

BAB I

PENDAHULUAN

I.1 Latar Belakang

Perkembangan teknologi informasi yang pesat telah memungkinkan komputer untuk meniru berbagai fungsi kognitif manusia, termasuk dalam bidang penglihatan. *Deep learning* sebagai cabang dari kecerdasan buatan, telah menjadi pendorong utama dalam kemajuan ini, terutama dalam pengolahan data visual. Dengan memanfaatkan arsitektur jaringan saraf tiruan yang kompleks, seperti *Convolutional Neural Networks* (CNN) dan *Recurrent Neural Networks* (RNN), *deep learning* mampu menghasilkan model yang secara signifikan meningkatkan kemampuan komputer dalam mengenali pola dan membuat keputusan berdasarkan data visual.

Deep learning memainkan peran sentral dalam *computer vision* dengan menyediakan algoritma yang dapat secara otomatis mempelajari fitur dari data tanpa memerlukan perancangan manual. CNN, misalnya, telah membuktikan keunggulannya dalam tugas-tugas seperti pengenalan wajah, klasifikasi gambar, dan segmentasi objek. Teknologi ini tidak hanya mengandalkan ekstraksi fitur statis tetapi juga mampu memahami konteks visual yang lebih kompleks dengan memproses data dalam berbagai lapisan non-linear (Zhou et al., 2021).

Kemajuan dalam deep learning juga telah memungkinkan pengembangan teknologi analisis citra dokumen (*document image analysis*). Bidang ini memanfaatkan *library* seperti OCR yang didukung oleh model *deep learning* untuk mengenali dan menginterpretasi informasi dalam citra dokumen. Sebagai contoh,

RNN sering digunakan untuk memproses data teks yang diekstraksi dari gambar, memungkinkan komputer untuk membaca dan memahami struktur teks secara efisien (Smith et al., 2020; Liu et al., 2019). Kombinasi CNN dan RNN sering diterapkan untuk tugas-tugas seperti pengenalan karakter yang kompleks dalam dokumen digital.

Pengenalan teks dokumen memiliki peran penting dalam mempermudah pengambilan informasi dari dokumen teks yang sering kali berisi data kritis. Dengan teknologi ini, informasi yang sebelumnya memerlukan waktu lama untuk diproses secara manual kini dapat diakses dengan cepat dan akurat. Hal ini berdampak positif pada kecepatan pengambilan keputusan, terutama oleh pimpinan yang memerlukan akses langsung ke data untuk mendukung strategi dan operasional organisasi. Dalam konteks ini, sistem berbasis deep learning dapat memberikan solusi yang andal untuk menyaring, menganalisis, dan menyajikan informasi relevan dari dokumen-dokumen yang kompleks, meningkatkan efisiensi dan efektivitas proses pengambilan keputusan.

Salah satu tantangan utama dalam analisis citra dokumen adalah kemampuan untuk menangani data dengan berbagai variasi, seperti perubahan pencahayaan, rotasi, atau noise. Deep learning telah menyediakan solusi yang andal dengan mengadopsi metode augmentasi data dan pelatihan model berbasis dataset besar. Pendekatan ini tidak hanya meningkatkan akurasi tetapi juga membuat sistem lebih tahan terhadap variabilitas dalam data dunia nyata (Gao et al., 2020).

Selain itu, aplikasi deep learning dalam bidang ini meluas ke berbagai sektor, termasuk kesehatan, keamanan, dan industri. Dalam perawatan kesehatan, model deep learning digunakan untuk menganalisis gambar medis, seperti

pemindaian MRI dan X-ray, untuk mendeteksi anomali secara otomatis. Di bidang keamanan, teknologi ini diterapkan dalam sistem pengawasan untuk mendeteksi aktivitas mencurigakan secara real-time. Bahkan dalam hiburan, deep learning mendukung teknologi seperti efek visual dan pengenalan wajah untuk personalisasi pengalaman pengguna.

Dengan kemampuan yang terus berkembang, *deep learning* telah membuktikan dirinya sebagai teknologi yang tidak hanya memperluas cakupan *computer vision* tetapi juga membawa dampak revolusioner dalam berbagai aplikasi dunia nyata. Pendekatan berbasis *deep learning* memberikan solusi yang lebih efisien, adaptif, dan akurat untuk mengatasi tantangan dalam pengolahan data visual dan analisis citra dokumen, menjadikannya fondasi utama dalam perkembangan teknologi informasi modern.

