

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan kecerdasan buatan dalam beberapa tahun terakhir mendorong munculnya AI-Generated Content (AIGC), yaitu konten digital yang dibuat secara otomatis oleh model generatif. Pada ranah pengolahan citra, kemajuan ini terutama didorong oleh *Denoising Diffusion Probabilistic Models* (DDPM) seperti Stable Diffusion, Midjourney, dan DALL-E yang mampu menghasilkan citra dengan kualitas visual tinggi dan detail yang kompleks [1, 2, 3]. Model-model tersebut tersedia dalam berbagai bentuk, mulai dari layanan komersial hingga model *open source* yang dapat dijalankan di perangkat lokal, sehingga proses pembuatan ilustrasi menjadi jauh lebih cepat dan mudah diakses oleh banyak orang.

Salah satu bentuk konten yang paling terdampak oleh kemunculan model *diffusion* adalah gambar anime dan ilustrasi karakter digital bergaya anime. Jenis citra ini banyak digunakan dalam industri gim, animasi, komik, dan media sosial, serta menjadi sumber penghasilan bagi ilustrator melalui sistem komisi atau penjualan karya. Dengan adanya model generatif, pengguna dapat menghasilkan citra anime hanya dengan memberikan deskripsi teks (*prompt*), tanpa memerlukan kemampuan menggambar secara manual. Di satu sisi, hal ini membuka peluang kreativitas baru, tetapi di sisi lain menimbulkan kekhawatiran terkait keaslian karya, plagiarisme, dan persaingan yang kurang adil antara karya buatan manusia dan hasil generasi AI.

Seiring meningkatnya kualitas citra yang dihasilkan, kemampuan manusia untuk membedakan citra buatan AI dari karya manusia semakin menurun. Corvi dkk. menunjukkan bahwa citra yang dihasilkan model *diffusion* modern sering kali sangat meyakinkan sehingga sulit dibedakan secara visual oleh pengamat manusia [1]. Studi lain mengenai deteksi citra AI dalam skenario daring juga menekankan bahwa mengandalkan penilaian manual saja tidak lagi cukup, terutama ketika jumlah citra yang beredar sangat besar dan generator baru terus bermunculan. Kondisi ini mendorong perlunya sistem deteksi otomatis yang dapat membantu mengidentifikasi apakah sebuah citra anime dibuat oleh manusia atau dihasilkan oleh model generatif.

Berbagai pendekatan telah dikembangkan untuk mendeteksi citra buatan AI. Salah satu pendekatan yang umum digunakan adalah klasifikasi berbasis *deep learning*, khususnya *Convolutional Neural Networks* (CNN). CNN telah menjadi standar untuk tugas klasifikasi citra karena kemampuannya mengekstraksi pola visual lokal dan global secara otomatis dari data citra [4]. Banyak penelitian sebelumnya menggunakan CNN untuk mendeteksi citra *deepfake*, manipulasi wajah, atau citra sintesis lain, dan menunjukkan hasil yang menjanjikan. Namun, mayoritas studi tersebut berfokus pada foto wajah manusia atau citra natural, bukan pada citra anime [5, 2, 3]. Hal ini menimbulkan pertanyaan apakah model yang dilatih pada jenis konten tersebut akan tetap efektif ketika diterapkan pada domain citra anime yang memiliki karakteristik visual berbeda, seperti garis tepi (*lineart*) yang tegas dan pewarnaan yang tidak selalu mengikuti pola dunia nyata.

