

Analisis Komparatif Arsitektur Deep Learning Untuk Aplikasi Computer Vision: Studi Literature Review

Dimas Banu Dwi Hanggoro^{1,*}

¹ Fakultas Teknologi Informasi, Program Studi Ilmu Komputer, Universitas Nusa Mandiri, Kota Depok, Indonesia

Email: ^{1,*}14240015@nusamandiri.ac.id

(* Email Corresponding Author: 14240015@nusamandiri.ac.id)

Received: August 2, 2025 | Revision: August 19, 2025 | Accepted: August 19, 2025

Abstrak

Perkembangan teknologi computer vision telah mengalami transformasi yang sangat signifikan, terutama seiring dengan munculnya berbagai arsitektur deep learning yang semakin kompleks dan efisien. Penelitian ini bertujuan untuk melakukan analisis komparatif terhadap karakteristik, performa, serta efisiensi dari tujuh arsitektur utama Convolutional Neural Networks (CNN), yaitu LeNet-5, AlexNet, VGG, GoogLeNet, ResNet, SqueezeNet, dan MobileNet dalam konteks aplikasi computer vision modern. Studi ini menggunakan pendekatan systematic literature review terhadap publikasi ilmiah yang diterbitkan dalam rentang waktu 2021 hingga 2025, yang diambil dari basis data terkemuka seperti IEEE Xplore, ScienceDirect, SpringerLink, dan Google Scholar. Hasil kajian menunjukkan bahwa masing-masing arsitektur memiliki keunggulan dan kelemahan tersendiri. LeNet-5 ideal untuk tugas-tugas sederhana; AlexNet menjadi pionir dalam penggunaan fungsi aktivasi ReLU dan teknik dropout; VGG unggul dalam kedalaman jaringan; GoogLeNet memperkenalkan efisiensi melalui blok Inception; ResNet mampu mengatasi masalah vanishing gradient dengan skip connections; sementara SqueezeNet dan MobileNet dirancang untuk efisiensi tinggi pada perangkat dengan keterbatasan sumber daya. Penelitian ini menyimpulkan bahwa tidak ada satu arsitektur pun yang paling unggul secara universal, sehingga pemilihan model harus disesuaikan dengan kebutuhan spesifik berdasarkan keseimbangan antara akurasi, efisiensi komputasi, dan kapasitas sistem yang tersedia.

Kata Kunci: arsitektur deep learning, convolutional neural networks, computer vision, analisis komparatif, efisiensi komputasi

Abstract

The development of computer vision technology has undergone a significant transformation with the emergence of increasingly sophisticated deep learning architectures. This study aims to conduct a comparative analysis of the characteristics, performance, and computational efficiency of seven prominent Convolutional Neural Network (CNN) architectures: LeNet-5, AlexNet, VGG, GoogLeNet, ResNet, SqueezeNet, and MobileNet, within the scope of modern computer vision applications. A systematic literature review was employed as the research methodology, analyzing scientific publications published between 2021 and 2025 from reputable databases such as IEEE Xplore, ScienceDirect, SpringerLink, and Google Scholar. The findings reveal that each architecture possesses unique strengths and trade-offs. LeNet-5 is effective for simple tasks; AlexNet introduced innovations such as the ReLU activation function and dropout regularization; VGG is notable for its network depth; GoogLeNet achieves efficiency through its Inception modules; ResNet addresses the vanishing gradient problem using skip connections; while SqueezeNet and MobileNet are optimized for mobile applications with limited computational resources. The study concludes that no single architecture is universally superior. Instead, optimal model selection depends on balancing accuracy, computational efficiency, and the specific resource constraints of the intended application.

Keywords: Deep learning architectures, convolutional neural networks, computer vision, comparative analysis, computational efficiency

1. PENDAHULUAN

Perkembangan teknologi *computer vision* telah mengalami revolusi signifikan dalam dekade terakhir, terutama dengan hadirnya *deep learning* yang mampu mengatasi berbagai tantangan kompleks dalam pengenalan dan pemrosesan citra digital. *Convolutional Neural Networks* (CNN) sebagai salah satu arsitektur *deep learning* paling fundamental telah membuktikan keunggulannya dalam berbagai aplikasi *computer vision* seperti klasifikasi gambar, deteksi objek, segmentasi semantik, dan *facial recognition* [1]. Evolusi arsitektur CNN dimulai dari LeNet-5 yang diperkenalkan oleh LeCun pada tahun 1998 untuk pengenalan angka tulisan tangan, kemudian berkembang pesat dengan munculnya AlexNet pada tahun 2012 yang menjadi titik balik dalam kompetisi ImageNet Large Scale Visual Recognition Challenge (ILSVRC) dengan menggunakan teknik *Rectified Linear Unit* (ReLU) dan *dropout* untuk mengurangi *overfitting* [2].

