BAB I
PENDAHULUAN

1.1. Latar Belakang Masalah

Optical Character Recognition (OCR) merupakan teknologi yang memungkinkan komputer untuk bisa mengenali karakter alfanumerik *(alphabetic and numeric*) yang ada dalam gambar atau dokumen yang dipindai, dan mengubahnya menjadi format yang bisa dibaca oleh komputer (Srivastava et al. 2022). Krutilla dan Riczu (2021) dalam penelitiannya menyimpulkan bahwa teknologi ini sangat penting dalam berbagai implementasi, mulai dari digitalisasi arsip hingga ke dalam sektor industri dan pendidikan. Penerapan *OCR* yang efektif telah terbutkti dapat mempercepat proses digitalisasi dokumen, mengurangi biaya operasional yang terkait dalam pengerjaan konvensional, serta meningkatkan aksesibilitas informasi.

Meskipun demikian, Kae et al. (2010) dalam penelitiannya menjelaskan bahwa variansi dan penerapan *specific-model OCR* dalam berbagai bidang masih sangat kurang. Mereka menemukan bahwa model *OCR* yang beredar dan digunakan oleh industri umumnya masih menggunakan *general-purpose OCR*, sehingga menghilangkan potensi peningkatan akurasi yang harusnya bisa didapatkan. Selain itu, Mereka juga melakukan pelatihan dengan menggunakan *initial output* dari model *general-purpose OCR Tesseract* yang dilatih kembali dengan menggunakan gaya *font* spesifik dari dokumen tersebut dan berhasil mengurangi kesalahan pada karakter yang disegmentasi dengan benar sebesar 34.1%. Dari kajian literatur inilah peneliti beranggapan bahwa diperlukan lebih banyak alternatif dan kajian analisis komparatif untuk terus meningkatkan kualitas dan variansi OCR, baik dalam model pembelajarannya, teknik prapemrosesan data, serta kualitas dan kuantitas dataset.

Penelitian Kae et al. (2010) menjadi salah satu dasar penting dalam pengembangan specific-model OCR karena menunjukkan bahwa penggunaan general-purpose OCR masih memiliki keterbatasan akurasi pada dokumen dengan karakteristik tertentu. Namun, mengingat rentang waktu penelitian yang cukup panjang hingga saat ini, permasalahan terkait akurasi OCR pada dokumen spesifik

masih menjadi perhatian dalam penelitian terkini. (Hegghammer 2022) menunjukkan bahwa performa OCR umum masih sangat dipengaruhi oleh noise dan karakteristik dokumen, sehingga pendekatan khusus seperti preprocessing dan penyesuaian model masih diperlukan. Selain itu, beberapa penelitian terbaru juga menunjukkan bahwa fine-tuning OCR terhadap font, bahasa, maupun domain dokumen tertentu mampu meningkatkan performa pengenalan karakter secara signifikan. Oleh karena itu, penelitian mengenai specific OCR masih relevan untuk dilakukan sebagai upaya meningkatkan akurasi dan robustness sistem OCR pada dokumen dengan karakteristik khusus.

Lanskap penelitian OCR modern kini didominasi oleh *Deep Learning,* yang telah membuktikan keunggulannya secara empiris atas metode tradisional. Sebagaimana yang telah ditunjukkan oleh (Perumal et al. 2024) dalam penelitiannya, perbandingan eksplisit antara algoritma tradisional dengan arsitektur *Deep Learning* modern seperti *Convolutional Recurrent Neural Network* (CRNN) menghasilkan peningkatan akurasi yang signifikan. Berdasar dari temuan ini, penelitian ini mengadopsi pendekatan *Deep Learning* dengan pertimbangan bahwa berbagai model *Deep Learning* telah terbukti berkontribusi besar terhadap peningkatan akurasi sistem OCR modern.

