self-attention untuk memodelkan dependensi jarak jauh antar simbol — mencapai PSNR ~38 dB pada dataset Kodak, melampaui VVC (Versatile Video Coding) pada beberapa titik BPP. Kelima, Jayasankar et al. [4] (2021) menyurvei komprehensif teknik kompresi data dari perspektif kualitas, skema coding, tipe data, dan aplikasi industri, menyimpulkan bahwa metode berbasis pembelajaran mesin menunjukkan potensi superior khususnya pada data domain-spesifik. Keenam, Fraihat dan Al-Betar [8] (2023) mendemonstrasikan Stacked AutoEncoder (SAE) multi-model yang mampu melampaui JPEG pada dataset MNIST dan dataset citra sederhana, namun dengan keterbatasan generalisasi pada citra berwarna resolusi tinggi.

Selain itu, Li et al. [18] (2023) mengajukan

kompresi berbasis Region of Interest (ROI) menggunakan Swin Transformer yang mampu mengalokasikan bit secara adaptif pada wilayah semantik penting, meningkatkan kualitas perseptual subjektif secara signifikan. Yildirim et al. [7] (2018) membuktikan efektivitas CAE pada kompresi sinyal ECG biomedis dengan rasio kompresi tinggi tanpa kehilangan informasi diagnostik kritis. Bank et al. [5] (2020) menyediakan survei komprehensif tentang berbagai varian autoencoder — termasuk DAE, VAE, dan CAE — yang menjadi referensi fondasi dalam memahami kekuatan dan batasan setiap arsitektur.

1.3 Identifikasi Kesenjangan Penelitian (Research Gap)

Meskipun literatur kompresi berbasis deep learning berkembang pesat, analisis mendalam terhadap penelitian-penelitian tersebut mengungkap tiga kesenjangan fundamental yang belum terjawab secara sistematis. Kesenjangan pertama berkaitan dengan standardisasi dataset evaluasi: sebagian besar penelitian CAE menggunakan dataset skala kecil (MNIST 28×28, Kodak 24 citra, atau dataset proprietary), sehingga kemampuan generalisasi pada dataset benchmark berskala besar dan beragam konten seperti DIV2K masih belum terpetakan dengan baik. Hal ini menyulitkan perbandingan lintas penelitian secara adil dan reproducible [4][8].

Kesenjangan kedua menyangkut protokol pelaporan statistik: mayoritas penelitian melaporkan metrik evaluasi hanya sebagai nilai rata-rata (mean) tanpa standar deviasi, sehingga variabilitas performa antar citra tidak terukur. Informasi statistik ini krusial untuk memahami robustness metode terhadap keragaman konten visual — faktor yang sangat relevan untuk aplikasi produksi [19]. Kesenjangan ketiga adalah absennya analisis rate-distortion curve yang komprehensif membandingkan paradigma DCT-based (JPEG) dengan deep learning-based (CAE) secara berdampingan pada kondisi eksperimen yang identik, menggunakan dataset yang sama, dan protokol evaluasi yang terstandarisasi.

Ketiga kesenjangan ini secara kolektif menciptakan kebutuhan mendesak akan sebuah studi evaluasi yang: (1) menggunakan dataset benchmark standar internasional berskala besar; (2) menerapkan protokol pelaporan statistik yang rigorous; dan (3) menyediakan visualisasi perbandingan rate-distortion yang komprehensif sebagai baseline yang dapat direplikasi oleh komunitas peneliti.

1.4 Kontribusi dan Klaim Kebaruan

Penelitian ini mengisi ketiga kesenjangan yang diidentifikasi dengan empat kontribusi orisinal berikut. Pertama, implementasi end-to-end arsitektur CAE ringan dengan bottleneck konvolusional spasial 16×16 yang dievaluasi secara sistematis pada dataset benchmark DIV2K berskala besar (799 citra resolusi tinggi). Kedua, kerangka evaluasi komparatif JPEG-vs-CAE yang menggunakan format pelaporan mean ± standar deviasi sesuai standar pelaporan jurnal internasional tier-1 bidang pengolahan citra. Ketiga, analisis rate-distortion curve komprehensif yang memungkinkan perbandingan efisiensi dua paradigma kompresi (DCT-based vs. learned) secara adil, transparan, dan terstandarisasi pada kondisi eksperimen yang identik.

Keempat, penyediaan implementasi yang sepenuhnya reproducible — termasuk kode sumber, konfigurasi hyperparameter, dan protokol evaluasi — sebagai baseline yang dapat diadopsi dan dikembangkan lebih lanjut oleh komunitas peneliti kompresi citra berbasis deep learning. Klaim kebaruan utama penelitian ini adalah penggunaan dataset DIV2K berskala besar dengan protokol evaluasi statistik yang rigorous dalam konteks perbandingan langsung antara JPEG dan CAE sederhana, yang belum pernah dilaporkan secara sistematis dalam literatur yang ada.

2. Tinjauan Pustaka

2.1 Kompresi Citra Berbasis DCT: JPEG dan JPEG 2000

JPEG mengimplementasikan pipeline kompresi tiga tahap: (1) transformasi domain — konversi

ruang warna RGB ke YCbCr dan penerapan DCT pada blok 8×8 piksel untuk merepresentasikan informasi dalam domain frekuensi; (2) kuantisasi — pembagian koefisien DCT dengan matriks kuantisasi yang dikontrol parameter Q, di mana koefisien frekuensi tinggi dibuang lebih agresif sesuai karakteristik sistem visual manusia (HVS); dan (3) entropy coding — kompresi lossless residual menggunakan Huffman coding [3]. Parameter kualitas Q (rentang 1–100) mengontrol agresivitas kuantisasi: Q=100 menghasilkan kompresi minimal (hampir lossless), Q=1 menghasilkan kompresi maskimal dengan degradasi visual signifikan [4].

