

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence*) dalam beberapa tahun terakhir telah mendorong kemajuan signifikan dalam kemampuan generatifnya, khususnya dalam pembuatan konten berbasis seni. Salah satu konten berbasis seni yang memiliki perkembangan pesat dari AI adalah generasi ilustrasi digital, dengan model generatif modern seperti *Generative Adversarial Networks* (GAN) dan model berbasis difusi yang menghasilkan ilustrasi digital dengan kualitas tinggi dan detail yang menyerupai karya manusia [1]. Kemudahan akses terhadap teknologi ini telah mendorong peningkatan penggunaan gambar hasil generasi AI dalam berbagai platform digital dan industri kreatif [2]. Namun, kemajuan tersebut juga menyebabkan batas antara karya yang dibuat oleh manusia dan yang dihasilkan oleh AI semakin sulit untuk dibedakan, sehingga menimbulkan berbagai kekhawatiran terkait orisinalitas, hak cipta, serta transparansi sumber konten [3].

Meskipun terdapat permasalahan dalam bidang ini, sebagian besar dari pendeteksi atau pengklasifikasi gambar lebih berfokus pada gambar dunia nyata [4]. Ilustrasi *anime* merupakan bagian dari industri kreatif yang mencakup porsi yang cukup besar [5]. Namun, karakteristik visual yang berbeda dari ilustrasi *anime* dibandingkan dengan gambar natural menyebabkan pendekatan yang digunakan pada citra dunia nyata tidak selalu dapat diterapkan secara langsung. Oleh karena itu, diperlukan metode khusus yang mampu menangani karakteristik unik dari ilustrasi *anime*, terutama dalam konteks klasifikasi gambar buatan AI dan manusia.

Urgensi permasalahan ini semakin nyata melalui rangkaian insiden yang terjadi di berbagai konvensi dan komunitas kreatif global sepanjang tahun 2025. Pada bulan September 2025, sebuah *booth* di Dragon Con, Atlanta, dievakuasi oleh pihak kepolisian setelah gagal membuktikan bahwa karya yang dijualnya merupakan hasil karya manusia, menyusul laporan dari pengunjung dan sesama *vendor* [6]. Insiden ini turut memicu respons dari penyelenggara konvensi lain; GalaxyCon mengumumkan larangan menyeluruh terhadap karya AI di seluruh rangkaian acaranya [7], sementara FanX di Salt Lake City melarang penjualan karya AI dari area *vendor* dengan alasan tidak memenuhi standar orisinalitas

[8]. Protes serupa juga terjadi di konvensi bertema *anime* seperti Animé Los Angeles, yang secara eksplisit melarang karya AI dari materi promosi, *exhibit hall*, dan *artist alley* [9]. Rangkaian insiden tersebut menunjukkan bahwa klaim keaslian karya telah menjadi permasalahan nyata dan berulang di komunitas kreatif, sementara deteksinya pada umumnya masih bergantung pada laporan komunitas dan investigasi manual, tanpa adanya metode verifikasi otomatis yang objektif.

Seiring dengan perkembangan dan kemudahan akses dari teknologi generasi ini, ilustrasi digital yang disebar dan dibuat semakin sulit dibedakan [10]. Kondisi ini menimbulkan tantangan bagi berbagai pihak, terutama para seniman, khususnya terkait perlindungan hak cipta atas karya mereka yang digunakan sebagai data pelatihan model generatif, yang oleh sebagian pihak bahkan dianggap sebagai bentuk pencurian [11]. Kekhawatiran ini tercermin dalam sebuah survei terhadap 203 seniman profesional, yang menemukan bahwa 79% di antaranya khawatir karyanya dimanfaatkan tanpa izin untuk melatih model generatif AI, dan 83% telah mengambil langkah protektif seperti penggunaan alat *anti-scraping* untuk mencegah penyalahgunaan karya mereka [12]. Oleh karena itu, terdapat kebutuhan akan suatu metode untuk dapat mengidentifikasi atau mengklasifikasikan apakah suatu ilustrasi merupakan hasil karya manusia atau buatan AI. Solusi ini diharapkan dapat membantu dan mendukung transparansi, menjaga keaslian karya, serta membantu moderasi dalam pengelolaan konten digital dalam bentuk ilustrasi.

Salah satu pendekatan yang dapat dilakukan adalah melakukan klasifikasi menggunakan model *deep learning*, seperti *EfficientNetV2* yang telah terbukti dalam kemampuannya dalam tugas klasifikasi gambar [13]. Pemilihan *EfficientNetV2* dibandingkan arsitektur yang lebih ringan seperti *MobileNet* mempertimbangkan cakupan penelitian terdahulu [14], yang menggunakan dataset berskala kecil dari satu generator AI dan secara eksplisit menyatakan perlunya pengembangan lebih lanjut pada dataset yang lebih besar. Penelitian ini merespons hal tersebut dengan memperluas skala dan keberagaman dataset, sehingga *EfficientNetV2* dipilih untuk kapasitas representasi fitur yang lebih tinggi dalam mengakomodasi variasi visual yang lebih beragam. *EfficientNetV2* sendiri adalah peningkatan dari *EfficientNet*, yang mengoptimalkan performa dengan menggunakan jumlah parameter yang lebih kecil serta membuat proses pelatihan data lebih cepat [15]. Arsitektur *EfficientNet* adalah berbasis *Convolutional Neural Network* (CNN), dengan pendekatan *compound scaling*. *Compound scaling* adalah metode penskalaan model, dengan meningkatkan *depth*, *width*, dan resolusi dari input (*image/input resolution*) yang meningkatkan efisiensi dan akurasi dari

EfficientNet [16].

