

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Deepfake merupakan bagian dari *Deep Learning* dengan pemanfaatan teknologi *Computer Vision* dan *Artificial Intelligence* (AI). *Deepfake* berpotensi melemahkan kepercayaan publik akan informasi yang beredar dalam lingkup sektor pemerintahan maupun masyarakat[1]. Berdasarkan hasil survey *VIDA Where's The Fraud – Protecting Indonesia Business from AI Generated Fraud*, Indonesia mengalami lonjakan sebesar 1550% pada kasus penipuan *deepfake* diantara tahun 2022 hingga 2023[2]. Oleh karna itu, pengembangan metode dalam deteksi *deepfake* menjadi kebutuhan mendesak dikarenakan memberikan dampak signifikan dalam penyebaran misinformasi dan disinformasi terutama di Indonesia sendiri dimana tingginya penggunaan sosial media berdampak rentannya terhadap hoaks dan misinformasi, serta dapat berdampak kepada isu sosial dan budaya.

Penelitian ini mengacu pada penelitian terdahulu yang menggabungkan ResNet50 dengan model *attention* untuk meningkatkan akurasi deteksi. Dalam penelitian tersebut, integrasi ResNet50 dengan *Long-Distance Attention Module* mencapai akurasi hingga 94.30% dan AUC 98.70%, menunjukkan bahwa integrasi mekanisme *attention* dapat meningkatkan performa klasifikasi secara signifikan[3]. Adapun juga penelitian sebelumnya yang menggabungkan model *Xception* dengan mekanisme *long-distance attention* untuk meningkatkan akurasi deteksi video *deepfake*. Dalam penelitian tersebut, model spasial-temporal yang diusulkan berhasil mencapai akurasi hingga 96.2% dan AUC sebesar 99.4% pada dataset FaceForensics++ HQ, serta menunjukkan performa unggul dalam mendeteksi berbagai jenis manipulasi wajah. Hasil ini menunjukkan bahwa integrasi mekanisme *attention* dalam domain spasial dan temporal dapat meningkatkan performa klasifikasi secara signifikan[4]. Dan juga penelitian ini juga mengacu pada studi sebelumnya yang menyoroti pentingnya *model attribution* dalam konteks video *deepfake*. Dalam penelitian tersebut, metode DMA-STA yang mengintegrasikan mekanisme *spatial* dan *temporal attention* dikembangkan untuk mengidentifikasi model pembuat *deepfake* secara akurat. Hasil eksperimen pada dataset DFDM menunjukkan bahwa metode ini mampu mencapai akurasi hingga

71.94% pada video berkualitas tinggi, menunjukkan efektivitas pendekatan berbasis *attention* dalam membedakan model generatif secara halus[5].

Penelitian ini bertujuan untuk meningkatkan matrix akurasi dalam mendeteksi gambar *deepfake* melalui pengembangan model berbasis *deep learning* menggunakan arsitektur ResNet50 yang dikombinasikan dengan CBAM. ResNet50 dipilih sebagai *backbone* model karena kemampuannya dalam mengekstraksi fitur mendalam dari gambar secara efisien melalui mekanisme *residual learning*. Namun, untuk lebih menyoroti informasi penting dalam citra, CBAM ditambahkan sebagai modul perhatian yang bekerja secara adaptif dalam dua tahap: *channel attention* dan *spatial attention*. Penambahan CBAM memungkinkan model untuk memfokuskan perhatian pada area atau fitur yang lebih relevan, sehingga meningkatkan performa klasifikasi[6]. Dataset yang digunakan terdiri dari dua kelas utama, yaitu gambar asli (*Real*) dan gambar palsu (*Deepfake*), yang kemudian dibagi ke dalam subset pelatihan, validasi, dan pengujian. Untuk mendukung stabilitas dan generalisasi model, digunakan teknik augmentasi citra serta algoritma AdamW sebagai *optimizer*. Setelah proses pelatihan selesai, evaluasi dilakukan menggunakan metrik seperti akurasi, *presisi*, *recall*, *F1-score*, dan AUC-ROC, disertai dengan visualisasi Grad-CAM. Visualisasi Grad-CAM berfungsi untuk menampilkan area dalam gambar yang paling memengaruhi keputusan model, sehingga memberikan transparansi dan interpretasi terhadap proses prediksi. Hasil yang diperoleh, diharapkan penambahan CBAM pada ResNet50 dapat meningkatkan fokus fitur terhadap wilayah-wilayah penting pada wajah yang mengindikasikan manipulasi, serta memberikan hasil evaluasi yang kompetitif dibandingkan model baseline tanpa CBAM.

Penelitian ini menawarkan pendekatan baru dengan mengombinasikan arsitektur ResNet50 sebagai model klasifikasi dan CBAM untuk pemfokusan pada objek detail seperti mata dan hidung dengan metode dengan dukungan visualisasi Grad-CAM untuk meningkatkan interpretabilitas prediksi dalam deteksi citra *deepfake*. Hasil dari pelatihan menunjukkan Hasil eksperimen menggunakan model **CBAM-ResNet50** menunjukkan bahwa model mampu mencapai akurasi pengujian sebesar 75.27%, dengan nilai precision 67.91%, recall 95.80%, dan F1-score 79.48% pada dataset yang digunakan. Nilai ini menggambarkan bahwa meskipun model cukup

baik dalam mendeteksi kelas positif (recall tinggi), namun tingkat presisi yang relatif rendah menyebabkan adanya prediksi positif yang keliru dalam jumlah cukup besar. Sementara itu, hasil pelatihan pada ResNet50 baseline menghasilkan akurasi pengujian sebesar 86.40%, dengan nilai precision 80.69%, recall 95.70%, F1-score 87.56%, dan AUC sebesar 94.62%.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah pada penelitian ini:

1. Bagaimana membangun model deteksi deepfake berbasis ResNet50 dan ResNet50 yang dimodifikasi dengan CBAM agar mampu memfokuskan area penting yang relevan pada citra?
2. Bagaimana performa kedua model tersebut setelah dievaluasi menggunakan metrik akurasi, precision, recall, F1-score, dan AUC?