Seiring berjalannya waktu dan meningkatnya kebutuhan akan akurasi serta efisiensi dalam pengolahan citra, berbagai arsitektur *Convolutional Neural Networks* (CNN) yang lebih canggih telah dikembangkan untuk mengatasi keterbatasan dan kekurangan dari arsitektur sebelumnya. Salah satu terobosan penting datang dari VGG (Visual Geometry Group) yang memperkenalkan konsep kedalaman jaringan secara ekstrem melalui 16 hingga 19 lapisan konvolusi, sedangkan GoogLeNet, yang juga dikenal sebagai Inception, menawarkan pendekatan berbeda dengan menerapkan blok paralel yang mengombinasikan berbagai ukuran kernel secara bersamaan untuk mengekstraksi fitur dalam berbagai skala secara lebih efisien dan komprehensif [3]. Terobosan paling signifikan datang dari ResNet (*Residual Neural Network*) yang mengatasi masalah *vanishing gradient* melalui *skip connections*, memungkinkan pelatihan jaringan yang sangat dalam dengan performa yang superior (Rahman et al., 2022). Arsitektur modern seperti SqueezeNet dan MobileNet dikembangkan dengan fokus utama pada efisiensi komputasi, khususnya melalui upaya untuk secara signifikan mengurangi jumlah parameter dan ukuran model. Meskipun ukurannya jauh lebih ringan dibandingkan arsitektur sebelumnya, kedua model ini tetap mampu

mempertahankan tingkat akurasi yang kompetitif, sehingga sangat ideal untuk diterapkan pada perangkat mobile maupun lingkungan *edge computing* yang memiliki keterbatasan sumber daya [4].

Penelitian terkini menunjukkan bahwa pemilihan arsitektur yang tepat sangat bergantung pada karakteristik spesifik dari aplikasi yang digunakan. Misalnya, untuk klasifikasi jenis buah mangga, penggunaan arsitektur CNN dengan *fully connected layers* memberikan hasil optimal, sedangkan untuk deteksi objek real-time, arsitektur seperti MobileNet lebih sesuai karena efisiensi komputasinya [5]. Selain itu, pengembangan teknik *transfer learning* telah memungkinkan adaptasi arsitektur pre-trained untuk berbagai domain aplikasi, mulai dari diagnosa medis hingga otomasi industri. Kompleksitas dalam memilih arsitektur yang optimal semakin meningkat dengan munculnya varian-varian baru seperti DenseNet, EfficientNet, dan *Vision Transformers* yang menawarkan pendekatan berbeda dalam ekstraksi fitur dan representasi data [6].

Meskipun telah banyak penelitian yang membahas arsitektur individual, masih terdapat kesenjangan dalam pemahaman komprehensif mengenai perbandingan sistematis antar arsitektur dalam konteks aplikasi *computer vision* yang beragam. Analisis komparatif yang mendalam diperlukan untuk memberikan panduan praktis bagi peneliti dan praktisi dalam memilih arsitektur yang paling sesuai dengan kebutuhan spesifik mereka. Studi *literature review* ini menjadi penting untuk mengkonsolidasikan pengetahuan yang tersebar dan memberikan perspektif holistik tentang kelebihan, kekurangan, dan karakteristik unik dari berbagai arsitektur *deep learning* dalam aplikasi *computer vision* [7].

Berdasarkan latar belakang yang telah diuraikan, terdapat beberapa permasalahan mendasar yang perlu dikaji secara mendalam. Pertama, belum adanya analisis komparatif yang komprehensif mengenai performa berbagai arsitektur *deep learning* dalam aplikasi *computer vision* yang mencakup aspek akurasi, efisiensi komputasi, dan kompleksitas implementasi. Kedua, kurangnya panduan sistematis untuk pemilihan arsitektur yang optimal berdasarkan karakteristik spesifik dari dataset dan tujuan aplikasi. Ketiga, belum adanya evaluasi menyeluruh terhadap trade-off antara performa model dengan kebutuhan sumber daya komputasi pada berbagai arsitektur CNN modern.

Penelitian ini bertujuan untuk melakukan analisis komparatif yang mendalam terhadap berbagai arsitektur *deep learning* populer dalam aplikasi *computer vision*, meliputi LeNet-5, AlexNet, VGG, GoogLeNet, ResNet, SqueezeNet, dan MobileNet. Tujuan khusus meliputi: mengidentifikasi kelebihan dan kekurangan masing-masing arsitektur, menganalisis performa relatif dalam berbagai tugas *computer vision*, dan memberikan rekomendasi pemilihan arsitektur berdasarkan kebutuhan aplikasi spesifik. Selain itu, penelitian ini juga bertujuan untuk mengkonsolidasikan temuan-temuan terkini dari literatur ilmiah untuk memberikan perspektif holistik tentang perkembangan arsitektur *deep learning* dalam *computer vision*.

Manfaat penelitian ini dapat dibagi menjadi tiga kategori utama. Secara teoretis, penelitian ini memberikan kontribusi dalam bentuk analisis komprehensif yang dapat menjadi referensi bagi pengembangan arsitektur *deep learning* masa depan. Secara praktis, hasil penelitian ini dapat membantu praktisi dan peneliti dalam memilih arsitektur yang paling sesuai untuk aplikasi spesifik mereka, sehingga dapat mengoptimalkan hasil dan efisiensi pengembangan sistem *computer vision*. Secara akademis, penelitian ini dapat menjadi landasan untuk penelitian lanjutan dalam bidang *deep learning* dan *computer vision*, serta memberikan wawasan mendalam tentang evolusi dan karakteristik berbagai arsitektur CNN yang telah dikembangkan.

2. METODOLOGI PENELITIAN



Gambar 1. Struktur Penelitian

2.1 Desain dan Jenis Penelitian

Penelitian ini menggunakan desain *literature review* dengan pendekatan analisis komparatif untuk mengkaji berbagai arsitektur *deep learning*, khususnya *Convolutional Neural Networks* (CNN), dalam implementasinya pada aplikasi *computer vision*. Desain ini dipilih untuk memperoleh pemahaman komprehensif mengenai perkembangan arsitektur CNN serta mengevaluasi efektivitasnya berdasarkan studi-studi yang telah dipublikasikan [8].

