Sibatik Journal
Jurnal Ilmiah Bidang Sosial, Ekonomi, Budaya, Teknologi, dan Pendidikan
E-ISSN: 2809-8544

ANALISIS GAMBAR WAJAH PALSU: MENDETEKSI KEASLIAN GAMBAR YANG DIMANIPULASI MENGGUNAKAN METODE *VARIATIONAL AUTOENCODER* DAN *FORENSICS DEEP NEURAL NETWORK*

ANALYSIS OF FAKE FACE IMAGES: DETECTING THE AUTHENTICITY OF MANIPULATED IMAGES USING VARIATIONAL AUTOENCODER METHODS AND DEEP NEURAL NETWORK FORENSICS

Regina Angelika Septi Rahayu[1], Handri Santoso[2]
Universitas Pradita, Indonesia
Email: regina.angelika@student.pradita.ac.id[1], handri.santoso@pradita.ac.id[2]

Abstract

Artificial intelligence (AI) is one of the technologies commonly used for automated computer systems. Artificial intelligence is designed to solve cognitive problems. The convenience provided by AI is sometimes misused, resulting in negative impacts on many people. One negative impact of AI technology misuse is deepfake. Deepfake is a technology used for image or video manipulation. The manipulation techniques used in deepfake are employed to alter images, such as faces, places, objects or even voices. Variational autoencoder (VAE) is a deep learning algorithm that can be used for facial manipulation. The result of the VAE process is an image obtained from the merging of original facial images during the training process. The new facial images generated from VAE training are called decoder images or manipulations. Both original facial images and manipulation facial images will be analyzed using the forensics deep neural network method. The analysis technique involves the use of error level analysis (ELA), which helps identify significant changes that occur in the images. Based on the testing results using both original and manipulated facial images, the applied method demonstrates the ability to detect real and fake images.

Keywords: *Artificial Intelligence (AI), Deepfake, Variational Autoencoder (VAE), Decoder, Forensics Deep Neural Network, Error Level Analysis (ELA)*

Abstrak

Artificial intelligence (AI) merupakan salah satu teknologi yang sering digunakan untuk sistem komputer otomatis. *Artificial intelligence* dirancang untuk melakukan pemecahan masalah kognitif. Kemudahan yang diberikan oleh AI terkadang sering disalahgunakan, sehingga menimbulkan dampak negatif bagi banyak orang. Dampak negatif yang ditimbulkan dari penyalahgunaan teknologi AI yaitu *deepfake*. *Deepfake* merupakan teknologi yang digunakan untuk manipulasi gambar atau video. Teknik manipulasi yang digunakan dalam *deepfake* digunakan untuk merubah gambar baik dari bentuk wajah, tempat, objek atau bahkan suara. *Variational autoencoder* (VAE) merupakan algoritma *deep learning* yang dapat digunakan untuk melakukan manipulasi wajah. Hasil dari proses VAE berupa gambar hasil dari penggabungan gambar wajah asli yang dilakukan selama proses pelatihan berlangsung. Gambar wajah baru yang dihasilkan dari proses pelatihan VAE disebut sebagai gambar decoder atau manipulasi. Gambar wajah asli dan gambar wajah manipulasi akan dianalisis menggunakan metode *forensics deep neural network*. Teknik analisis dilakukan menggunakan perhitungan dari *error level analysis* (ELA), teknik ini digunakan untuk mengetahui seberapa besar perubahan yang terjadi pada gambar. Berdasarkan dari hasil pengujian predikai yang dilakukan dengan menggunakan gambar wajah asli dan manipulasi menunjukan jika metode yang diterapkan dapat mendeteksi gambar *real* dan *fake.*

Kata kunci: *Artificial Intelligence (AI), Deepfake, Variational Autoencoder (VAE), Decoder, Forensics Deep Neural Network, Error Level Analysis (ELA)*

Regina Angelika Septi Rahayu[1], Handri Santoso[2]
DOI: https://doi.org/10.54443/sibatik.v2i9.1312

PENDAHULUAN

Artificial Intelligence (AI) adalah salah satu jenis kemajuan teknologi yang dapat memberikan banyak perubahan dan kemudahan pada kehidupan manusia. AI sering diimplementasikan pada teknologi pintar yang menghasilkan sistem otomatisasi. Tidak hanya dapat memberikan kemudahan dalam kehidupan, AI juga dapat menimbulkan sebuah ancaman dan masalah yang dapat merugikan manusia. Salah satu teknologi AI yang dapat memberikan dampak negatif jika disalahgunakan adalah *deepfake*. *Deepfake* merupakan teknologi AI yang sering digunakan untuk memanipulasi sebuah gambar atau video. Teknik manipulasi ini mengacu pada perubahan video dan gambar baik dari segi wajah, latar belakang, ekspresi wajah bahkan suara. Dampak negatif yang dapat ditimbulkan dari tindakan penyalahgunaan deepfake adalah penipuan, pemerasan pencurian data diri, pornografi, dan penyebaran informasi palsu.

Beberapa tahun terakhir, banyak lembaga pemerintahan yang telah menggunakan teknologi *image forensic*, seperti lembaga akses dan hukum. Penerapan image forensic pada lembaga akses dan hukum sering digunakan untuk mendeteksi keaslian dan keakuratan dari gambar dan video yang telah dimanipulasi. *Deep learning* merupakan metode yang digunakan untuk proses pendeteksian gambar dan video, karena *deep learning* dapat mendeteksi perubahan jaringan saraf gambar wajah. Kegiatan pendeteksian *deepfake* menggunakan metode *deep learning* telah mengalami peningkatan sekitar 99,8%. Selain lembaga akses dan hukum, terdapat beberapa perusahan teknologi yang telah menerapkan pendeteksian AI deepfake menggunakan deep learning yaitu Microsoft, IBM, dan Amazon.

Meskipun metode pendeteksian tersebut sudah diterapkan di beberapa lembaga dan perusahaan teknologi, namun hasil akhir dari proses yang dilakukan masih belum sesuai dengan harapan hasil terbaik. Hal ini mengacu pada beberapa elemen perubahan yang tidak dapat dideteksi dengan baik seperti jenis kulit dan gender, yang disebabkan oleh tolak ukur yang salah. Kesalahan dari perhitungan tolak ukur ini disebabkan oleh penerapan algoritma yang tidak sesuai. Penerapan algoritma yang sesuai dapat memberikan kewaspadaan dan karakteristik kekurangan model, sehingga hasil yang didapatkan akan lebih baik dan memiliki tingkat akurasi yang tinggi.

TINJAUAN PUSTAKA

Deep Learning

Deep learning adalah sub bagian khusus dari *machine learning*; yang merepresentasikan pembelajaran data dan lapisan. Arti kata '*deep*' dalam *deep learning* berarti sebuah pemahaman yang mendalam dalam pencapaian sebuah pendekatan, yang berkaitan dengan gagasan lapisan representasi yang berurutan. Nama lain dari representasi lapisan ini adalah pembelajaran representasi hierarkis. Pada proses paparan data pelatihan model, sering melibatkan puluhan bahkan ratusan lapisan representasi yang dipelajari secara otomatis. Sementara pada *machine learning*, representasi lapisan hanya dilakukan pada satu atau dua representasi data, sehingga sering disebut dengan *shallow learning*.