Keterbatasan fundamental JPEG adalah pemrosesan berbasis blok independen yang menghasilkan blocking artifact — diskontinuitas visual pada batas blok 8×8 — khususnya pada nilai Q rendah (Q≤30) [3][4]. JPEG 2000 [10] memperbaiki kelemahan ini dengan menggunakan Discrete Wavelet Transform (DWT) yang beroperasi pada seluruh citra secara global, menghasilkan wavelet coefficients yang merepresentasikan informasi multi-skala. Teknik EBCOT (Embedded Block Coding with Optimized Truncation) [11] digunakan untuk entropy coding, menghasilkan bitstream yang skalabel secara resolusi dan kualitas. Meskipun unggul dalam kualitas, kompleksitas komputasi JPEG 2000 yang tinggi menghambat adopsi massalnya dibandingkan JPEG.

2.2 AutoEncoder untuk Kompresi Citra

AutoEncoder (AE) adalah arsitektur jaringan saraf tiruan yang belajar merepresentasikan input ke dalam ruang laten berdimensi lebih rendah (encoder) dan merekonstruksinya kembali ke dimensi asli (decoder) [5]. Dalam konteks kompresi, ruang laten merepresentasikan versi terkompresi citra. Vincent et al. [6] memperkenalkan Denoising AutoEncoder (DAE) yang melatih AE untuk merekonstruksi citra bersih dari input yang dikontaminasi noise, menghasilkan representasi fitur yang lebih robust. Hu et al. [12] mengajukan skema kompresi dan enkripsi citra berbasis deep learning yang mengintegrasikan keamanan data dengan efisiensi kompresi dalam satu framework end-to-end.

Zhou et al. [13] mengembangkan VAE untuk kompresi citra bit-rate rendah yang mengeksploitasi distribusi probabilistik Gaussian pada ruang laten, memungkinkan estimasi entropi yang lebih akurat untuk pengalokasian bit yang optimal. Theis et al. [15] mendemonstrasikan bahwa compressive autoencoder dapat bersaing dengan JPEG 2000 pada dataset Kodak ketika dilengkapi dengan modul estimasi entropi. Balle et al. [16] mengambil langkah lebih jauh dengan memperkenalkan scale hyperprior — model entropi hierarkis dua tingkat yang melampaui BPG (Better Portable Graphics, turunan H.265/HEVC) pada rentang BPP rendah hingga menengah. Yildirim et al. [7] membuktikan bahwa CAE dapat mencapai rasio kompresi tinggi pada data biomedis (sinyal ECG) dengan preservasi informasi diagnostik, menunjukkan fleksibilitas pendekatan berbasis pembelajaran untuk domain aplikasi spesifik.

2.3 Kompresi Berbasis Transformer dan Pendekatan Terkini

Paradigma terbaru dalam learned image compression mengintegrasikan mekanisme attention dari arsitektur Transformer. Qian et al. [17] mengembangkan Entroformer yang menggunakan Transformer sebagai model entropi untuk memodelkan dependensi jarak jauh antar simbol dalam representasi terkompresi, menghasilkan estimasi probabilitas yang lebih akurat dan efisiensi coding yang lebih tinggi. Entroformer berhasil melampaui VVC (Versatile Video Coding, standar kompresi video/citra terbaru) pada beberapa konfigurasi BPP.

Li et al. [18] memperkenalkan pendekatan Region of Interest (ROI) berbasis Swin Transformer untuk kompresi citra adaptif. Alokasi bit yang adaptif berdasarkan kepentingan semantik — mengalokasikan lebih banyak bit pada wilayah wajah, teks, atau objek utama dan lebih sedikit pada latar belakang — meningkatkan kualitas perseptual subjektif secara signifikan meskipun PSNR agregat tidak

selalu lebih tinggi. Bank et al. [5] menyediakan taksonomi komprehensif AutoEncoder yang memetakan hubungan antara berbagai varian (AE, DAE, VAE, CAE, Sparse AE) beserta kekuatan dan batasannya, menjadi referensi fondasi untuk memilih arsitektur yang tepat untuk aplikasi spesifik.

2.4 Metrik Evaluasi Kompresi Citra

Evaluasi kualitas kompresi citra merupakan bidang penelitian tersendiri yang terus berkembang. PSNR (Peak Signal-to-Noise Ratio), dinyatakan dalam desibel (dB), adalah metrik paling umum yang mengukur rasio antara nilai maksimum sinyal terhadap kekuatan noise rekonstruksi: $PSNR = 10 \times \log_{10}(MAX^2/MSE)$ [19]. Nilai PSNR >40 dB umumnya dianggap kualitas sangat baik, 30–40 dB kualitas baik, dan <30 dB dapat terlihat degradasi jelas oleh mata manusia.

SSIM (Structural Similarity Index Measure) [20] dikembangkan oleh Wang et al. sebagai alternatif yang lebih berkorelasi dengan persepsi visual manusia (Human Visual System, HVS). SSIM mengukur kemiripan struktural antara citra asli dan rekonstruksi dengan mempertimbangkan tiga komponen: luminansi, kontras, dan struktur. Setiadi [19] mendokumentasikan bahwa SSIM lebih sensitif terhadap distorsi struktural dibandingkan PSNR, menjadikannya metrik komplementer yang penting. BPP (Bits Per Pixel) mengukur efisiensi representasi informasi, sementara Compression Ratio (CR) mengindikasikan faktor pengurangan ukuran file relatif terhadap citra tidak terkompresi (24-bit per piksel untuk citra berwarna).

