Jurnal Rekayasa Sistem Informasi dan Teknologi
Volume 3, No 1 Agustus 2025
e-ISSN : 3025-888X

ATRIBUSI *GAMBAR* SINTETIK STYLEGAN2 MENGGUNAKAN *ARTIFICIAL FINGERPRINT* BERBASIS CNN

Alvin Yusuf Riziq[1], Sri Mulyono[2]
Universitas Islam Sultan Agung[1,2]
alvinyusufriziq@gmail.com[1*] sri.m@unissula.ac.id[2]

Received: 23-07- 2025	Revised: 08-08-2025	Approved: 15-08-2025

ABSTRAK

Model generatif seperti StyleGAN2 mampu menghasilkan gambar sintetik yang menyerupai gambar nyata, namun menimbulkan potensi pelanggaran hak cipta karena sering dilatih menggunakan data tanpa izin pemilik. Penelitian ini bertujuan untuk mengembangkan metode atribusi gambar dengan menyisipkan artificial fingerprint secara tersembunyi ke dalam dataset wajah sebelum digunakan dalam pelatihan model generatif. Metode yang digunakan berbasis Convolutional Neural Network (CNN) dengan dua komponen utama: encoder untuk menyisipkan fingerprint dan decoder untuk mendeteksinya kembali. Dataset yang digunakan adalah FFHQ beresolusi 128×128 piksel, dan model dilatih selama 10 epoch menggunakan algoritma Adam. Evaluasi dilakukan menggunakan Binary Cross Entropy (BCE) untuk mengukur akurasi deteksi fingerprint dan Mean Squared Error (MSE) untuk menilai kualitas visual gambar. Hasil menunjukkan bahwa metode ini berhasil menyisipkan fingerprint secara imperseptibel (MSE < 0.01) dan mengekstraksinya kembali dengan tingkat akurasi sangat tinggi (bitwise accuracy > 99%). Pendekatan ini memberikan kontribusi teknis dalam sistem atribusi otomatis dan perlindungan hak cipta pada pengembangan AI generatif.
Kata kunci: fingerprint, atribusi, CNN, StyleGAN2, hak cipta.

PENDAHULUAN

Perkembangan pesat teknologi kecerdasan buatan generatif, khususnya Generative Adversarial Networks (GAN) seperti StyleGAN2, telah menghadirkan paradigma baru dalam sintesis gambar digital. Model-model ini kini mampu menghasilkan citra wajah manusia yang sangat fotorealistik (Yu et al., 2021), sehingga sulit dibedakan dari foto asli. Penerapan luas model generatif ini menimbulkan kekhawatiran serius terkait pelanggaran hak cipta dan penyalahgunaan data (Wißmann et al., 2024). Banyak model generatif dilatih menggunakan dataset citra berskala besar yang sering kali tidak berlisensi sah (Franceschelli & Musolesi, 2022), sehingga keluaran model tersebut dapat sangat mirip dengan data pelatihan, menimbulkan risiko plagiarisme visual dan klaim kepemilikan konten (Wang et al., 2023).

Masalah utama dalam konteks ini adalah ketiadaan sistem atribusi yang andal untuk melacak asal-usul citra hasil generatif ke data atau model pembangunnya (Cassia et al., 2025). Tanpa mekanisme atribusi otomatis, penggunaan karya pihak lain tanpa izin sulit diidentifikasi, sehingga potensi pelanggaran royalti, plagiasi, dan kesulitan penegakan hukum hak cipta meningkat (Zhong et al., 2023). Penelitian-penelitian awal mencoba mengandalkan noise yang disebut artificial fingerprint statistik yang tertinggal pada gambar GAN untuk atribusi. Misalnya, Yu et al. (2019) menunjukkan bahwa setiap model GAN meninggalkan artificial fingerprint unik pada citra yang dihasilkannya (Yu et al., 2019). Namun, teknik deteksi pasif seperti ini hanya mampu mengidentifikasi jejak setelah gambar dihasilkan, dan terbukti rentan terhadap variasi arsitektur, fine-tuning, maupun evolusi model generatif (Song et al., 2024).