Regina Angelika Septi Rahayu[1], Handri Santoso[2]
DOI: https://doi.org/10.54443/sibatik.v2i9.1312

Artificial Intelligence (AI)

Artificial Intelligence (AI) merupakan bidang ilmu komputer yang dirancang khusus untuk memecahkan masalah kognitif, berkaitan dengan ilmu pembelajaran, pemecahan masalah, dan pola penalaran. AI seringkali dikaitkan dengan robotika atau adegan futuristic robot fisik ilmiah. Prof. Pedro Domingos mendefinisikan AI sebagai 'lima suku' *machine learning* yaitu; Simbolis, Koneksionisme, Evolusioner, Bayesian, dan Analogi. *Artificial Intelligence* (AI) digunakan untuk proses otomatisasi semua tugas intelektual yang pada umumnya dikerjakan oleh manusia. Oleh sebab itu AI merupakan bidang umum dari *machine learning* dan *deep learning* yang memiliki cakupan luas dan melibatkan banyak pendekatan.

Deepfake

Deepfake berasal dari kata '*deep learning*' dan '*fake*' yang berarti hasil manipulasi menggunakan kecerdasan buatan (AI). *Deepfake* menjadi salah satu contoh kemajuan teknologi kecerdasan buatan yang memiliki potensi implementasi luas. Teknik yang diterapkan adalah mengubah ekspresi wajah seseorang dalam sebuah gambar asli untuk dapat menghasilkan gambar baru atau palsu dengan ekspresi wajah yang berbeda dari gambar aslinya. Konsep perubahan gambar wajah seseorang dengan orang lain dilakukan dengan menggunakan algoritma *deep learning*, dengan hasil yang didapatkan sangat realistis dan tidak dapat dibedakan dengan mata manusia. Manipulasi yang dilakukan *deepfake* tidak hanya dapat dilakukan pada perubahan ekspresi wajah manusia, tetapi juga perubahan tempat, hewan, objek, suara dan lain-lain.

Variational Autoencoder (VAE)

Variational Autoencoder (VAE) merupakan bagian dari algoritma *autoencoder*. Algoritma *autoencoder* merupakan proses dari *neural network* yang terdiri dari dua bagian yaitu *encoder network*, representasi data input dengan dimensi tinggi menjadi vector representasi dimensi rendah; *decoder network*, mengkompresi vector representasi yang diterima menjadi domein aslinya. Pada *autoencoder* perubahan yang terjadi pada gambar tidaklah banyak dan tidak memiliki banyak varian. Sehingga dalam penelitian ini algoritma yang digunakan adalah *variational autoencoder*. Pada prosesnya, *variational autoencoder* tidak memiliki perbedaan dengan *autoencoder* yang membedakan adalah dalam perhitungan matematis yaitu pada *encoder* dan *loss function* agar hasil yang didapatkan bervariasi.

Deep Neural Network (DNNs)

Deep Neural Network (DNNs) terdiri dari serangkaian lapisan. Setiap lapisan berisi unit, yang terhubung dengan lapisan unit sebelumnya, melalui serangkaian *weights*. Lapisan umum yang terdapat dalam DNNs adalah lapisan padat, lapisan ini menghubungkan semua unit lapisan secara langsung ke setiap unit dan lapisan sebelumnya. Lapisan di setiap unit DNNs dapat mempresentasikan aspek input asli yang akurat. Menariknya DNNs dapat

Regina Angelika Septi Rahayu[1], Handri Santoso[2]
DOI: https://doi.org/10.54443/sibatik.v2i9.1312

menemukan lapisan *weights* dengan prediksi yang akurat di setiap lapisan. Proses ini disebut dengan *training the network.*

Forensics Deep Learning

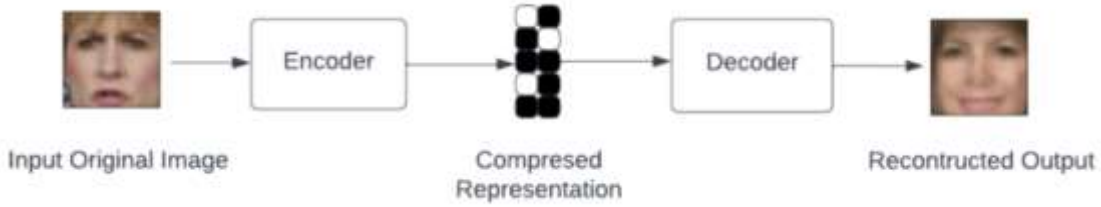
Forensics deep learning merupakan proses yang digunakan untuk mengetahui klasifikasi dari gambar asli dan manipulasi. Proses klasifikasi yang dilakukan dengan menggunakan metode *forensics* digunakan untuk mengetahui seberapa besar *error level analysis* (ELA). *Error level analysis* merupakan teknik untuk mengetahui seberapa besar gambar asli mengalami perubahan dengan menghitung perbedaan rata-rata nilai *luminance* dan *chrominance,* dengan hasil gambar konversi dominan hitam dan titik putih pada area gambar asli maupun manipulasi.

METODE

Penelitian ini menggunakan metode kuantitatif, dalam proses pengerjaanya. Penelitian ini membahas mengenai masalah atau fenomena sosial yaitu *deepfake,* dengan melakukan uji coba analisis menggunakan *forensic deep neural network.* Data yang digunakan dalam penelitian ini diperoleh dari *source* kumpulan data *public Kaggle,* yaitu dataset gambar wajah asli. Sementara untuk dataset gambar wajah manipulasi akan diperoleh dari hasil proses *image generated.*

Metode Pengolahan Data

A. Proses Image Generated



Gambar 1 Diagram Proses Image Generated

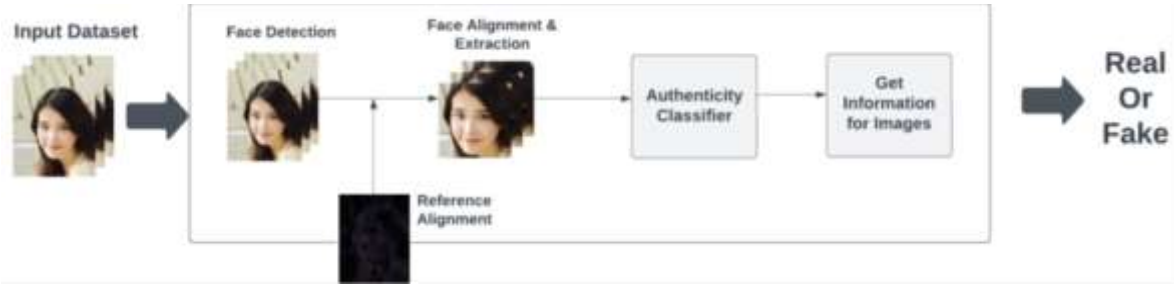
Data gambar wajah asli yang telah diunduh dari Kaggle akan digunakan untuk membuat model gambar wajah manipulasi dengan menggunakan algoritma variational autoencoder. Berikut Langkah-langkah yang akan dilakukan:

1. Persiapan data: Dataset yang akan digunakan harus dipersiapkan dan sesuai dengan format gambar RGB dan ukuran gambar harus sama yaitu 85 x 85 pixel.
2. Training data: Dataset yang telah dipersiapkan, akan dilatih dengan menggunakan model variational autoencoder.
3. Encoding: Tahap ini terdiri dari mengambil data gambar wajah asli dan mengkonversinya menjadi representasi vektor di dalam ruang fitur.
4. Decoding: Tahap ini merupakan proses mengambil representasi vektor dan mengkonversinya kembali menjadi gambar wajah baru.