1.3. Tujuan dan Manfaat

1. Penelitian ini dibatasi hanya pada penggunaan arsitektur ResNet50 dan modifikasinya dengan CBAM, tanpa membandingkan dengan modul attention lain.
2. Evaluasi model hanya dilakukan pada dataset yang digunakan dalam penelitian ini dengan metrik akurasi, precision, recall, F1-score, dan AUC.

Tabel 1. 1 Keterkaitan tujuan, pengujian & kesimpulan

No.	Tujuan	Pengujian
1	Meningkatkan kinerja dan pemfokusan fitur deteksi <i>deepfake</i> .	Pengujian akurasi pada dataset DFDC.
2	Penambahan CBAM untuk meningkatkan akurasi.	Dengan pelatihan menggunakan 8000 dataset dengan train 70%, validasi 10% dan test.
3	Melihat apakah kinerja CBAM dapat meningkatkan akurasi.	CBAM di sisipkan di tengah layer ResNet50 dengan penambahan AdamW
4	Membandingkan Kinerja CBAM pada 6000 Dataset	Dalam hasil pengujian ditemukan hasil dari <i>Pecision: 67.91%, Accuracy: 75.27% Recall: 95.80 dan F1-Score: 79.48</i>

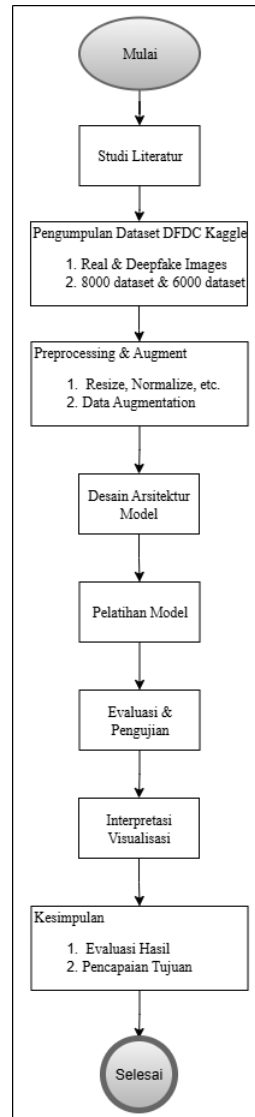
2.1. Batasan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah pada penelitian ini:

1. Penelitian ini dibatasi hanya untuk penggunaan arsitektur ResNet50 yang dioptimalkan dengan CBAM untuk meningkatkan akurasi deteksi *deepfake* dan pemfokusan pada fitur yang relevan pada citra.
2. Penelitian ini dibatasi hanya pada penggunaan CBAM untuk peningkatan akurasi.

2.2. Metode Penelitian

Metode penelitian yang digunakan dalam tugas akhir ini merupakan *mixed method* dari metode kualitatif dimana pengumpulan informasi dan penelitian terdahulu untuk melihat gambaran dan gap serta kuantitatif.



Gambar 1. 1 Diagram Flowchart Penelitian

Tahapan penelitian diawali dengan pengumpulan teori dan studi Pustaka terkait arsitektur CNN, khususnya model ResNet50 dan metode *attention* CBAM, serta teknik interpretabilitas Grad-CAM. Setelah pemahaman konseptual diperoleh, dilakukan proses perancangan dan implementasi model menggunakan framework TensorFlow/Keras. Dataset *deepfake* diklasifikasikan menjadi dua kelas *Real* (wajah asli) dan *Deepfake* (wajah yang dimanipulasi), kemudian dilakukan *preprocessing* serta augmentasi untuk meningkatkan keragaman data. Model dibangun dengan *backbone* ResNet50, dilengkapi modul CBAM untuk memperkuat fitur penting dalam proses ekstraksi. Selanjutnya, dilakukan pelatihan dan pengujian model menggunakan data yang telah disiapkan. Evaluasi performa yang diperoleh melalui pengamatan, percobaan, atau penemuan dan kuantitatif

menggunakan metrik akurasi, precision, recall, F1-score, ROC AUC, dan *confusion matrix*. Visualisasi hasil melalui Grad-CAM bertujuan memberikan interpretasi visual terhadap fokus area prediksi model. Penelitian ini dilakukan secara berulang menggunakan pendekatan eksperimen dan simulasi, hingga diperoleh hasil terbaik yang mendukung pencapaian tujuan deteksi *deepfake* secara optimal.

2.3. Jadwal Pelaksanaan

Tabel 1. 2 Jadwal Pelaksanaan Tugas Akhir

No.	Deskripsi Tahapan	Bulan 1	Bulan 2	Bulan 3	Bulan 4	Bulan 5	Bulan 6
1	Studi Literatur						
2	Pengumpulan Data						
3	Pembentukan Struktur Model						
4	Pengujian						
5	Pengkajian Hasil						
6	Penyusunan Laporan/Buku TA						