Untuk mengatasi keterbatasan tersebut, muncul pendekatan atribusi aktif melalui *artificial fingerprinting*. Metode ini sengaja menyisipkan artificial fingerprint tersembunyi ke dalam data pelatihan model generative (Yu et al., 2021). Sebagai contoh,

Yu et al. (2021) mengusulkan arsitektur encoder-decoder berbasis CNN untuk menyematkan fingerprint imperseptibel pada citra pelatihan GAN. Artificial fingerprint ini kemudian terbawa ke dalam model dan muncul kembali di semua citra sintetis yang dihasilkannya, sehingga sumber gambar dapat diatribusi secara langsung dengan mendeteksi artificial fingerprint tersebut (Yu et al., 2021). Metode artificial fingerprinting ini terbukti hanya menimbulkan perubahan minimal pada kualitas visual dan secara otomatis menyederhanakan tugas pendeteksian serta atribusi *deepfake*.

Merespons isu di atas, penelitian ini difokuskan pada pengembangan dan evaluasi metode artificial fingerprint berbasis CNN untuk citra wajah sintetis yang dihasilkan oleh StyleGAN2. Tujuannya adalah menyisipkan informasi atribusi ke dalam dataset pelatihan secara imperseptibel, sekaligus memastikan fingerprint tersebut dapat diekstraksi kembali dari citra hasil dengan akurasi tinggi. Dengan demikian, penggunaan data pelatihan dapat terlacak secara tersembunyi namun tepercaya, mendukung penegakan hak cipta dan akuntabilitas di era AI generatif. Adapun rumusan masalah penelitian ini adalah:

1. Bagaimana menyisipkan artificial fingerprint ke dalam citra wajah beresolusi tinggi menggunakan jaringan CNN berbasis arsitektur U-Net?
2. Bagaimana akurasi model decoder dalam mengekstraksi kembali fingerprint dari citra hasil encoding dengan tetap mempertahankan kualitas visual?

Dengan pendekatan ini, diharapkan dapat dikembangkan sistem atribusi otomatis yang tahan terhadap berbagai distorsi gambar, sebagai upaya perlindungan hak cipta dan tanggung jawab penggunaan data di era AI generatif.

METODE PENELITIAN

Dalam penelitian ini, metode yang digunakan adalah *Convolutional Neural Network* (CNN) dengan arsitektur yang menyerupai U-Net, sebagaimana diterapkan dalam model watermarking generative (Chen et al., 2025). Arsitektur U-Net terdiri atas dua komponen utama, yaitu *encoder* dan *decoder* (Ince et al., 2025). *Encoder* berfungsi untuk menyisipkan informasi *artificial fingerprint* ke dalam citra, sedangkan *decoder* bertugas mengekstraksi kembali informasi tersebut melalui proses konvolusi CNN guna mendeteksi *artificial fingerprint* yang telah disisipkan sebelumnya (Fang et al., 2023). Adapun tahapan yang harus dilakukan dalam penelitian ini antara lain:

Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari *dataset* publik Flickr-Faces-HQ (FFHQ), yakni Kumpulan 70.000 citra wajah beresolusi tinggi yang secara luas dimanfaatkan dalam pelatihan model generatif seperti StyleGAN (Matuzevičius, 2024). *Dataset* ini menyediakan beragam citra wajah manusia dengan variasi usia, ras, ekspresi, dan atribut lainnya, sehingga sangat sesuai untuk kebutuhan pelatihan model *Generative Adversarial Network* (GAN).

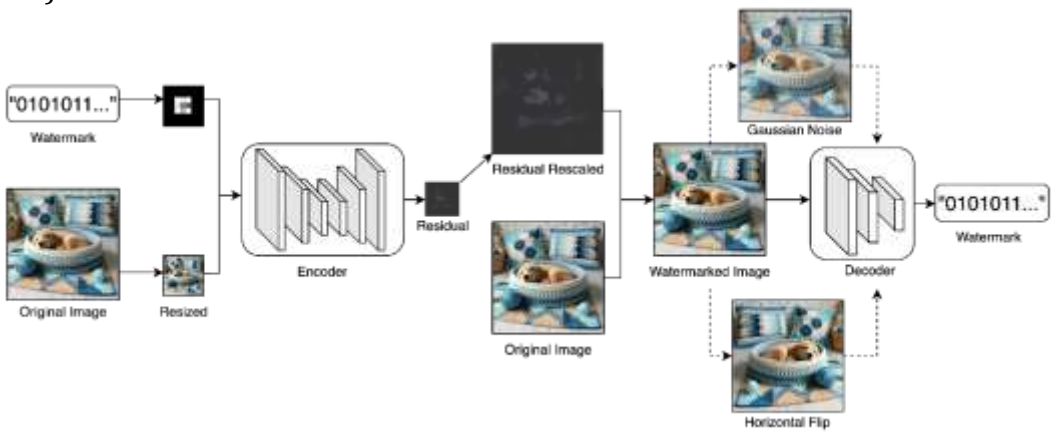
Dataset FFHQ bersifat sumber terbuka (*open source*) dan tersedia secara bebas untuk keperluan penelitian akademik melalui repositori resmi NVIDIA di GitHub. Dengan demikian, penelitian ini tidak memerlukan proses pengumpulan data primer melalui observasi langsung maupun eksperimen di lapangan.