Di antara beragam arsitektur *Deep Learning*, Convolutional Neural Network (CNN) telah menjadi fondasi utama untuk tugas-tugas pemrosesan gambar, termasuk segmentasi dan ekstraksi fitur visual. Lebih lanjut, perkembangan dalam arsitektur CNN telah melahirkan varian-varian baru yang lebih baru. Dalam konteks inilah, *Residual Network* (*ResNet*) diperkenalkan sebagai salah satu solusi paling efektif. He et al. (2015) dalam penelitiannya membuktikan bahwa konsep *residual block* pada model *ResNet* terbukti bisa mengatasi masalah *vanishing gradient,* yang merupakan tantangan kritis dalam pelatihan jaringan syaraf yang sangat dalam. *ResNet* memperkenalkan konsep inovatif berupa *residual block* yang memungkinkan pelatihan jaringan dengan puluhan hingga ratusan layer secara efektif. Karena kedalaman arsitektur inilah yang memungkinkan model untuk mempelajari representasi fitur yang semakin hierarkis, abstrak, dan kompleks. Fitur-fitur kompleks inilah yang bisa membaca fitur fitur alfabetikal mulai dari garis dan sudut sederhana hingga bentuk karakter yang spesifik dan kontekstual. Hal

inilah yang dibutuhkan untuk mengenali teks dengan akurasi tinggi dalam beragam font, ukuran, gaya, dan kondisi citra yang bervariasi. Atas dasar itulah penulis mengimplementasikan *ResNet* pada studi kali ini.

Tentu saja *ResNet* juga memilki beberapa variansi, salah satunya adalah *ResNet50* dan *ResNet101* yang memilki jumlah lapisan yang berbeda sesuai namanya *ResNet50* dengan 50 lapisan dan *ResNet101* dengan 101 lapisan. Karena *printed-document format* memiliki beberapa ragam font dengan beberapa detail yang sangat kecil, dua model inilah yang penulis pilih karena dirasa memiliki lapisan yang cukup banyak sehingga detail fitur dari berbagai macam gaya font, sehingga model bisa melakukan prose *features extraction* dari setiap detail secara menyeluruh. Karena *ResNet* merupakan model yang sangat berat dalam segi pelatihan dan evaluasi, jadi perancangan model ini harus dilakukan seefisien mungkin untuk bisa mendapatkan hasil yang optimal. Selain teknik prapemrosesan dan model arsitektur, kualitas kuantitas dataset juga bisa mempengaruhi kualitas program OCR yang dibuat.

Guna menguji efektivitas ResNet dalam tugas pengenalan karakter, penelitian ini memerlukan dataset yang merepresentasikan tantangan OCR sesuai dari studi kasus yang kita bahas sebelumnya, yaitu *printed-document format*. Untuk itu dataset Chars74K dipilih karena kemampuannya memenuhi kriteria ini, dengan cakupannya yang luas atas berbagai font dan gaya natural. Dataset ini diperkenalkan oleh (de Campos et al., n.d.) dalam penelitian mereka "Character Recognition in Natural Images". Dataset ini berisi 62.992 gambar dengan 62 kelas (huruf dan angka) yang mencakup *font* dan gaya yang sangat luas yang umum ditemui dalam dokumen digital dan hasil pindaian. Karakteristik inilah yang membuat dataset ini menjadi faktor yang ideal untuk melatih dan mengevaluasi model dalam mengenali karakter dari berbagai macam format dokument tercetak.

Perlu diperhatikan dengan skalabilitas dataset ini, yang memiliki lebih dari 74.000 gambar dalam format PNG dan tersebar dalam 62 kelas menghadirkan tantangan komputasi tersendiri. Khususnya dalam kecepatan dan proses *loading data* selama waktu pelatihan berlangsung. Meskipun demikian, kompleksitas ini justru merepresentasikan skenario nyata dalam pelatihan model OCR, sekaligus

menjadi kesempatan untuk membandingkan kinerja model secara keseluruhan. Dari latar belakang yang telah dijelaskan, beberapa permasalahan kunci yang menjadi fokus penelitian ini adalah.

