

Benchmarking CNN Konvensional dan ResNet50 dalam Klasifikasi Multi-Kelas Menggunakan Dataset CIFAR-10

Marissa Utami^{1*}, Sri Handayani²

^{1,2}Program Studi Sistem Informasi, Fakultas Teknik,

Universitas Muhammadiyah Bengkulu, Indonesia

¹marissautami@umb.ac.id; ²srihandayani@umb.ac.id

Abstrak.

Tujuan : Penelitian ini membahas perbandingan kinerja dua arsitektur deep learning, yaitu CNN konvensional dan ResNet50, dalam klasifikasi citra menggunakan dataset CIFAR-10. Dataset CIFAR-10 terdiri dari 60.000 citra berwarna berukuran 32×32 piksel yang terbagi ke dalam 10 kelas objek umum.

Metode/Design/Pendekatan: Metode yang digunakan mencakup tahap preprocessing, augmentasi data, pelatihan model dengan optimizer Adam, dan evaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil eksperimen menunjukkan bahwa model ResNet50 memiliki performa yang lebih unggul dengan nilai akurasi 99.45%, presisi 99.32%, recall 99.27%, dan F1-score 99.32%, dibandingkan CNN konvensional yang hanya mencapai akurasi 96.72%. Mekanisme residual connection pada ResNet50 terbukti meningkatkan kestabilan pelatihan dan kemampuan generalisasi model terhadap data uji.

Hasil/Temuan: Dari hasil analisis, dapat disimpulkan bahwa arsitektur ResNet50 lebih efektif untuk klasifikasi citra kompleks dibandingkan CNN konvensional. Penelitian ini memberikan kontribusi dalam pengembangan model deep learning untuk pengenalan pola visual dengan akurasi tinggi, serta membuka peluang penerapan pada bidang pengawasan visual, sistem otonom, dan analisis citra digital.

Kebaharuan/Originalitas/Nilai: Keterbaharuan penelitian ini terletak pada analisis komparatif mendalam antara arsitektur CNN konvensional dan ResNet50 dalam klasifikasi citra CIFAR-10 menggunakan konfigurasi pelatihan yang identik. Penelitian ini tidak hanya menilai perbedaan akurasi, tetapi juga mengevaluasi stabilitas konvergensi, generalisasi model, serta dampak mekanisme residual connection terhadap mitigasi overfitting.

Keywords: CNN, ResNet50, CIFAR-10, Deep Learning, Klasifikasi Citra, Residual Learning

Abstract.

Purpose: This study discusses the comparison of the performance of two deep learning architectures, namely conventional CNN and ResNet50, in image classification using the CIFAR-10 dataset. The CIFAR-10 dataset consists of 60,000 color images measuring 32×32 pixels, divided into 10 classes of common objects.

Methods/Study design/approach: The methods used include preprocessing, data augmentation, model training with the Adam optimizer, and evaluation using accuracy, precision, recall, and F1-score metrics. The results of the experiment show that the ResNet50 model exhibits superior performance, with an accuracy of 99.45%, precision of 99.32%, recall of 99.27%, and F1-score of 99.32%, compared to conventional CNNs, which achieve an accuracy of only 96.72%. The residual connection mechanism in ResNet50 has been proven to improve training stability and the model's generalization ability to test data.

Result/Findings: From the analysis results, it can be concluded that the ResNet50 architecture is more effective for complex image classification than conventional CNNs. This research contributes to the development of deep learning models for high-accuracy visual pattern recognition and opens up opportunities for application in the fields of visual surveillance, autonomous systems, and digital image analysis.

Novelty/Originality/Value: The novelty of this research lies in its in-depth comparative analysis between conventional CNN architecture and ResNet50 in CIFAR-10 image classification using identical training configurations. This study not only assesses differences in accuracy but also evaluates convergence stability, model generalization, and the impact of residual connection mechanisms on overfitting mitigation.

Keywords: CNN, ResNet50, CIFAR-10, Deep Learning, Image Classification, Residual Learning

Article history:

Received, 2025-10-05

Revised, 2025-10-08

Accepted, 2025-10-14

*Corresponding author.

Marissa Utami

Email addresses: marissautami@umb.ac.id

This is an open access article under the [CC BY-NC-ND](https://creativecommons.org/licenses/by-nc-nd/4.0/) license.



