BAB I
PENDAHULUAN

1.1 Latar Belakang

Deteksi Video *deepfake* adalah cara untuk mengekstrak dan mengklasifikasikan antara video asli atau palsu yang dibuat menggunakan *Artificial intelligence* (AI) (Nguyen *et al.*, 2019; Cao and Gong, 2021). Video deepfake menimbulkan distorsi dan kekhawatiran karena dapat digunakan dalam berbagai tujuan tidak baik seperti pencemaran nama baik, menciptakan bias politik, sabotase, intimidasi, eksploitasi, propaganda palsu, pembajakan dan kegiatan kejahatan lainnya (Afchar *et al.*, 2019; Karandikar *et al.*, 2020). Dengan adanya metode deteksi deepfake dapat membantu konsumen konten digital untuk mengatasi disinformasi dan membedakan antara video asli dan palsu (Lyu, 2020; Rahul *et al.*, 2020). Namun, perkembangan teknologi khususnya pada perangkat lunak berbasis *machine learning* telah mempercepat dan memudahkan proses membuat video *deepfake* terlihat asli namun meninggalkan beberapa jejak manipulasi (Lyu, 2020; Thaw *et al.*, 2020). Di sisi lain, pengembangan metode deteksi deepfake dibatasi oleh jumlah dan kualitas dataset, proses deteksi, dan inkonsisten *Performance Evaluation yaitu* model tidak dapat mendeteksi video palsu yang dimanipulasi menggunakan *tools video editing* (Lyu, 2020). Oleh karena itu diperlukan teknik untuk mendeteksi video *deepfake*.

Beberapa peneliti menggunakan *Artificial intelligence* (AI) untuk mendeteksi video deepfake. Li et al (2018) mengusulkan metode deteksi video *deepfake* berdasarkan inkonsitensi yang dilihat dalam aspek fisik atau fisiologis yaitu tidak memiliki kedipan mata yang wajar, menggunakan *Convolutional Neural Networks* (CNN) dengan keberhasilan yang ditunjukan oleh nilai *Receiver Operating Characteristic* (ROC) yaitu 0.98. Selanjutnya, Yang *et al* (2019) berkontribusi untuk mendeteksi video *deepfake* berdasarkan pose kepala yang tidak koheren menggunakan *Support Vector Machine* (SVM) yang

menghasilkan skor Area Under ROC (AUROC) yaitu 0.843. Namun kekurangan dari teknik ini yaitu ketergantungan pada jumlah dataset yang mempengaruhi hasil deteksi (Li and Lyu, 2018).

Untuk mengatasi masalah tersebut, pada penelitian berikutnya Li *et al* (2020) berkontribusi memberikan dataset video *deepfake* dalam skala besar yaitu Celeb-DF untuk mengurangi kesenjangan dalam kualitas dataset *deepfake*. Dengan menggunakan dataset Celeb-DF, de Lima *et al* (2020) berkontribusi mendeteksi *deepfake* menggunakan *Spatiotemporal Convolutional Network* (SCN) yang menunjukan skor Area Under ROC (AUROC) yaitu metode RCN sebesar 74,87% dengan akurasi 76,25%, R2Plus1D sebesar 99,43 dengan akurasi 98,07 dan R3D sebesar 99,73 dengan akurasi 98,26. Namun, proses analisis dengan mendeteksi artefak statistik dalam video *per-frame* mengakibatkan informasi hasil deteksi tergantung dari kualitas *frame* video (D. Li *et al.*, 2020; de Lima *et al.*, 2020). Marissa Koopman, et al (2018) mengusulkan teknik deteksian deepfake pada frame gambar dengan menggunakan analisis *Photo-Response Non Uniformity* (PRNU) untuk memeriksa pola PRNU antara setiap frame dengan kesuksesan nilai pada cut-of 0,05 menghasilkan 3,8% false positive rate, dan 0% false negative rate, yang berarti analisis ini dapat digunakan untuk mendeteksi deepfake karena analisis pola sensor, khususnya *Photo-Response Non-Uniformity* (PRNU), merupakan salah satu teknik forensik paling *powerful* untuk gambar digital (Fridrich, 2009).

Berdasarkan uraian yang telah dijelaskan di atas, penelitian ini menggunakan *Spatiotemporal Convolutional Network* dan *Photo-Response Non Uniformity* untuk mendeteksi video deepfake dengan judul "**Deteksi Deepfake Menggunakan *Spatiotemporal Convolutional Network* dan *Photo-Response Non-Uniformity***".

1.2 Masalah Penelitian

Berdasarkan latar belakang yang telah dijelaskan di atas, masalah yang ingin dicari solusinya dalam penelitian ini dibagi dalam dua bagian, yaitu identifikasi masalah dan rumusan masalah.

1.2.1 Identifikasi Masalah

Sesuai dengan latar belakang di atas, masalah yang terdapat di dalam penelitian ini yaitu:

1. Video *deepfake* terus berkembang dan semakin terlihat asli menyebabkan disinformasi sehingga perlu model untuk mendeteksi video *deepfake* dari beragam dataset dan teknik terbaru.
2. Adanya ketergantungan pada jumlah dan kualitas dataset, jumlah *frame*, dan jenis video yang dideteksi sehingga menyebabkan rendahnya nilai akurasi.
3. Adanya inkonsisten *performance evaluation* sehingga model tidak dapat mendeteksi video palsu yang dimanipulasi menggunakan *tools video editing*.