Dalam upaya meningkatkan performa dari CNN, *Convolutional Block Attention Module* diciptakan sebagai modul tambahan opsional yang mudah diintegrasikan ke model CNN [17]. CBAM ini merupakan mekanisme *attention* yang secara adaptif memberikan fokus melalui dua *sub-module*, yaitu *channel attention* dan *spatial attention*. Kedua komponen *attention* ini bekerja secara berurutan untuk memberikan fokus atau perhatian ke fitur yang relevan dari *input*. Modul ini telah digunakan dalam beberapa penelitian sebelumnya, yang terbukti meningkatkan performa dari *EfficientNetV2* [18], *DeepConvNet* [19], dan *Res4net* [20].

Berbagai penelitian telah dilakukan dalam usaha untuk memberikan solusi terhadap masalah ini. [21] membahas tentang pendeteksi gambar buatan AI berbasis *Convolutional Neural Network (CNN)*, yang masih mudah mengklasifikasikan antara gambar buatan AI dan yang bukan. Penemuan dari riset tersebut juga dibahas di [22], yang menyatakan bahwa ada "sidik jari" dalam gambar yang dibuat oleh GAN untuk mempermudah klasifikasi. Sedangkan untuk tugas pendeteksian dari model berbasis difusi, [14] juga menunjukkan hasil yang mirip, dengan model yang berhasil mengklasifikasi semua gambar AI dengan benar, namun dengan dataset yang lebih kecil dan model yang lebih ringan, seperti *MobileNet*. Penelitian ini dilakukan dengan menggunakan model *EfficientNetV2*, dengan penambahan lapisan CBAM.

Penelitian ini mencakup klasifikasi gambar dari model *EfficientNetV2* untuk gambar *anime*, dengan implementasi CBAM pada tahap akhir pemrosesan gambar. Sistem ini dirancang dengan tujuan mengklasifikasi gambar *anime* berdasarkan buatan AI atau manusia. Penelitian ini juga bertujuan untuk mengevaluasi performa model *EfficientNetV2* dalam mengklasifikasi gambar dengan tipe *anime*, serta pengaruh modul CBAM terhadap model dalam klasifikasi dengan berbagai metrik evaluasi. Hasil yang diperoleh diharapkan dapat dimanfaatkan untuk memberikan kontribusi dalam klasifikasi gambar buatan AI dan manusia.

1.2 Rumusan Masalah

Berdasarkan latar belakang, rumusan masalah dari penelitian ini adalah:

1. Bagaimana kinerja model *EfficientNetV2* dalam melakukan klasifikasi gambar anime?

2. Bagaimana mengimplementasikan modul CBAM terhadap model *EfficientNetV2* dan apa pengaruhnya terhadap performa model dalam mengklasifikasi gambar anime?

1.3 Batasan Permasalahan

Dalam penelitian, batasan permasalahan yang ada sebagai berikut:

1. Dataset yang digunakan merupakan gambar tipe anime yang telah melalui proses praproses tertentu, dengan sumber utama dari Danbooru, dan model AI pembuat gambar yang beragam.
2. Model klasifikasi yang digunakan dalam penelitian ini adalah *EfficientNetV2* sebagai model dasar.
3. Penelitian ini hanya berfokus pada tugas klasifikasi gambar anime, tanpa membahas tugas lain seperti deteksi objek atau segmentasi, atau bentuk media lain seperti video.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Menganalisis kinerja model *EfficientNetV2* dalam melakukan klasifikasi gambar anime.
2. Menambahkan dan Menganalisis pengaruh penambahan modul CBAM terhadap kinerja model *EfficientNetV2*.

1.5 Manfaat Penelitian

Adapun manfaat yang diharapkan dari penelitian ini adalah:

1. Menjadi referensi bagi peneliti lain dalam pengembangan model klasifikasi gambar berbasis *deep learning*, terutama yang menggunakan *EfficientNetV2*.
2. Memberikan alternatif pendekatan dalam meningkatkan performa model klasifikasi gambar menggunakan modul perhatian.

1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN
Berisi latar belakang permasalahan klasifikasi gambar *anime* buatan AI dan manusia, rumusan masalah, batasan, tujuan, dan manfaat penelitian yang berfokus pada *EfficientNetV2* dengan modul CBAM.
- Bab 2 LANDASAN TEORI
Berisi teori-teori yang mendukung penelitian, seperti *deep learning*, *EfficientNetV2*, CBAM, metode evaluasi, serta penelitian terdahulu yang relevan.
- Bab 3 METODOLOGI PENELITIAN
Berisi tahapan penelitian, dataset yang digunakan, proses praproses data, perancangan model yang ditambahkan modul CBAM, serta proses evaluasi kinerja model.
- Bab 4 HASIL DAN DISKUSI
Berisi hasil implementasi, evaluasi model, serta analisis terhadap performa model yang dihasilkan.
- Bab 5 KESIMPULAN DAN SARAN
Berisi kesimpulan mengenai performa *EfficientNetV2* dan kontribusi modul CBAM dalam klasifikasi gambar *anime* buatan AI dan manusia, beserta saran pengembangan penelitian selanjutnya.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A