Perancangan Model CNN

Model yang dikembangkan terdiri dari dua komponen utama, yaitu encoder dan decoder, yang dirancang untuk menyisipkan dan mengekstrak artificial fingerprint secara tersembunyi dari citra wajah menggunakan pendekatan steganografi berbasis

CNN.

Model yang dikembangkan menggunakan arsitektur encoder–decoder U-Net yang dimodifikasi untuk menyisipkan dan mengekstrak *artificial fingerprint* secara tersembunyi. Pada encoder, terdapat jalur *downsampling* (beberapa blok konvolusional) yang secara bertahap mengekstraksi fitur spasial dari citra input, dan jalur *upsampling* (transpose convolutions) yang merekonstruksi kembali citra dengan bantuan *skip connections* untuk menjaga detail resolusi tinggi (Zeng et al., 2023)(Wang, Byrnes, et al., 2023). Arsitektur U-Net ini efektif dalam mengekstraksi fitur frekuensi tinggi citra, sesuai dengan kebutuhan steganografi (penyisipan data tersembunyi) (Ji et al., 2025).



Gambar 1. Metode artificial fingerprinting yang menggunakan encoder dan decoder (Xu et al., 2025)

a. **Encoder** menggunakan arsitektur yang menyerupai U-Net, terdiri dari dua jalur utama: downsampling dan upsampling. Jalur downsampling terdiri atas beberapa blok konvolusional yang mengekstraksi fitur spasial dari citra input secara bertahap, sementara jalur upsampling merekonstruksi kembali citra dengan memanfaatkan *skip connection* untuk menjaga informasi resolusi tinggi. Artificial fingerprint berupa vektor biner diproyeksikan menjadi representasi spasial, kemudian dikombinasikan dengan citra input melalui channel-wise concatenation. Output encoder berupa residual image yang ditambahkan ke citra asli menghasilkan *stego image* yang secara visual tetap identik namun mengandung fingerprint tersembunyi.
b. **Decoder** bertugas mengekstraksi kembali artificial fingerprint dari gambar hasil generate. Arsitekturnya bersifat konvolusional murni dan terdiri dari dua tahap utama: blok konvolusional untuk ekstraksi fitur dan reduksi spasial, serta fully connected layer untuk memetakan fitur ke dalam vektor biner fingerprint. Pelatihan dilakukan secara end-to-end dengan fungsi kehilangan utama berupa Binary Cross Entropy (BCE), serta metrik evaluasi tambahan seperti Mean Squared Error (MSE) dan bitwise accuracy.

Desain sistem ini memungkinkan penyisipan fingerprint yang imperseptibel secara visual namun tetap dapat diekstraksi dengan akurasi tinggi, mendukung proses atribusi otomatis pada citra sintetik hasil model generatif seperti StyleGAN2.

Pelatihan Model

Model encoder dan decoder dilatih menggunakan dataset FFHQ beresolusi 128×128 piksel yang terdiri dari sekitar 70.000 citra wajah. Pelatihan dilakukan selama 10 epoch

dengan tujuan mengoptimalkan kemampuan encoder dalam menyisipkan artificial fingerprint ke dalam gambar, serta kemampuan decoder dalam mengekstraksi fingerprint tersebut secara akurat.
Proses pelatihan dilakukan dengan pasangan input berupa citra dan vektor fingerprint biner. Encoder menyisipkan fingerprint ke dalam gambar, sedangkan decoder mencoba memprediksi kembali fingerprint dari hasil gambar tersebut. Optimisasi dilakukan menggunakan algoritma Adam dengan gabungan dua fungsi kerugian utama:

- Binary Cross Entropy (BCE) untuk mengevaluasi akurasi deteksi fingerprint
- Mean Squared Error (MSE) untuk memastikan kualitas visual gambar tetap terjaga

Kedua metrik ini digunakan secara simultan untuk menjaga keseimbangan antara imperseptibilitas penyisipan dan akurasi deteksi. Nilai BCE dan MSE yang rendah menunjukkan bahwa sistem berhasil menyisipkan fingerprint secara tersembunyi tanpa menurunkan kualitas citra, serta dapat mengenalinya kembali secara konsisten (Tallam, 2025).