1.2.2 Rumusan Masalah

Adapun rumusan masalah dalam penelitian ini adalah menggabungkan model *Spatiotemporal Convolutional Network dan Photo-Response Non-Uniformity* (PRNU) untuk dapat mendeteksi video deepfake dengan akurasi tinggi.

1.3 Tujuan dan Manfaat Penelitian

Tujuan dari penelitian ini adalah menghasilkan model yang digunakan untuk mendeteksi video deepfake dengan akurasi tinggi. Sedangkan, manfaat dari penelitian ini yaitu sebagai berikut:

1. Model ini dapat digunakan dan bisa diterapkan pada perangkat yang ada saat ini untuk mendeteksi video deepfake.

2. Model ini dapat digunakan untuk mendeteksi video palsu selain hasil dari *artificial intelligence* (AI).
3. Hasil penelitian ini dapat dijadikan referensi untuk penelitian lebih lanjut.

1.4 Batasan Masalah

Adapun yang menjadi batasan masalah pada penelitian ini adalah sebagai berikut:

1. Penelitian ini menggunakan *dataset* video Celeb-DF V2 dari Li *et al* (2020). Dalam *dataset* ini terdapat 590 video asli, dan 5639 video palsu. *Dataset* ini tersedia secara umum dan dapat diakses melalui link http://www.cs.albany.edu/~lsw/celeb-deepfakeforensics.html. Kemudian, FaceForensics++ terdiri yang terdiri dari 363 video asli yang berasal dari 28 aktor dalam 16 adegan berbeda dan lebih dari 3000 video yang telah dimanipulasi menggunakan deepfake dan mask biner yang sesuai (https://github.com/ondyari/FaceForensics). Dan, Face-HQ merupakan dataset dengan resolusi tinggi yang memiliki 4000 file gambar (https://github.com/NVlabs/ffhq-dataset).
2. Penelitian ini menggunakan dataset video asli dan palsu dengan batasan manipulasi dengan teknik *re-contextualizing*, *lookalikes*, *face swapping*, *face replacement*, *face reenactment*, dan *face generation*.
3. Penelitian ini menggunakan batasan minimal dimensi file video dataset (144 × 256, 3) pixel RGB.
4. Penelitian ini menyajikan model yang berfokus pada pendeteksian bagian wajah, tidak termasuk audio sehingga video akan dibersihkan terlebih dahulu dengan memangkas bagian wajah di setiap frame.

1.5 Metodologi Penelitian

Metodologi yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Studi Literatur

Pada tahap ini dilakukan pengumpulan bahan referensi yang berhubungan dengan penelitian, seperti: *deepfake*, deteksi video *deepfake*, metode *Spatiotemporal Convolutional Network* (SCN), analisis *Photo-Response Non-Uniformity* (PRNU) dan bahan referensi lain terkait dengan penelitian mengenai deteksi video *deepfake* dari beberapa penelitian sebelumnya.

2. Tahap Analisis

   Pada tahap ini dilakukan proses untuk mengidentifikasi data yang dibutuhkan, masalah dan tantangan yang harus diselesaikan dan menjelaskan solusi yang diusulkan untuk menyelesaikan masalah dan tantangan yang ada. Proses dalam *machine learning* akan digambarkan dalam bentuk flowchart.

3. Perancangan Model

   Perancangan model dimulai dari mengumpulkan dataset video asli/palsu, menerapkan preprocessing data dengan mengekstrak video dalam bentuk frame, kemudian mendeteksi dan memangkas frame yang terdeteksi wajah. *Spatiotemporal Convolutional Network* menggunakan *ResNext CNN* untuk feature extraction dan LSTM untuk klasifikasi kemudian PRNU untuk analisis pola PRNU pada input video yang akan dideteksi.

4. Pengujian

   a. Membandingkan model dari masing-masing dataset untuk melihat nilai akurasi yang dihasilkan.
   b. Membandingkan metode *Spatiotemporal Convolutional Network* dengan PRNU dan tanpa PRNU untuk menguji hasil deteksi video *deepfake*.
   c. Melakukan pengujian terhadap model yang dihasilkan untuk mendeteksi video *deepfake* dengan beragam jenis dan kualitas video.

5. Menarik kesimpulan dari hasil pengujian
6. Menyusun laporan Tesis.

1.6 Sistematika Penulisan

Sistematika penulisan laporan penelitian ini terdiri dari 5 bab, dimana secara garis besar masing-masing bab membahas hal – hal berikut ini. Bab 1 Pendahuluan, berisi penjelasan umum, masalah dan solusi yang akan dilakukan penelitian. Bab 2 berisi studi literatur dan tinjauan singkat terkait masalah dan metode yang berhubungan dengan penelitian yang akan dilakukan. Bab 3 Metodologi Penelitian, berisi identifikasi masalah, langkah-langkah dari metode yang diusulkan, data yang digunakan, alat-alat penelitian dan metode analisis. Bab 4 Hasil dan Pengujian, berisi hasil dari sistem yang dibangun dan analysis berdasarkan pengujian yang dilakukan. Bab 5 Kesimpulan dan Saran, berisi kesimpulan yang diperoleh dari hasil dan pengujian penelitian yang sudah dilakukan dan saran untuk hasil yang lebih baik dalam penelitian yang sejenis.