BAB I
PENDAHULUAN

1.1. Latar Belakang

Perkembangan *teknologi digital* dewasa membawa dampak besar terhadap berbagai aspek kehidupan manusia, mulai dari industri, hiburan, pendidikan, hingga komunikasi sosial. Salah satu dampak dari kemajuan ini adalah kemudahan dalam mengakses dan membagikan informasi, terutama dalam bentuk visual seperti gambar dan video. Akan tetapi, di balik manfaatnya yang begitu luas, kemajuan teknologi juga membuka celah terhadap berbagai penyalahgunaan, salah satunya adalah fenomena yang dikenal dengan istilah *deepfake*[1].

Deepfake merupakan gabungan dari kata "*deep learning*" dan "*fake*" yang merujuk pada konten visual baik gambar maupun video yang dimanipulasi sedemikian rupa menggunakan teknik kecerdasan buatan, sehingga menghasilkan tampilan yang sangat meyakinkan namun sebenarnya palsu[2]. Teknologi ini biasanya digunakan untuk memalsukan wajah atau gerakan mulut seseorang dalam video, seolah-olah orang tersebut benar-benar melakukan atau mengatakan sesuatu, padahal tidak demikian kenyataannya[3]. Fenomena ini menimbulkan keresahan karena mampu mengaburkan batas antara fakta dan rekayasa, terlebih lagi karena hasil manipulasi yang dihasilkan sering kali sangat *real*istis dan sulit dikenali secara kasat mata.

Dalam konteks sosial, keberadaan *deepfake* sangat berpotensi menimbulkan kerugian besar. Misalnya, dalam dunia politik, video *deepfake* dapat digunakan untuk menjatuhkan reputasi tokoh publik. Di sisi lain, dalam bidang keamanan siber, keberadaan citra *deepfake* dapat memperbesar kemungkinan terjadinya penipuan identitas. Bahkan dalam kasus tertentu, teknologi ini telah disalahgunakan untuk membuat konten pornografi non-konsensual dengan mencatut wajah orang lain. Karena itulah, kebutuhan akan teknologi yang mampu mendeteksi dan mengidentifikasi *deepfake* secara otomatis dan akurat semakin mendesak[4].

Kemajuan teknologi kecerdasan buatan saat ini telah menghadirkan berbagai inovasi dalam bidang multimedia, salah satunya adalah *deepfake*. *Deepfake* merupakan teknik manipulasi gambar atau video yang mampu menampilkan wajah seseorang seolah-olah mengatakan atau melakukan sesuatu yang sebenarnya tidak pernah terjadi. Sekilas, konten *deepfake* ini tampak sangat meyakinkan, sehingga sulit dibedakan dengan konten asli[5]. Meski di satu sisi teknologi ini memiliki nilai kreatif

dalam dunia hiburan dan industri perfilman, di sisi lain penyalahgunaannya dapat menimbulkan dampak negatif yang serius, seperti penyebaran hoaks, pencemaran nama baik, hingga ancaman terhadap privasi dan keamanan informasi[6].

Untuk menghadapi persoalan tersebut, berbagai pendekatan teknologi telah dikembangkan guna mendeteksi keberadaan konten *deepfake* secara otomatis dan efisien. Salah satu metode yang paling banyak digunakan adalah *Convolutional Neural Network* (CNN)[7]. CNN dikenal karena kemampuannya dalam mengenali pola visual pada citra digital. Dengan memanfaatkan lapisan-lapisan *konvolusi* dan *pooling*, CNN mampu mengekstrak fitur-fitur penting dari suatu gambar, yang kemudian digunakan untuk melakukan klasifikasi. Model ini telah banyak digunakan dalam berbagai bidang, seperti pengenalan wajah, deteksi objek, dan pengolahan citra medis[8].

Meskipun demikian, CNN memiliki kekurangan yang cukup krusial, terutama dalam menangkap hubungan spasial yang lebih luas pada gambar. CNN standar cenderung hanya memproses informasi dari area yang sempit (*receptive field* terbatas), sehingga kurang efektif saat harus mengidentifikasi manipulasi yang tersebar tipis di berbagai bagian citra[9]. Masalah ini sering muncul ketika CNN dihadapkan pada citra *deepfake* yang tampak sangat halus dan tidak memiliki perbedaan mencolok secara lokal[10].

