

BAB I

PENDAHULUAN

1.1 Latar Belakang

Klasifikasi gambar merupakan bidang fundamental dalam visi komputer. Hal ini banyak diimplementasikan dalam berbagai industri. Beberapa industri populer yang banyak menggunakan klasifikasi gambar diantaranya adalah otomotif, ritel, keamanan, dan kesehatan (Tanveer et al., 2020). Selain itu, beberapa contoh penggunaan klasifikasi gambar pada bidang kesehatan yaitu dengan mengambil data gambar biomedis dari pasien dapat membantu tenaga spesialis atau ahli. Gambar tersebut dapat digunakan dan dilatih di dalam sistem pintar untuk digunakan dalam membantu menentukan penyakit berdasarkan gambar-gambar tersebut setelah dianalisis (Tchito Tchapgba et al., 2021). Contoh lain penerapan klasifikasi gambar pada bidang bisnis dapat digunakan untuk menganalisis citra satelit sehingga dapat mengkategorikan jenis bisnis yang cocok (Suvon et al., 2023). Implementasi dari klasifikasi gambar juga mencakup berbagai bidang lainnya. Dengan demikian, klasifikasi gambar merupakan bidang dasar dalam visi komputer dengan berbagai implementasi dalam beberapa bidang, seperti di bidang kesehatan yaitu klasifikasi biomedis. Secara keseluruhan, klasifikasi gambar merupakan bidang yang berhubungan dengan kecerdasan buatan, khususnya dalam subbidang pembelajaran mesin dari jaringan saraf dalam (*Deep Neural Network*). Beberapa pendekatan jaringan saraf dalam untuk mengklasifikasikan gambar diantaranya adalah arsitektur konvolusional kemudian berkembang dengan adanya

arsitektur *visual geometry group*, *selective kernel*, *convolutional block attention module*, dan *batch normalization*.

Jaringan Saraf Dalam (*Deep Neural Network/DNN*) merupakan subbidang dari pembelajaran mesin yang banyak digunakan untuk klasifikasi gambar. Dalam DNN, sejumlah besar data digunakan untuk menciptakan suatu prediksi. Dalam DNN, khususnya klasifikasi gambar, banyak menggunakan Jaringan Saraf Konvolusional (*Convolutional Neural Network/CNN*) sebagai bagian dari model jaringan saraf (Krishna & Kumar Kalluri, 2019). Meskipun CNN merupakan metode yang revolusioner untuk klasifikasi gambar, metode tersebut masih memiliki beberapa kekurangan. Pertama, CNN akan menghasilkan komputasi yang tinggi pada gambar yang besar, yang berdampak pada kebutuhan perangkat keras yang tinggi. Kedua, dalam proses konvolusi, pengaruh pada informasi secara menyeluruh akan terlewatkan karena konvolusi merupakan pemrosesan informasi secara lokal (Sun et al., 2020; Zheng et al., 2022). Ketiga, *kernel* yang dimiliki pada CNN memiliki *kernel* yang sama. Oleh karena itu, CNN memiliki kelemahan melihat objek dengan skala yang berbeda (Li et al., 2019). Dengan demikian, dapat disimpulkan bahwa DNN merupakan subbidang dari pembelajaran mesin, dan CNN merupakan metode yang banyak digunakan untuk studi kasus klasifikasi gambar. Arsitektur CNN yang sering digunakan untuk mengklasifikasikan gambar diantaranya *visual geometry group*, jaringan saraf residual, dan sebagainya.

Visual Geometry Group (VGG) merupakan arsitektur yang memanfaatkan *Convolutional Neural Network* (CNN) untuk klasifikasi gambar. VGG dikembangkan dan berhasil menyelesaikan permasalahan klasifikasi gambar dalam

kompetisi *ImageNet Challenge* 2014, dengan menempati posisi pertama dan kedua pada trek lokalisasi dan klasifikasi dengan mendapatkan eror 23.7% dengan 138 juta parameter. Arsitektur VGG memanfaatkan filter konvolusi (3x3), yang memberikan peningkatan signifikan dibandingkan konfigurasi sebelumnya. Selain itu, arsitektur ini memanfaatkan kedalaman arsitektur, dengan menggunakan hingga 16-19 lapisan untuk mencapai hasil terbaik (Simonyan & Zisserman, 2014). Meskipun memiliki prestasi yang luar biasa, arsitektur ini tetap menghadapi permasalahan yang umum dalam CNN diantaranya komputasi tinggi, keterbatasan ekstraksi fitur, dan *kernel* yang tetap. Selain itu, kedalaman arsitektur VGG menimbulkan permasalahan baru, yaitu nilai yang fluktuatif, yang menyebabkan proses pelatihan menjadi tidak efektif dan efisien. Selanjutnya, untuk memperkaya ekstraksi fitur pada CNN digunakanlah metode *attention*.

