BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan pesat teknologi *Generative AI* saat ini telah memungkinkan terciptanya citra sintetik dengan tingkat realisme visual yang sangat tinggi dan sulit dibedakan oleh mata manusia [1]. Salah satu bentuk implementasinya, yakni visualisasi yang diciptakan, dimanipulasi, atau dianalisis melalui proses algoritma yang canggih. Penciptaan citra sintetik yang sangat mirip dengan gambar asli ini dimungkinkan oleh kemajuan model generatif seperti *Generative Adversarial Networks* (GAN), *Variational Autoencoders* (VAE), dan model difusi. Keberadaan teknologi pembuat citra tersebut kini telah dimanfaatkan secara luas dalam berbagai bidang, mulai dari seni digital, augmentasi data pelatihan, *Augmented Reality* (AR), hingga pembuatan konten media sosial [2]. Meskipun sangat bermanfaat, kemiripan visual yang nyaris sempurna antara citra buatan dan aslinya ini menimbulkan tantangan serius terkait verifikasi keaslian. Kemajuan pesat teknologi gambar yang dihasilkan AI telah mencapai titik dimana membedakan gambar asli dan sintetik menjadi tantangan bagi manusia [3]. Hal ini didukung pada penelitian [4] terkait konten deepfake pada pemilihan politik di kanada. Hasil penelitian tersebut menunjukkan dengan banyaknya media sintetik saat ini yang mudah diakses dan sulit dibedakan keasliaannya menimbulkan risiko lebih tinggi untuk menyesatkan pengguna. Selain itu *deepfake* juga memicu perdebatan mengenai keaslian dan kepemilikan kreatif, seperti kontroversi ketika sebuah gambar hasil *Generative AI* memenangkan penghargaan seni. Kondisi tersebut secara tidak langsung menyebabkan penurunan tingkat kepercayaan publik di berbagai platform, mulai dari media sosial dan konten daring hingga media massa tradisional [5].

Untuk mengatasi masalah diatas, beberapa penelitian melakukan pendekatan mendeteksi *deepfake* menggunakan *deep learning*. Pada Penelitian [6] menggunakan *Vision Transformer* (ViT) untuk mendeteksi *deepfake* dan menghasilkan sebuah dataset OPENFAKE. Pembuatan dataset OPENFAKE didorong oleh kebutuhan mendesak untuk mengatasi ancaman misinformasi yang semakin canggih, terutama di rana politik. Hasil dari penelitian tersebut menghasilkan akurasi 99%. Namun model ViT masih kurang dapat diinterprestasikan dan membutuhkan sumber daya yang besar serta dataset masif agar dapat belajar dengan baik [7]. Oleh karna itu, model ini tidak cocok digunakan pada kasus dengan ketersediaan data yang terbatas [6]. Sebagai solusi untuk mengatasi tingginya beban kerja

perangkat keras tanpa mengorbankan kualitas deteksi, arsitektur model ResNet-50 diajukan sebagai algoritma klasifikasi biner andal yang lebih efisien. Keandalan arsitektur ini didukung oleh penelitian [8] yang mendeteksi citra sintetik menggunakan model ResNet-50 di integrasikan dengan *Squeeze-and-Excitation (SE) Block* mencapai akurasi 96% pada dataset AI vs. Human-Generated Images. Selain itu, performa tinggi juga dibuktikan oleh penelitian [9] untuk mendeteksi *deepfake* menggunakan ResNet-50 mencapai 91% menggunakan *Modified Squeeze-and-Excitation* (MSE) pada data h FaceForensics++. Meskipun demikian, masih memiliki keterbatasan dalam memfokuskan perhatian pada bagian-bagian citra yang mengandung artefak atau pila manipulasi halus, sehingga beberapa informasi penting berpontensi kurang dimanfaatkan selama proses ekstraksi fitur [10].

Salah satu pendekatan yang dapat digunakan untuk mengatasi keterbatasan tersebut adalah dengan mengintegrasikan *attention mechanism* ke dalam arsitektur model ResNet50. *Convolutional Block Attention Module* (CBAM) merupakan *attention mechanism* yang dirancang untuk mengarahkan model agar berfokus secara selektif pada fitur-fitur yang paling informatif dalma suatu citra dengan penambahan kompleksitas komputasi yang relatif kecil. Pada penelitian [11] melakukan evaluasi arsitektur CNN pada dataset CIFAR10 menunjukkan bahwa Model Resnetx+CBAM memiliki akurasi unggul dibandingkan dengan model CNN based dan CNN+SE. Berdasarkan penjelasan diatas, maka penelitian ini mengusulkan model deteksi citra sintetik menggunakan arsitektur model ResNet-50 yang diintegrasikan dengan CBAM untuk meningkatkan kemampuan model dalam mengenali karakteristik citra sintetik. Pada pengujian ini menggunakan dataset berkarakteristik seperti OPENFAKE, yaitu citra yang tidak hanya berfokus pada potret manusia, melainkan juga mencakup elemen visual kompleks seperti pemandangan, furnitur, dan latar belakang, sebagai upaya untuk memecahkan rumusan masalah tersebut. Berdasarkan uraian diatas, maka pada penelitian ini diusulkan judul "Deteksi Citra Sintetik Menggunakan ResNet-50 dengan Convolutional Block Attention Module".

1.2 Rumusan Masalah

Berdasarkan latar belakang, rumusan masalah dalam penelitian ini adalah sulitnya mendeteksi citra asli dan sinteksi menggunakan sumber daya komputasi yang terbatas.

1.3 Tujuan

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Mengimplementasikan algoritma klasifikasi biner untuk deteksi citra sintetik menggunakan ResNet-50 yang diintegrasikan dengan CBAM.
2. Menyediakan alternatif model deteksi citra yang efisien secara perangkat keras.

1.4 Manfaat

Manfaat dari penelitian ini adalah sebagai berikut:

1. Memberikan literatur terkain efektivitas penerapan CBAM pada arsitektur ResNet50 untuk deteksi cintra sintetik.
2. Menjadi referesni bagi penelitina selanjutnya untuk mengembangkan deteksi citra sintetik pada dataset yang memiliki karakteristik.

1.5 Ruang Lingkup

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini berfokus pada penggunaan arsitektur ResNet-50 yang dimodifikasi dengan mengintegrasikan modul atensi CBAM pada setiap blok *bottleneck* untuk meningkatkan fokus pada fitur spasial dan saluran (*channel*).
2. Dataset yang digunakan adalah dataset sekunder OPENFAKE dari platform Kaggle, yang mencakup citra asli dan citra sintetik dengan menggunakan dataset 8,522. https://www.kaggle.com/datasets/vijithsai/fake-real-images
3. Masalah dibatasi pada klasifikasi biner, yaitu membedakan citra ke dalam dua kategori: citra asli (*Real*) yang dihasilkan manusia tanpa manipulasi algoritma, dan citra sintetik (*Fake*) hasil produksi model AI.
4. Tahapan *preprocessing* data dibatasi pada pembersihan citra dari konten tidak pantas (*NSFW filtering*), pembagian dataset dengan rasio 80:10:10 (data latih, validasi, dan uji), penyesuaian dimensi resolusi (*resize*) menjadi 224 × 224 piksel, normalisasi menggunakan nilai standar ImageNet, serta penerapan augmentasi data secara dinamis yang hanya diberlakukan pada data latih.
5. Keberhasilan penelitian diukur melalui parameter *Confusion Matrix* (Akurasi, Presisi, *Recall*, dan F1-Score) serta analisis visual Grad-CAM untuk memverifikasi area fokus perhatian model pada citra.