

BAB I

PENDAHULUAN

1.1 Latar belakang masalah

Setiap manusia pada dasarnya selalu membutuhkan serta bergantung pada informasi. Dalam era yang serba cepat dan praktis ini, teknologi yang menunjang penyebaran informasi terus dikembangkan dan setiap orang memiliki aksesibilitas yang mudah dan bebas terhadap sebagian besar teknologi itu. Contoh yang paling nyata dan umum adalah internet, dikarenakan banyak sekali konten positif serta negatif yang beredar tanpa adanya filter sehingga memungkinkan terjadinya penyalahgunaan terhadap keberadaan konten - konten tersebut. Pornografi adalah contoh dari sekian banyak konten negatif yang tersebar di internet. Definisi dari pornografi adalah materi yang menciptakan pikiran seksual, berisi gambar eksplisit atau deskripsi tindakan seksual yang melibatkan alat kelamin [1]. Keberadaannya dapat menimbulkan berbagai permasalahan sosial terutama kepada para remaja. Terdapat ulasan dari berbagai penelitian [2] yang menunjukkan bahwa pornografi dapat mengganggu kesehatan perkembangan seksual pada anak modern.

Telah banyak penelitian terkait dengan cara untuk menyaring suatu konten pornografi [3], [4] dan [5]. Namun inovasi dan berbagai pengembangan tetap diperlukan agar metode yang telah dirancang mampu menghasilkan keluaran yang lebih optimal. Dalam masa awal penelitian tentang konten pornografi, Basilio *et al* [3] membuat metode pendeteksi citra pornografi dengan memperhatikan banyaknya komposisi kulit dalam suatu citra. Tetapi penentuan konten pornografi yang mengacu pada komposisi warna kulit akan memiliki tingkat akurasi yang rendah jika *background* dari objek dalam citra tersebut mempunyai komposisi warna serupa dengan kulit yang asli. Tidak semua warna kulit juga dapat terdeteksi dikarenakan hanya warna kulit yang berada di antara ketentuan *threshold* saja yang akan terdeteksi. Dalam penelitian lain oleh Nuraisha *et al* [4], dilakukan pendeteksian pornografi dengan menggunakan K-NN sebagai metode klasifikasinya, tetapi penerapan metode ini tidak terlalu cocok untuk menangani data dengan *noise* yang banyak sehingga terbatas dalam segi karakteristik *dataset* yang dapat dijadikan *input training*.

Seiring dengan perkembangan teknologi, *deep learning* mulai digunakan untuk melakukan klasifikasi citra termasuk pornografi ataupun berbagai pengolahan video berbasis *computer vision* [6], [7], [8], dan [9] yang salah satu metodenya merupakan *Convolutional Neural Network* (CNN) [5]. Dalam tugas akhir ini, penulis menggunakan metode CNN untuk mengklasifikasikan citra pornografi guna mengatasi permasalahan pada beberapa metode sebelumnya [3], [4], dikarenakan CNN dapat menerima *input dataset* dengan karakteristik lebih umum atau dengan banyak *noise* sekalipun. Dalam penelitian lain CNN juga terbukti memiliki performansi yang sangat baik pada kasus *Computer Vision*, dalam hal klasifikasi citra [10].

CNN termasuk dalam *supervised training method* sehingga bisa mengetahui keterkaitan akan *input* dan *output* dari banyak *sample* [5]. Untuk dapat membentuk pola-pola keterkaitan tersebut, banyak parameter dari CNN yang harus dipelajari, sehingga dibutuhkan juga jumlah *dataset* yang besar [5]. CNN memiliki struktur jaringan mendalam yang setiap levelnya merepresentasikan suatu fitur dan setiap fitur yang lebih tinggi merupakan abstrak dari fitur dibawahnya serta terdiri atas tiga tipe *layer* yaitu *convolutional layer*, *pooling layer* dan *full connection layer* [5]. Metode CNN memiliki banyak arsitektur struktur jaringan serta selalu mengalami perkembangan guna meningkatkan performansi dari sistem yang dirancang, beberapa contoh arsitekturnya seperti VGG-16 [10] dan ZFNet [11].