Metode *attention* merupakan teknik yang digunakan untuk menyelesaikan permasalahan pada model atau arsitektur *Convolutional Neural Network* (CNN) yang dalam atau lebar (Zheng et al., 2022). Pendekatan *attention* dapat menyelesaikan permasalahan mengenai kebutuhan komputasi yang besar dan hilangnya informasi pada CNN. *Attention* dalam visi komputer merupakan suatu metode yang diadaptasi dari cara kerja otak manusia. Penerapan metode *attention* pada berbagai bidang visi komputer telah menunjukkan peningkatan yang signifikan (Sun et al., 2020). Dengan demikian, dapat disimpulkan bahwa metode *attention* merupakan cara yang efektif untuk menangani permasalahan pada CNN. Pada penelitian ini, penggunaan *Convolutional Block Attention Module* (CBAM) digunakan untuk menyelesaikan permasalahan pada CNN.

Metode *Convolutional Block Attention Module* (CBAM) merupakan modul jaringan *feed-forward* yang mudah diintegrasikan ke dalam model berbasis *Convolutional Neural Network* (CNN). Modul ini menggabungkan pendekatan modul spasial dan kanal. Dengan desain ini, diperoleh modul yang umum, ringan, dan mudah diintegrasikan pada arsitektur berbasis CNN. Metode CBAM ini telah tervalidasi terhadap beberapa *dataset* tolok ukur seperti *Imagenet-1K*, *MS COCO*, dan lainnya. Pada penelitian Woo dkk. menunjukkan hasil eror 21.07% pada *dataset ImageNet-1K* dengan jumlah parameter 48.96 juta. Selain itu, setiap eksperimen yang dilakukan pada *dataset* tersebut mendapat peningkatan yang konsisten (Woo et al., 2018). Adapun kesenjangan dalam penelitian sebelumnya, salah satunya pada penelitian F.Wang dkk. yang menggunakan *Residual Attention Network* dengan modul *encoder-decoder* (F. Wang et al., 2017). Penelitian yang dilakukan oleh Woo dkk. menguraikan proses yang mempelajari *attention* kanal dan *attention* spasial secara terpisah. Hal ini mengakibatkan komputasi yang lebih sedikit dan *overhead parameter* yang lebih rendah, sehingga menciptakan modul yang mudah diintegrasikan dengan model berbasis CNN (Woo et al., 2018). Oleh karena itu, metode CBAM dapat membantu memperkaya ekstraksi fitur dan mudah diintegrasikan dengan CNN. Selain itu, terdapat permasalahan lain yang dialami oleh CNN yaitu ukuran *kernel* yang tetap. Untuk menyelesaikan permasalahan tersebut, pada penelitian ini digunakan metode *selective kernel*.

Metode *Selective Kernel* (SK) merupakan teknik yang digunakan untuk mengatasi permasalahan pada CNN yang memiliki kernel atau bidang persepsi tetap (Li et al., 2019). Sebagaimana telah disebutkan sebelumnya, salah satu

permasalahan utama pada CNN adalah *kernel* yang tetap, yang menyebabkan model tidak mampu mengenali objek dengan skala yang berbeda secara efektif (Li et al., 2019). Dengan arsitektur yang diajukan oleh Li dkk. mendapat hasil yang baik dengan eror 17.33% pada *dataset* CIFAR-100 dengan parameter 27.7 juta (Li et al., 2019). Dalam penelitian Szegedy dkk. terdapat kesenjangan dalam kemampuan mengagregasi informasi skala pada *kernel* CNN, seperti yang terlihat pada model *Inception* (Szegedy et al., 2016). Pada penelitian Li dkk. mengembangkan pendekatan *nonlinier* untuk mengagregasi informasi dari berbagai *kernel*, memungkinkan model beradaptasi dengan *kernel* atau bidang persepsi yang berbeda. Hasil eksperimen menunjukkan bahwa model yang dihasilkan mampu melampaui *state-of-the-art* pada *dataset* CIFAR dan *Imagenet* (Li et al., 2019). Dengan demikian metode SK dapat digunakan untuk membuat *kernel* menjadi adaptif. Selain itu, terdapat permasalahan lain dalam arsitektur VGG yaitu nilai fluktuatif. Oleh karena itu, pada laporan ini akan menggunakan metode *batch normalization*.