Evaluasi Model

Evaluasi performa model dilakukan dengan menggunakan dua metrik utama, yaitu Binary Cross Entropy (BCE) Loss dan Mean Squared Error (MSE) Loss, yang masing-masing merepresentasikan aspek akurasi deteksi fingerprint dan kualitas visual gambar.
BCE Loss digunakan untuk mengukur akurasi ekstraksi artificial fingerprint oleh decoder. Karena fingerprint direpresentasikan sebagai vektor biner, metrik ini dinilai paling tepat untuk mengevaluasi kesesuaian antara hasil prediksi dan ground truth. Nilai BCE yang rendah menunjukkan bahwa sistem berhasil mengenali kembali fingerprint secara akurat meskipun gambar telah melalui proses encoding yang kompleks (Terven et al., 2025).
Sementara itu, MSE Loss digunakan untuk menilai kemiripan visual antara gambar asli dan gambar hasil encoding. Nilai MSE yang rendah mengindikasikan bahwa penyisipan fingerprint berlangsung secara imperseptibel, tanpa menyebabkan degradasi visual yang dapat terdeteksi oleh mata manusia (Padhi et al., 2024)(Zhao et al., 2021). Hal ini krusial dalam konteks *invisible watermarking*, di mana kualitas gambar harus tetap terjaga.
Kedua metrik ini digunakan secara bersamaan untuk memastikan bahwa sistem tidak hanya mampu menyisipkan jejak digital secara tersembunyi, tetapi juga dapat mengenalinya kembali dengan akurasi tinggi. Efektivitas metode dinilai berhasil apabila kedua nilai loss menunjukkan hasil minimal secara konsisten.

HASIL DAN PEMBAHASAN

Dataset yang digunakan dalam penelitian ini adalah Flickr-Faces-HQ (FFHQ), yaitu kumpulan citra wajah beresolusi tinggi yang secara luas digunakan dalam pelatihan model generatif seperti StyleGAN. Dataset ini terdiri dari sekitar 70.000 gambar wajah manusia dengan variasi usia, ekspresi, pencahayaan, dan latar belakang yang beragam, sehingga sangat representatif untuk pelatihan model generatif maupun sistem penyisipan fingerprint berbasis deep learning.

Gambar 2. Dataset gambar FFHQ

Gambar 2 menampilkan contoh subset dari dataset FFHQ dengan resolusi yang telah disesuaikan sebesar 128×128 piksel. Karena dataset ini dirancang secara khusus untuk keperluan pelatihan model GAN, format dan kualitas data sudah sesuai dengan standar input yang dibutuhkan oleh model. Oleh karena itu, tidak dilakukan preprocessing tambahan terhadap gambar sebelum digunakan dalam proses pelatihan encoder, decoder, maupun model StyleGAN2.

Hasil Modeling

a. Encoder

Model encoder yang dikembangkan memiliki struktur menyerupai arsitektur U-Net, dengan penyesuaian untuk mendukung proses penyisipan artificial fingerprint ke dalam citra. Proses dimulai dengan mengubah vektor fingerprint berdimensi satu menjadi tensor dua dimensi berukuran 16×16, yang kemudian di-*upsample* hingga mencapai dimensi 128×128 agar sesuai dengan ukuran citra input. Fingerprint tersebut kemudian dikombinasikan secara channel-wise dengan gambar asli, menghasilkan tensor gabungan berukuran (B, 2, 128, 128) yang menjadi input utama jaringan.
Arsitektur encoder terdiri dari dua jalur utama:

- Encoder path (downsampling) bertugas mengekstraksi fitur dari tensor gabungan melalui beberapa lapisan konvolusional bertingkat dengan operasi strided convolution, menghasilkan representasi spasial yang lebih padat dan semantik.
- Decoder path (upsampling) berfungsi merekonstruksi fitur yang telah diekstrak dengan meningkatkan kembali dimensi spasial secara bertahap menggunakan teknik upsampling. Proses ini dibantu dengan *skip connection* dari jalur encoder untuk menjaga detail spasial resolusi tinggi.