Dalam CNN, arsitektur sangat berperan dalam menentukan performa sistem. Penulis menggunakan arsitektur ResNet-50 [12] sebagai arsitektur jaringan CNN dalam tugas akhir ini. Klasifikasi yang dilakukan dalam tugas akhir ini dibagi ke dalam dua kelas citra yaitu *non-pornographic* dan *pornographic*. Besar harapan tugas akhir ini dapat digunakan sebagai acuan dalam penelitian selanjutnya, terutama dalam kasus klasifikasi citra.

1.2 Rumusan masalah

Berdasarkan deskripsi latar belakang dan penelitian terkait, maka dirumuskan beberapa masalah pada tugas akhir ini yaitu:

1. Metode sebelumnya yang digunakan dalam klasifikasi citra pornografi masih memiliki permasalahan dalam keterbatasan citra *dataset* yang menjadi *input*.
2. Bagaimana cara melakukan klasifikasi citra pornografi menggunakan metode CNN?
3. Pengaruh perubahan konfigurasi *learning rate*, *epochs* serta *batch size* terhadap performansi model sistem.

1.3 Tujuan dan manfaat

Berdasarkan rumusan masalah diatas, tujuan dari penelitian ini adalah:

1. Merancang suatu sistem klasifikasi citra pornografi yang dapat mempelajari dan memprediksi *dataset* dengan karakteristik yang lebih umum dan beragam.
2. Merancang suatu sistem klasifikasi citra pornografi menggunakan metode CNN dengan arsitektur ResNet-50.
3. Menganalisis performansi dari metode CNN dalam klasifikasi berdasarkan konfigurasi dari *learning rate*, *epochs* serta *batch size*.

Adapun manfaat dari penelitian ini adalah:

1. Menghilangkan keterbatasan ragam karakteristik *dataset* yang menjadi *input* dari model sistem.
2. Mengetahui kinerja dari metode CNN dalam klasifikasi citra pornografi.

1.4 Batasan masalah

Guna mempersempit lingkup permasalahan agar pengerjaan dapat terfokus pada penyelesaian masalah, maka dibuat batasan masalah sebagai berikut:

1. Menggunakan arsitektur ResNet-50.
2. Menggunakan *NPDI Dataset* sebagai *dataset* yang terdiri dari 400 *pornographic videos* dan 400 *non-pornographic videos* serta menghasilkan 16,727 citra setelah proses ekstraksi dari video tersebut.

3. *Dataset* yang digunakan untuk kelas *non-pornographic* merupakan citra yang terdiri dari sekumpulan objek manusia, benda atau kejadian alam, dan untuk kelas *pornographic* terdiri dari gambar eksplisit yang menampilkan bagian tubuh secara terbuka dengan komposisi warna kulit yang dominan dalam satu *frame*.
4. *Learning rate*, *epochs* serta *batch size* adalah *hyperparameter* yang dikonfigurasi untuk skema pengujian.
5. Menggunakan bahasa pemrograman Python dan perangkat lunak pemrograman Spyder.
6. Menggunakan komputer dengan spesifikasi:
 - *Processor* : Intel Core i5 9400f
 - *OS* : Windows 10
 - *RAM* : 16 GB DDR4
 - *SSD* : 256 GB
 - *GPU* : NVIDIA GeForce GTX 1660 TI

1.5 Metode penelitian

Metode yang disusun untuk menyelesaikan penelitian ini terdiri dari beberapa tahapan sebagai berikut:

1. Studi literatur
Melakukan studi literatur dengan mencari, memahami dan menganalisis jurnal, artikel maupun web yang berkaitan dengan CNN, ResNet-50, klasifikasi dan materi pendukung tugas akhir lainnya.
2. Konsultasi dengan dosen pembimbing
Melakukan diskusi dengan dosen pembimbing berkaitan dengan pemilihan metode yang digunakan untuk sistem beserta langkah-langkah dasar guna perancangan sistem.
3. Perancangan sistem
Melakukan perancangan model ResNet-50 beserta sistem klasifikasi dan analisisnya menggunakan bantuan piranti lunak Spyder. Dalam merancang sistem ditentukan parameter-parameter yang akan dikonfigurasi dalam

model. Setelah itu model yang sudah terkonfigurasi dilatih dengan *dataset* yang telah dipersiapkan.