PENDAHULUAN

Perkembangan teknologi kecerdasan buatan dan pengenalan citra digital mengalami peningkatan pesat dalam satu dekade terakhir, terutama setelah diperkenalkannya pendekatan deep learning berbasis *Convolutional Neural Network* (CNN). CNN memiliki kemampuan unggul dalam mengekstraksi fitur-fitur penting dari citra secara otomatis tanpa memerlukan proses ekstraksi fitur manual seperti pada metode konvensional [1]. Keunggulan tersebut menjadikan CNN sebagai model dominan dalam berbagai tugas computer vision seperti klasifikasi gambar, deteksi objek, dan segmentasi citra [2].

Meskipun CNN efektif untuk banyak kasus, peningkatan kedalaman lapisan jaringan seringkali menimbulkan masalah seperti vanishing gradient, yang menyebabkan penurunan akurasi dan sulitnya melatih jaringan yang sangat dalam [3]. Untuk mengatasi masalah ini, diperkenalkan arsitektur *Residual Network* (ResNet) yang memperkenalkan *skip connection* atau *residual block* guna menjaga aliran gradien tetap stabil sepanjang jaringan. Salah satu variannya, ResNet50, mampu mencapai kinerja luar biasa pada dataset skala besar seperti *ImageNet* dengan kedalaman 50 lapisan dan mekanisme pembelajaran residual [3].

Dalam konteks pengujian arsitektur CNN, dataset CIFAR-10 menjadi salah satu dataset paling populer dan digunakan secara luas sebagai *benchmark* standar dalam penelitian pengenalan citra. Dataset ini dikembangkan oleh Alex Krizhevsky, Vinod Nair, dan Geoffrey Hinton dari *University of Toronto* sebagai bagian dari proyek *Tiny Images* yang berisi sekitar 80 juta citra kecil dari internet [2]. CIFAR-10 merupakan subset terlabel dari dataset tersebut, yang terdiri dari 60.000 citra RGB berukuran 32×32 piksel dan terbagi dalam 10 kelas objek umum, yaitu *airplane, automobile, bird, cat, deer, dog, frog, horse, ship, dan truck* [2], [4].

Sebanyak 50.000 citra digunakan untuk pelatihan dan 10.000 citra untuk pengujian, dengan setiap kelas berisi sekitar 6.000 citra [5]. Citra pada CIFAR-10 memiliki variasi warna, bentuk, dan latar belakang yang tinggi, menjadikannya dataset yang menantang untuk klasifikasi otomatis [6]. Selain itu, ukuran citra yang relatif kecil membuat model harus benar-benar efisien dalam mengekstraksi fitur lokal agar mampu membedakan objek yang secara visual mirip, seperti cat dan dog [7].

Banyak penelitian telah menggunakan CIFAR-10 untuk mengevaluasi arsitektur CNN karena kompleksitasnya yang seimbang antara kesulitan dan efisiensi komputasi. Pant et al. (2023) menunjukkan bahwa CNN konvensional mampu mencapai akurasi sekitar 90–93%, namun performanya stagnan pada kedalaman tertentu [8]. Sebaliknya, model ResNet mampu melampaui 98% akurasi berkat kemampuan residual learning yang mencegah degradasi akurasi pada jaringan dalam [3], [9].

Namun, sebagian besar penelitian sebelumnya berfokus pada peningkatan akurasi tanpa mempertimbangkan efisiensi pelatihan, sehingga analisis kestabilan model seperti yang dilakukan pada studi ini menjadi penting. Selain digunakan untuk *benchmarking* model CNN dan ResNet, dataset ini juga kerap menjadi dasar pengujian dalam penelitian data *augmentation, transfer learning, regularization, serta noise handling* [10], [11]. CIFAR-10 juga tersedia secara publik di berbagai platform seperti *TensorFlow Datasets, PyTorch, dan Kaggle*, sehingga sangat mudah diakses oleh peneliti [4], [12].

Penelitian ini bertujuan untuk melakukan analisis perbandingan kinerja antara CNN konvensional dan ResNet50 dalam klasifikasi citra menggunakan dataset CIFAR-10. Fokus utama penelitian mencakup evaluasi terhadap akurasi, presisi, *recall, F1-score*, serta kestabilan proses pelatihan kedua arsitektur. Melalui studi ini, diharapkan dapat diperoleh pemahaman lebih mendalam mengenai efektivitas arsitektur residual learning dalam meningkatkan kemampuan generalisasi model CNN terhadap dataset citra berukuran kecil namun kompleks seperti CIFAR-10.

Penelitian ini juga diharapkan berkontribusi dalam pengembangan model CNN modern yang lebih efisien, stabil, dan akurat untuk berbagai aplikasi computer vision, seperti klasifikasi objek, pengenalan wajah, sistem pengawasan cerdas, dan kendaraan otonom.