Selain analisis di domain piksel (spasial), beberapa penelitian menunjukkan bahwa artefak khas dari model generatif justru lebih jelas terlihat di domain frekuensi. Discrete Cosine Transform (DCT) adalah salah satu teknik transformasi yang banyak digunakan dalam pemrosesan citra untuk merepresentasikan citra dalam bentuk koefisien frekuensi [6]. DCT memisahkan komponen frekuensi rendah yang merepresentasikan struktur global dan komponen frekuensi tinggi yang menyimpan detail halus. Frank dkk. menunjukkan bahwa operasi upsampling pada arsitektur generatif dapat meninggalkan pola artefak grid yang konsisten di frekuensi tinggi, dan pola ini dapat dimanfaatkan untuk mengenali citra hasil generator AI [7]. Penelitian lain juga mengeksplorasi pembelajaran di domain frekuensi untuk meningkatkan generalisasi model dalam tugas pengenalan citra dan deteksi manipulasi [8, 9].

Pendekatan yang menggabungkan informasi spasial dan frekuensi semakin banyak dilirik. Ju dkk. mengusulkan metode *Global and Local Feature Fusion* (GLFF) yang menggabungkan fitur global dan lokal untuk mendeteksi citra yang disintesis AI dan menunjukkan bahwa kombinasi beberapa jenis fitur dapat meningkatkan kemampuan deteksi terhadap berbagai generator [5]. Sementara itu, *benchmark* berskala besar seperti GenImage dan GenDet digunakan untuk menguji kemampuan model deteksi citra AI dalam skenario lintas generator dan distribusi data yang beragam [2]. Studi-studi tersebut menyoroti pentingnya generalisasi dan ketangguhan (*robustness*) model deteksi ketika dihadapkan pada berbagai jenis model generatif dan variasi pasca-pemrosesan.

Di sisi lain, masih sedikit penelitian yang secara khusus mengkaji deteksi citra buatan AI pada domain citra anime. Padahal, citra anime memiliki gaya artistik

yang sangat bervariasi serta proses pembuatan yang berbeda dengan foto kamera. Perbedaan ini berpotensi memengaruhi pola artefak di domain frekuensi maupun pola yang ditangkap CNN di domain spasial. Oleh karena itu, dibutuhkan penelitian yang secara khusus berfokus pada klasifikasi citra anime hasil generator AI dan karya manusia, dengan mempertimbangkan kombinasi fitur spasial dan frekuensi.

Salah satu *dataset* yang secara khusus dikembangkan untuk domain tersebut adalah AnimeDL-2M. *Dataset* ini berisi citra anime nyata, citra yang dimanipulasi sebagian, dan citra yang sepenuhnya dihasilkan AI, sehingga relevan untuk penelitian deteksi citra AI pada domain anime [10]. Berdasarkan latar belakang tersebut, tugas akhir ini berfokus pada analisis representasi frekuensi berbasis DCT dan FFT pada CNN untuk klasifikasi gambar anime buatan AI dan karya manusia. CNN digunakan untuk mengekstraksi fitur spasial dari citra, sementara DCT dan FFT digunakan untuk memperoleh representasi frekuensi yang dapat mengungkap artefak halus yang ditinggalkan model generatif. Eksperimen awal membandingkan model ResNet18 RGB, ResNet18+DCT *Grayscale*, dan ResNet18+FFT *Grayscale*. Setelah hasil awal dianalisis, penelitian ini juga menambahkan eksperimen lanjutan berupa ResNet18+DCT RGB dan ResNet18+FFT RGB untuk melihat apakah representasi frekuensi yang dipertahankan pada masing-masing kanal warna dapat memengaruhi performa klasifikasi. Kinerja model dievaluasi menggunakan metrik klasifikasi seperti akurasi, *precision*, *recall*, dan *F1-score* agar penilaian performa lebih komprehensif dan tidak hanya bergantung pada satu metrik saja [11]. Diharapkan, penelitian ini dapat memberikan kontribusi dalam pengembangan metode deteksi citra AI pada domain ilustrasi digital, sekaligus menjadi dasar bagi penelitian lanjutan yang lebih luas.