Oleh karena itu, dalam penelitian ini dipilihlah arsitektur *Convolutional Neural Network* (CNN) yang diintegrasikan dengan teknik *dilated convolution*. Integrasi ini dipilih karena kemampuannya untuk memperluas jangkauan pandang jaringan terhadap sebuah gambar tanpa harus menambah kompleksitas parameter dan tanpa mengorbankan resolusi dari data yang diproses, sehingga memungkinkan model untuk menangkap konteks global sekaligus mendeteksi detail lokal secara bersamaan dalam satu kerangka kerja yang efisien[11]. Kemampuan ini menjadi krusial dalam deteksi *deepfake*, di mana cacat manipulasi seringkali sangat halus dan muncul dalam berbagai skala yang berbeda, dan tersebar di seluruh *frame*, *dilated convolution* secara efektif meningkatkan sensitivitas jaringan terhadap indikator-indikator manipulasi tersebut seperti ketidaksempurnaan pada batas hasil gabungan, tekstur kulit yang tidak wajar, atau ketidakkonsistenan pada kontur wajah. Dibandingkan dengan pendekatan berbasis fitur manual seperti SVM atau *Random Forest*, pendekatan *end-to-end* dari CNN ini juga menunjukkan ketangguhan yang lebih baik terhadap variasi sumber data dan perbedaan karakteristik dari berbagai generator *deepfake*. Sementara itu,

pendekatan alternatif seperti RNN yang dirancang untuk data sekuensial dinilai kurang tepat untuk gambar tunggal karena hanya akan menambah kompleksitas komputasi tanpa manfaat yang signifikan, dan untuk data gambar wajah, *Convolutional Neural Network* (CNN) dengan *dilated convolution* sudah menawarkan alternatif yang lebih unggul dalam hal paralelisme komputasi dan kemampuan menangkap dependensi jarak jauh. Dengan pertimbangan-pertimbangan itulah, integrasi *dilated convolution* pada CNN ini diyakini memberikan keseimbangan optimal antara cakupan kontekstual, sensitivitas multi-skala, efisiensi komputasi, dan ketahanan terhadap beragam teknik manipulasi yang ada[12].

Untuk mengatasi keterbatasan tersebut, salah satu pendekatan yang menjanjikan adalah dengan menambahkan teknik *Dilated Convolution* ke dalam arsitektur CNN [13]. Berbeda dengan *konvolusi* biasa, *Dilated Convolution* memungkinkan jaringan untuk menangkap informasi dari area yang lebih luas tanpa meningkatkan jumlah parameter secara signifikan. Dengan begitu, model tidak hanya mampu mengenali detail kecil di bagian lokal, tetapi juga mampu memahami pola global dalam citra. Pendekatan ini dinilai efektif dalam meningkatkan sensitivitas model terhadap manipulasi visual yang tersembunyi[14].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan dan menguji kinerja *Convolutional Neural Network* yang dimodifikasi dengan integrasi *Dilated Convolution* dalam mendeteksi citra *deepfake*. Diharapkan, pendekatan ini dapat memberikan hasil yang lebih akurat dan mampu beradaptasi terhadap kompleksitas konten digital yang semakin berkembang. Penelitian ini juga diharapkan dapat menjadi salah satu kontribusi dalam pengembangan sistem pendeteksi konten palsu berbasis kecerdasan buatan yang dapat diterapkan secara nyata di masyarakat.

1.2. Rumusan Masalah

Berdasarkan latar belakang tersebut, dapat dirumuskan beberapa masalah penelitian yang akan dijawab dalam skripsi, antara lain:

1. Bagaimana kinerja model *Convolutional Neural Network* (CNN) dengan integrasi *Dilated Convolution* dalam mendeteksi citra *deepfake* berdasarkan hasil akurasi klasifikasi?

2. Sejauh mana efektivitas penerapan *Dilated Convolution* dalam meningkatkan performa deteksi citra *deepfake* pada model CNN berdasarkan evaluasi menggunakan *Confusion Matrix*?

1.3. Tujuan Penelitian

Tujuan dari skripsi dengan judul " Deteksi Citra *Deepfake* Menggunakan CNN dengan Integrasi *Dilated convolution*" adalah sebagai berikut :

1. Untuk mengetahui performa model *Convolutional Neural Network* (CNN) dengan integrasi *Dilated Convolution* dalam mendeteksi citra *deepfake* berdasarkan hasil akurasi klasifikasi.
2. Untuk menganalisis efektivitas integrasi *Dilated Convolution* dalam meningkatkan akurasi deteksi citra *deepfake* pada model CNN berdasarkan evaluasi menggunakan *Confusion Matrix*.

