

BAB I

PENDAHULUAN

Bab pendahuluan ini membahas lebih lanjut mengenai latar belakang dilakukannya penelitian ini, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan yang akan digunakan dalam penyusunan dokumen tugas akhir Penerapan *Transfer Learning* dengan Arsitektur Xception untuk Klasifikasi Citra Wajah *Deepfake*.

1.1 Latar Belakang

Perkembangan teknologi kecerdasan buatan yang pesat dan masif telah membawa perubahan besar dalam hal pengolahan citra digital (Tang dkk., 2020). Hal ini terlihat pada penerapannya dalam berbagai bidang seperti pengenalan wajah, klasifikasi dan deteksi objek, analisis citra medis dalam dunia kedokteran, hingga dalam bidang keamanan digital perbankan. Namun, seiring dengan pemanfaatan teknologi kecerdasan buatan, khususnya dalam hal pengolahan citra, terdapat juga beberapa masalah yang dapat menimbulkan pro dan kontra di masyarakat. Salah satu hal yang cukup menjadi perhatian luas saat ini adalah perkembangan *deepfake* yang merupakan hasil dari manipulasi citra atau video berbasis kecerdasan buatan yang memungkinkan pembuatan konten palsu menjadi tampak realistis dengan menggunakan bantuan model generatif seperti *Variational Autoencoders* atau *Generative Adversarial Networks* (Coccomini dkk., 2021).

Perkembangan *deepfake* tidak lepas dari kecanggihan beberapa teknologi yang ada dibelakangnya. Teknologi yang umum dikenal memiliki andil besar dalam pembuatan citra *deepfake* adalah GANs dan VAEs. *Generative Adversarial Networks* (GANs) merupakan model generatif yang terdiri dari *generator* dan *discriminator*, yang saling bersaing untuk menghasilkan data baru yang menyerupai data asli. *Generator* berperan mempelajari distribusi data, sedangkan *discriminator* bertugas membedakan antara data asli dan data hasil manipulasi (Wodajo & Atnafu, 2021). Sementara itu, teknologi lain seperti *Variational Autoencoders* (VAEs) digunakan untuk menangkap informasi penting dari data dalam bentuk representasi berdimensi rendah, yang bisa dimanipulasi untuk menghasilkan data baru. VAEs terdiri dari *encoder* yang menghasilkan *mean* dan *varians*, serta *decoder* yang membangkitkan data baru dengan menambahkan *noise Gaussian* (Liu dkk., 2023).

Seiring dengan perkembangannya, *deepfake* memberikan kontribusi positif di berbagai industri yang meliputi dunia hiburan, pendidikan, media sosial, hingga *e-commerce*. Industri hiburan menggunakan *deepfake* untuk menghidupkan kembali adegan klasik dari aktor yang sudah lama tiada, sehingga tampak nyata dan menghadirkan suasana masa lalu secara realistis melalui manipulasi citra, video, maupun suara artifisial (Westerlund, 2019). Selain itu, *deepfake* juga bermanfaat untuk menyampaikan pesan *awareness campaign* seperti yang dilakukan pada kampanye kesadaran malaria global pada tahun 2019, yang menampilkan David Beckham dapat berbicara dalam berbagai bahasa (Spitz, 2024). Dalam konteks komunikasi daring, teknologi *deepfake* memungkinkan terciptanya *telepresence* yang lebih realistis, mendorong peningkatan kualitas komunikasi dan hubungan interpersonal secara daring. Dalam industri *e-commerce*, *deepfake* digunakan untuk mendorong potensi bisnis melalui dunia *marketing*, di mana para pemilik *brand* mampu memvisualisasikan produk pakaian yang dijual melalui model virtual (Mukta dkk., 2023).