1.2 Rumusan Masalah

1. Bagaimana efektivitas model Convolutional Neural Networks (CNN) dalam mengklasifikasikan citra anime hasil generator AI dan karya manusia?
2. Bagaimana pengaruh integrasi fitur frekuensi berbasis Discrete Cosine Transform (DCT) dan Fast Fourier Transform (FFT) terhadap performa model CNN?
3. Bagaimana perbandingan kinerja model ResNet18 RGB, ResNet18+DCT *Grayscale*, ResNet18+FFT *Grayscale*, ResNet18+DCT RGB, dan ResNet18+FFT RGB dalam mengklasifikasikan citra anime buatan AI

dan karya manusia berdasarkan metrik akurasi, *precision*, *recall*, *F1-score*, *confusion matrix*, waktu pelatihan, dan waktu inferensi?

1.3 Batasan Permasalahan

1. Jenis gambar yang digunakan dibatasi pada gambar anime dan ilustrasi karakter digital, bukan foto dunia nyata atau gambar realistik sepenuhnya.
2. *Dataset* yang digunakan merupakan subset dari AnimeDL-2M yang telah dikurasi ke dalam dua kelas, yaitu *ai* dan *human*. Penelitian ini tidak membahas perbandingan performa berdasarkan jenis generator AI, jenis manipulasi, atau identitas pembuat gambar.
3. Model yang digunakan dibatasi pada arsitektur CNN berbasis ResNet18 sebagai model klasifikasi RGB, ResNet18+DCT *Grayscale*, ResNet18+FFT *Grayscale*, ResNet18+DCT RGB, dan ResNet18+FFT RGB.
4. Evaluasi performa model mencakup metrik akurasi, *precision*, *recall*, *F1-score*, *confusion matrix*, waktu pelatihan, dan waktu inferensi.
5. Penelitian tidak membahas aspek etika, legalitas, atau dampak sosial penggunaan ilustrasi AI, dan hanya berfokus pada proses klasifikasinya.

1.4 Tujuan Penelitian

1. Mengembangkan model klasifikasi yang mampu membedakan gambar anime buatan AI dan gambar anime karya manusia.
2. Menganalisis pengaruh penggunaan fitur frekuensi berbasis Discrete Cosine Transform (DCT) dan Fast Fourier Transform (FFT), baik dalam bentuk *grayscale* maupun berbasis kanal RGB, terhadap performa Convolutional Neural Networks (CNN).
3. Mengevaluasi performa model klasifikasi berdasarkan akurasi, *precision*, *recall*, *F1-score*, *confusion matrix*, waktu pelatihan, dan waktu inferensi untuk mengetahui efektivitas metode yang digunakan.

1.5 Manfaat Penelitian

1. Bagi industri kreatif, penelitian ini dapat membantu mengidentifikasi karya yang dihasilkan AI, terutama untuk verifikasi keaslian karya digital.
2. Bagi komunitas akademik, penelitian ini dapat menjadi referensi pengembangan metode klasifikasi gambar yang menggabungkan CNN dengan fitur frekuensi.
3. Memberikan kontribusi metodologis terkait analisis pemanfaatan fitur frekuensi berbasis DCT dan FFT sebagai informasi tambahan pada model *deep learning* untuk tugas klasifikasi gambar.

1.6 Sistematika Penulisan

Sistematika penulisan laporan ini disusun sebagai berikut.

1. Bab 1 Pendahuluan berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.
2. Bab 2 Landasan Teori berisi konsep dasar yang mendukung penelitian, penelitian terdahulu, serta *research gap* dan posisi penelitian.
3. Bab 3 Metodologi Penelitian menjelaskan alur penelitian, *dataset*, pra-pemrosesan data, ekstraksi fitur DCT dan FFT, perancangan model, skenario pelatihan, dan skenario evaluasi.
4. Bab 4 Hasil dan Diskusi menyajikan hasil pelatihan, hasil pengujian, perbandingan performa model, analisis *confusion matrix*, pembahasan hasil, serta keterbatasan penelitian.
5. Bab 5 Simpulan dan Saran berisi simpulan yang disesuaikan dengan tujuan penelitian serta saran untuk pengembangan penelitian selanjutnya.