1.2. Identifikasi Masalah

Berdasarkan latar belakang masalah di atas, permasalah yang dapat diidentifikasi adalah yaitu:

1) Karena *ResNet101* dan *ResNet50* merupakan model yang cukup berat, ditambah dengan dataset Chars74K yang termasuk dataset yang cukup besar dengan kebutuhan augmentasi data yang harus disesuaikan ulang dengan model nantinya. Jika perangkat keras tidak cukup memadai, model yang lebih dalam seperti *ResNet101* akan mengalami keterbatasan dalam waktu pelatihan dan inferensi.
2) Chars74K berisi berbagai jenis gaya penulisan alfanumerik dengan *format* document yang memiliki tingkat kerumitan yang beragam. Tentunya harus dilakukan proses *preprocessing,* agar data yang digunakan dalam pelatihan dan inferensi bisa menghasilkan karakteristik yang sama.
3) Chars74K merupakan dataset yang berisi 62.992 gambar dengan 62 kelas dan tiap gambarnya memiliki resolusi (128, 128) pixel dengan beragam font dan style dalam bentuk alfanumerik *printed document*. Jaditiap kelas dalam dataset ini memiliki 1.016 gambar, meskipun keseluruhan data memiliki jumlah yang banyak tapi jumlah data per kelas memiliki data yang tidak cukup untuk dilakukan pelatihan. Tentu saja harus dilakukan proses augmentasi data agar data yang dihasilkan bisa cukup untuk model yang akan dilatih dan memiliki karakteristik yang lebih beragam.
4) Data inferensi yang kita dapatkan merupakan hasil dari program deteksi dan klasifikasi yang sudah dibuat. Pertimbangan *noise*, pencahayaan, distorsi perspektif dan lain lainnya membuat data ini tentunya belum layak untuk digunakan sebagai data inferensi.

1.3. Pembatasan Masalah

Berdasarkan latar belakang dan identifikasi masalah yang telah dijabarkan, terdapat beberapa pembatasan masalah yang dapat diidentifikasi, yaitu:

1.3.1. Dataset

1) Penelitian ini hanya menggunakan dataset Chars74K untuk pelatihan dan pengujian model yang digunakan dalam tugas pengenalan karakter. Dataset ini berisi semua data gambar alfanumerik yang digunakan dalam penelitian.
2) Fokus penelitian ini hanya pada printed character, dan tidak mencakup karakter lain seperti karakter tulisan tangan.

1.3.2. Model

1) Penelitian ini hanya menggunakan model *ResNet50* dan *ResNet101* dalam mengukur kinerja model OCR pada dataset Chars74K.
2) Penelitan ini tidak menggunakan *pre-trained weight* dalam melatih modelnya.

1.3.3. Proses pelatihan

1) Proses pelatihan menggunakan hyperparameter yang telah diatur seperti jumlah *epoch*s, *learning rate* dan ukuran *batch* gambar tanpa menggunakan hyperparameter lainnya yang lebih kompleks.
2) Data augmentation dilakukan pada dataset untuk meningkatkan jumlah dan variansi data dalam pelatihan, namun ada beberapa pembatasan pada manipulasi variansi data.
3) Data *preprocessing* dilakukan pada dataset sebelum dilakukan pelatihan untuk menghasilkan karakteristik yang sama antara data inferensi dan data pelatihan.

1.3.4. Kinerja

1) Analisis kinerja yang diukur terbatas pada pengukuran akurasi dan loss model dalam melakukan pengenalan karakteristik pada dataset Chars74K.
2) Analisis kinerja juga mencakup waktu inferensi, waktu pelatihan.

1.3.5. Evaluasi

Evaluasi hanya dilakukan perbandingan hasil kuantitatif model tanpa melibatkan uji statistik lanjut atau analisis varians yang lebih kompleks.