Namun, di balik potensinya yang cukup besar, kemunculan *deepfake* juga dapat menimbulkan tantangan yang tidak dapat diabaikan. Salah satu kekhawatiran utama adalah penyalahgunaan teknologi ini untuk menciptakan konten palsu yang dapat merusak reputasi individu atau perusahaan, seperti citra atau video *deepfake* untuk penipuan atau disinformasi. Menurut Kaur dkk. (2024), *deepfake* merupakan salah satu dari lima jenis penipuan identitas tertinggi pada tahun 2023. *Deepfake* juga memiliki risiko cukup besar dalam hal pemerasan, intimidasi, dan disrupsi politik yang dapat membahayakan demokrasi dan keamanan nasional (Martínez dkk., 2024). Hal ini terlihat dalam beberapa kasus perpolitikan di Amerika Serikat yang mana *deepfake* dimanfaatkan untuk memanipulasi opini publik terhadap tokoh-tokoh penting dalam dunia politik (Mukta dkk., 2023). Oleh karena itu, klasifikasi *deepfake* menjadi hal yang mendesak mengingat potensi penyalahgunaannya yang dapat mengancam masyarakat dan membahayakan keamanan informasi digital (Elpeltagy dkk., 2023). Hal itulah yang menjadi dasar bagi beberapa ilmuwan dan akademisi untuk melakukan penelitian terhadap *deepfake* guna menghasilkan model yang *robust* dan memiliki akurasi tinggi serta efektif dalam pengidentifikasian *deepfake*.

Penelitian mengenai klasifikasi *deepfake* telah banyak dilakukan beberapa tahun belakangan. Alkurdi dkk. (2024) menggunakan arsitektur Xception pada dataset *FaceForensics* dan mencapai akurasi hingga 99,69%. Penelitian yang dilakukan oleh Joshi & Nivethitha (2024) memperoleh akurasi 93,01% dengan *transfer learning* menggunakan

arsitektur Xception pada 90.000 citra. Sementara itu, Kim dkk. (2019) menunjukkan bahwa Xception lebih unggul dibandingkan arsitektur lain seperti ShallowNet dan VGG16 dalam klasifikasi wajah *deepfake* dengan akurasi 62%. Beberapa penelitian tersebut menunjukkan bahwa arsitektur Xception memiliki performa yang kuat dalam klasifikasi konten *deepfake*. Meskipun arsitektur Xception telah banyak digunakan dalam klasifikasi citra *deepfake*, sebagian besar penelitian sebelumnya belum mengevaluasi secara sistematis pengaruh strategi pelatihan seperti *hyperparameter tuning*, *transfer learning*, dan *fine-tuning* terhadap performa model. Oleh karena itu, penelitian ini bertujuan mengisi celah tersebut dengan mengevaluasi pengaruh kombinasi strategi pelatihan dan regularisasi L2 dan *dropout* menggunakan *transfer learning* dengan arsitektur Xception. Penelitian ini diharapkan dapat memberikan kontribusi melalui model klasifikasi *deepfake* yang akurat dan *robust* di tengah pesatnya perkembangan teknologi GANs.

1.2 Rumusan Masalah

Dari berbagai permasalahan yang telah dijelaskan dalam latar belakang tersebut, dapat dijabarkan rumusan masalah, yaitu bagaimana tingkat performa akurasi optimal yang diperoleh dari penerapan *transfer learning* dengan arsitektur Xception dalam melakukan klasifikasi citra wajah *deepfake* melalui berbagai skenario eksperimen yang dilakukan.

1.3 Tujuan dan Manfaat

Tujuan yang ingin diperoleh dari penelitian ini adalah untuk mencari performa dan efektivitas penerapan *transfer learning* dengan arsitektur Xception dalam melakukan klasifikasi citra wajah asli dan *deepfake* berdasarkan parameter optimal yang diperoleh melalui strategi *hyperparameter tuning*, *transfer learning*, *fine-tuning*, dan regularisasi pada dataset dengan variasi data yang kompleks.

Penelitian ini diharapkan dapat memberikan manfaat dalam perkembangan ilmu pengetahuan dan perlindungan bagi masyarakat, utamanya dalam upaya memerangi misinformasi atau konten *hoax* yang disebabkan oleh perkembangan *deepfake*, sehingga dapat meminimalisir terjadinya perpecahan di masyarakat.