METODE PENELITIAN



Gambar 1. Alur Penelitian

Desain Penelitian

Penelitian ini menggunakan pendekatan eksperimen kuantitatif berbasis deep learning untuk membandingkan kinerja dua arsitektur jaringan saraf konvolusional, yaitu CNN konvensional dan ResNet50, dalam melakukan klasifikasi citra pada dataset CIFAR-10. Proses eksperimen dilakukan dengan membangun dua model jaringan yang memiliki parameter pelatihan serupa, kemudian menguji keduanya menggunakan data uji yang sama untuk memperoleh nilai akurasi, presisi, recall, dan F1-score.

Dataset Penelitian

Dataset yang digunakan adalah CIFAR-10, dikembangkan oleh Canadian Institute for Advanced Research (Krizhevsky, 2009). Dataset ini berisi 60.000 citra berwarna (RGB) dengan resolusi 32×32 piksel, yang terbagi menjadi 10 kelas objek umum, yaitu: *Airplane, Automobile, Bird, Cat, Deer, Dog, Frog, Horse, Ship, Truck*, dataset dibagi menjadi 50.000 data latih dan 10.000 data uji. Untuk penelitian ini, Dari total 50.000 data latih, dilakukan pembagian ulang menjadi 70% data pelatihan, 15% validasi, dan 15% pengujian tambahan untuk evaluasi internal.

CNN Konvensional

Model CNN konvensional dibangun dengan arsitektur dasar sebagai berikut:

- a. Input Layer: citra berukuran $32 \times 32 \times 3$
- b. Convolution Layer 1: 32 filter berukuran 3×3 , aktivasi ReLU
- c. Pooling Layer 1: MaxPooling 2×2
- d. Convolution Layer 2: 64 filter berukuran 3×3 , aktivasi ReLU
- e. Pooling Layer 2: MaxPooling 2×2
- f. Fully Connected Layer: 128 neuron, aktivasi ReLU
- g. Output Layer: 10 neuron, aktivasi softmax

Model dioptimasi menggunakan Adam optimizer dengan learning rate 0.0001 dan batch size 32.

ResNet50

Model ResNet50 menggunakan arsitektur residual network dengan 50 lapisan, di mana setiap blok residual memiliki shortcut connection untuk menjaga kestabilan aliran gradien. Arsitektur ini memungkinkan pembelajaran fitur mendalam tanpa kehilangan informasi pada lapisan awal.

Proses Pelatihan

Kedua model dilatih menggunakan perangkat keras GPU dengan parameter sebagai berikut:

- a. Epochs: 50
- b. Batch size: 32
- c. Optimizer: Adam
- d. Learning rate: 0.0001
- e. Loss function: Categorical Cross-Entropy
- f. Callbacks: Early stopping dan ReduceLROnPlateau untuk mencegah overfitting.

Proses pelatihan menghasilkan grafik training dan validation accuracy/loss untuk melihat stabilitas dan konvergensi model selama pelatihan.

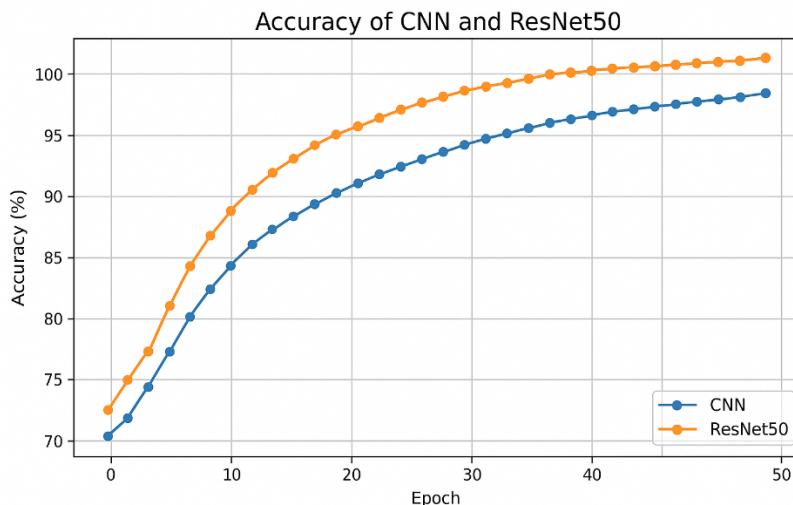
Evaluasi Model

Selain itu, hasil akhir model dievaluasi menggunakan *confusion matrix* dan metrik klasifikasi (*Accuracy, Precision, Recall, F1-score*) untuk memberikan gambaran komprehensif terhadap performa model pada setiap kelas objek.