Regina Angelika Septi Rahayu[1], Handri Santoso[2]
DOI: https://doi.org/10.54443/sibatik.v2i9.1312

5. Face Generation: Tahap ini menggunakan representasi vector untuk menghasilkan gambar wajah manipulasi, dengan hasil yang akan memiliki nilai berbeda pada setiap dimensi vector.
6. Evaluation: Evaluasi dilakukan dengan menggunakan metrik seperti mean squared error (MSE) atau peak signal-to-noise ratio (PSNR) untuk membandingkan gambar wajah asli dengan gambar wajah manipulasi yang dihasilkan oleh metode variational autoencoder.

B. Proses Image Analysis



Gambar 2 Diagram Proses Image Analysis

Data yang telah didapatkan dari proses image generate akan diolah menggunakan deep learning DNNs dan metode forensic untuk mendapatkan informasi yang berkaitan dengan gambar tersebut. Berikut langkah-langkah pengolahan data yang akan dilakukan:

1. Persiapan Data: Dataset yang digunakan adalah dataset gambar wajah asli dan gambar wajah manipulasi. Data tersebut harus diolah dan dipre-proses sebelum digunakan untuk proses pelatihan.
2. Ekstraksi Fitur: Fitur wajah dari data yang diuji akan diekstrak menggunakan algoritma deep neural network seperti Cov2D, MaxPool2.
3. Perbandingan Fitur: Fitur yang diekstrak dari gambar yang diuji akan dibandingkan dengan fitur referensi yang diambil dari gambar wajah asli.
4. Klasifikasi: Hasil dari perbandingan fitur wajah tersebut akan diklasifikasikan sebagai asli atau palsu dengan menggunakan algoritma klasifikasi seperti SVM atau Random Forest.
5. Verifikasi: Hasil dari tahap ini akan dikonfirmasi kaskuratannya dengan mengevaluasi tingkat keaslian dari model DNNs yang digunakan dan hasil digital forensic dari gambar tersebut.

HASIL DAN PEMBAHASAN

Analisis Proses Pembuatan Gambar Wajah Manipulasi dengan Menggunakan Algoritma Variational Autoencoder (VAE)

Proses pembuatan gambar wajah manipulasi dilakukan dengan menggunakan sekumpulan dataset gambar wajah asli. Gambar wajah asli diambil dari sumber kumpulan dataset Kaggle. Dataset yang digunakan memiliki 202.599 data gambar, data tersebut akan

Regina Angelika Septi Rahayu[1], Handri Santoso[2]
DOI: https://doi.org/10.54443/sibatik.v2i9.1312

dibagi menjadi 80% data train dan 20% data val. Diagram proses dapat dilihat pada metode penelitian proses *image generated.*

Gambar wajah asli yang diinput akan diubah ukurannya menjadi 85 x 85 pixel, hal ini bertujuan untuk membuat ukuran gambar yang presisi dan memudahkan proses training. Training model akan dilakukan dengan menggunakan algoritma *variational autoencoder*, algoritma ini merupakan sub bagian dari algoritma *autoencoder*. *Reparameterization trick*, proses ini dilakukan untuk operasi sampling gambar dalam model generatif. Proses operasi sampling ini dapat dilakukan dengan menggunakan rumus:

‘z = mean + exp(log variance / 2) * epsilon’

Keterangan:

- Z = representasi laten yang ingin dihasilkan dalam proses sampling.
- Mean = nilai rata-rata dari distribusi normal, digunakan untuk pemetaan data dalam ruang laten.
- Log_variance = logaritma dari sampling baku (variance) distribusi normal yang menggambarkan variasi data dalam ruang laten.
- Epsilon = variable acak dengan distribusi normal standar, yaitu nilai rata-rata 0 dan simpangan baku 1.

Sehingga diperoleh gradien yang dapat digunakan untuk pelatihan model dengan metode pemetaan stokastik.

Proses pemetaan data input ke dalam ruang laten yang digunakan dalam proses generasi dan rekonstruksi disebut dengan proses encoder. Encoder dibangun dengan menggunakan lapisan konvolusi untuk mengekstrak fitur dan menghasilkan distribusi laten. Hasil dari proses encoder dapat dilihat di tabel berikut:

Table 1 Tabel Encoder Proses VAE

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	[(None, 218, 178, 3)]	0	[]
conv2d_22 (Conv2D)	(None, 218, 178, 32)	896	['input_1[0][0]']
batch_normalization (BatchNormalization)	(None, 218, 178, 32)	128	['conv2d_22[0][0]']

max_pooling2d (MaxPooling2D)	(None, 109, 89, 32)	0	['batch_normalization[0][0]']
conv2d_23 (Conv2D)	(None, 109, 89, 64)	18496	['max_pooling2d[0][0]']
batch_normalization_1 (BatchNormalization)	(None, 109, 89, 64)	256	['conv2d_23[0][0]']
max_pooling2d_1 (MaxPooling2D)	(None, 55, 45, 64)	0	['batch_normalization_1[0][0]']
conv2d_24 (Conv2D)	(None, 55, 45, 128)	73856	['max_pooling2d_1[0][0]']
batch_normalization_2 (BatchNormalization)	(None, 55, 45, 128)	512	['conv2d_24[0][0]']
max_pooling2d_2 (MaxPooling2D)	(None, 28, 23, 128)	0	['batch_normalization_2[0][0]']
conv2d_25 (Conv2D)	(None, 28, 23, 256)	295168	['max_pooling2d_2[0][0]']
batch_normalization_3 (BatchNormalization)	(None, 28, 23, 256)	1024	['conv2d_25[0][0]']
max_pooling2d_3 (MaxPooling2D)	(None, 14, 12, 256)	0	['batch_normalization_3[0][0]']

Regina Angelika Septi Rahayu[1], Handri Santoso[2]
DOI: https://doi.org/10.54443/sibatik.v2i9.1312

flatten (Flatten)	(None, 43008)	0	['max_pooling2d_3[0][0]']
flatten_1 (Flatten)	(None, 43008)	0	['max_pooling2d_3[0][0]']
z_mean (Dense)	(None, 512)	22020608	['flatten[0][0]']
z_log_sigma (Dense)	(None, 512)	22020608	['flatten_1[0][0]']
batch_normalization_4 (BatchNormalization)	(None, 512)	2048	['z_mean[0][0]']
batch_normalization_5 (BatchNormalization)	(None, 512)	2048	['z_log_sigma[0][0]']
lambda (Lambda)	(None, 512)	0	['batch_normalization_4[0][0]','batch_normalization_5[0][0]']
Total params: 44,435,648 Trainable params: 44,432,640 Non-trainable params: 3,008			

Berdasarkan hasil encoder yang ditampilkan dalam tabel diatas dapat diketahui jika perubahan dimensi layer konvolusi dilakukan dengan menggunakan Cov2D, kemudian diikuti dengan normalisasi batch dan operasi MaxPooling untuk mengurangi dimensi. Selanjutnya dilakukan perhitungan pada variabel laten dengan mengubah output lapisan konvolusi menjadi flatten, diikuti dengan lapisan Dense yang menghasilkan nilai rata-rata dan log varian, lapisan tersebut juga diikuti dengan normalisasi batch. Dalam perhitungan Lambda nilai yang digunakan merupakan nilai ‘z’ pada proses perhitungan *reparameterization trick,* untuk menerapkan sampling dari distribusi laten.