1.4 Ruang Lingkup

Adapun ruang lingkup yang terdapat dalam penelitian klasifikasi citra wajah *deepfake* menggunakan *transfer learning* dengan arsitektur Xception adalah sebagai berikut:

1. Penelitian ini hanya difokuskan untuk klasifikasi *deepfake* pada dataset citra wajah, dan bukan pada dataset berbasis video ataupun audio.
2. Dataset yang digunakan dalam penelitian ini adalah dataset citra wajah yang berasal dari Kaggle (Link: <https://www.kaggle.com/dagnelies/deepfake-faces>) dengan judul “Deepfake Faces”, yang berisi 79.341 citra *fake* dan 16.293 citra *real* dengan dimensi ukuran citra 224x224.
3. *Convolutional Neural Network* yang digunakan dalam penelitian ini adalah arsitektur Xception yang sebelumnya sudah dilatih atau *pre-trained* pada dataset besar *ImageNet* sebagai dasar untuk melakukan *transfer learning* dan kemudian dilanjutkan dengan proses *fine-tuning* Xception.
4. *Batch size* yang digunakan dalam penelitian ini adalah *batch size* 32.

1.5 Sistematika Penulisan

Berikut adalah penjelasan singkat mengenai sistematika penulisan dalam penyusunan tugas akhir Penerapan *Transfer Learning* dengan Arsitektur Xception untuk Klasifikasi Citra Wajah *Deepfake*:

BAB I PENDAHULUAN

Bab pendahuluan ini membahas secara lebih detail mengenai latar belakang dilakukannya penelitian, rumusan masalah, tujuan dan manfaat, batasan ruang lingkup penelitian, serta sistematika penulisan yang akan digunakan dalam penyusunan dokumen tugas akhir Penerapan *Transfer Learning* dengan Arsitektur Xception untuk Klasifikasi Citra Wajah *Deepfake*.

BAB II LANDASAN TEORI

Bab ini mengulas berbagai penelitian terdahulu serta landasan teoritis yang menjadi dasar penelitian, meliputi pembahasan mengenai konsep *deepfake*, teknologi *Generative Adversarial Networks* (GANs), *undersampling*, teknik *preprocessing* dan augmentasi data, serta klasifikasi citra. Selain itu, bab ini juga menjelaskan prinsip dasar *transfer learning*,

arsitektur Xception, serta struktur *Convolutional Neural Network* (CNN) beserta komponen-komponen penting di dalamnya. Topik lain yang turut diuraikan mencakup algoritma optimisasi, fungsi *loss*, strategi pelatihan lanjutan seperti *hyperparameter tuning*, *feature extraction*, dan *fine-tuning*, metrik evaluasi, serta *tools* dan *library* yang digunakan dalam pelaksanaan penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini menyajikan penjelasan mengenai rancangan alur penelitian yang dilakukan, mulai dari proses pengumpulan dan pembagian dataset, tahap *preprocessing* dan augmentasi data, hingga pembangunan arsitektur Xception. Selanjutnya, dijabarkan pula tahapan pelatihan model yang mencakup proses pencarian *hyperparameter* optimal, *transfer learning*, penerapan *fine-tuning* dengan regularisasi, serta evaluasi performa model menggunakan berbagai metrik seperti akurasi, presisi, *recall*, dan *F1-score*.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menguraikan secara mendalam spesifikasi lingkungan pengembangan dan perangkat yang digunakan dalam proses pelatihan dan pengujian model, serta menjelaskan detail dataset yang digunakan dalam penelitian, termasuk proses pembagian data, *preprocessing*, dan augmentasi citra. Selain itu, dibahas pula implementasi pembangunan arsitektur Xception, serta analisis terhadap hasil pelatihan dan evaluasi performa model berdasarkan skenario pengujian yang telah dilakukan.

BAB V PENUTUP

Bab ini menyajikan kesimpulan dari keseluruhan pembahasan pada bab-bab sebelumnya, serta saran untuk pengembangan lebih lanjut terkait penelitian klasifikasi citra wajah *deepfake* menggunakan *transfer learning* dengan arsitektur Xception.