Batch Normalization (BN) merupakan teknik yang diterapkan pada setiap lapisan jaringan saraf dalam (*deep neural network*) untuk menormalkan aktivasi di tengah lapisan yang digunakan. Metode ini cenderung meningkatkan efektivitas dan efisiensi proses pelatihan model pembelajaran mendalam (Bjorck et al., 2018). Adapun kesenjangan dalam penelitian sebelumnya, salah satunya pada penelitian oleh Keskar dkk. ditemukan masalah divergensi pada awal pelatihan, dimana sebagian kecil aktivasi (biasanya di lapisan dalam) mengalami "eksplorisif" (Keskar et al., 2016). Solusi umum untuk masalah ini adalah dengan memperkecil laju

pembelajaran. Namun, dengan pendekatan *batch normalization*, tidak lagi diperlukan penggunaan laju pembelajaran yang kecil (Bjorck et al., 2018). Dengan demikian, teknik *batch normalization* ini mampu mengatasi masalah fluktuasi pada model yang dalam. Selain itu, terdapat masalah lain pada pengembangan yaitu waktu pengembangan.

Masalah utama dalam pembelajaran mendalam adalah lamanya proses pelatihan. Seiring dengan meningkatnya kompleksitas dan akurasi model pembelajaran mendalam, kebutuhan komputasi dan memori berkembang dengan pesat, sering kali melampaui kemampuan perkembangan perangkat keras GPU (Mittal & Vaishay, 2019). Sebagai contoh, pelatihan model ResNet-50 selama 100 *epoch* dengan *dataset Imagenet* menggunakan GPU M40 memerlukan waktu hingga 14 hari (You et al., 2017). Oleh karena itu, diperlukan perangkat keras GPU yang lebih canggih untuk dapat melatih model secara efektif dan efisien.

Penelitian yang dirujuk oleh artikel ilmiah ini adalah studi yang dilakukan oleh S.-H.Wang dkk. Penelitian tersebut mengembangkan model berbasis VGG dengan pendekatan *attention* menggunakan metode CBAM (S.-H. Wang et al., 2022). Namun, terdapat beberapa celah penelitian yang perlu ditelaah lebih lanjut dalam studi tersebut, antara lain mengenai penerapan spesifik CBAM, nilai fluktuatif pada model yang dalam, serta performa model ketika dihadapkan pada masalah klasifikasi yang lebih umum.

Berdasarkan penjabaran sebelumnya, terdapat beberapa permasalahan pada arsitektur VGG, seperti pada *Convolutional Neural Network* (CNN) yang memiliki keterbatasan dalam hal komputasi besar dan dapat menyebabkan informasi secara

menyeluruh terlewatkan, karena konvolusi merupakan pemrosesan informasi secara lokal. Dengan kata lain, CNN memiliki keterbatasan dalam mengekstraksi fitur pada gambar. Selain itu, CNN memiliki ukuran *kernel* yang tetap sehingga tidak mampu mengenali objek dengan skala yang berbeda. Selanjutnya, arsitektur VGG yang dalam menghadapi permasalahan pada nilai yang fluktuatif, yang menyebabkan proses konvergensi melambat dan meningkatkan risiko *overfitting* selama proses generalisasi. Oleh karena itu, diperlukan upaya untuk meningkatkan efisiensi proses pelatihan model yang memiliki komputasi tinggi tersebut.

Penelitian ini akan berkontribusi pada pengembangan model dengan pendekatan *attention* dan *selective kernel*, yang telah menunjukkan hasil yang baik pada jaringan *deep learning* (Li et al., 2019; F. Wang & Tax, 2016). Laporan ini akan menggunakan metode *attention* CBAM karena metode tersebut sederhana, efektif, dan merupakan modul yang umum sehingga dapat diintegrasikan dengan model CNN dengan mudah dan untuk mengatasi fluktuasi pada arsitektur dalam digunakan metode *batch normalization* (Woo et al., 2018; Bjorck et al., 2018). Selain itu, penelitian ini bertujuan untuk melakukan eksperimen terhadap model arsitektur VGG. Melalui eksperimen tersebut, diharapkan model dapat dijalankan dengan komputasi yang lebih rendah serta mencapai *state-of-the-art* jika terjadi peningkatan performa yang signifikan.