Penggunaan citra dokumen memiliki tantangan tersendiri dikarenakan beberapa kelemahannya seperti ketergantungan terhadap kualitas gambar, pengenalan karakter yang kurang sempurna, sering kali tidak menyertakan konteks visual yang diperlukan, keterbatasan dalam pencarian dan indeks, kesulitan dalam pemrosesan bahasa, serta kesulitan dalam pemeliharaan (Dessy Tri, 2022).

Penelitian ini menggunakan *deep learning* untuk menganalisis dan memproses data teks dari citra dokumen dengan baik. *Deep learning* menggunakan model jaringan saraf tiruan yang terdiri dari banyak lapisan (layer) yang saling terhubung untuk mempelajari pola kompleks dari data input. Hal ini sangat membantu dalam hal pengenalan citra dokumen yang sangat kompleks. *Deep learning* memiliki beberapa arsitektur yang sering digunakan untuk pengenalan citra. Pada penelitian ini akan membandingkan 2 jenis arsitektur jaringan saraf

tiruan, yaitu *Recurrent Neural Network (RNN)* dan *Convolutional Neural Network (CNN)*.

Convolutional Neural Network (CNN) awalnya dirancang untuk pemrosesan citra, namun belakangan ini juga menunjukkan efektivitasnya dalam pemrosesan teks. CNN menggunakan operasi konvolusi untuk mengekstraksi fitur dari teks, seperti pengenalan pola dalam urutan kata. Dalam pemrosesan citra, CNN secara efektif menangkap fitur spasial dan pola lokal melalui lapisan konvolusi yang bertingkat, yang memungkinkan model ini untuk mengenali variasi dalam gaya huruf, ukuran, dan orientasi teks. Kemampuan ini membuat CNN sangat kuat dalam menghadapi distorsi dan noise yang sering ditemukan dalam citra teks. Dengan memanfaatkan filter konvolusi, CNN dapat mengekstraksi fitur penting sekaligus meminimalkan noise yang tidak relevan, memungkinkan pemrosesan yang lebih efisien dan akurat (He et al., 2020; Liu et al., 2019).

CNN juga dirancang untuk mengoptimalkan penggunaan sumber daya komputasi dengan mengurangi jumlah parameter melalui penggunaan filter berbagi berat dan operasi pooling. Ini membuat pelatihan model lebih cepat dan penggunaan memori lebih efisien dibandingkan dengan model jaringan saraf lainnya. Kinerja unggul CNN dalam berbagai kondisi pencahayaan, latar belakang, dan orientasi teks menjadikannya pilihan yang sangat baik untuk aplikasi dunia nyata di mana variabilitas citra tidak dapat dihindari (Tan & Le, 2020; Liu et al., 2021).

CNN terus berkembang menjadi metode yang sangat efisien dalam pemrosesan teks. CNN tidak hanya mampu mengenali pola visual tetapi juga telah diterapkan dengan sukses untuk analisis teks dengan mengekstraksi fitur-fitur

seperti kata atau frasa kunci dalam urutan teks. Dengan menerapkan operasi konvolusi, CNN dapat mengidentifikasi pola-pola penting dalam teks, seperti pengenalan entitas atau deteksi sentimen, yang sangat berguna dalam pemrosesan bahasa alami (Zhang & Wallace, 2019). CNN telah menunjukkan kemampuannya dalam menghadapi tantangan variabilitas dalam pengenalan teks, termasuk perbedaan gaya tulisan, ukuran, dan orientasi. Arsitektur berlapis dari CNN memungkinkan pemrosesan dari fitur sederhana ke fitur yang lebih kompleks secara bertahap, menjadikannya alat yang sangat ampuh dalam aplikasi OCR dan pengenalan karakter yang membutuhkan akurasi tinggi dalam kondisi yang berubah-ubah (Gehring et al., 2017).