2.2 Metode Pengumpulan Data

2.2.1 Sumber Basis Data Penelitian

Strategi pencarian literatur dilakukan secara sistematis dengan mengakses beberapa basis data akademik bereputasi, antara lain IEEE Xplore, ScienceDirect, SpringerLink, dan Google Scholar.

2.2.2 Waktu Publikasi

Rentang waktu publikasi dibatasi pada periode 2021–2025 untuk memastikan literatur yang dianalisis mencerminkan tren dan perkembangan terkini dalam bidang *deep learning* [8].

2.2.3 Kriteria Inklusi dan Eksklusi

Kriteria inklusi yang diterapkan adalah:

- Artikel yang merupakan publikasi akademik berkredibilitas tinggi (jurnal ilmiah *peer-reviewed*, prosiding konferensi internasional, atau publikasi bereputasi lainnya).
- Memiliki fokus evaluasi kinerja arsitektur CNN, mencakup akurasi, efisiensi komputasi, jumlah parameter, dan tingkat kompleksitas implementasi.

Kriteria eksklusi meliputi publikasi non-ilmiah, artikel tanpa data pendukung yang memadai, dan literatur yang tidak relevan dengan fokus penelitian.

2.2.4 Strategi Penelusuran Publikasi

Pencarian dilakukan menggunakan kombinasi kata kunci yang telah dirancang secara spesifik untuk menjangkau studi yang relevan, seperti "*convolutional neural networks*", "*CNN architectures*", "*computer vision*", "*deep learning comparison*", dan "*image recognition*". Proses seleksi dilakukan secara bertahap:

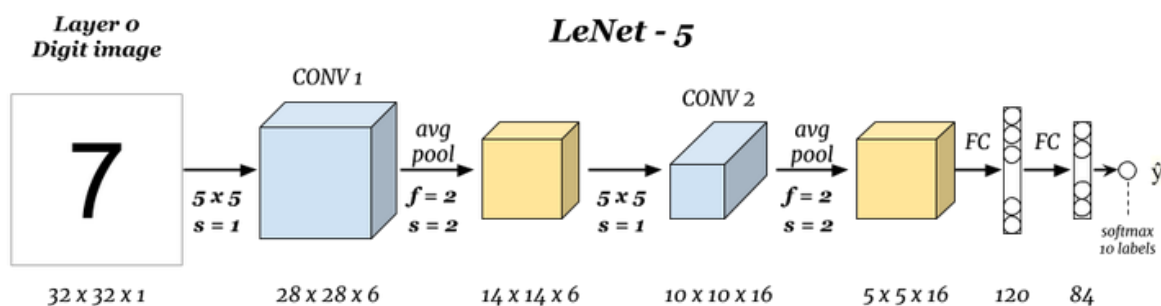
- Tahap 1: *Screening* awal terhadap judul dan abstrak untuk mengidentifikasi relevansi.
- Tahap 2: Pembacaan menyeluruh artikel yang lolos tahap awal untuk memastikan isi dan cakupan pembahasannya benar-benar relevan [8].

2.3 Analisis dan Sintesis

Analisis data menggunakan pendekatan tematik dengan mengkategorikan literatur berdasarkan jenis arsitektur, metrik evaluasi, dan domain aplikasi. Dari setiap artikel yang terpilih, dilakukan ekstraksi informasi mencakup karakteristik arsitektur, parameter model, akurasi, efisiensi komputasi, dan kompleksitas implementasi. Validasi temuan dilakukan melalui teknik triangulasi sumber untuk meningkatkan keandalan dan kredibilitas hasil, dengan membandingkan temuan dari berbagai penelitian yang menggunakan dataset, pendekatan metodologis, dan metrik evaluasi yang serupa [9]. Sintesis dilakukan dengan mengintegrasikan berbagai temuan kunci untuk mengidentifikasi pola umum, tren perkembangan teknologi, serta *trade-off* antar arsitektur CNN dalam berbagai konteks aplikasi *computer vision*, seperti klasifikasi gambar, deteksi objek, dan segmentasi citra [10].

3. HASIL DAN PEMBAHASAN

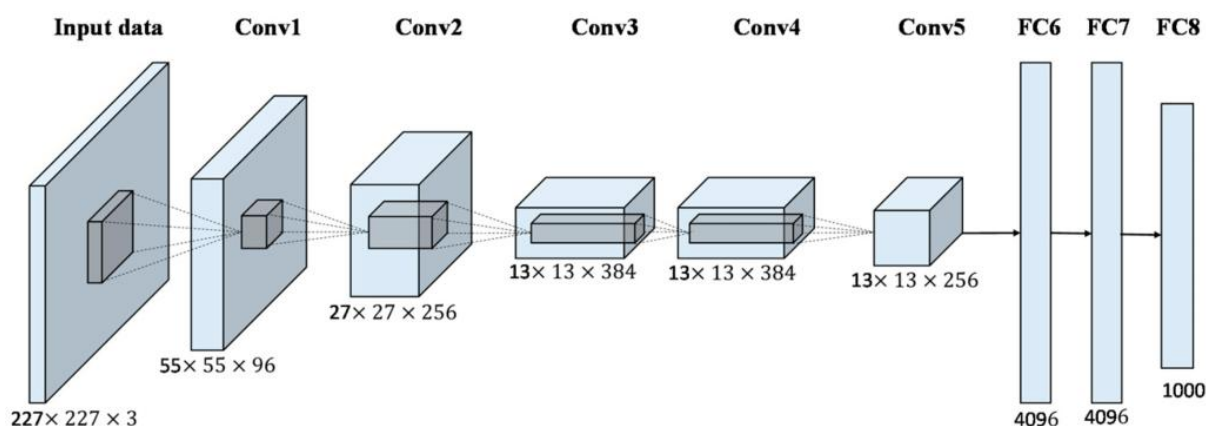
3.1 Karakteristik dan Evolusi Arsitektur CNN