1.2 Rumusan Masalah

Pada rumusan penelitian tertuang pada Tabel 1. 1.

Tabel 1. 1 Rumusan masalah

Research Problems (RP)		Literature Supports	Research Questions (RQ)	
RP1	Model CNN mempunyai kelemahan pada informasi ekstraksi fitur yang masih terbatas.	J. Sun, J. Jiang, and Y. Liu, “An Introductory Survey on Attention Mechanisms in Computer Vision Problems,” in <i>Proceedings - 2020 6th International Conference on Big Data and Information Analytics, BigDIA 2020</i> , Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 295–300. doi: 10.1109/BigDIA51454.2020.00054.	RQ1	Bagaimana CBAM memperkaya ekstraksi fitur pada model?
RP2	Proses konvergensi yang lambat dan masih terdapatnya <i>overfitting</i> pada proses generalisasi.	Bjorck, J., Gomes, C., Selman, B., & Weinberger, K. Q. (n.d.). <i>Understanding Batch Normalization</i> .	RQ2	Bagaimana <i>batch normalization</i> dapat mempercepat proses konvergensi dan meningkatkan generalisasi pada model?
RP3	<i>Kernel</i> CNN statis tidak dapat mengenali objek secara tepat pada skala yang berbeda.	Li, X., Wang, W., Hu, X., & Yang, J. (n.d.). Selective Kernel Networks.	RQ3	Bagaimana <i>selective kernel</i> dapat mengenali objek dengan tepat pada skala yang berbeda?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah dapat diuraikan tujuan penelitian yang akan dilaksanakan:

- Melakukan modifikasi arsitektur VGG dengan penambahan modul CBAM, BN, dan SK.
- Perbandingan hasil evaluasi performa model VGG yang sudah dimodifikasi.

1.4 Manfaat Penelitian

Dengan penelitian ini diperoleh manfaat secara teoritis yaitu:

- Memperluas wawasan keilmuan khususnya dalam visi komputer.

- b) Mengembangkan metode untuk berbagai keperluan khususnya dalam implementasi keilmuan visi komputer.
- c) Secara akademik dapat memahami perilaku dan mengintegrasikan dari algoritma yang diteliti dalam hal ini model VGG atau CNN, *selective kernel*, BN, dan pendekatan *attention* CBAM.

Dengan penelitian ini diperoleh manfaat secara praktis yaitu:

- d) Secara praktis pengguna dapat memiliki metode yang lebih kokoh untuk berbagai keperluan dibidang visi komputer seperti mendeteksi komponen dalam kendaraan.
- e) Secara praktis pengguna dapat memiliki metode yang lebih akurat dan umum untuk berbagai keperluan dibidang visi komputer seperti mendeteksi penyakit pada tanaman.

1.5 Batasan Masalah

Batasan masalah penelitian ini yaitu:

- a) Penelitian ini berfokus pada pengembangan arsitektur dengan basis VGG-Net dengan metode *selective kernel*, CNN, *batch normalization*, dan CBAM.
- b) Lingkungan pengembangan penelitian ini akan digunakan pada lingkungan Google Colab dengan GPU T4.
- c) Batasan dari *dataset* CIFAR-10 berukuran 32x32 dalam 10 kelas dan CIFAR-100 berukuran 32x32 dalam 100 kelas.

1.6 Sistematika Penulisan

Laporan ini disusun dengan sistematika penulisan yang dirancang untuk memastikan fokus pada pokok permasalahan tanpa terganggu oleh isu-isu lainnya. Sistematika penulisan laporan ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini mencakup Latar Belakang, Rumusan Masalah, Tujuan Penelitian, Manfaat Penelitian, Batasan Masalah, dan Sistematika Penulisan.

BAB II LANDASAN TEORI

Bab ini berisi penjabaran penelitian-penelitian yang dirujuk serta dasar teori yang menjadi acuan dalam penelitian ini.

BAB III METODOLOGI

Bab ini berisi langkah-langkah yang dilakukan selama proses penyusunan laporan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi hasil dan pembahasan dari laporan.

BAB V KESIMPULAN DAN SARAN

Bab ini membahas kesimpulan dan saran berdasarkan hasil laporan yang telah disusun.