BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi kecerdasan buatan telah melahirkan *deepfake*, sebuah inovasi berbasis deep learning yang menggunakan algoritma seperti *GANs* dan *autoencoder*. Teknologi ini bekerja dengan mempelajari ekspresi wajah dan gerakan target untuk menciptakan manipulasi visual yang sangat realistis, sehingga hasilnya sulit dibedakan dari rekaman asli [1].

Di Indonesia, *deepfake* menghadirkan ancaman serius terhadap privasi, terutama melalui penyebaran konten pornografi non-konsensual. Namun, upaya penegakan hukum sering terkendala oleh kekosongan norma (*rechtsvakuüm*), di mana regulasi seperti UU ITE belum mengatur manipulasi visual secara spesifik. Oleh karena itu, pengembangan sistem deteksi yang handal menjadi solusi teknis yang mendesak untuk memitigasi dampak ini [2].

Mendeteksi *deepfake* pada skenario dunia nyata memiliki tantangan tersendiri akibat degradasi kualitas, seperti kompresi dan noise saat video diunggah ke media sosial. Kinerja model deteksi sering menurun signifikan saat menghadapi modifikasi data tersebut, sehingga diperlukan arsitektur yang tangguh (*robust*) serta metode pra-pemrosesan yang mampu memperbaiki kualitas fitur citra [3].

Sebagai solusi arsitektur, XceptionNet dipilih karena keunggulannya dalam menggantikan modul Inception standar dengan depthwise separable convolutions. Arsitektur ini terbukti lebih efisien dalam penggunaan parameter dan menghasilkan performa ekstraksi fitur yang lebih tinggi. [4].

Selain pemilihan model, tahapan pra-pemrosesan seperti penyelarasan wajah dan Contrast Limited Adaptive Histogram Equalization (*CLAHE*) memegang peranan krusial. Teknik ini berfungsi meningkatkan detail tekstur dan kontras lokal pada wajah, yang terbukti secara signifikan membantu model dalam mempelajari fitur anomali dan meningkatkan akurasi diagnose [5].

Berdasarkan tantangan tersebut, penelitian ini bertujuan membangun sistem deteksi *deepfake* dengan mengimplementasikan arsitektur XceptionNet sebagai *feature extractor* utama. Dengan menggabungkan model yang efisien dan teknik pra-pemrosesan yang tepat, sistem ini diharapkan mampu mengatasi masalah degradasi kualitas konten dalam ranah forensik digital [3].

Eskalasi penyalahgunaan teknologi *Deepfake* kian mengkhawatirkan dalam beberapa tahun terakhir. Nurdin dan Nugraha [6] menyoroti bahwa *Deepfake* telah berevolusi menjadi ancaman serius bagi keamanan nasional, di mana konten audio-visual yang dimanipulasi dapat digunakan untuk menyebarkan disinformasi yang merusak stabilitas negara. Dari perspektif hukum di Indonesia, Prayoga dan Tuasikal menegaskan bahwa penyebaran konten Deepfake yang merugikan dapat dikategorikan sebagai tindak pidana siber yang memerlukan penanganan forensik digital yang presisi [7].

Tantangan deteksi semakin kompleks ketika video dimanipulasi dengan teknik canggih yang sulit dikenali mata telanjang. Singh dan Dhumane (2025) dalam tinjauan terbarunya menyebutkan bahwa metode deteksi konvensional seringkali gagal menghadapi Deepfake generasi baru yang memanfaatkan *Generative Adversarial Networks* (*GANs*) yang terus diperbarui [8]. Selain itu, Kaur [9] menambahkan bahwa tantangan utama saat ini adalah mendeteksi video pada platform media sosial yang telah mengalami kompresi tinggi, sehingga menghilangkan jejak artefak visual yang biasanya menjadi petunjuk bagi detector. Oleh karena itu, diperlukan pendekatan berbasis *Deep Learning* yang lebih tangguh, seperti penggunaan arsitektur CNN modern yang dikombinasikan dengan teknik perbaikan citra.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang masalah, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana mengimplementasikan metode Convolutional Neural Network (CNN) dengan arsitektur XceptionNet untuk mendeteksi artefak spasial pada citra wajah hasil rekayasa deepfake?