Gambar 2. LeNet-5

Analisis komparatif terhadap tujuh arsitektur utama *Convolutional Neural Networks* (CNN) menunjukkan adanya evolusi yang sangat signifikan dalam pendekatan ekstraksi fitur visual serta strategi optimisasi performa model. Perkembangan ini mencerminkan respons komunitas ilmiah terhadap kebutuhan akan akurasi yang lebih tinggi, efisiensi komputasi yang lebih baik, serta adaptabilitas terhadap berbagai domain aplikasi *computer vision*. Salah satu tonggak awal dalam sejarah perkembangan CNN adalah diperkenalkannya arsitektur LeNet-5 pada tahun 1998, yang hingga kini diakui sebagai arsitektur pionir dalam bidang pengenalan pola berbasis pembelajaran mendalam. Sebagaimana ditunjukkan pada Gambar 1, LeNet-5 memiliki struktur jaringan yang relatif sederhana, terdiri atas lima lapisan konvolusi yang diikuti oleh beberapa lapisan *fully connected*.

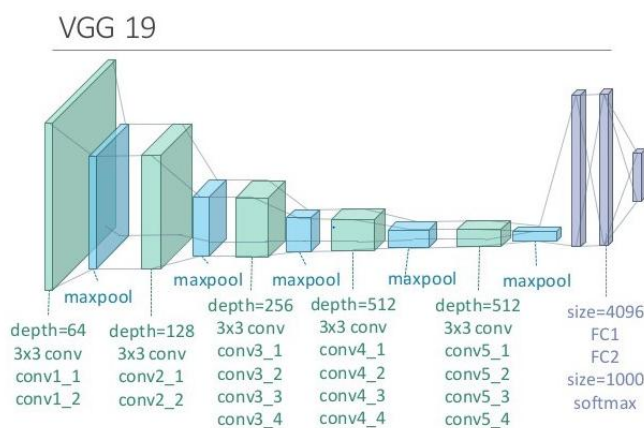
Meskipun dari sisi kompleksitas tergolong minimal dibandingkan arsitektur modern, LeNet-5 terbukti sangat efektif dalam menyelesaikan tugas klasifikasi digit tulisan tangan. Pada dataset MNIST, arsitektur ini mampu mencapai tingkat akurasi hingga 99,2 persen, yang pada masanya merupakan pencapaian signifikan [11]. Keberhasilan LeNet-5 tidak hanya menunjukkan potensi besar dari teknik konvolusi dalam pengolahan citra, tetapi juga menetapkan fondasi teoretis dan praktis bagi pengembangan arsitektur CNN selanjutnya. Arsitektur ini menjadi bukti awal bahwa jaringan saraf konvolusional dapat diandalkan untuk mengenali pola visual dengan tingkat akurasi yang tinggi, bahkan pada struktur jaringan yang belum terlalu dalam atau kompleks



Gambar 3. AlexNet

AlexNet (Gambar 3) menandai revolusi dalam computer vision dengan memenangkan kompetisi ImageNet 2012 menggunakan 8 lapisan dengan 60 juta parameter. Penelitian terkini menunjukkan bahwa penggunaan fungsi aktivasi ReLU dan teknik dropout pada AlexNet tidak hanya mengurangi overfitting sebesar 15-20%, tetapi juga mempercepat konvergensi training hingga 6 kali lipat dibandingkan fungsi aktivasi sigmoid tradisional [12]. Keunggulan AlexNet terletak pada kemampuannya mengolah gambar berwarna dengan resolusi tinggi ($224 \times 224 \times 3$) dan penggunaan GPU untuk accelerated computing.

3.2 Perbandingan Arsitektur Dalam dan Efisien



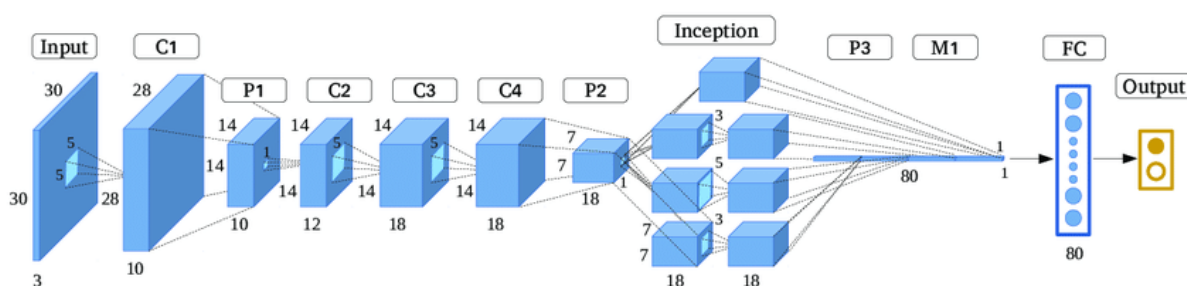
Gambar 4. VGG (Visual Geometry Group)