Recurrent Neural Network (RNN) adalah arsitektur jaringan saraf tiruan yang dirancang khusus untuk memproses data berurutan, seperti teks. RNN memiliki sifat memori jangka panjang (long-term memory) yang memungkinkan mereka untuk mengingat informasi dari masa lalu dalam pemrosesan teks. RNN, terutama varian seperti *Long Short-Term Memory (LSTM)* dan *Gated Recurrent Unit (GRU)*, sangat cocok untuk memodelkan data sekuensial. Dalam konteks pengenalan citra teks, ini memungkinkan RNN untuk memahami urutan karakter dalam teks dan hubungan antar karakter (Li Y et al., 2019). RNN mampu menangani urutan teks dengan panjang yang bervariasi tanpa perlu mengubah arsitektur model, berguna dalam pengenalan teks di mana panjang kata atau kalimat dapat sangat bervariasi (Khan M.A et al., 2020). RNN dapat dengan mudah diintegrasikan dengan CNN untuk membentuk model hibrida yang sangat kuat. Dalam pendekatan ini, CNN bertanggung jawab untuk ekstraksi fitur spasial dari data input seperti gambar atau video, sedangkan RNN digunakan untuk memproses

urutan temporal dari fitur-fitur yang telah diekstraksi oleh CNN, memungkinkan model untuk memahami konteks temporal dan pola berurutan dalam data. Kombinasi ini sangat efektif dalam berbagai aplikasi seperti pengenalan teks dan analisis video, di mana memahami urutan dan konteks temporal sama pentingnya dengan pengenalan pola visual (Hu et al., 2021; Xiao et al., 2019). Dalam pengenalan teks, memahami konteks karakter adalah kunci. RNN memiliki keunggulan dalam mengasimilasi konteks dari urutan karakter, yang membantu dalam meningkatkan akurasi pengenalan teks terutama dalam bahasa alami yang kompleks (Liu C et al., 2019).

Beberapa penelitian sebelumnya telah melakukan analisis terhadap pengenalan teks pada citra dokumen. Seperti yang dilakukan oleh (G.R. Hemanth et al., 2023) melakukan penelitian dan mendapatkan hasil bahwa jaringan saraf tiruan yang dilatih menggunakan dataset berupa tulisan tangan mendapatkan hasil bahwa RNN dapat memproses input yang lebih besar meskipun memiliki daya komputasi yang lebih rendah. Di sisi lain, jaringan saraf tiruan CNN membutuhkan data yang lebih besar untuk pelatihan.

(Mayur Bhargab Bora et al., 2019) mengemukakan hasil penelitiannya bahwa *Optical Character Recognition* (OCR) dapat diimplementasikan dengan menggunakan CNN, yang merupakan arsitektur jaringan saraf tiruan yang populer, berhasil mengambil gambar optik karakter sebagai input dan menghasilkan karakter yang sesuai sebagai output.

(Kasyfi Ivanedra & Metty Mustikasari, 2019) mengemukakan, RNN memiliki nilai akurasi yang lebih baik dibandingkan dengan jaringan saraf tiruan sederhana karena bobot yang dihitung akan lebih akurat mendekati persamaan

setiap kata. Jenis RNN yang digunakan adalah *LSTM* (*Long Short Term Memory*) untuk menutupi kekurangan pada RNN yang tidak dapat menyimpan memori untuk dipilah dan menambahkan mekanisme *Attention* agar setiap kata dapat lebih fokus pada konteks. Dalam penelitiannya, ia menguji performa sistem menggunakan *Precision*, *Recall*, dan *F-Measure* dengan membandingkan hasil ringkasan yang dihasilkan oleh sistem dan ringkasan yang dibuat oleh manusia. Dataset yang digunakan adalah data artikel berita dengan jumlah total artikel sebanyak 4515 buah artikel. Pengujian dibagi berdasarkan data dengan menggunakan *Stemming* dan dengan teknik *Non-stemming*. Nilai rata-rata *recall* artikel berita *Non-stemming* adalah sebesar 41%, *precision* sebesar 81%, dan *F-measure* sebesar 54,27%. Sedangkan nilai rata-rata *recall* artikel berita dengan teknik *stemming* sebesar 44%, *precision* sebesar 88%, dan *F-measure* sebesar 58,20 %.