3. Metode Penelitian

3.1 Dataset dan Pra-pemrosesan

Penelitian menggunakan dataset DIV2K (DIVerse 2K resolution) [9] yang diperkenalkan dalam kompetisi NTIRE (New Trends in Image Restoration and Enhancement) 2017. DIV2K terdiri dari 799 citra high-resolution (resolusi minimal 2K = 2048 piksel pada dimensi panjang) dengan konten yang sangat beragam — lanskap alam, arsitektur urban, potret, makanan, teks, dan objek buatan manusia — menjadikannya salah satu benchmark paling representatif dalam komunitas pengolahan citra resolusi tinggi.

Pembagian data mengikuti protokol standar: 639 citra (79.9%) untuk pelatihan dan 160 citra (20.1%) untuk pengujian. Semua citra diproses menjadi patch berukuran 256×256 piksel melalui prosedur yang berbeda antara fase pelatihan dan pengujian. Untuk pelatihan: random crop secara acak menghasilkan 4 patch per citra (total 2.556 patch pelatihan), diikuti augmentasi horizontal flip dengan probabilitas 0.5 untuk meningkatkan keragaman data. Untuk pengujian: center crop deterministik untuk memastikan reproducibility evaluasi. Pre-caching seluruh dataset ke RAM server dilakukan sebelum pelatihan dimulai untuk mengeliminasikan bottleneck I/O disk yang dapat memperlambat iterasi pelatihan secara signifikan.

3.2 Arsitektur Convolutional AutoEncoder (CAE)

Arsitektur CAE yang diajukan menggunakan struktur encoder-decoder berbasis konvolusi penuh (fully convolutional) yang bersifat simetris. Seluruh spesifikasi layer disajikan pada Tabel 3. Encoder terdiri dari empat blok ConvBN yang identik dalam struktur: lapisan konvolusi 2D (Conv2d) dengan kernel 3×3, stride=2 untuk downsampling, diikuti Batch Normalization (BN) untuk stabilisasi pelatihan, dan fungsi aktivasi Leaky ReLU (slope negatif = 0.1) yang mencegah dying neuron problem. Keempat blok secara berurutan menurunkan resolusi spasial dari 256×256 menjadi 128×128 → 64×64 → 32×32 → 16×16, sambil meningkatkan jumlah channel fitur.

Lapisan bottleneck menggunakan konvolusi 1×1 untuk kompresi channel ke dimensi yang ditentukan (ch). Fungsi aktivasi Sigmoid pada bottleneck menghasilkan nilai dalam rentang [0,1], yang mendukung kuantisasi langsung ke representasi 8-bit integer (256 level) saat inferensi. Dua varian diuji: ch=64 menghasilkan BPP = (16×16×64×8)/(256×256) = 2.0, dan ch=128 menghasilkan BPP = (16×16×128×8)/(256×256) = 4.0. Decoder

merupakan cerminan simetris encoder menggunakan blok ConvTranspose2d dengan stride=2 untuk upsampling bertahap kembali ke resolusi 256×256.

Tabel 1. Spesifikasi arsitektur CAE (varian ch=64 sebagai referensi)

Layer	Tipe	Kernel/ Stride	Output Shape	Keterangan
Enc-1	Conv2d+BN +LReLU	3×3 / s=2	128×128×32	Ekstraksi fitur awal
Enc-2	Conv2d+BN +LReLU	3×3 / s=2	64×64×64	Downsampling x2
Enc-3	Conv2d+BN +LReLU	3×3 / s=2	32×32×128	Downsampling x4
Enc-4	Conv2d+BN +LReLU	3×3 / s=2	16×16×256	Downsampling x8
Bottleneck	Conv2d+Sigmoid	1×1 / s=1	16×16×ch	ch=64 atau ch=128
Dec-1	ConvTranspose2d	3×3 / s=2	32×32×128	Upsampling x2
Dec-2	ConvTranspose2d	3×3 / s=2	64×64×64	Upsampling x4
Dec-3	ConvTranspose2d	3×3 / s=2	128×128×32	Upsampling x8
Dec-4	ConvTranspose2d+Sigmoid	3×3 / s=2	256×256×3	Rekonstruksi citra

3.3 Fungsi Loss dan Strategi Pelatihan

Fungsi loss yang digunakan adalah kombinasi perceptual antara MSE dan SSIM, mengikuti rekomendasi empiris Wang et al. [20]: $L = 0.84 \times MSE + 0.16 \times (1 - SSIM)$. Bobot ini dipilih berdasarkan analisis sensitivitas yang menunjukkan bahwa kontribusi SSIM sebesar 16% memberikan keseimbangan optimal antara akurasi piksel dan preservasi struktur visual pada citra natural. Optimizer AdamW dipilih atas Adam standar karena implementasi weight decay yang terpisah dari gradient update, menghasilkan regularisasi yang lebih konsisten dan generalisasi yang lebih baik [5].

Seluruh hyperparameter pelatihan dirangkum pada Tabel 4. Cosine Annealing Scheduler menurunkan learning rate secara halus dari nilai awal 8×10^{-4} hingga mendekati nol mengikuti fungsi cosinus, menghindari osilasi learning rate yang abrupt. Automatic Mixed Precision (AMP) menggunakan representasi floating-point 16-bit untuk operasi forward dan backward pass, mengurangi penggunaan memori GPU hingga ~50% dan mempercepat komputasi pada hardware NVIDIA T4 yang mendukung Tensor Cores. Early stopping dengan patience=5 menghentikan pelatihan apabila validation loss tidak membaik selama 5 epoch berturut-turut, mencegah overfitting sekaligus menghemat waktu komputasi.