Tabel 1.
Struktur Model Encoder

Tipe Layer	Input Shape	Output Shape
Input	(B, 2, 128, 128)	(B, 2, 128, 128)
Conv 1	(B, 2, 128, 128)	(B, 32, 128, 128)
Conv 2	(B, 32, 128, 128)	(B, 32, 64, 64)
Conv 3	(B, 32, 64, 64)	(B, 64, 32, 32)
Conv 4	(B, 64, 32, 32)	(B, 128, 16, 16)

Tipe Layer	Input Shape	Output Shape
Conv 5	(B, 128, 16, 16)	(B, 256, 8, 8)
Up 6 + Skip 4	(B, 256, 8, 8)	(B, 128, 16, 16)
Up 7 + Skip 3	(B, 128, 16, 16)	(B, 64, 32, 32)
Up 8 + Skip 2	(B, 64, 32, 32)	(B, 32, 64, 64)
Up 9 + Skip 1	(B, 32, 64, 64)	(B, 32, 128, 128)
Conv 10	(B, 32, 128, 128)	(B, 32, 128, 128)
Final Output	(B, 32, 128, 128)	(B, 1, 128, 128)

Output akhir dari jaringan adalah *residual image* yang kemudian ditambahkan secara elemen-per-elemen dengan citra input, menghasilkan gambar tersisip yang secara visual mirip dengan gambar asli namun mengandung fingerprint tersembunyi. Rincian struktur arsitektur encoder ditunjukkan pada Tabel 1.

b. Decoder

Model decoder memiliki arsitektur yang lebih sederhana, bersifat konvolusional murni dan fokus pada proses ekstraksi fingerprint dari gambar tersisip. Jaringan diawali dengan serangkaian lapisan konvolusi dan aktivasi ReLU untuk menangkap pola spasial dari citra input. Setelah itu, fitur hasil ekstraksi diratakan dan diproses oleh dua lapisan fully connected.

Tabel 2.
Struktur Model Decoder

Tipe Layer	Input Shape	Output Shape
Input	(1, 1, 128, 128)	(1, 32, 64, 64)
Conv2d + ReLU	(1, 32, 64, 64)	(1, 32, 64, 64)
Conv2d + ReLU	(1, 32, 64, 64)	(1, 64, 32, 32)
Conv2d + ReLU	(1, 64, 32, 32)	(1, 64, 32, 32)
Conv2d + ReLU	(B, 64, 32, 32)	(B, 128, 16, 16)
Conv2d + ReLU	(B, 128, 16, 16)	(B, 256, 8, 8)
Conv2d + ReLU	(B, 256, 8, 8)	(B, 128, 16, 16)
Conv2d + ReLU	(B, 128, 16, 16)	(B, 64, 32, 32)
Up 8 + Skip 2	(B, 64, 32, 32)	(B, 32, 64, 64)
Up 9 + Skip 1	(B, 32, 64, 64)	(B, 32, 128, 128)
Conv 10	(B, 32, 128, 128)	(B, 32, 128, 128)
Final Output	(B, 32, 128, 128)	(B, 1, 128, 128)

Lapisan dense pertama (FC1) bertugas melakukan pemetaan representasi fitur ke dimensi laten, sementara lapisan kedua (FC2) menghasilkan vektor fingerprint yang diprediksi. Decoder ini tidak menggunakan *skip connection* dan dirancang untuk memaksimalkan akurasi deteksi fingerprint dalam format biner. Struktur lengkap decoder dapat dilihat pada Tabel 2.

Hasil Pelatihan Model

Model Encoder dan Decoder dilatih menggunakan dataset citra wajah FFHQ beresolusi 128×128 piksel selama 10 epoch. Proses pelatihan menggunakan kombinasi dua fungsi kerugian, yaitu Binary Cross Entropy (BCE) untuk mengukur akurasi ekstraksi artificial fingerprint, serta Mean Squared Error (MSE) untuk mengevaluasi perbedaan visual antara citra asli dan citra hasil penyisipan.