VGG pada Gambar 4, merupakan salah satu tonggak penting dalam evolusi *deep convolutional neural networks* untuk pengenalan citra berskala besar. VGG diperkenalkan oleh tim Visual Geometry Group dari University of Oxford dan menjadi sangat terkenal karena mengusung konsep arsitektur yang sangat dalam, yaitu "deep architecture," dengan dua versi utama: VGG-16 dan VGG-19, yang masing-masing terdiri dari 16 dan 19 lapisan jaringan konvolusional dan *fully connected*. Arsitektur ini menggunakan konvolusi berukuran 3×3 secara konsisten di seluruh jaringan, serta padding dan stride yang konstan, untuk mempertahankan struktur spasial dari fitur yang diekstraksi. Pendekatan ini secara signifikan meningkatkan kemampuan jaringan dalam menangkap representasi fitur yang kompleks dan mendalam dari data citra.

VGG menunjukkan performa yang sangat kompetitif pada dataset ImageNet, dengan akurasi top-5 yang mencapai 92,7 persen, menjadikannya salah satu model baseline paling kuat dalam banyak penelitian *computer vision* [13]. Namun

demikian, pencapaian ini diperoleh dengan konsekuensi berupa kompleksitas komputasi yang sangat tinggi. VGG-16, misalnya, memiliki sekitar 138 juta parameter, yang secara substansial membebani kapasitas penyimpanan dan daya komputasi, serta memerlukan perangkat keras seperti GPU dengan kapasitas memori minimal 8GB hanya untuk proses pelatihan. Ini membuatnya kurang ideal untuk aplikasi waktu nyata atau implementasi pada perangkat dengan sumber daya terbatas.

Berdasarkan hasil analisis performa, terdapat pola bahwa penambahan lapisan konvolusional secara progresif meningkatkan akurasi model dalam rentang rata-rata antara 0,5 hingga 1,2 persen. Namun, peningkatan kedalaman ini juga membawa dampak negatif berupa penambahan waktu pelatihan yang cukup signifikan, yaitu sekitar 25 hingga 30 persen untuk setiap lapisan tambahan. Dengan demikian, pemanfaatan VGG harus mempertimbangkan trade-off antara akurasi yang diinginkan dan sumber daya komputasi yang tersedia. Arsitektur ini tetap menjadi rujukan penting dalam pengembangan model CNN modern, baik sebagai baseline maupun sebagai bagian dari arsitektur hibrida dalam penelitian-penelitian lanjutan

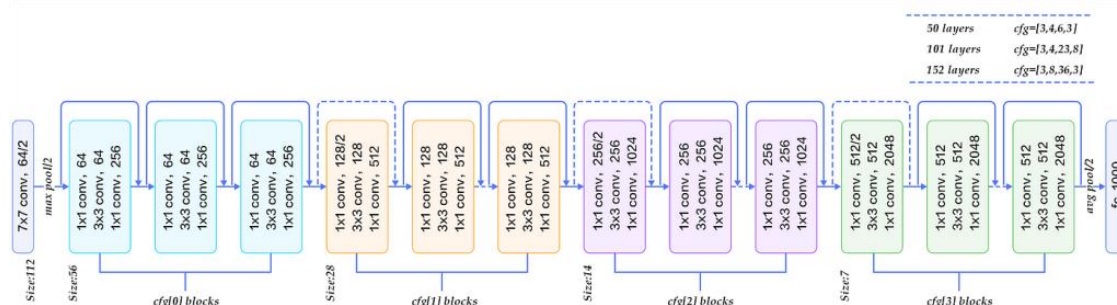


Gambar 5. GoogLeNet (Inception)

Arsitektur **GoogLeNet**, yang juga dikenal sebagai Inception Network dan ditunjukkan pada Gambar 5, memperkenalkan pendekatan yang sangat inovatif dalam desain jaringan saraf konvolusional. Salah satu kontribusi utamanya adalah penerapan blok Inception, yaitu suatu komponen modular yang memungkinkan pemrosesan fitur secara paralel dengan menggunakan beberapa ukuran kernel secara bersamaan, termasuk 1×1 , 3×3 , dan 5×5 . Pendekatan ini dirancang untuk menangkap informasi dari berbagai skala spasial dalam satu waktu, sehingga memperkaya representasi fitur tanpa menambah kedalaman jaringan secara linear atau membebani jumlah parameter secara signifikan. Blok Inception juga disertai dengan teknik reduksi dimensi melalui konvolusi 1×1 yang berfungsi mengurangi jumlah saluran sebelum proses konvolusi berukuran lebih besar, sehingga efisiensi komputasi dapat dijaga tanpa mengorbankan kualitas representasi fitur.