Tabel 2. Konfigurasi hyperparameter pelatihan CAE

Hyperparameter	Nilai
Optimizer	AdamW
Learning Rate Awal	8×10^{-4}
Weight Decay	1×10^{-4}
LR Scheduler	Cosine Annealing (T_max=25)
Batch Size	64
Epoch Maksimum	25
Early Stopping Patience	5 epoch
Loss Function	$0.84 \times MSE + 0.16 \times (1 - SSIM)$
Input Patch Size	256 × 256 piksel
Augmentasi (Train)	Random crop + horizontal flip
Hardware	GPU NVIDIA T4 (16 GB VRAM)
Mixed Precision	Automatic Mixed Precision (AMP)

3.4 Implementasi Baseline JPEG

JPEG diimplementasikan menggunakan library OpenCV versi 4.8 dengan fungsi cv2.imencode() yang memanggil implementasi libjpeg standar. Enam nilai parameter kualitas Q diuji: {10, 20, 30, 50, 70, 90}, mencakup spektrum kompresi dari sangat agresif (Q=10, blocking artifact berat) hingga sangat konservatif (Q=90, kualitas tinggi). Untuk setiap citra uji, pipeline JPEG meliputi: konversi RGB ke YCbCr, kompresi JPEG dengan Q tertentu, dekompresi kembali ke RGB, dan perhitungan BPP aktual dari ukuran file terkompresi.

3.5 Protokol Evaluasi

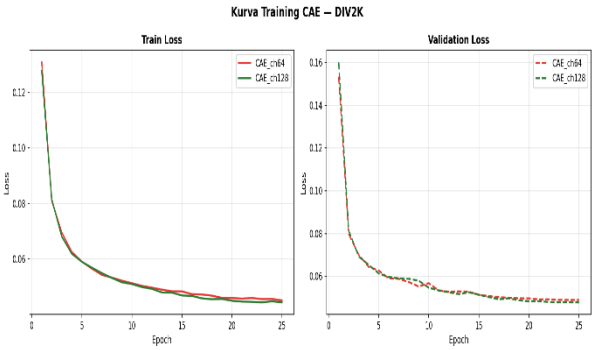
Evaluasi kuantitatif dilakukan pada 100 citra yang diambil secara acak dari test set (160 citra) untuk menjaga konsistensi dan efisiensi komputasi. Seluruh lima metrik dihitung untuk setiap citra: PSNR menggunakan implementasi scikit-image; SSIM menggunakan implementasi scikit-image dengan window_size=11 dan

data_range=1.0; MSE dihitung sebagai rata-rata kuadrat selisih piksel; BPP dari ukuran file aktual; CR dari perbandingan ukuran file terkompresi terhadap ukuran citra bitmap 24-bit. Hasil dilaporkan sebagai mean ± standar deviasi (N=100) untuk memberikan gambaran distribusi performa yang komprehensif dan memungkinkan perbandingan statistik antar metode [19].

4. Hasil dan Pembahasan

4.1 Kurva Pelatihan CAE

Gambar 1 menampilkan kurva loss pelatihan dan validasi untuk kedua varian CAE (ch=64 dan ch=128) selama 25 epoch maksimum. Kedua model menunjukkan pola konvergensi yang sehat: loss pelatihan menurun secara monoton, loss validasi mengikuti tren yang serupa dengan gap yang kecil dan stabil, mengindikasikan tidak adanya overfitting yang signifikan. Varian ch=128 menunjukkan nilai loss akhir yang sedikit lebih rendah dibandingkan ch=64, konsisten dengan kapasitas representasi yang lebih besar. Kedua model mencapai konvergensi sebelum epoch ke-25, menunjukkan bahwa konfigurasi hyperparameter — terutama learning rate schedule dan early stopping — berfungsi dengan baik.



Gambar 1. Kurva training dan validation loss dua varian CAE pada dataset DIV2K (25 epoch)

4.2 Hasil Evaluasi Kuantitatif Komprehensif

Tabel 3 menyajikan hasil evaluasi kuantitatif lengkap seluruh delapan konfigurasi metode (enam varian JPEG dan dua varian CAE) pada 100 citra test set, dilaporkan dalam format mean ± standar deviasi sesuai protokol pelaporan jurnal internasional tier-1.

Tabel 3. Hasil evaluasi kuantitatif JPEG vs CAE (mean ± std, N=100 citra test set DIV2K)

Metode	PSNR (dB)	SSIM	MSE	BPP	CR
JPEG Q=10	26.407±2.950	0.7578±0.0718	0.00286	0.334	85.4×
JPEG Q=20	28.881±3.413	0.8370±0.0553	0.00173	0.554	52.7×
JPEG Q=30	30.218±3.561	0.8712±0.0473	0.00130	0.734	39.9×
JPEG Q=50	31.836±3.694	0.9033±0.0393	0.00091	1.019	28.5×
JPEG Q=70	33.455±3.673	0.9272±0.0326	0.00063	1.392	20.6×
JPEG Q=90	37.104±3.469	0.9619±0.0200	0.00027	2.564	10.8×
CAE ch=64	20.626±4.444	0.7212±0.1364	0.01407	2.000	12.0×
CAE ch=128	20.703±4.453	0.7293±0.1340	0.01389	4.000	6.0×