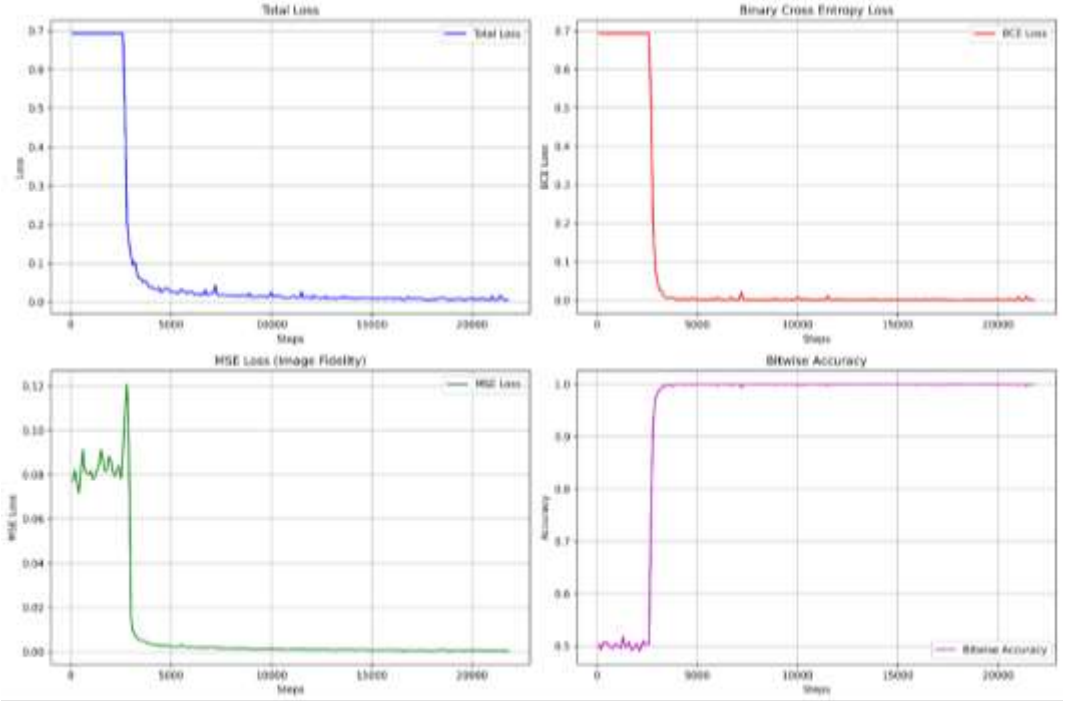
Penggunaan BCE bertujuan untuk memastikan bahwa fingerprint dapat diprediksi secara akurat dalam format biner oleh model decoder. Sementara itu, MSE

digunakan untuk menjaga agar proses penyisipan tidak menyebabkan distorsi visual yang signifikan, memastikan imperseptibilitas fingerprint dalam citra hasil encoding.

Selain dua fungsi kerugian tersebut, sistem juga dievaluasi menggunakan metrik tambahan berupa bitwise accuracy, yaitu persentase bit artificial fingerprint yang berhasil diprediksi dengan benar dibandingkan total bit. Metrik ini memberikan informasi yang lebih detail tentang ketepatan ekstraksi pada tingkat bit, yang sangat penting dalam konteks atribusi digital berbasis fingerprint.

Hasil Evaluasi Model

Evaluasi performa sistem dilakukan dengan mengamati tren perubahan pada metrik Binary Cross Entropy (BCE) Loss, Mean Squared Error (MSE) Loss, Total Loss (gabungan BCE dan MSE), serta *bitwise accuracy* selama proses pelatihan. Keempat metrik ini memberikan gambaran menyeluruh terhadap efektivitas sistem dalam menyisipkan dan mengekstraksi artificial fingerprint secara tersembunyi, tanpa mengorbankan kualitas visual citra.



Gambar 3. Grafik evaluasi model berdasarkan metrik Binary Cross Entropy (BCE) Loss, Mean Squared Error (MSE) Loss, Total Loss, dan Bitwise Accuracy.

Pada gambar 3, terlihat pada fase awal pelatihan, nilai BCE dan MSE relatif tinggi, menandakan bahwa model masih berada dalam tahap eksplorasi parameter. BCE berkisar di angka 0,7 dan menunjukkan penurunan signifikan setelah langkah ke-3000, mengindikasikan bahwa model mulai mampu mengenali pola artificial fingerprint secara konsisten. Sementara itu, MSE berada di kisaran 0,07–0,09 pada awalnya, lalu menurun secara bertahap, menandakan peningkatan kemampuan sistem dalam menjaga kualitas visual gambar.

