

BAB I PENDAHULUAN

1.1 Latar Belakang

Perkembangan kecerdasan buatan dalam bidang pengolahan citra dan video telah mendorong kemajuan signifikan pada teknologi manipulasi visual, salah satunya adalah deepfake yang mampu menghasilkan video wajah sintesis dengan tingkat kemiripan yang sangat tinggi sehingga sulit dibedakan dari video asli oleh pengamatan manusia. Fenomena ini menimbulkan kekhawatiran karena deepfake berpotensi disalahgunakan untuk penyebaran informasi palsu, penipuan digital, pencemaran nama baik, serta ancaman terhadap keamanan dan kepercayaan publik. Sejumlah penelitian menyatakan bahwa peningkatan kualitas deepfake membuat metode deteksi konvensional berbasis analisis statis atau satu frame menjadi kurang efektif, sehingga dibutuhkan pendekatan yang mampu memahami dinamika visual video secara lebih menyeluruh[1]

Seiring berkembangnya teknik manipulasi wajah, berbagai studi menemukan bahwa proses pembuatan deepfake sering menghasilkan ketidakkonsistenan temporal antar-frame, seperti pergerakan ekspresi wajah yang tidak alami, transisi gerakan yang tidak mulus, serta perubahan pola visual mikro yang sulit terdeteksi secara spasial saja. Ketidakwajaran tersebut lebih mudah diamati melalui analisis hubungan urutan waktu antar-frame dibandingkan pendekatan frame-by-frame[2], menunjukkan bahwa pembelajaran fitur spatio-temporal menggunakan jaringan 3D CNN mampu menangkap perbedaan pola gerakan pada video wajah palsu. Temuan serupa juga disampaikan oleh[3] memanfaatkan informasi gerak antar-frame untuk mengidentifikasi anomali visual pada video deepfake.

Untuk mengatasi keterbatasan metode statis, berbagai pendekatan berbasis pembelajaran mendalam mulai mengintegrasikan ekstraksi fitur spasial dan temporal secara bersamaan. Beberapa penelitian menggabungkan Convolutional Neural Network (CNN) sebagai ekstraktor fitur visual dengan Recurrent Neural

Network (RNN), khususnya Gated Recurrent Unit (GRU), untuk memodelkan hubungan temporal antar-frame.[4] serta penelitian lanjutan oleh MDPI Applied Sciences (2024) membuktikan bahwa arsitektur CNN–GRU mampu meningkatkan performa deteksi deepfake karena model tidak hanya memahami bentuk wajah, tetapi juga dinamika gerakannya. Pendekatan fusi spatio-temporal juga diperkuat melalui penggunaan attention mechanism dan information fusion untuk menangkap ketidakkonsistenan visual yang kompleks[5], [6]

Berdasarkan berbagai kajian tersebut, peluang penelitian terbuka lebar untuk mengembangkan sistem deteksi deepfake yang berfokus pada ekstraksi fitur temporal mendalam dari pola visual antar-frame. Penelitian ini diarahkan pada implementasi *deep temporal feature extraction* dengan memanfaatkan CNN sebagai ekstraktor fitur spasial pada setiap frame wajah dan GRU sebagai pemodel hubungan temporal antar-frame. Pendekatan ini dipilih karena terbukti efektif dalam menangkap inkonsistensi visual yang muncul akibat manipulasi video wajah, sekaligus relevan dengan tren penelitian terkini dalam deteksi deepfake berbasis video tanpa ketergantungan pada informasi audio [1], [2], [3], [7]

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan di atas, maka dapat dirumuskan permasalahan penelitian sebagai berikut:

1. Bagaimana penerapan metode *deep temporal feature extraction* dalam mendeteksi deepfake melalui analisis pola visual antar-frame pada video wajah?
2. Seberapa efektif kombinasi Convolutional Neural Network (CNN) dan Gated Recurrent Unit (GRU) dalam mempelajari hubungan temporal antar-frame untuk membedakan video deepfake dan video asli?

1.3 Batasan Masalah

Mengingat luasnya cakupan teknologi *Deep Learning* pada pengolahan video, peneliti membatasi ruang lingkup penelitian agar pembahasan lebih terarah dan mendalam. Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini dibatasi pada deteksi deepfake berbasis video wajah, sehingga objek yang dianalisis hanya berupa wajah manusia yang muncul pada video.
2. Analisis deepfake dalam penelitian ini hanya menggunakan informasi visual antar-frame, tanpa melibatkan informasi audio atau sinkronisasi suara.
3. Metode yang digunakan dalam penelitian ini dibatasi pada pendekatan *deep temporal feature extraction* dengan arsitektur Convolutional Neural Network (CNN) sebagai ekstraktor fitur spasial dan Gated Recurrent Unit (GRU) sebagai pemodel hubungan temporal antar-frame.
4. Proses ekstraksi fitur visual dilakukan pada urutan frame hasil *face cropping*, sehingga area selain wajah tidak menjadi bagian dari analisis.
5. Tahapan *preprocessing* visual dibatasi pada teknik dasar peningkatan kualitas citra, seperti *face cropping*, penyesuaian ukuran citra, dan peningkatan kontras, tanpa melakukan manipulasi visual lanjutan.