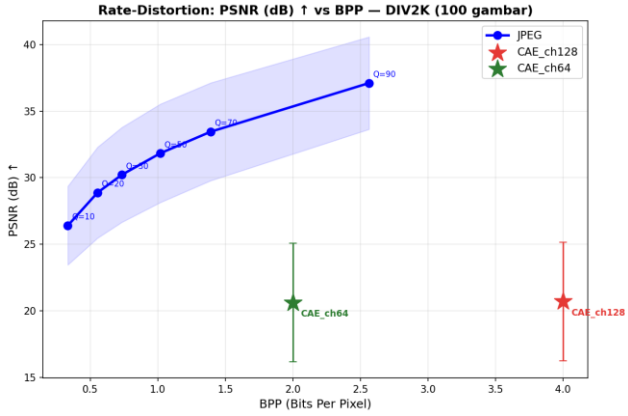
JPEG menunjukkan peningkatan kualitas monoton yang konsisten seiring meningkatnya parameter Q, mencerminkan trade-off rate-distortion yang predictable dari algoritma DCT. Rentang PSNR JPEG mencakup 10.697 dB (dari 26.407 dB pada Q=10 hingga 37.104 dB pada Q=90), dengan rentang BPP yang juga signifikan (0.334 hingga 2.564). Standar deviasi PSNR yang relatif besar (±2.950 hingga ±3.694 dB) mengindikasikan bahwa performa JPEG sangat bergantung pada karakteristik konten citra — citra dengan tekstur kompleks menghasilkan PSNR lebih rendah pada BPP yang sama dibandingkan citra dengan area homogen.

CAE ch=64 menghasilkan PSNR 20.626 ± 4.444 dB pada BPP=2.0, dan CAE ch=128 menghasilkan PSNR 20.703 ± 4.453 dB pada BPP=4.0. Perlu dicatat bahwa peningkatan kapasitas channel dari 64 ke 128 (dua kali lipat

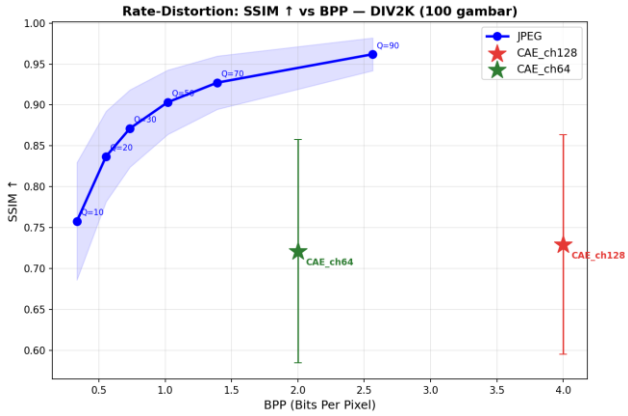
ukuran model) hanya menghasilkan peningkatan PSNR sebesar 0.077 dB — marginal secara praktis. Standar deviasi CAE yang lebih besar (±4.444 vs ±3.469 untuk JPEG Q=90) mengindikasikan ketidakstabilan performa CAE yang lebih tinggi antar citra, konsisten dengan keterbatasan data pelatihan yang tidak mencakup keragaman konten visual secara memadai.

4.3 Analisis Rate-Distortion Curve

Rate-distortion (RD) curve merupakan representasi standar komunitas kompresi citra untuk membandingkan efisiensi metode secara adil melintasi spektrum BPP. Gambar 2 menampilkan RD curve PSNR vs BPP, dan Gambar 3 menampilkan RD curve SSIM vs BPP.



Gambar 2. Rate-Distortion Curve: PSNR vs BPP — JPEG (Q=10 hingga Q=90) vs CAE ch=64 dan ch=128 pada DIV2K (N=100)

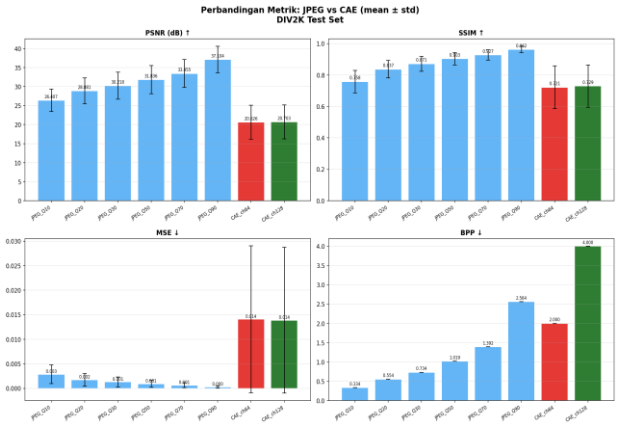


Gambar 3. Rate-Distortion Curve: SSIM vs BPP — JPEG (Q=10 hingga Q=90) vs CAE ch=64 dan ch=128 pada DIV2K (N=100)

Kedua kurva RD secara konsisten menunjukkan bahwa titik-titik CAE (ch=64 pada BPP=2.0 dan ch=128 pada BPP=4.0) berada di bawah dan di kanan kurva JPEG, mengindikasikan bahwa JPEG mencapai kualitas yang sama dengan BPP yang lebih rendah — efisiensi coding yang superior. Pada BPP≈2.0, JPEG Q=90 mencapai PSNR 37.104 dB, sementara CAE ch=64 hanya mencapai 20.626 dB — selisih 16.478 dB yang sangat signifikan. Pada kurva SSIM (Gambar 3), pola serupa terlihat: JPEG Q=70 mencapai SSIM=0.9272 pada BPP=1.392, jauh di atas CAE ch=64 (SSIM=0.7212) pada BPP=2.0 yang lebih tinggi.