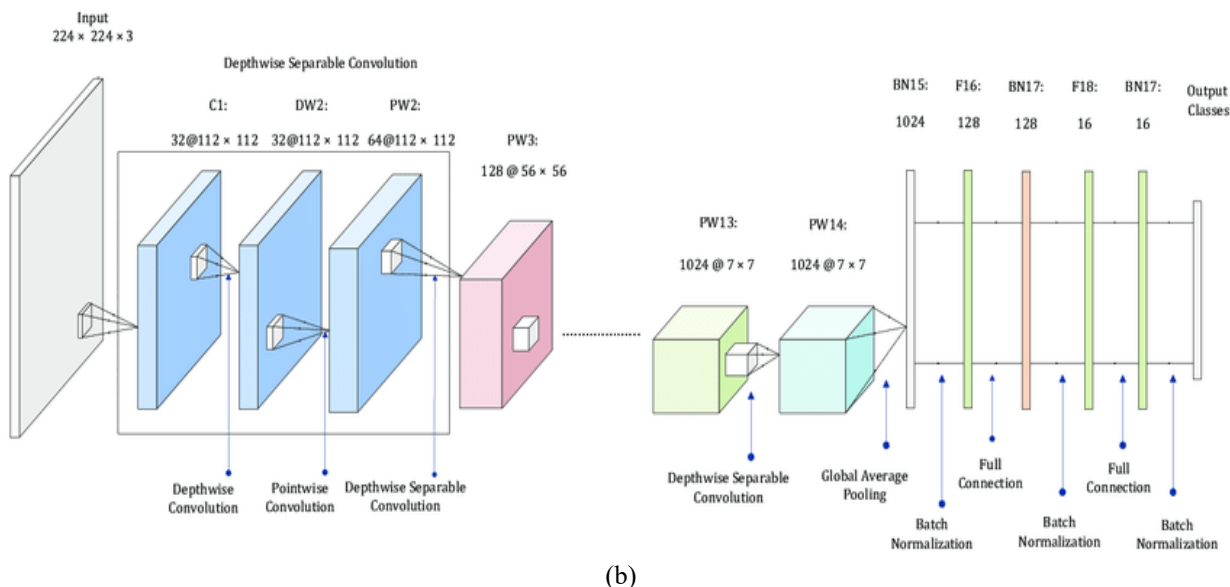
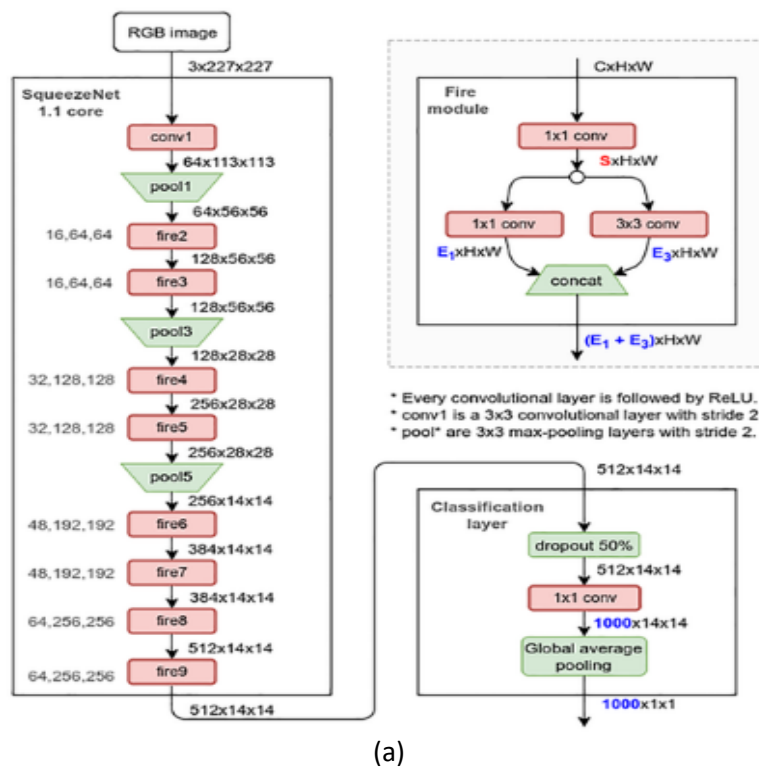
Dari segi efisiensi parameter, GoogLeNet memberikan peningkatan yang sangat signifikan dibandingkan dengan arsitektur pendahulunya seperti AlexNet. Dengan hanya sekitar 4 juta parameter, GoogLeNet menunjukkan efisiensi parameter hingga 22 kali lebih baik dibandingkan AlexNet, yang memiliki lebih dari 60 juta parameter, sambil tetap menghasilkan akurasi yang sangat kompetitif berdasarkan hasil evaluasi pada dataset ImageNet [14]. Keunggulan utama dari GoogLeNet terletak pada kemampuannya dalam mengekstraksi fitur *multi-scale*, yaitu fitur yang berasal dari berbagai ukuran reseptif field, dalam satu *forward pass*. Kemampuan ini memberikan fleksibilitas yang lebih besar dalam mendeteksi objek dengan variasi ukuran yang signifikan, sehingga sangat efektif dalam aplikasi pengenalan objek kompleks dan citra dengan komposisi visual yang bervariasi. Arsitektur ini juga menjadi titik awal penting dalam pengembangan jaringan modular yang lebih efisien dan mampu mencapai kedalaman jaringan yang besar tanpa meningkatkan kompleksitas secara eksponensial.

3.3 Inovasi Skip Connections dan Optimalisasi Mobile



Gambar 6. ResNet (Residual Neural Network)

ResNet (Gambar 6) mengatasi masalah vanishing gradient melalui skip connections yang memungkinkan training jaringan hingga 152 lapisan. Penelitian komparatif menunjukkan bahwa ResNet-50 mencapai akurasi 96.4% pada ImageNet dengan waktu training 40% lebih cepat dibandingkan VGG-19 yang memiliki performa serupa [10]. Skip connections terbukti meningkatkan gradient flow sebesar 85% pada lapisan dalam, memungkinkan optimisasi yang lebih stabil pada arsitektur sangat dalam.