Nilai total loss, sebagai gabungan dari kedua fungsi kerugian tersebut, memperlihatkan penurunan serupa dan stabilisasi pada titik rendah mendekati nol, yang mengindikasikan tercapainya konvergensi model. Hal ini menunjukkan keberhasilan sistem dalam menyelaraskan dua tujuan utama: akurasi deteksi fingerprint dan minimnya distorsi visual.

Metrik bitwise accuracy juga menunjukkan peningkatan signifikan. Awalnya berkisar 49–51%, mendekati tebakan acak, namun setelah langkah ke-3000 mengalami lonjakan hingga mendekati 100%. Hal ini menunjukkan bahwa hampir seluruh bit fingerprint berhasil diprediksi secara tepat oleh decoder. Kinerja ini menunjukkan bahwa sistem tidak hanya mampu mempelajari distribusi probabilistik fingerprint, tetapi juga merekonstruksinya secara presisi pada tingkat granular bit.

Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa arsitektur yang diusulkan berhasil menyisipkan artificial fingerprint secara imperseptibel dan mengenalinya kembali secara akurat, mendukung penerapan sistem dalam skenario atribusi digital yang menuntut presisi tinggi dan kualitas citra yang tetap terjaga.

KESIMPULAN

Penelitian ini berhasil mengembangkan sebuah sistem artificial fingerprinting berbasis Convolutional Neural Network (CNN) untuk atribusi gambar sintetik yang dihasilkan oleh StyleGAN2. Sistem terdiri dari dua komponen utama, yaitu encoder untuk menyisipkan artificial fingerprint secara tersembunyi ke dalam citra, dan decoder untuk mengekstraksi kembali fingerprint tersebut dalam format biner.

Model dilatih selama 10 epoch menggunakan dataset FFHQ beresolusi 128×128 piksel, dan menunjukkan performa optimal setelah melewati sekitar 3.000 iterasi. Evaluasi akhir menunjukkan nilai Binary Cross Entropy (BCE) sebesar 0,0071, serta bitwise accuracy sebesar 0,9996 pada epoch ke-10. Nilai tersebut mengindikasikan bahwa fingerprint dapat dikenali kembali dengan akurasi sangat tinggi, tanpa menimbulkan distorsi visual yang signifikan, sebagaimana ditunjukkan oleh nilai Mean Squared Error (MSE) yang rendah.

Hasil ini menunjukkan bahwa sistem yang diusulkan efektif dalam menyisipkan dan mengenali jejak digital secara imperseptibel, serta memiliki potensi untuk diterapkan dalam skenario atribusi digital dan perlindungan hak kekayaan intelektual, khususnya dalam konteks pengembangan dan distribusi konten berbasis AI generatif.

DAFTAR PUSTAKA

Cassia, M., Guarnera, L., Casu, M., Zangara, I., & Battiato, S. (2025). Deepfake Forensic Analysis: Source Dataset Attribution and Legal Implications of Synthetic Media Manipulation. *arXiv preprint*. http://arxiv.org/abs/2505.11110v1

Chen, Y., Vice, J., Akhtar, N., Haldar, N. A. H., & ... (2025). Image Watermarking of Generative Diffusion Models. *arXiv preprint arXiv …*, *14*(8), 1–11. https://arxiv.org/abs/2502.10465

Fang, H., Jia, Z., Qiu, Y., Zhang, J., Zhang, W., & Chang, E. C. (2023). De-END: Decoder-Driven Watermarking Network. *IEEE Transactions on Multimedia*, *25*, 7571–7581. https://doi.org/10.1109/TMM.2022.3223559

Franceschelli, G., & Musolesi, M. (2022). Copyright in generative deep learning. *Data and Policy*, *4*(3), 1–18. https://doi.org/10.1017/dap.2022.10

Ince, S., Kunduracioglu, I., Algarni, A., Bayram, B., & Pacal, I. (2025). Deep learning for cerebral vascular occlusion segmentation: A novel ConvNeXtV2 and GRN-integrated U-Net framework for diffusion-weighted imaging. *Neuroscience*, *574*(April), 42–53. https://doi.org/10.1016/j.neuroscience.2025.04.010

Ji, P., Zhang, Y., & Lv, Z. (2025). Edge-Guided Dual-Stream U-Net for Secure Image Steganography. *Applied Sciences (Switzerland)*, *15*(8). https://doi.org/10.3390/app15084413