Regina Angelika Septi Rahayu[1], Handri Santoso[2]
DOI: https://doi.org/10.54443/sibatik.v2i9.1312

Decoder merupakan tahapan yang dilakukan untuk menghasilkan rekonstruksi gambar dari laten setelah encoder. Hasil decoder akan dihubungkan dengan encoder yang telah dideklarasikan sebelumnya untuk pembelajaran dan generasi data baru pada proses VAE. Hasil dari proses decoder dapat dilihat dari tabel berikut:

Table 2 Tabel Decoder Proses VAE

Layer (type)	Output Shape	Param #
decoder_input (InputLayer)	[(None, 512)]	0
dense (Dense)	(None, 43008)	22063104
reshape (Reshape)	(None, 14, 12, 256)	0
up_sampling2d (UpSampling2D)	(None, 28, 24, 256)	0
cropping2d (Cropping2D)	(None, 28, 23, 256)	0
conv2d_transpose (Conv2DTranspose)	(None, 28, 23, 256)	590080
batch_normalization_6 (BatchNormalization)	(None, 28, 23, 256)	1024
up_sampling2d_1 (UpSampling 2D)	(None, 56, 46, 256)	0
cropping2d_1 (Cropping2D)	(None, 55, 45, 256)	0
conv2d_transpose_1 (Conv2DTranspose)	(None, 55, 45, 128)	295040

Regina Angelika Septi Rahayu[1], Handri Santoso[2]
DOI: https://doi.org/10.54443/sibatik.v2i9.1312

batch_normalization_7 (BatchNormalization)	(None, 55, 45, 128)	512
up_sampling2d_2 (UpSampling2D)	(None, 110, 90, 128)	0
cropping2d_2 (Cropping2D)	(None, 109, 89, 128)	0
conv2d_transpose_2 (Conv2DTranspose)	(None, 109, 89, 64)	73792
batch_normalization_8 (BatchNormalization)	(None, 109, 89, 64)	256
up_sampling2d_3 (UpSampling2D)	(None, 218, 178, 64)	0
conv2d_transpose_3 (Conv2DTranspose)	(None, 218, 178, 32)	18464
batch_normalization_9 (BatchNormalization)	(None, 218, 178, 32)	128
conv2d_transpose_4 (Conv2DTranspose)	(None, 218, 178, 3)	867
Total params: 23,043,267 Trainable params: 23,042,307 Non-trainable params: 960		

Berdasarkan hasil decoder yang ditampilkan diatas dapat diketahui jika lapisan Dense digunakan untuk mengubah input vektor dari distribusi laten menjadi vektor dengan ukuran yang sesuai untuk diubah kembali menjadi gambar. Ukuran vektor output dihitung dengan mengalikan dimensi dari output lapisan terakhir dalam encoder. Operasi upsampling dan cropping dilakukan untuk mengembalikan dimensi tensor yang diubah menjadi ukuran yang lebih besar. Operasi Conv2DTranspose dilakukan untuk operasi konvolusi transposisi untuk menghasilkan rekonstruksi gambar dari distribusi laten. Setiap lapisan proses ini diikuti oleh proses normalisasi batch untuk untuk memperbaiki stabilitas dan konvergensi model.

Regina Angelika Septi Rahayu[1], Handri Santoso[2]
DOI: https://doi.org/10.54443/sibatik.v2i9.1312

Dalam *variational autoencoder*, perhitungan model vae menjadi tahap terakhir yang digunakan untuk menyempurnakan gambar manipulasi yang dilakukan. Perhitungan model ini dilakukan dengan menghubungkan output distribusi laten 'z' dari encoder ke input decoder. Model vae yang dihasilkan akan digunakan untuk pelatihan dan evaluasi proses vae secara keseluruhan. Berikut tabel hasil dari model vae yang telah dilakukan:

Table 3 Tabel Model VAE

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 218, 178, 3)]	0
encoder (Functional)	[(None, 512),(None, 512),(None, 512)]	44435648
decoder (Functional)	(None, 218, 178, 3)	23043267
Total params: 67,478,915 Trainable params: 67,474,947 Non-trainable params: 3,968		

Proses *face generation*, menggunakan representasi vektor untuk menghasilkan gambar wajah manipulasi, dengan hasil yang memiliki nilai berbeda pada setiap dimensi vektor. Perhitungan yang dilakukan adalah untuk mengetahui total loss dari *reconstruction loss* dan *KL divergence loss*. Untuk mendapatkan hasil dari *reconstruction loss* rumus yang dapat digunakan adalah:

'reconstruction_loss = mse(K.flatten(input_img), K.flatten(output))'

Fungsi *MeanSequaredError* (MSE) adalah fungsi yang digunakan untuk menghitung selisih kuadrat rata-rata antara gambar asli dan output gambar rekonstruksi. Sementara untuk *KL divergence loss* perhitungannya dilakukan dengan menggunakan rumus:

'kl_loss = -0.5 * K.sum (1 + z_log_sigma - K.square(z_mean) - K.exp(z_log_sigma), axis=-1)'

Perhitungan *KL divergence loss* dilakukan berdasarkan distribusi laten. Dimana proses utama perhitungan terjadi pada proses

'-0.5*sum(1 + log(sigma^2) - mu^2 - sigma^2)'

Regina Angelika Septi Rahayu[1], Handri Santoso[2]
DOI: https://doi.org/10.54443/sibatik.v2i9.1312

Keterangan:

- Sum(...) = Operasi Penjumlahan dari semua elemen yang terdapat dalam tanda kurung. Penjumlahan akan dilakukan pada dimensi laten (untuk menghitung total loss persample dalam bath).
- ‘sigma^2’ = Vektor yang berisi nilai varians dari distribusi laten dalam VAE.
- ‘log(sigma^2)’ = Operasi logaritma yang diaplikasikan dari setiap elemen dari ‘sigma^2’.
- ‘mu’ = Vektor yang berisi nilai rata-rata dari distribusi laten dalam VAE.