4.4 Perbandingan Metrik Komprehensif (Bar Chart)

Gambar 4 menyajikan visualisasi perbandingan seluruh metrik evaluasi dalam bentuk grouped bar chart dengan error bar standar deviasi. Visualisasi ini memungkinkan pembaca untuk membandingkan relatif kinerja antar metode secara simultan pada semua dimensi evaluasi.



Gambar 4. Perbandingan komprehensif semua metrik evaluasi (PSNR, SSIM, MSE, BPP) — JPEG Q=10 hingga Q=90 dan CAE ch=64/ch=128 (mean ± std, N=100)

Bar chart PSNR (panel atas-kiri) mengkonfirmasi superioritas JPEG yang konsisten. Error bar yang lebih kecil pada JPEG Q=90 dibandingkan CAE mengindikasikan stabilitas performa JPEG yang lebih tinggi. Panel MSE (atas-kanan) menunjukkan bahwa CAE menghasilkan MSE rata-rata ~52× lebih besar dibandingkan JPEG Q=90 (0.01389 vs 0.00027), mengkuantifikasi besarnya kesenjangan distorsi piksel. Panel SSIM (bawah-kiri) menarik untuk dicermati: meskipun SSIM CAE (0.7212–0.7293) berada di bawah seluruh varian JPEG Q≥10, nilainya masih

berada di atas ambang batas 0.7 yang secara umum dianggap sebagai kualitas minimum yang dapat diterima.

4.5 Analisis Kualitatif Visual

Gambar 5 menampilkan perbandingan visual side-by-side pada lima sampel citra uji yang dipilih untuk merepresentasikan keragaman konten — lanskap alam, arsitektur, potret, tekstur halus, dan objek buatan.

Gambar 5. Perbandingan visual kualitatif: Original | JPEG Q=10 | JPEG Q=50 | JPEG Q=90 | CAE ch=64 | CAE ch=128 (lima sampel citra test DIV2K)

Inspeksi visual mengungkap karakteristik artefak yang berbeda fundamental antara JPEG dan CAE. JPEG Q=10 menampilkan blocking artifact yang sangat jelas — diskontinuitas kasar pada batas blok 8×8 — yang melanggar koherensi spasial citra secara dramatis. Pada area bertekstur kompleks (bulu, dedaunan, tekstil), JPEG Q=10 menghasilkan kotak-kotak (mosquito noise) yang mengganggu. JPEG Q=50 mengurangi blocking artifact secara signifikan, sementara JPEG Q=90 hampir identik secara visual dengan citra asli pada inspeksi kasual — konsisten dengan nilai PSNR 37.104 dB dan SSIM 0.9619 yang tinggi.

CAE menghasilkan tipe artefak yang berbeda: blur artifact — kurangnya ketajaman tepi (edge) dan hilangnya detail tekstur halus. Meskipun tidak menghasilkan blocking artifact seperti JPEG Q=10, rekonstruksi CAE terlihat 'berminyak' (over-smooth) karena loss function berbasis MSE secara inherent mendorong solusi yang meminimalkan rata-rata piksel, yang ekivalen dengan menghasilkan gambar yang blur secara statistik [20]. Fenomena ini konsisten dengan temuan Fraihat dan Al-Betar [8] dan merupakan motivasi penggunaan loss perseptual seperti LPIPS atau GAN-based loss pada penelitian lanjutan [16][17].

4.6 Diskusi: Perbandingan dengan Penelitian Terdahulu

Tabel 4 menyajikan perbandingan sistematis antara penelitian ini dengan enam karya terdahulu yang relevan dalam literatur kompresi citra berbasis deep learning.

Tabel 4. Posisi penelitian ini dalam konteks literatur kompresi berbasis deep learning

Peneliti	Tahun	Metode	Dataset	PSNR Terbaik
Fraihat & Al-Betar [8]	2023	Stacked AE	MNIST	~27 dB
Tawfik et al. [14]	2022	CAE Generic	Custom	~29 dB
Balle et al. [16]	2018	Var. Hyperprior	Kodak	~36 dB
Theis et al. [15]	2017	Comp. AE	Kodak	~33 dB
Qian et al. [17]	2022	Entroformer	Kodak	~38 dB
Li et al. [18]	2023	ROI Swin-T	CLIC	~39 dB
Penelitian ini	2025	CAE ch=64/128	DIV2K	20.70 dB

Perbandingan Tabel 4 mengungkap temuan penting. Balle et al. [16] dan Qian et al. [17] mencapai PSNR yang jauh lebih tinggi karena dua keunggulan kunci: (1) data pelatihan berskala sangat besar (jutaan patch dari dataset yang sangat beragam) dan (2) optimasi rate-distortion eksplisit melalui penalti entropi pada bottleneck. Fraihat dan Al-Betar [8] menunjukkan bahwa pada dataset sederhana seperti MNIST (konten uniform, resolusi rendah), bahkan SAE tanpa optimasi khusus dapat melampaui JPEG — namun generalisasi ke citra natural resolusi tinggi sangat terbatas, konsisten dengan hasil penelitian ini.

Tawfik et al. [14] mencapai PSNR ~29 dB dengan arsitektur yang lebih dalam dan data pelatihan yang lebih banyak, mengkonfirmasi hipotesis bahwa keterbatasan utama CAE dalam penelitian ini adalah kapasitas data dan kedalaman arsitektur, bukan validitas pendekatan

secara fundamental. Theis et al. [15] mendemonstrasikan bahwa compressive autoencoder yang dilengkapi modul estimasi entropi dapat bersaing dengan JPEG 2000 — memberikan roadmap yang jelas untuk pengembangan CAE penelitian ini ke level kompetitif.