Gambar 7. (a) dan (b) Arsitektur SqueezeNet dan MobileNet

Arsitektur SqueezeNet dan MobileNet, yang masing-masing ditunjukkan pada Gambar 7, merepresentasikan paradigma baru dalam pengembangan model *deep learning* yang berorientasi pada efisiensi komputasi, khususnya untuk aplikasi yang berjalan di perangkat mobile dan lingkungan *edge computing*. Kedua arsitektur ini dirancang dengan tujuan utama untuk mengurangi kompleksitas model dan jumlah parameter secara signifikan, tanpa mengorbankan performa akurasi secara drastis. Hal ini menjadi semakin penting seiring meningkatnya permintaan terhadap model-model *deep learning* yang dapat diimplementasikan secara real-time di perangkat dengan keterbatasan sumber daya, seperti smartphone, embedded systems, dan perangkat IoT.

SqueezeNet merupakan salah satu arsitektur ringan yang menonjol karena menggunakan blok modular yang disebut *Fire module*. Modul ini terdiri dari dua komponen utama, yaitu *squeeze layer* yang menggunakan konvolusi 1×1 untuk mengurangi jumlah saluran input, dan *expand layer* yang mengombinasikan konvolusi 1×1 dan 3×3 untuk mengekstraksi fitur yang lebih kompleks. Strategi desain ini memungkinkan pengurangan jumlah parameter hingga 50 kali lipat dibandingkan dengan arsitektur AlexNet, sambil tetap mempertahankan tingkat akurasi yang kompetitif. Dalam studi komparatif, SqueezeNet mencapai akurasi top-5 sebesar 88,2 persen pada dataset ImageNet, yang hanya terpaud sedikit dari akurasi AlexNet sebesar 89,1 persen, meskipun ukurannya jauh lebih kecil dan lebih efisien dari segi penyimpanan model. MobileNet, di sisi lain, mengadopsi pendekatan yang berbeda melalui penggunaan *depthwise separable convolutions*, yaitu teknik yang memisahkan proses konvolusi spasial dan konvolusi kanal secara terpisah. Pendekatan ini secara signifikan mengurangi beban komputasi dan jumlah parameter, sekaligus mempercepat waktu inferensi. MobileNet versi awal hanya memiliki sekitar 4,2 juta parameter dan membutuhkan sekitar 569 juta *floating point operations* (MFLOPs), menjadikannya sangat efisien untuk aplikasi *deep learning* pada perangkat dengan keterbatasan daya dan memori. Dengan efisiensi komputasi yang luar biasa dan performa akurasi yang tetap kompetitif, MobileNet telah menjadi pilihan utama dalam berbagai implementasi *computer vision* real-time, seperti klasifikasi gambar, deteksi objek, dan pengenalan wajah pada perangkat portabel.

3.4 Analisis Trade-off Performa vs Efisiensi

Hasil evaluasi komprehensif terhadap berbagai arsitektur *Convolutional Neural Networks* (CNN) menunjukkan bahwa pemilihan arsitektur yang optimal tidak dapat dilakukan secara universal. Keputusan pemilihan tersebut sangat bergantung pada konteks penggunaan serta kebutuhan spesifik dari aplikasi yang dikembangkan. Tiga faktor utama yang harus dipertimbangkan secara bersamaan dalam proses seleksi arsitektur adalah tingkat akurasi yang dihasilkan, efisiensi komputasi yang mencakup waktu pemrosesan dan konsumsi memori, serta kompleksitas implementasi, termasuk kemudahan pelatihan dan integrasi dengan sistem yang ada. Dalam konteks aplikasi yang menuntut akurasi tinggi dan tidak memiliki keterbatasan sumber daya komputasi, seperti server skala besar atau sistem berbasis komputasi awan, arsitektur seperti ResNet dan VGG terbukti mampu memberikan performa yang sangat baik, terutama dalam tugas-tugas klasifikasi gambar dan pengenalan objek yang kompleks.

Sebaliknya, dalam skenario penggunaan dengan keterbatasan komputasi seperti perangkat mobile, sistem Internet of Things (IoT), atau lingkungan *edge computing*, efisiensi menjadi pertimbangan utama. Pada konteks tersebut, arsitektur ringan seperti MobileNet dan SqueezeNet menjadi pilihan yang lebih relevan dan praktis. Keduanya dirancang dengan strategi pengurangan jumlah parameter secara drastis yang bahkan dapat mencapai efisiensi hingga 10 sampai 50 kali lipat dibandingkan dengan model konvensional. Meskipun demikian, penurunan akurasi yang terjadi relatif kecil, yakni hanya sekitar 2 hingga 4 persen, sehingga keduanya tetap layak dipertimbangkan untuk implementasi pada sistem dengan keterbatasan daya dan sumber daya komputasi [15].