1.4. Perumusan Masalah

Jadi masalah yang bisa dirumuskan dari penjelasan sebelumnya adalah bagaimana analisis komparatif kinerja model *ResNet50* dan *ResNet101* dalam pengenalan karakter menggunakan dataset Chars74K dalam uji statistik hipotesis?

1.5. Tujuan Penelitian

Tujuan utama penelitian ini adalah untuk melakukan analisis komparatif kinerja model ResNet50 dan ResNet101 pada dataset Chars74K. Fokus perbandingan mencakup aspek akurasi, waktu komputasi selama pelatihan, dan kecepatan inferensi. Melalui perbandingan ini, diharapkan dapat diketahui kelebihan dan kekurangan masing-masing model dalam konteks pengenalan karakter. Selain itu, penelitian ini juga akan mengeksplorasi beberapa teknik prapemrosesan data untuk mengoptimalkan kinerja model. Secara keseluruhan, hasil penelitian ini diharapkan dapat menjadi referensi dan dasar pertimbangan dalam pemilihan model serta teknik pra-pemrosesan untuk pengembangan sistem OCR yang lebih akurat dan efisien di masa depan.

1.6. Manfaat Penelitian

1.6.1. Manfaat Teoritis

Manfaat teoritis yang dihasilkan dari penelitian ini adalah.

1) Penelitian ini diharapkan bisa memberikan kontribusi sebagai referensi baru dalam bidang pengolahan citra komputer, terutama dalam pengembangan program OCR. Dengan membandingkan dan menganalisis model *ResNet50* dan *ResNet101*, penelitian ini diharapkan bisa memperkaya dan bisa memberikan sudut pandang baru mengenai penerapan *deep learning* dalam tugas pengenalan karakter dengan konteks khusus penggunaan OCR dalam *format printed document*, dengan menggunakan dataset Chars74K.
2) Penelitian ini bisa memberikan wawasan lebih mengenai kedalaman arsitektur *ResNet* terhadap akurasi, kecepatan inferensi dan penggunaan sumber daya komputer terhadap penggunaan dataset Chars74K, serta dapat membuka peluang untuk pengembangan model - model serupa yang lebih optimal di masa yang akan datang.

1.6.2. Manfaat Praktis

Manfaat praktis yang dihasilkan dari penelitian ini adalah.

1) Hasil penelitian ini dapat digunakan untuk pengembangan sistem OCR dalam memilih arsitektur serta dataset yang paling sesuai untuk aplikasi pengenalan karakter, terutama untuk objek yang memiliki karasteristik sama seperti dataset Chars74K.
2) Penelitian ini diharapkan bisa memberikan rekomendasi bagi praktisi industri dan pendidikan agar bisa mendapatkan opsi lain untuk bisa menerapkan program ini sebagai cara meningkatkan produktifitas.

1.6.3. Manfaat Untuk Penelitian Selanjutnya

Manfaat dari penelitian ini yang bisa digunakan untuk penelitian selanjutnya adalah.

1) Penelitian ini diharapkan bisa memberikan wawasan baru mengenai teknik-teknik prapemrosesan gambar, *masking* serta informasi mengenai berbagai hal lainnya yang bisa jadi pertimbangan krusial dalam pengembangan program pengenalan karakter.
2) Penelitian ini dapat menjadi referensi untuk penelitian lebih lanjut dalam bidang pengenalan karakter menggunakan model *deep learning*, khususnya mengenai perbandingan arsitektur *ResNet* dengan model-model lainnya.
3) Hasil dari penelitian ini dapat membuka peluang untuk pengembangan model yang lebih optimal lagi dengan mempertimbangkan faktor efisiensi, sumber daya komputasi serta akurasi yang lebih tinggi yang dapat meningkatkan kualitas OCR dalam berbagai implementasi.