Perhitungan ini dilakukan pada setiap sampel dalam batch selama proses training berlangsung. Total loss dalam proses ini dilakukan dengan cara menggabungkan kedua komponen loss yaitu *reconstruction loss* dan *KL divergence loss,* dimana nilai ‘0.0001’ digunakan sebagai bobot *KL divergence loss.*

Kompilasi model dilakukan menggunakan tiga model (Encoder. Decoder, dan Model vae), dengan menggunakan optimizer ‘rmsprop’ dan fungsi loss yang telah didefinisikan sebelumnya. Optimizer ‘rmsprop’ digunakan untuk meningkatkan kecepatan *learning rate* dengan konvergen yang lebih cepat. Proses pelatihan dilakukan menggunakan generator data (train), dengan jumlah langkah per epoch dilakukan sebanyak besaran data atau ketentuan, dan melakukan validasi menggunakan generator data (validasi) dengan jumlah validasi dilakukan sebanyak 500 kali. Proses training dapat dilakukan sesuai dengan durasi epoch training yang diinginkan.

Gambar 3 Hasil Proses Variational Autoencoder

Gambar diatas merupakan hasil dari proses pelatihan *variational autoencoder*, dimana gambar tersebut merupakan gambar wajah asli yang telah digabungkan menjadi satu sehingga nantinya akan menghasilkan gambar wajah baru. Gambar wajah baru yang dihasilkan disebut dengan gambar hasil decoder.

Regina Angelika Septi Rahayu[1], Handri Santoso[2]
DOI: https://doi.org/10.54443/sibatik.v2i9.1312

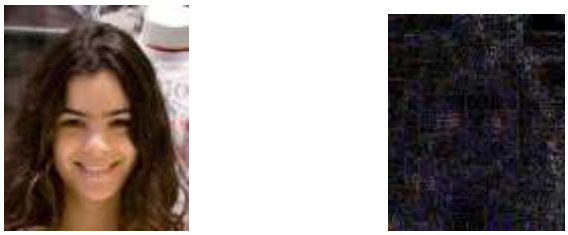
Gambar 4 Gambar Wajah Asli dan Gambar Wajah Decoder

Gambar hasil decoder yang diperoleh akan digunakan untuk melakukan proses analisis tingkat akurasi keaslian sebuah gambar.

Analisis Tingkat Akurasi Keaslian Gambar Wajah Menggunakan Metode Forensics Deep Neural Network (DNNs)

Proses analisis dilakukan dengan menggunakan dua jenis dataset yaitu dataset gambar wajah asli dan gambar wajah manipulasi hasil dari proses *variational autoencoder*. Pre-proses data merupakan tahapan yang dilakukan sebelum dataset dilatih dan dianalisis. Tahapan pra-proses dilakukan untuk proses ekstraksi data, menyamakan ukuran gambar dan persiapan untuk proses pembuatan *error level analysis.*

Error level Analysis (ELA), teknik pemetaan yang digunakan untuk mengetahui seberapa besar gambar mengalami perubahan. Hasil dari proses ini berupa gambar yang dominan berwarna hitam dengan bintik putih yang mengikuti bentuk pola gambar.



Gambar 5 Hasil ELA Proses Gambar Wajah Asli

Gambar 6 Hasil ELA Proses Gambar Wajah Manipulasi

Gambar di menunjukan perubahan dari gambar biasa menjadi gambar ELA, konversi gambar yang dilakukan menggunakan tingkatan kualitas JPEG sebesar 90. Hasil gambar ELA tersebut akan melewati proses pra-pelatihan, dengan dilakukan perubahan pada ukuran gambar 85 x 85-pixel menjadi array NumPy, perataan array gambar menjadi satu dimensi dan normalisasi nilai pixel gambar sehingga berada pada rentang 0 – 1.0.

‘X = [] # ELA converted images

Y = [] # 0 for fake, 1 for real’

Regina Angelika Septi Rahayu[1], Handri Santoso[2]
DOI: https://doi.org/10.54443/sibatik.v2i9.1312

Untuk mendapatkan hasil akurasi yang sesuai, dataset gambar wajah asli dan wajah manipulasi yang digunakan akan diproses dengan melakukan iterasi pada setiap berkas gambar yang terdapat dalam direktori. Path dalam direktori dan nama berkas gambar akan diubah untuk mendapatkan data path yang lebih lengkap. Gambar yang telah dipersiapkan dalam bentuk array kan ditambahkan ke dalam daftar 'X', serta menambahkan label 1/0 ke dalam daftar 'Y' dimana label 1 menandakan bahwa gambar tersebut positif sedangkan label 0 menandakan bahwa gambar tersebut negatif.

'X_train, X_val, Y_train, Y_val = train_test_split(X, Y, test_size = 0.2, random_state=5)'

Data yang telah diinputkan ke dalam label 'X' dan 'Y' akan dibagi menjadi set train dan validation menggunakan fungsi 'train_test_split' dari modul 'sklearn.model_selection'. Parameter 'test-size = 0.2' digunakan untuk menentukan proporsi data yang akan digunakan sebagai data validasi, dalam hal ini data yang digunakan adalah 20%. 'random_state=5' digunakan untuk menghasilkan pembagian data yang konsisten setiap kali program dijalankan.

Proses pembuatan model *neural network* dilakukan dengan menggunakan *Sequential* API dari Keras. Hasil dari proses pembuatan arsitektur model *neural network* terdiri dari lapisan konvolusi, max pooling, dropout, dan dense.

Table 4 Tabel Arsitektur Model Neural Network

Layer (type)	Output Shape	Param #
conv2d_36 (Conv2D)	(None, 81, 81, 32)	2432
conv2d_37 (Conv2D)	(None, 77, 77, 32)	25632
max_pooling2d_7 (MaxPooling2D)	(None, 38, 38, 32)	0
dropout_14 (Dropout)	(None, 38, 38, 32)	0
flatten_7 (Flatten)	(None, 46208)	0
dense_14 (Dense)	(None, 256)	11829504
dropout_15 (Dropout)	(None, 256)	0
dense_15 (Dense)	(None, 2)	514
Total params: 11,858,082 Trainable params: 11,858,082 Non-trainable params: 0		

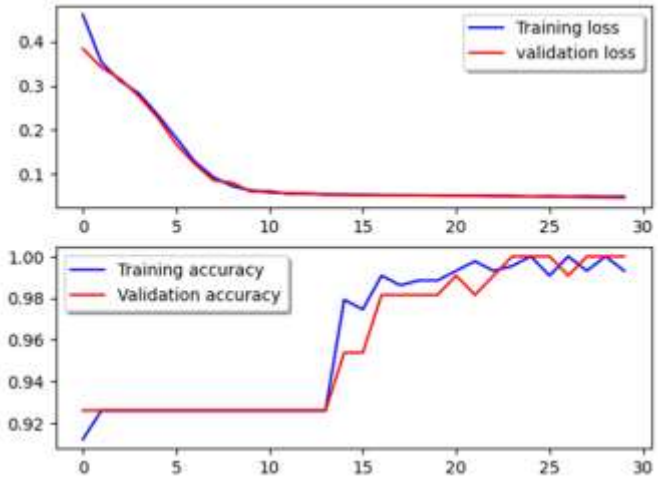
Berdasarkan table di atas, diketahui jika dalam proses pembuatan arsitektur model perlu adanya penambahan dua lapisan konvolusi *Conv2D* ke dalam mode. Lapisan ini memiliki 32 filter dengan ukuran kernel 5 x 5, menggunakan aktivasi ReLU. Lapisan *MaxPooling2D* merupakan operasi pooling yang digunakan untuk mengurangi ukuran representasi spasial dengan ukuran pool 2 x 2. Penambahan lapisan *Dropout* bertujuan untuk mencegah terjadinya overfitting degan mengatur koneksi aca kantar lapisan. Lapisan *Flatten* digunakan untuk mengubah tensor multidimensi menjadi vektor satu dimensi yang akan

Regina Angelika Septi Rahayu[1], Handri Santoso[2]
DOI: https://doi.org/10.54443/sibatik.v2i9.1312

menjadi input untuk lapisan berikutnya. Penambahan lapisan *Dense* pertama dilakukan ke dalam model dengan 265-unit neuron dan menggunakan aktivasi ReLU, sedangkan lapisan *Dense* kedua dilakukan dengan menambahkan 2 unit neuron yang menghasilkan output klasifikasi penggunaan aktivasi softmax.