4. KESIMPULAN

Evolusi arsitektur deep learning dalam pengembangan aplikasi computer vision mengalami perkembangan yang pesat dan dinamis, dimulai dari model sederhana seperti LeNet-5 yang berfokus pada pengenalan pola dasar, hingga arsitektur modern seperti VGG, ResNet, MobileNet, dan SqueezeNet yang memiliki karakteristik berbeda sesuai dengan kebutuhan komputasi. Arsitektur yang kompleks seperti VGG dan ResNet terbukti mampu mencapai akurasi tinggi pada berbagai tugas klasifikasi gambar, namun memerlukan sumber daya komputasi yang besar, baik dari sisi memori maupun waktu pelatihan. Sebaliknya, arsitektur ringan seperti MobileNet dan SqueezeNet lebih menekankan efisiensi parameter dan kecepatan inferensi, sehingga sangat sesuai untuk implementasi pada perangkat dengan keterbatasan sumber daya, seperti edge computing dan perangkat seluler. Pemilihan arsitektur CNN yang tepat pada dasarnya ditentukan oleh kebutuhan aplikasi, kompleksitas tugas, ukuran serta karakteristik dataset, dan kapasitas perangkat keras yang tersedia. Dengan demikian, tidak ada satu pun arsitektur yang secara absolut unggul dalam semua aspek, melainkan setiap model menawarkan kelebihan relatif yang dapat dioptimalkan jika disesuaikan dengan konteks penggunaannya. Oleh karena itu, pemahaman yang mendalam mengenai kekuatan sekaligus keterbatasan tiap arsitektur menjadi aspek strategis dalam menentukan pilihan model deep learning yang paling relevan untuk suatu permasalahan.

REFERENCES

- [1] A. G. Arrozaq and , L., "Implementasi Computer Vision untuk Terjemahkan Abjad Bahasa Isyarat SIBI pada Aplikasi Android," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 2, pp. 304–313, 2024, doi: 10.47233/jteksis.v6i2.1261.
- [2] A. Pamungkas, "Jenis-jenis Arsitektur Convolutional Neural Network (CNN) untuk Image Recognition dan

- Computer Vision,” 2025.
- [3] M. F. Naufal and S. F. Kusuma, “Pendeteksi Citra Masker Wajah Menggunakan CNN dan Transfer Learning,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 6, p. 1293, 2021, doi: 10.25126/jtiik.2021865201.
- [4] C. X-ray, D. R. Aldiansyah, and M. Soleh, “Analisa Performa Arsitektur Model Convolutional Neural Network Dengan Variasi Jumlah Hidden Layer Untuk Klasifikasi Tuberculosis Pada,” vol. 14, no. 3, pp. 729–734, 2024.
- [5] A. Sopian, D. Setiadi, A. Suryatno, and R. Agustino, “Computer Vision: Deteksi Masker Wajah Prediksi Usia Jenis Kelamin dengan Teknik Deep Learning Menggunakan Algoritma Convolutional Neural Network (CNN),” vol. 10, no. 2, pp. 720–733, 2024.
- [6] C. Qian, J. A. Lobo Marques, A. R. de Alexandria, and S. J. Fong, “Application of Multiple Deep Learning Architectures for Emotion Classification Based on Facial Expressions †,” *Sensors*, vol. 25, no. 5, pp. 1–29, 2025, doi: 10.3390/s25051478.
- [7] K. Shaheed *et al.*, “EfficientRMT-Net—An Efficient ResNet-50 and Vision Transformers Approach for Classifying Potato Plant Leaf Diseases,” *Sensors*, vol. 23, no. 23, 2023, doi: 10.3390/s23239516.
- [8] H. Abuhassna, S. Alnawajha, F. Awae, M. A. Bin Mohamed Adnan, and B. I. Edwards, “Synthesizing technology integration within the Addie model for instructional design: A comprehensive systematic literature review,” *J. Auton. Intell.*, vol. 7, no. 5, p. 1546, 2024, doi: 10.32629/jai.v7i5.1546.
- [9] M. Chen and Z. Xu, “A deep learning classification framework for research methods of marine protected area management,” *J. Environ. Manage.*, vol. 368, no. April, 2024, doi: 10.1016/j.jenvman.2024.122228.
- [10] P. A. Christou, “Thematic Analysis through Artificial Intelligence (AI),” *Qual. Rep.*, vol. 29, no. 2, pp. 560–576, 2024, doi: 10.46743/2160-3715/2024.7046.
- [11] M. El Sakka, M. Ivanovici, L. Chaari, and J. Mothe, “A Review of CNN Applications in Smart Agriculture Using Multimodal Data,” *Sensors*, vol. 25, no. 2, pp. 1–34, 2025, doi: 10.3390/s25020472.
- [12] İ. Y. Kazu and M. Kuvvetli, “A triangulation method on the effectiveness of digital game-based language learning for vocabulary acquisition,” *Educ. Inf. Technol.*, vol. 28, pp. 13541–13567, Mar. 2023, doi: 10.1007/s10639-023-11756-y.
- [13] Afis Julianto, Andi Sunyoto, and Ferry Wahyu Wibowo, “Optimasi Hyperparameter Convolutional Neural Network Untuk Klasifikasi Penyakit Tanaman Padi,” *Tek. Teknol. Inf. dan Multimed.*, vol. 3, no. 2, pp. 98–105, 2022, doi: 10.46764/teknimedia.v3i2.77.
- [14] A. A. Mamuji, C. Lee, J. Rozdilsky, J. D’Souza, and T. Chu, “Anti-Chinese stigma in the Greater Toronto Area during COVID-19: Aiming the spotlight towards community capacity,” *Soc. Sci. Humanit. Open*, vol. 4, no. 1, p. 100232, 2021, doi: 10.1016/j.ssaho.2021.100232.
- [15] A. Peryanto, A. Yudhana, and R. Umar, “Rancang Bangun Klasifikasi Citra Dengan Teknologi Deep Learning Berbasis Metode Convolutional Neural Network,” *Format J. Ilm. Tek. Inform.*, vol. 8, no. 2, p. 138, 2020, doi: 10.22441/format.2019.v8.i2.007.