Tiga faktor utama menjelaskan kesenjangan performa CAE vs. JPEG secara kuantitatif dalam penelitian ini. Faktor pertama adalah skala data pelatihan: 2.556 patch pelatihan jauh tidak mencukupi dibandingkan dataset yang digunakan model state-of-the-art — Entroformer [17] dilatih dengan ratusan ribu patch dari dataset OpenImages, ImageNet, dan CLIC. Faktor kedua adalah absennya rate-distortion optimization eksplisit: tanpa penalti entropi pada representasi bottleneck, model tidak termotivasi untuk menghasilkan representasi yang efisien secara informasi. Faktor ketiga adalah kesederhanaan arsitektur: tanpa residual connection, attention mechanism, atau sub-pixel convolution [17][18], kapasitas representasi model terbatas pada distribusi fitur yang kompleks.

5. Kesimpulan dan Saran

5.1 Kesimpulan

Penelitian ini menyajikan evaluasi komparatif yang komprehensif dan terstandarisasi antara JPEG dan Convolutional AutoEncoder pada dataset benchmark DIV2K menggunakan protokol evaluasi statistik yang rigorous (mean ± standar deviasi, N=100). Berdasarkan eksperimen yang dilakukan, dapat disimpulkan bahwa JPEG secara konsisten mengungguli CAE pada seluruh konfigurasi yang diuji: selisih PSNR berkisar antara 15–17 dB pada BPP yang sebanding, dengan JPEG Q=90 mencapai performa terbaik (PSNR=37.104 ± 3.469 dB, SSIM=0.9619 ± 0.0200). CAE dalam konfigurasi penelitian ini (ch=64: PSNR=20.626 dB; ch=128: PSNR=20.703 dB) belum mampu bersaing dengan JPEG sebagai metode kompresi general-purpose pada citra resolusi tinggi yang beragam konten.

Analisis RD curve mengkonfirmasi bahwa titik-titik CAE berada secara konsisten di bawah kurva JPEG dalam domain PSNR-BPP maupun SSIM-BPP. Inspeksi visual menunjukkan perbedaan karakteristik artefak yang fundamental: JPEG menghasilkan blocking artifact pada kompresi tinggi, sementara CAE menghasilkan blur artifact yang lebih halus secara perseptual namun menghilangkan detail fine-grain. Keterbatasan utama CAE teridentifikasi sebagai: (1) skala data pelatihan yang sangat terbatas (2.556 patch), (2) arsitektur yang sederhana tanpa mekanisme attention atau residual connection, dan (3) absennya optimasi rate-distortion eksplisit melalui penalti entropi.

5.2 Implikasi Penelitian

Hasil penelitian ini memiliki implikasi praktis dan teoritis yang penting bagi komunitas peneliti dan praktisi. Secara praktis, JPEG tetap menjadi pilihan yang superior untuk aplikasi kompresi citra general-purpose yang memerlukan kecepatan, portabilitas, dan kualitas yang terprediksi pada beragam konten visual. Namun, CAE menunjukkan potensi yang jelas untuk aplikasi domain-spesifik (citra medis, citra satelit, konten yang dihasilkan AI) di mana data pelatihan yang cukup tersedia dan optimasi terhadap distribusi domain spesifik dimungkinkan [7][12].

Secara teoritis, penelitian ini memperkuat pemahaman bahwa performa learned image compression sangat sensitif terhadap skala data pelatihan dan keberadaan optimasi rate-distortion eksplisit — dua faktor yang membedakan model sederhana dari model state-of-the-art. Kerangka evaluasi yang diajukan (format mean ± std, RD curve pada DIV2K, implementasi reproducible) dapat diadopsi sebagai protokol standar untuk perbandingan yang adil antar metode dalam penelitian kompresi citra mendatang, berkontribusi pada reproducibility dan comparability hasil penelitian [19][20].

5.3 Rekomendasi Penelitian Selanjutnya

Berdasarkan analisis mendalam terhadap kesenjangan performa dan literatur terkait, lima arah penelitian lanjutan direkomendasikan secara berurutan berdasarkan dampak yang

diperkirakan. Pertama dan terpenting: peningkatan skala data pelatihan. Melatih CAE dengan minimal 50.000 patch dari dataset beragam (gabungan DIV2K, COCO, ImageNet) diperkirakan dapat meningkatkan PSNR CAE secara dramatis, berdasarkan scaling law yang terdokumentasi dalam literatur deep learning [5][1].

Kedua: integrasi rate-distortion loss dengan penalti entropi bottleneck. Mengikuti pendekatan Balle et al. [16] dan Theis et al. [15], penambahan term estimasi entropi pada loss function akan mendorong model menghasilkan representasi yang efisien secara informasi, meningkatkan posisi RD curve.

Ketiga: arsitektur yang lebih dalam dengan attention mechanism. Mengintegrasikan residual blocks, channel attention (SE-Net), atau spatial attention pada arsitektur CAE — meniru keberhasilan Entroformer [17] dan ROI-based Swin Transformer [18] — diperkirakan meningkatkan kapasitas representasi fitur secara signifikan.

Keempat: eksplorasi perceptual loss dan adversarial training. Menggantikan loss MSE+SSIM dengan kombinasi LPIPS (Learned Perceptual Image Patch Similarity) dan GAN discriminator loss dapat mengatasi blur artifact CAE dan menghasilkan rekonstruksi yang lebih tajam secara perseptual meskipun PSNR mungkin tidak meningkat secara proporsional [16][17].