Proses melatih model *neural network* dilakukan menggunakan data *train* dan data validasi. Dalam proses melatih hal pertama yang harus dilakukan adalah menentukan *learning rate* awal yang akan digunakan dalam *optimizer*. *Optimizer* dilakukan menggunakan penurunan *learning rate (decay)* yang dihitung berdasarkan jumlah epochs. Kompilasi model dilakukan dengan menggunakan hasil *optimizer, loss function,* dan matrik evaluasi untuk menentukan akurasi. Data latih (X_train dan Y_train) digunakan untuk melatih model, sedangkan data validasi (X_val dan Y_val) digunakan untuk mengevaluasi model pada setiap epochs. Selama proses pelatihan, objek *EarlyStopping* digunakan untuk menghentikan pelatihan lebih awal jika tidak ada peningkatan metrik validasi selama proses epochs belangsung.

Setelah proses pelatihan model *neural network* dilakukan dan proses sudah berhasil, hasil akurasi data (latih dan validasi) dapat dilihat dari kurva *loss* dan *accuracy* data berikut:



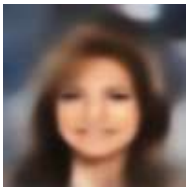
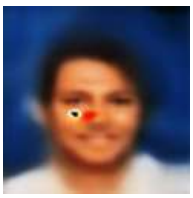
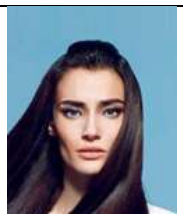
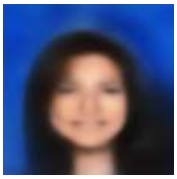
Gambar 7 Kurva Akurasi dan Loss Model Neural Network

Analisis Hasil Akhir dari Penerapan Metode Forensics Deep Neural Network Terhadap Penelitian

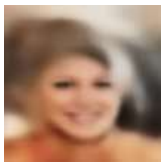
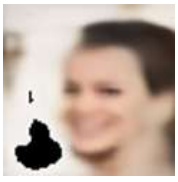
Berdasarkan hasil dari proses pelatihan model *neural network* dapat diketahui seberapa besar persentase dari prediksi tingkat keaslian dari gambar yang dianalisis. Berikut ini hasil prediksi presentase keaslian gambar, gambar yang digunakan adalah sepuluh gambar wajah asli dan sepuluh gambar wajah manipulasi hasil dari proses *variational autoencoder.*

Regina Angelika Septi Rahayu[1], Handri Santoso[2]
DOI: https://doi.org/10.54443/sibatik.v2i9.1312

Table 5 Tabel Hasil Presentase Prediksi Keaslian Gambar Wajah

No	Jenis Gambar	Gambar	Hasil Prediksi
1	Asli		Class: real Confidence: 100.00
	Manipulasi		Class: fake Confidence: 54.62
2	Asli		Class: real Confidence: 100.00
	Manipulasi		Class: fake Confidence: 99.39
3	Asli		Class: real Confidence: 99.50
	Manipulasi		Class: real Confidence: 86.51
4	Asli		Class: real Confidence: 100.00
	Manipulasi		Class: fake Confidence: 99.70

Regina Angelika Septi Rahayu[1], Handri Santoso[2]
DOI: https://doi.org/10.54443/sibatik.v2i9.1312

5	Asli		Class: real Confidence: 100.00
	Manipulasi		Class: fake Confidence: 50.38
6	Asli		Class: real Confidence: 100.00
	Manipulasi		Class: fake Confidence: 89.26
7	Asli		Class: real Confidence: 100.00
	Manipulasi		Class: fake Confidence: 99.97
8	Asli		Class: real Confidence: 100.00
	Manipulasi		Class: fake Confidence: 50.37

Regina Angelika Septi Rahayu[1], Handri Santoso[2]
DOI: https://doi.org/10.54443/sibatik.v2i9.1312

9	Asli		Class: real Confidence: 100.00
	Manipulasi		Class: fake Confidence: 50.46
10	Asli		Class: real Confidence: 100.00
	Manipulasi		Class: fake Confidence: 93.17

Berdasarkan hasil pengujian presentasi prediksi keaslian gambar, diketahui jika masih terdapat kesalahan atau *error* yang terjadi, dimana gambar wajah manipulasi terdeteksi sebagai ‘Real’. Kesalahan tersebut disebabkan karena terjadinya perubahan pada saat proses iterasi direktori dan berkas, dimana seharusnya hasil proses iterasi yang dilakukan menghasilkan nilai ‘500’ untuk direktori gambar wajah asli dan ‘540’ untuk direktori gambar wajah manipulasi.

Uji tingkat akurasi gambar wajah tidak hanya dilakukan dengan menggunakan gambar wajah hasil manipulasi VAE. Perhitungan prediksi tingkat akurasi juga dilakukan dengan menggunakan gambar wajah yang telah diedit secara manual dengan menggunakan Canva editor dan DPI converter. Uji coba dilakukan dengan menggunakan tiga gambar yang telah di edit menggunakan Canva editor dan tiga gambar yang telah diubah menggunakan DPI converter.

Table 6 Tabel Hasil Presentasi Prediksi Gambar Menggunakan Canva Editor

No	Jenis Gambar	Gambar	Hasil Prediksi
1	Asli		Class: real Confidence: 100.00

Regina Angelika Septi Rahayu[1], Handri Santoso[2]
DOI: https://doi.org/10.54443/sibatik.v2i9.1312

	Manipulasi		Class: real Confidence: 51.26
2	Asli		Class: real Confidence: 100.00
	Manipulasi		Class: real Confidence: 50.03
3	Asli		Class: real Confidence: 99.50
	Manipulasi		Class: real Confidence: 50.11

Table 7 Tabel Hasil Persentase Prediksi Gambar Menggunakan DPI Converter

1	Asli		Class: real Confidence: 98.24
	Manipulasi		Class: real Confidence: 77.29
2	Asli		Class: real Confidence: 99.06

Regina Angelika Septi Rahayu[1], Handri Santoso[2]
DOI: https://doi.org/10.54443/sibatik.v2i9.1312

	Manipulasi		Class: real Confidence: 93.44
3	Asli		Class: real Confidence: 99.96
	Manipulasi		Class: real Confidence: 96.33

Berdasarkan hasil pengujian prediksi tingkat keaslian gambar yang dilakukan dengan menggunakan gambar hasil Canva editor pada table 6 dan gambar DPI converter pada table 7, menunjukkan jika gambar manipulasi tersebut terdeteksi sebagai gambar ‘real’. Hal ini disebabkan karena gambar wajah manipulasi yang diujikan tidak mengalami perubahan yang signifikan (perubahan bentuk wajah, mata, hidung, ekspresi dll), namun perubahan yang terjadi pada gambar manipulasi tersebut terjadi pada perubahan kualitas gambar, sehingga gambar tersebut dapat dikatakan sebagai gambar asli.