Matuzevičius, D. (2024). Diverse Dataset for Eyeglasses Detection: Extending the Flickr-Faces-HQ (FFHQ) Dataset. *Sensors*, *24*(23). https://doi.org/10.3390/s24237697

Padhi, S. K., Tiwari, A., & Ali, S. S. (2024). Deep Learning-Based Dual Watermarking for Image Copyright Protection and Authentication. *IEEE Transactions on Artificial Intelligence*, *5*(12), 6134–6145. https://doi.org/10.1109/TAI.2024.3485519

Song, H. J., Khayatkhoei, M., & Abdalmageed, W. (2024). ManiFPT: Defining and Analyzing Fingerprints of Generative Models. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 10971–10981. https://doi.org/10.1109/CVPR52733.2024.01026

Tallam, K. (2025). *Embedding Trust at Scale: Physics-Aware Neural Watermarking for Secure and Verifiable Data Pipelines*. 1–21. http://arxiv.org/abs/2506.12032

Terven, J., Cordova-Esparza, D. M., Romero-González, J. A., Ramírez-Pedraza, A., & Chávez-Urbiola, E. A. (2025). A comprehensive survey of loss functions and metrics in deep learning. *Artificial Intelligence Review*, *58*(7). https://doi.org/10.1007/s10462-025-11198-7

Wang, Z., Byrnes, O., Wang, H., Sun, R., Ma, C., Chen, H., Wu, Q., & Xue, M. (2023). Data Hiding With Deep Learning: A Survey Unifying Digital Watermarking and Steganography. *IEEE Transactions on Computational Social Systems*, *10*(6), 2985–2999. https://doi.org/10.1109/TCSS.2023.3268950

Wang, Z., Lyu, L., Chen, C., Zeng, Y., & Ma, S. (2023). Where Did I Come From? Origin Attribution of AI-Generated Images. *Advances in Neural Information Processing Systems*, *36*(NeurIPS).

Wißmann, A., Zeiler, S., Nickel, R. M., & Kolossa, D. (2024). Whodunit: Detection and Attribution of Synthetic Images by Leveraging Model-specific Fingerprints. *ACM International Conference Proceeding Series*, 65–72. https://doi.org/10.1145/3643491.3660280

Xu, R., Hu, M., Lei, D., Li, Y., Lowe, D., Gorevski, A., Wang, M., Ching, E., & Deng, A. (2025). InvisMark: Invisible and Robust Watermarking for AI-generated Image Provenance. *Proceedings - 2025 IEEE Winter Conference on Applications of Computer Vision, WACV 2025*, 909–918. https://doi.org/10.1109/WACV61041.2025.00098

Yu, N., Davis, L., & Fritz, M. (2019). Attributing fake images to GANs: Learning and analyzing GAN fingerprints. *Proceedings of the IEEE International Conference on Computer Vision*, *2019-Octob*, 7555–7565. https://doi.org/10.1109/ICCV.2019.00765

Yu, N., Skripniuk, V., Abdelnabi, S., & Fritz, M. (2021). Artificial Fingerprinting for Generative Models: Rooting Deepfake Attribution in Training Data. *Proceedings of the IEEE International Conference on Computer Vision*, 14428–14437. https://doi.org/10.1109/ICCV48922.2021.01418

Zeng, L., Yang, N., Li, X., Chen, A., Jing, H., & Zhang, J. (2023). Advanced Image Steganography Using a U-Net-Based Architecture with Multi-Scale Fusion and Perceptual Loss. *Electronics (Switzerland)*, *12*(18), 1–18. https://doi.org/10.3390/electronics12183808

Zhao, X., Liu, H., Fan, W., Liu, H., Tang, J., & Wang, C. (2021). AutoLoss: Automated Loss Function Search in Recommendations. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 3959–3967. https://doi.org/10.1145/3447548.3467208

Zhong, H., Chang, J., Yang, Z., Wu, T., Mahawaga Arachchige, P. C., Pathmabandu, C., & Xue, M. (2023). Copyright Protection and Accountability of Generative AI: Attack,

Watermarking and Attribution. In *ACM Web Conference 2023 - Companion of the World Wide Web Conference, WWW 2023* (Vol. 1, Nomor 1). Association for Computing Machinery. https://doi.org/10.1145/3543873.3587321