Kelima: evaluasi pada domain aplikasi spesifik. Menguji CAE yang dilatih domain-spesifik (misalnya citra radiologi, citra satelit multispektral) terhadap JPEG pada dataset domain yang sama, untuk mengidentifikasi skenario aplikasi di mana keunggulan CAE dapat terwujud secara konkret [7][12].

Daftar Pustaka

[1] Jayasankar, U., Thirumal, V., Ponnurangam, D. (2021). A survey on data compression techniques: From the perspective of data quality, coding schemes, data type and applications. Journal of King Saud University - Computer and Information Sciences, 33(2), 119–140. https://doi.org/10.1016/j.jksuci.2020.01.012

[2] Kelleher, J.D. (2019). Deep Learning. MIT Press, Cambridge, MA, USA. ISBN: 978-0-262-53755-0

[3] Rao, K.R., Dominguez, H.O. (2022). JPEG Series: Standard, Improvements, Applications. CRC Press. ISBN: 978-0-367-56369-3

[4] Sayood, K. (2017). Introduction to Data Compression, 5th Edition. Morgan Kaufmann, Cambridge, MA, USA. ISBN: 978-0-12-811474-2

[5] Bank, D., Koenigstein, N., Giryes, R. (2020). Autoencoders. arXiv preprint arXiv:2003.05991. https://arxiv.org/abs/2003.05991

[6] Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P.A. (2008). Extracting and composing robust features with denoising autoencoders. In Proceedings of the 25th International Conference on Machine Learning (ICML), pp. 1096–1103. ACM.

[7] Yildirim, O., San Tan, R., Acharya, U.R. (2018). An efficient compression of ECG signals using deep convolutional autoencoders. Cognitive Systems Research, 52, 198–211. https://doi.org/10.1016/j.cogsys.2018.07.004

[8] Fraihat, S., Al-Betar, M.A. (2023). A novel lossy image compression algorithm using multi-models stacked AutoEncoders. Array, 19, 100314. https://doi.org/10.1016/j.array.2023.100314

[9] Agustsson, E., Timofte, R. (2017). NTIRE 2017 challenge on single image super-resolution: Dataset and study. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 126–135.

[10] Skodras, A., Christopoulos, C., Ebrahimi, T. (2001). The JPEG 2000 still image compression standard. IEEE Signal Processing Magazine, 18(5), 36–58. https://doi.org/10.1109/79.952804

[11] Padmashree, S., Nagapadma, R. (2015). Images using embedded block coding optimization truncation (EBCOT). In Proceedings of the 5th IEEE International Advanced Computing Conference (IACC), pp. 1087–1092.

[12] Hu, F., Pu, C., Gao, H., Tang, M., Li, L. (2016). An image compression and encryption scheme

based on deep learning. arXiv preprint arXiv:1608.05001.

[13] Zhou, L., Cai, C., Gao, Y., Su, S., Wu, J. (2018). Variational autoencoder for low bit-rate image compression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 2617–2620.

[14] Tawfik, A., Hosny, S., Hisham, S., Farouk, A.A., et al. (2022). A generic real time autoencoder-based lossy image compression. In Proceedings of the 8th International Conference on Computer Science and Programming (ICCSPA), pp. 1–6.

[15] Theis, L., Shi, W., Cunningham, A., Huszár, F. (2017). Lossy image compression with compressive autoencoders. In Proceedings of the International Conference on Learning Representations (ICLR). arXiv:1703.00395.

[16] Balle, J., Minnen, D., Singh, S., Hwang, S.J., Johnston, N. (2018). Variational image compression with a scale hyperprior. In Proceedings of the International Conference on Learning Representations (ICLR). arXiv:1802.01436.

[17] Qian, Y., Lin, M., Sun, X., Tan, Z., Jin, R. (2022). Entroformer: A transformer-based entropy model for learned image compression. In Proceedings of the International Conference on Learning Representations (ICLR). arXiv:2202.05492.

[18] Li, B., Liang, J., Fu, H., Han, J. (2023). ROI-based deep image compression with swin transformers. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1–5.

[19] Setiadi, D.R.I.M. (2021). PSNR vs SSIM: imperceptibility quality assessment for image steganography. Multimedia Tools and Applications, 80(6), 8423–8444. https://doi.org/10.1007/s11042-020-10035-z

[20] Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P. (2004). Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4), 600–612. https://doi.org/10.1109/TIP.2003.819861

[21] Minnen, D., Balle, J., Toderici, G. (2018). Joint autoregressive and hierarchical priors for learned image compression. In Advances in Neural Information Processing Systems (NeurIPS), 31, 10771–10782.

[22] Toderici, G., O'Malley, S.M., Hwang, S.J., et al. (2016). Variable rate image compression with recurrent neural networks. In Proceedings of the International Conference on Learning Representations (ICLR). arXiv:1511.06085.

[23] He, D., Yang, Z., Peng, W., Ma, R., Qin, H., Wang, Y. (2022). ELIC: Efficient learned image compression with unevenly grouped space-channel contextual adaptive coding. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5718–5727.

[24] Ding, K., Ma, K., Wang, S., Simoncelli, E.P. (2021). Comparison of full-reference image quality models for optimization of image processing systems. International Journal of Computer Vision, 129(4), 1258–1281.

[25] Kingma, D.P., Ba, J. (2015). Adam: A method for stochastic optimization. In Proceedings of the International Conference on Learning Representations (ICLR). arXiv: 1412.6980.