6. Dataset yang digunakan dalam penelitian ini dibatasi sebanyak 2.000 data video wajah, yang terdiri dari 1.000 video deepfake (fake) dan 1.000 video asli (real).

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan, tujuan yang akan dicapai oleh peneliti dalam penelitian ini adalah:

1. Untuk menerapkan metode *deep temporal feature extraction* dalam mendeteksi deepfake melalui analisis pola visual antar-frame pada video wajah.
2. Untuk menganalisis tingkat efektivitas kombinasi Convolutional Neural Network (CNN) dan Gated Recurrent Unit (GRU) dalam mempelajari hubungan temporal antar-frame guna membedakan video deepfake dan video asli.

1.5 Manfaat Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan, tujuan yang akan dicapai oleh peneliti dalam penelitian ini adalah:

1. Manfaat Teoritis

Penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan keilmuan di bidang kecerdasan buatan dan pengolahan citra digital, khususnya dalam kajian deteksi deepfake berbasis video. Hasil penelitian ini dapat memperkaya referensi terkait penerapan *deep temporal feature extraction* dalam menganalisis pola visual antar-frame pada video wajah. Selain itu, penelitian ini diharapkan dapat menjadi dasar ilmiah bagi penelitian selanjutnya yang berfokus pada pemodelan hubungan temporal menggunakan kombinasi Convolutional Neural Network (CNN) dan Gated Recurrent Unit (GRU) dalam deteksi manipulasi visual. Untuk Menganalisis kualitas citra luaran model untuk memvalidasi stabilitas deteksi dan ketajaman fitur wajah (*visual fidelity*) yang dihasilkan pada setiap urutan *frame* secara sekuensial.

2. Manfaat Praktis

Secara praktis, hasil penelitian ini diharapkan dapat dimanfaatkan sebagai dasar pengembangan sistem deteksi deepfake berbasis visual yang lebih efektif, khususnya pada video wajah. Model yang dihasilkan dapat digunakan sebagai pendukung dalam upaya verifikasi keaslian konten video digital pada berbagai bidang, seperti media digital, keamanan informasi, dan forensik multimedia. Selain itu, pendekatan deteksi yang hanya mengandalkan analisis visual antar-frame diharapkan dapat menjadi alternatif solusi yang lebih sederhana dan efisien tanpa ketergantungan pada informasi audio.

1.6 Sistematika Penulisan

1.6.1 BAB I PENDAHULUAN

Bab ini berisi uraian mengenai latar belakang penelitian, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, serta sistematika penulisan skripsi. Bab ini memberikan gambaran umum mengenai alasan dan ruang lingkup penelitian yang dilakukan.

1.6.2 BAB II TINJAUAN PUSTAKA

Bab ini memuat kajian pustaka yang berkaitan dengan penelitian, meliputi penelitian terdahulu yang relevan serta dasar-dasar teori yang mendukung penelitian. Pembahasan mencakup konsep deepfake, pengolahan citra dan video, deep learning, ekstraksi fitur spasial dan temporal, serta metode Convolutional Neural Network (CNN) dan Gated Recurrent Unit (GRU) yang digunakan dalam penelitian.

1.6.3 BAB III METODE PENELITIAN

Bab ini menjelaskan metode penelitian yang digunakan, meliputi gambaran umum objek penelitian, desain penelitian, tahapan pengolahan data, proses preprocessing video wajah, perancangan arsitektur model deep temporal feature extraction, serta skema pelatihan dan pengujian model. Pada bab ini juga dijelaskan dataset yang digunakan serta teknik evaluasi kinerja sistem.

1.6.4 BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi sistem deteksi deepfake berbasis analisis pola visual antar-frame. Pembahasan meliputi hasil pelatihan dan pengujian model, analisis performa deteksi, serta interpretasi hasil yang diperoleh berdasarkan tujuan penelitian. Evaluasi dilakukan untuk mengetahui efektivitas metode CNN-GRU dalam mendeteksi deepfake pada video wajah.

1.6.5 BAB V PENUTUP

Bab ini berisi kesimpulan yang diperoleh dari hasil penelitian serta saran untuk pengembangan penelitian selanjutnya. Kesimpulan disusun berdasarkan tujuan dan hasil pembahasan, sedangkan saran diberikan sebagai bahan pertimbangan untuk peningkatan metode atau perluasan ruang lingkup penelitian di masa mendatang.