2. Apakah penerapan tahapan pra-pemrosesan menggunakan metode Contrast Limited Adaptive Histogram Equalization (CLAHE) serta augmentasi data spasial secara signifikan dapat memperjelas jejak manipulasi pada dataset video berkompresi tinggi (c23)?
3. Bagaimana perbandingan performa (Akurasi, Presisi, Recall, dan F1-Score) antara model baseline (Skenario 1) dengan model yang dioptimasi menggunakan strategi Partial Unfreeze, Dropout (0.5), dan L2 Regularization (Skenario 2)?

1.3 Batasan Masalah

Untuk menjaga fokus penelitian agar tidak meluas dan tetap relevan dengan tujuan yang ingin dicapai, peneliti menetapkan batasan-batasan masalah sebagai berikut:

1. Lingkup Data (Data Scope):
   A. Objek penelitian difokuskan pada media visual berupa citra wajah digital (frame-based).
   B. Parameter yang diabaikan elemen audio/suara dan analisis temporal (urutan waktu antar-frame dalam video) tidak dimasukkan dalam penelitian ini, karena fokus utama adalah deteksi manipulasi visual spasial pada setiap frame.
2. Metode dan Algoritma (Method):
   A. Variabel yang diteliti: Arsitektur utama yang digunakan adalah Xception (Extreme Inception). Arsitektur lain seperti Inception V3 atau VGG16 hanya digunakan sebagai pembanding (baseline) untuk mengukur efektivitas Xception.
   B. Teknik Deep Learning dibatasi pada ranah Supervised Learning menggunakan dataset berlabel.
3. Teknik Pra-pemrosesan (Preprocessing Variables):
   A. Variabel preprocessing yang akan diuji pengaruhnya secara spesifik adalah perbaikan kontras menggunakan metode CLAHE

(Contrast Limited Adaptive Histogram Equalization) dan penyelarasan wajah (Face Alignment) berbasis landmark wajah.

B. Teknik preprocessing lain di luar kedua metode tersebut tidak diterapkan dalam penelitian ini.

4. Asumsi Penelitian:

A. Diasumsikan bahwa setiap citra masukan memiliki setidaknya satu wajah yang dapat dideteksi oleh sistem pendeteksi wajah standar (seperti MTCNN atau Haar Cascade). Citra tanpa wajah atau dengan oklusi (tertutup) ekstrem berada di luar lingkup deteksi sistem ini.

B. Kualitas citra masukan diasumsikan beragam, mulai dari kualitas tinggi hingga yang mengalami degradasi (kompresi/noise), untuk menguji ketangguhan model.

1.4 Tujuan Penelitian

Mengacu pada permasalahan yang telah dirumuskan, tujuan utama yang hendak dicapai oleh peneliti dalam penelitian ini adalah:

1. Mengimplementasikan dan memvalidasi kinerja model *Deep Learning* berbasis arsitektur XceptionNet untuk klasifikasi citra *deepfake*, serta menganalisis efektivitas strategi *Partial Unfreeze* dalam meningkatkan stabilitas model (mencegah *overfitting*) dan akurasi dibandingkan dengan strategi pelatihan standar (*baseline*).
2. Menganalisis pengaruh penerapan teknik preprocessing khususnya perbaikan kontras (*CLAHE*) dan penyelarasan wajah (*Face Alignment*) terhadap peningkatan ketangguhan (*robustness*) model saat mendeteksi citra yang mengalami penurunan kualitas akibat kompresi atau gangguan noise.