4. Pengujian

Menguji sistem yang telah dirancang dan dilatih dengan melakukan simulasi percobaan klasifikasi terhadap data baru yang belum pernah dilihat oleh sistem sebelumnya.

5. Analisis hasil uji

Memperhatikan keluaran yang dihasilkan oleh sistem serta melakukan analisis terhadap hasilnya guna melakukan evaluasi dan optimalisasi terhadap sistem jika masih ditemukan kekurangan. Analisis dapat dilakukan dengan memperhatikan parameter dalam model serta dapat dibantu dengan beberapa metode analisis seperti *confusion matrix*.

6. Penyusunan buku tugas akhir

Mendokumentasikan seluruh hasil kerja kedalam bentuk buku tugas akhir sekaligus di dalamnya tercakup kesimpulan dari tugas akhir yang dibuat.

1.6 Sistematika penulisan

Sistematika penulisan laporan tugas akhir ini adalah sebagai berikut:

- Bab 2 TINJAUAN PUSTAKA

Bab ini memaparkan tentang dasar teori yang dipakai dalam melakukan perancangan, *training* serta evaluasi dari model sistem.

- Bab 3 MODEL SISTEM DAN PERANCANGAN

Bab ini memaparkan tentang sistematika pembentukan model sistem dari tahap awal hingga dapat mengeluarkan hasil beserta berbagai konfigurasi yang akan diterapkan dalam model sistem.

- Bab 4 HASIL DAN ANALISIS

Bab ini memaparkan hasil dari proses *training* dan pengujian model yang telah dibuat terhadap data baru serta analisis dari hasil tersebut.

- Bab 5 SIMPULAN DAN SARAN

Bab ini memaparkan tentang simpulan saran dari tugas akhir yang dikerjakan.

1.7 Jadwal pelaksanaan

Guna menjadi acuan dan evaluasi terhadap tahap pengerjaan tugas akhir, berikut adalah jadwal pelaksanaan kegiatan yang dijalankan, dilengkapi dengan *milestone*:

Tabel 1.1 Jadwal pelaksanaan.

| No. | Tahap | Durasi | Tanggal Selesai | Milestone |
|-----|--------------------------------------|----------|------------------|---|
| 1. | Studi metode | 4 Minggu | 24 Oktober 2019 | Melakukan studi literatur CNN dan ResNet-50. |
| 2. | Perancangan Sistem | 8 Minggu | 19 Desember 2019 | Merancang model sistem klasifikasi ResNet-50. |
| 3. | Pengumpulan <i>dataset</i> | 3 Minggu | 9 Januari 2020 | Mengumpulkan dan melakukan cek terhadap <i>dataset</i> yang memenuhi kriteria sistem. |
| 4. | <i>Training</i> dan Pengujian Sistem | 4 Minggu | 6 Februari 2020 | Melatih model sistem dirancang dengan <i>dataset</i> serta melakukan simulasi untuk menguji performansi model sistem. |
| 5. | Evaluasi | 4 Minggu | 5 Maret 2020 | Mengevaluasi dan modifikasi model sistem jika diperlukan adanya optimalisasi. |
| 6. | Penyusunan Buku Tugas Akhir | 4 Minggu | 2 April 2020 | Menyusun buku tugas akhir dan penarikan kesimpulan terhadap tugas akhir. |
| 7. | Penyusunan Jurnal | 3 Minggu | 23 April 2020 | Menyusun jurnal dari tugas akhir untuk kepentingan publikasi. |