Selain untuk mengetahui tingkat persentase akurasi keaslian menggunakan data gambar. Proses *forensics deep neural network* juga digunakan untuk melakukan analisis perbandingan pada informasi gambar, dengan tujuan untuk meningkatkan keabsahan dari hasil analisis yang didapatkan. Dalam analisis perbandingan metadata dilakukan dengan menggunakan modul PIL (pillow) dan metadata EXIF. Modul PIL digunakan untuk membaca metadata dan informasi dari sebuah gambar, sedangkan metadata EXIF digunakan untuk mendapatkan data dari metadata gambar.

Table 8 Tabel Perbedaan Metadata Gambar Asli dan Gambar Palsu

Perbedaan Metadata		
No	**Gambar Asli**	**Gambar Palsu**
1	'file_name' : '000049.jpg', 'width' : 178, 'height' : 218, 'format' : 'JPEG', 'mode' : 'RGB', 'is_animated': False, 'frames' : 1,	'file_name' :'decoder_output_28.png', 'width' : 85, 'height' : 85, 'format' : 'PNG', 'mode' : 'RGBA', 'is_animated': False, 'frames' : 1,

Regina Angelika Septi Rahayu[1], Handri Santoso[2]
DOI: https://doi.org/10.54443/sibatik.v2i9.1312

	'created' :'2023-05-25T10:45:00', 'modified' :'2023-05-19T05:08:57', 'accessed' :'2023-05-25T10:45:00', 'bit_depth' : 'N/A',	'created':'2023-07-28T11:58:14.532090', 'modified' :'2023-06-15T14:46:36', 'accessed':'2023-07-28T12:01:59.844728', 'bit_depth' : 'N/A',
2	'file_name' : '000058.jpg', 'width' : 178, 'height' : 218, 'format' : 'JPEG', 'mode' : 'RGB', 'is_animated': False, 'frames' : 1, 'created' :'2023-05-25T10:44:52', 'modified' :'2023-05-19T05:08:57', 'accessed' :'2023-05-25T10:44:52', 'bit_depth' : 'N/A',	'file_name': 'decoder_output_0.png', 'width' : 85, 'height' : 85, 'format' : 'PNG', 'mode' : 'RGBA', 'is_animated': False, 'frames' : 1, 'created':'2023-07-28T11:58:14.528090', 'modified' :'2023-06-15T12:48:28', 'accessed':'2023-07-28T12:01:59.844728', 'bit_depth' : 'N/A',

Regina Angelika Septi Rahayu[1], Handri Santoso[2]
DOI: https://doi.org/10.54443/sibatik.v2i9.1312

3	'file_name' : '000046.jpg', 'width' : 178, 'height' : 218, 'format' : 'JPEG', 'mode' : 'RGB', 'is_animated': False, 'frames' : 1, 'created' :'2023-05-25T10:44:54', 'modified' :'2023-05-19T05:08:57', 'accessed' :'2023-05-25T10:44:54', 'bit_depth' : 'N/A',	'file_name':'decoder_output_23.png', 'width' : 85, 'height' : 85, 'format' : 'PNG', 'mode' : 'RGBA', 'is_animated': False, 'frames' : 1, 'created' :'2023-07-28T11:58:14.531090', 'modified':'2023-06-15T14:23:24', 'accessed' :'2023-07-28T12:01:59.844728', 'bit_depth' : 'N/A',
4	'file_name' : '000055.jpg', 'width' : 178, 'height' : 218, 'format' : 'JPEG', 'mode' : 'RGB', 'is_animated': False, 'frames' : 1, 'created' :'2023-05-25T10:44:52', 'modified' :'2023-05-19T05:08:57', 'accessed' :'2023-05-25T10:44:52', 'bit_depth' : 'N/A',	'file_name' :'decoder_output_19.png', 'width' : 85, 'height' : 85, 'format' : 'PNG', 'mode' : 'RGBA', 'is_animated': False, 'frames' : 1, 'created' :'2023-07-28T11:58:14.530090', 'modified':'2023-06-15T13:59:48', 'accessed' :'2023-07-28T12:01:59.844728', 'bit_depth' : 'N/A',

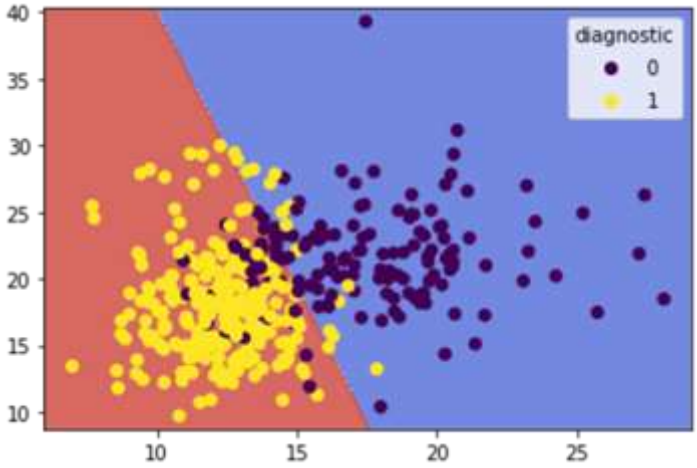
Regina Angelika Septi Rahayu[1], Handri Santoso[2]
DOI: https://doi.org/10.54443/sibatik.v2i9.1312

5	'file_name' : '000052.jpg', 'width' : 178, 'height' : 218, 'format' : 'JPEG', 'mode' : 'RGB', 'is_animated': False, 'frames' : 1, 'created':'2023-05-25T10:45:00', 'modified':'2023-05-19T05:08:57', 'accessed':'2023-05-25T10:45:00', 'bit_depth' : 'N/A',	'file_name' :'decoder_output_21.png', 'width' : 85, 'height' : 85, 'format' : 'PNG', 'mode' : 'RGBA', 'is_animated': False, 'frames' : 1, 'created':'2023-07-28T11:58:14.531090', 'modified':'2023-06-15T14:23:24', 'accessed':'2023-07-28T12:01:59.844728', 'bit_depth' : 'N/A',

Table 8 di atas, menunjukkan perbedaan isi metadata dari lima gambar asli dan lima gambar palsu. Perbedaan metadata yang sangat terlihat dan dapat digunakan sebagai patokan dalam analisis tingkat akurasi adalah ukuran gambar, data waktu gambar dibuat (download), data waktu modifikasi gambar, dan data waktu gambar tersebut diakses. Pada gambar wajah asli diketahui jika gambar memiliki ukuran 178 x 218, sedangkan gambar wajah palsu memiliki ukuran yang lebih kecil dengan ukuran 85 x 85. Selain itu gambar wajah asli memiliki mode warna RGB sedangkan gambar palsu memiliki mode warna RGBA.