1.5 Manfaat Penelitian

Berdasarkan tujuan yang telah ditetapkan, hasil dari penelitian ini diharapkan dapat memberikan kontribusi positif yang dapat dimanfaatkan baik secara teoritis maupun praktis, sebagai berikut:

1.5.1 Manfaat Teoritis

Secara keilmuan, penelitian ini diharapkan dapat memperkaya khazanah pengetahuan di bidang Teknik Komputer, khususnya pada ranah Visi Komputer (*Computer Vision*) dan Forensik Digital:

1. Memberikan bukti empiris mengenai ketangguhan arsitektur XceptionNet sebagai ekstraktor fitur dalam mendeteksi manipulasi wajah (*deepfake*) pada kondisi video kualitas rendah, serta memvalidasi peran teknik *preprocessing CLAHE* dalam mempertegas artefak spasial agar lebih mudah dikenali oleh model.
2. Menyumbangkan analisis teknis mengenai efektivitas metode pra-pemrosesan (*preprocessing*) gabungan yaitu perbaikan kontras (*CLAHE*) dan penyelarasan wajah dalam meningkatkan akurasi deteksi pada citra yang mengalami degradasi kualitas (seperti kompresi atau *noise*), sehingga dapat menjadi referensi bagi pengembangan model *Deep Learning* yang lebih *robust* di masa depan.

1.5.2 Manfaat Praktis

Secara aplikatif, hasil penelitian ini diharapkan dapat memberikan manfaat langsung bagi berbagai pemangku kepentingan:

1. Bagi Penegak Hukum & Ahli Forensik, sistem berfungsi sebagai perangkat pendukung (*supporting tool*) verifikasi keaslian bukti digital, krusial untuk mengatasi kendala pembuktian visual dalam kasus konten manipulatif seperti hoaks atau pornografi non-konsensual.

2. Bagi Masyarakat, berkontribusi pada perlindungan privasi dan keamanan digital dengan menyediakan dasar teknis deteksi penyalahgunaan identitas, guna meminimalisir risiko kerugian akibat disinformasi dan pencemaran nama baik berbasis *deepfake*.
3. Bagi Pengembang sistem, Menjadi panduan teknis dalam merancang filter keamanan konten yang adaptif terhadap gangguan kualitas data (seperti kompresi dan noise) pada skenario dunia nyata

1.6 Sistematika Penulisan

BAB I PENDAHULUAN

Bab ini menjelaskan dasar pemikiran dilakukannya penelitian, yang mencakup latar belakang masalah mengenai maraknya penyebaran video manipulasi wajah (*deepfake*) dan dampaknya. Selain itu, bab ini juga memuat rumusan masalah, batasan masalah (fokus pada arsitektur XceptionNet), tujuan penelitian, manfaat penelitian, serta sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini menguraikan rangkuman penelitian terdahulu yang relevan dengan deteksi *deepfake*, serta memaparkan landasan teori yang digunakan sebagai acuan dalam penelitian. Teori yang dibahas meliputi konsep *Deep Learning*, *Convolutional Neural Network* (*CNN*), Transfer Learning, detail arsitektur XceptionNet, teknologi *Deepfake*, serta metrik evaluasi performa model.

BAB III METODE PENELITIAN

Bab ini menjelaskan tahapan dan metode yang dilakukan dalam menyelesaikan masalah penelitian. Di dalamnya terdapat uraian mengenai alur penelitian, teknik pengumpulan dataset, tahapan pra-pemrosesan data (*preprocessing*), perancangan arsitektur model menggunakan XceptionNet, serta skenario pengujian yang akan dilakukan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini memuat hasil dari implementasi sistem yang telah dirancang. Pembahasan mencakup proses pelatihan (*training*) model, analisis grafik akurasi dan loss, hasil pengujian model terhadap data uji, serta analisis kinerja model XceptionNet dalam mendeteksi video *deepfake* berdasarkan metrik evaluasi yang telah ditentukan.

BAB V PENUTUP

Bab ini berisi kesimpulan yang diperoleh dari hasil analisis dan pengujian sistem secara keseluruhan, apakah model berhasil memenuhi tujuan penelitian. Bab ini juga memuat saran-saran yang konstruktif untuk pengembangan penelitian selanjutnya agar sistem deteksi *deepfake* dapat menjadi lebih baik.