CNN sangat efektif dalam memahami struktur karakter atau kata pada tulisan tangan dengan cara membantu ekstraksi otomatis fitur-fitur yang berbeda dan menjadikan CNN sebagai pendekatan yang paling cocok untuk memecahkan masalah pengenalan tulisan tangan. Sistem yang dibangun untuk mendeteksi tulisan dengan format yang berbeda. Perkembangan tulisan tangan semakin canggih, dimana ditemukan berbagai macam karakter tulisan tangan seperti digit, angka, tulisan kursif, simbol, dan skrip termasuk bahasa Inggris dan bahasa lainnya (Rohini G. Khalkar et al., 2021).

Penggunaan RNN, dapat memberikan hasil yang lebih baik pada dataset teks Arab. Lapisan layer di atas RNN memungkinkan model untuk mengidentifikasi area fokus dari urutan input, sehingga proses pembelajaran menjadi lebih cepat dan efisien. Dalam tahap prapemrosesan, mereka membuat dataset baru dari gambar

berisi satu kata dari Alif dan Activ, yang membantu mengurangi tingkat kesalahan baris sebesar 85% hingga 87%. Model yang mereka buat memberikan hasil yang lebih baik dibandingkan model sederhana berbasis CNN, RNN, dan hibrida CNN-RNN (Hanan Butt et al., 2021).

Penelitian sudah sampai tahap implementasi model dengan menggabungkan Model CNN dan RNN pada dokumen Resi dan Kuitansi dengan memperkuat penerapan teknik augmentasi untuk meningkatkan generalisasi model dalam proses pelatihan dan evaluasi sehingga mendapatkan hasil akurasi sebesar 83% dengan rasio data latih dan uji adalah 95:5. CNN dan RNN merupakan teknik pembelajaran mendalam yang telah terbukti efisien dalam mengolah dan mengenali gambar-gambar kompleks, termasuk gambar-gambar tulisan tangan dengan berbagai variasi bentuk dan ukuran (Yafi Cahyono Adi et al., 2024).

I.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka rumusan masalah dalam penelitian ini adalah :

1. Bagaimana performa model CNN?
2. Bagaimana performa model CRNN?
3. Bagaimana performa kedua model tersebut saat diimplementasikan dalam bahasa pemrograman python?

I.3 Tujuan

Penelitian ini dilakukan dengan tujuan untuk

1. Mendapatkan performa model CNN.

2. Mendapatkan performa model CRNN.
3. Mendapatkan hasil dan performa kedua model tersebut saat diimplementasikan kedalam bahasa pemrograman python.

I.4 Manfaat

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Memberikan informasi mendalam terhadap perbandingan performa model CNN dan CRNN untuk pengenalan teks pada citra dokumen.
2. Menjadi referensi untuk penelitian lain yang memiliki topik yang sama dengan penelitian ini.

I.5 Batasan Masalah

Berdasarkan perumusan masalah di atas terdapat batasan masalah sebagai berikut:

1. Penelitian ini akan membandingkan 2 algoritma *Neural Network* yaitu *Convolutional Neural Network* (CNN) dan gabungan (*hybrid*) CNN-RNN (CRNN).
2. Penelitian ini menggunakan *library OCR (Optical Character Recognition)* dalam membantu melakukan pengenalan teks pada citra dokumen.
3. Dataset yang akan digunakan dalam penelitian ini adalah dataset MNIST yang diperoleh dari *Tensorflow.org* pada laman <https://www.tensorflow.org/datasets/catalog/mnist> dan dokumen

sample.pdf berisi teks untuk implementasi hasil.

4. Jumlah data yang digunakan adalah 70.000 data dengan data latih sebanyak 60.000 data dan data uji sebanyak 10.000 data.

I.6 Sistematika Pembahasan

Penyusunan tesis ini menggunakan kerangka pembahasan yang terbentuk dalam susunan bab yang dapat dijelaskan sebagai berikut:

BAB I : PENDAHULUAN

Bab ini berisikan tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian serta sistematika penulisan.

BAB II : TINJAUAN PUSTAKA

Bab ini menjelaskan teori-teori dan referensi penelitian yang diperoleh dari berbagai sumber, seperti buku dan artikel penelitian, yang meliputi teori tentang model Deep learning CNN dan CRNN serta penelitian-penelitian terkait.

BAB III : METODOLOGI PENELITIAN

Bab ini berisikan tentang data yang digunakan dan metode yang digunakan di dalam pelaksanaan penelitian.