Forensik VGG16 merupakan model yang dari Tensorflow yang digunakan untuk pembuatan dan pelatihan *neural network*, serta tugas-tugas vision termasuk dalam klasifikasi gambar. Klasifikasi gambar yang dilakukan adalah untuk mengetahui seberapa banyak persebaran gambar asli dan gambar palsu yang dimiliki. Data yang digunakan untuk proses klasifikasi menggunakan data dari metadata gambar asli dan gambar palsu. Metadata akan dipre-proses dan dilatih dengan menggunakan model forensik VGG16. *Support Vector Machine* (SVM) adalah metode dalam *supervised learning* yang dapat digunakan untuk klasifikasi gambar.

Regina Angelika Septi Rahayu[1], Handri Santoso[2]
DOI: https://doi.org/10.54443/sibatik.v2i9.1312



Gambar 5 Diagram SVM Persebaran Metadata Gambar Asli dan Gambar Palsu

Pada gambar 5 diatas, merupakan gambar diagram SVM yang menampilkan hasil dari proses klasifikasi yang dilakukan pada gambar asli dan gambar palsu. Pada diagram diatas gambar wajah asli diberi tanda dengan plot warna kuning (label 1), sedangkan untuk gambar wajah palsu diberi tanda plot warna ungu (label 0).

PENUTUP

Kesimpulan

Terdapat beberapa point kesimpulan dari penelitian ini yaitu:

1. *Deepfake* adalah teknologi AI yang dapat melakukan teknik manipulasi gambar, gambar yang dihasilkan oleh *deepfake* memiliki tingkat realisitas yang tinggi sehingga sulit dibedakan oleh mata manusia.
2. *Variational autoenoder* merupakan algoritma *deep learning* yang dapat digunakan untuk membuat gambar wajah manipulasi, gambar hasil dari proses VAE berupa gambar wajah baru yang merupakan gabungan dari gambar wajah asli.
3. Gambar *decoder* hasil dari proses VAE digunakan untuk proses analisis tingkat akurasi keaslian gambar.
4. Proses analisis dilakukan dengan menggunakan metode *forensic neural network*, dengan menentukan ELA yang terdapat pada gambar wajah asli dan gambar wajah manipulasi.
5. *Error level analysis* (ELA) teknik pemetaan gambar yang bertujuan untuk mengetahui perubahan yang terjadi pada gambar, hasil yang didapatkan berupa gambar yang dominan hitam dengan bintik putih membentuk pola gambar.
6. Berdasarkan hasil pengujian presentasi prediksi yang dilakukan menggunakan sepuluh gambar wajah asli dan sepuluh gambar wajah manipulasi, terdapat satu error atau kesalahan yang terjadi, dimana gambar wajah palsu terdeteksi sebagai ‘Real’.
7. Berdasarkan hasil pengujian presentasi prediksi yang dilakukan dengan menggunakan tiga gambar hasil Canva editor dan tiga gambar hasil DPI converter, mendapatkan hasil analisis gambar manipulasi tersebut terdeteksi sebagai ‘Real’.

8. Hasil dari proses menyimpan dan mendapatkan metadata gambar asli dan palsu dapat digunakan untuk menjadi tolak ukur dalam penentuan akurasi keaslian sebuah gambar.
9. Hasil pengujian persentase prediksi menunjukkan jika metode ini dapat mendeteksi tingkat akurasi keaslian dengan baik.

Saran

Hasil dari proses analisis keaslian gambar wajah menunjukan jika metode yang digunakan dapat mendeteksi tingkat akurasi dan keaslian gambar melalui hasil prediksi *real* dan *fake.* Namun masih terdapat beberapa kekurangan dari metode yang diterapkan dalam penelitian ini.

Beberapa saran untuk pengembangan metode analisis ini yaitu dengan melakukan pembuatan gambar manipulasi VAE menggunakan metode lain seperti GAN atau AE. Untuk proses keakuratan gambar dapat dilengkapi dengan adanya pengambilan informasi gambar menggunakan metode *digital forensics* agar hasil akurasi menjadi lebih akurat dan dipercaya.

DAFTAR PUSTAKA

Candra, P. N., & Prapanca, A. (2020). Klasifikasi Gambar Asli dan Manipulasi Menggunakan Error Level Analysis (ELA) Sebagai Proses Komputasi Metode Convolutional Neural Network (CNN). *Journal of Informatics and Computer Science (JINACS)*, *2*(01), 9–18. https://doi.org/10.26740/jinacs.v2n01.p9-18

Das, A., Viji, K. S. A., & Sebastian, L. (2022, July 29). A survey on deepfake video detection techniques using deep learning. *2022 Second International Conference on Next Generation Intelligent Systems (ICNGIS)*. http://dx.doi.org/10.1109/icngis54955.2022.10079802

Das, S., Seferbekov, S., Datta, A., Islam, Md. S., & Amin, Md. R. (2021, October). Towards Solving the DeepFake Problem : An Analysis on Improving DeepFake Detection using Dynamic Face Augmentation. *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. http://dx.doi.org/10.1109/iccvw54120.2021.00421

Erhan, D., Szegedy, C., Toshev, A., & Anguelov, D. (2014, June). Scalable object detection using deep neural networks. *2014 IEEE Conference on Computer Vision and Pattern Recognition*. http://dx.doi.org/10.1109/cvpr.2014.276

Foster, D. (2019). *Generative deep learning: Teaching machines to paint, write, compose, and play*. "O'Reilly Media, Inc."

Keras: The high-level API for TensorFlow. (n.d.). *TensorFlow*. Retrieved June 25, 2023, from https://www.tensorflow.org/guide/keras

Kim, E., & Cho, S. (2021). Exposing fake faces through deep neural networks combining content and trace feature extractors. *IEEE Access*, *9*, 123493–123503. https://doi.org/10.1109/access.2021.3110859

Li, J. (n.d.). *CelebFaces attributes (celeba) dataset*. Kaggle. Retrieved June 25, 2023, from https://www.kaggle.com/jessicali9530/celeba-dataset

Malik, A., Kuribayashi, M., Abdullahi, S. M., & Khan, A. N. (2022). DeepFake detection for human face images and videos: A survey. *IEEE Access*, *10*, 18757–18775. https://doi.org/10.1109/access.2022.3151186

shaft49. (2020, September 9). real vs fake images (casia dataset). *Kaggle*. https://www.kaggle.com/code/shaft49/real-vs-fake-images-casia-dataset

Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A Survey of face manipulation and fake detection. *Information Fusion*, *64*, 131–148. https://doi.org/10.1016/j.inffus.2020.06.014

Yavuzkilic, S., Sengur, A., Akhtar, Z., & Siddique, K. (2021). Spotting deepfakes and face manipulations by fusing features from multi-stream cnns models. *Symmetry*, *13*(8), 1352. https://doi.org/10.3390/sym13081352