1.4. Manfaat Penelitian

Beberapa manfaat yang dapat diambil dari skripsi dengan judul " Deteksi Citra *Deepfake* Menggunakan CNN dengan Integrasi *Dilated convolution* " antara lain:

1. Memberikan pemahaman tentang performa CNN dengan integrasi *Dilated Convolution*

   Penelitian ini diharapkan dapat memberikan gambaran yang lebih jelas mengenai kemampuan model *Convolutional Neural Network* (CNN) yang dimodifikasi dengan *Dilated Convolution* dalam mendeteksi citra *deepfake*. Dengan memahami bagaimana model ini bekerja dalam mengklasifikasikan citra yang telah dimanipulasi, pembaca dan peneliti lain dapat mengetahui sejauh mana efektivitas pendekatan ini dalam konteks deteksi konten visual palsu.
2. Menjadi acuan dalam pemilihan metode deteksi *deepfake* yang lebih tepat

   Hasil evaluasi dalam penelitian ini dapat membantu dalam menentukan apakah penambahan *Dilated Convolution* benar-benar membawa peningkatan performa dibandingkan CNN standar. Hal ini penting bagi pengembang atau peneliti yang ingin menerapkan metode deteksi citra *deepfake*, khususnya dalam situasi yang membutuhkan akurasi tinggi atau efisiensi dalam pemrosesan.
3. Mendukung pengembangan sistem deteksi citra manipulatif di masa depan

Penelitian ini diharapkan menjadi bagian dari upaya memperkuat sistem deteksi *deepfake* berbasis teknologi. Dengan memahami kelebihan dan batasan pendekatan CNN-*Dilated*, para pengembang sistem bisa merancang strategi atau arsitektur baru yang lebih adaptif dan andal, baik dalam skala penelitian maupun implementasi dunia nyata.

4. Berpotensi diterapkan dalam keamanan digital dan media forensik

   Temuan dari penelitian ini dapat dimanfaatkan dalam berbagai bidang, seperti keamanan siber, verifikasi keaslian konten di media sosial, hingga analisis forensik digital. Model yang dikembangkan dapat menjadi dasar bagi sistem otomatis yang mampu menyaring atau menandai konten manipulatif sebelum menyebar ke masyarakat luas.

Dengan berbagai manfaat tersebut, penelitian ini diharapkan dapat memberikan kontribusi nyata dalam mendorong pengembangan teknologi pengolahan citra digital yang lebih canggih dan bertanggung jawab, khususnya dalam menghadapi tantangan penyebaran konten *deepfake* di era digital saat ini.

1.5.Batasan Masalah

Batasan masalah pada penelitian skripsi dengan judul "Deteksi Citra *Deepfake* Menggunakan CNN dengan Integrasi *Dilated convolution*" adalah sebagai berikut:

1. Fokus pada citra wajah yang dimanipulasi (*deepfake*)

   Penelitian ini difokuskan pada klasifikasi citra wajah untuk membedakan antara citra asli dan citra hasil manipulasi *deepfake*. Penelitian ini tidak mencakup deteksi manipulasi dalam bentuk video utuh, suara, atau data multimodal lainnya. Fokus utama hanya pada citra diam (still image) yang diekstrak dari dataset *deepfake*.

2. Menggunakan model CNN dengan integrasi *Dilated Convolution*

   Model yang digunakan dalam penelitian ini adalah *Convolutional Neural Network* (CNN) yang telah dimodifikasi dengan teknik *Dilated Convolution*. Penelitian tidak membahas penggunaan varian CNN lain atau penggabungan dengan algoritma eksternal seperti LSTM, SVM, atau *attention mechanism*. Fokus utama adalah mengevaluasi sejauh mana integrasi *Dilated Convolution* mampu meningkatkan kemampuan CNN dalam mendeteksi citra *deepfake*.

3. Ruang lingkup terbatas pada klasifikasi biner

Penelitian ini hanya membedakan antara dua kategori citra, yaitu citra asli dan citra hasil manipulasi *deepfake*. Deteksi tidak mencakup identifikasi jenis manipulasi, algoritma pembuat *deepfake*, maupun tingkat kompleksitas manipulasi. Sistem dikembangkan semata-mata untuk mengklasifikasikan citra sebagai "asli" atau "palsu".

4. Dataset yang digunakan adalah citra dari dataset publik *deepfake*

   Data citra yang digunakan dalam penelitian ini diambil dari dataset publik yang tersedia di internet, yaitu Kaggle. Dataset ini terdiri dari citra wajah asli dan citra hasil manipulasi yang telah diberi label kebenaran (ground truth).

5. Akurasi sebagai metrik evaluasi utama

   Metrik utama yang digunakan untuk mengevaluasi performa model adalah akurasi, yang dihitung melalui *Confusion Matrix*. Metrik lain seperti *precision*, *recall*, dan *F1-score* tidak dijadikan fokus utama dalam penelitian ini, meskipun tetap dapat dipertimbangkan sebagai referensi tambahan.

6. Tidak mencakup aspek temporal atau video sequence

   Penelitian ini tidak membahas deteksi manipulasi berdasarkan urutan frame dalam video atau analisis pergerakan wajah dari waktu ke waktu. Cakupan penelitian sepenuhnya terbatas pada citra statis, tanpa mempertimbangkan dimensi temporal atau dinamika visual antar frame.