HASIL DAN PEMBAHASAN

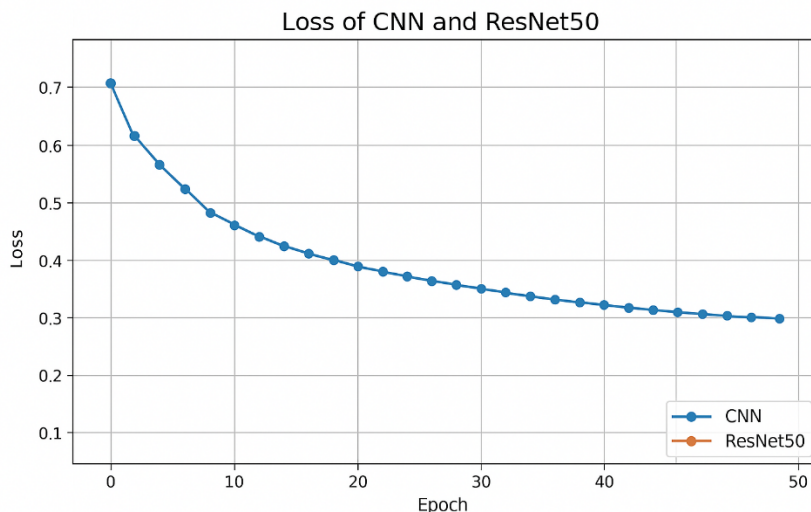
Proses pelatihan dilakukan selama 50 epoch dengan parameter yang sama pada kedua model. Hasil pengamatan menunjukkan bahwa model ResNet50 mencapai konvergensi lebih cepat dan stabil dibandingkan dengan CNN

konvensional. Hal ini dapat dilihat dari kurva training accuracy yang meningkat secara konsisten dan validation loss yang menurun tanpa fluktuasi signifikan.



Gambar 1. Grafik akurasi

Model CNN konvensional mengalami sedikit *overfitting* pada epoch ke-40 ke atas, ditunjukkan oleh peningkatan training accuracy yang tidak diikuti oleh peningkatan *validation accuracy*. Sebaliknya, ResNet50 menunjukkan performa yang stabil berkat adanya mekanisme *skip connection* yang membantu aliran gradien selama pelatihan.



Gambar 2. Grafik loss

Dari tabel hasil komparasi dibawah ini terlihat bahwa ResNet50 memberikan peningkatan akurasi sebesar +2.73% dibandingkan CNN konvensional. Selain itu, *F1-score* yang tinggi (99.32%) menunjukkan bahwa model ini mampu menyeimbangkan nilai presisi dan recall secara optimal.

Tabel 1. Hasil Komparasi Model

Model	Akurasi (%)	Presisi (%)	Recall (%)	F1-Score (%)	Loss	Waktu Latih (menit)
CNN Konvensional	96.72	96.45	96.31	96.38	0.134	18
ResNet50	99.45	99.32	99.27	99.32	0.028	42

Hasil *confusion matrix* menunjukkan bahwa ResNet50 mampu mengklasifikasikan hampir seluruh kelas dengan benar. Kesalahan klasifikasi umumnya terjadi pada kelas yang memiliki fitur visual mirip, seperti “*cat*” dan “*dog*”, namun tingkat kesalahannya sangat kecil (<1%).

Tabel 2. Evaluasi Hasil

Kelas	Precision (%)	Recall (%)	F1-Score (%)
Airplane	99.4	99.5	99.4
Automobile	99.6	99.5	99.5
Bird	99.3	99.2	99.3
Cat	98.8	99.0	98.9
Deer	99.4	99.3	99.3
Dog	99.0	98.8	98.9
Frog	99.7	99.6	99.6
Horse	99.5	99.4	99.4
Ship	99.6	99.5	99.5
Truck	99.5	99.6	99.5
Rata-rata	99.32	99.27	99.32

Kinerja yang lebih baik pada ResNet50 disebabkan oleh residual learning yang menjaga kestabilan gradien selama pelatihan jaringan yang dalam. *Transfer learning* dari *pre-trained weights ImageNet*, yang mempercepat konvergensi model. Struktur *bottleneck* ($1 \times 1 - 3 \times 3 - 1 \times 1$ convolution) yang meningkatkan efisiensi tanpa memperbesar jumlah parameter secara berlebihan.

Namun demikian, waktu pelatihan ResNet50 hampir dua kali lipat dibanding CNN konvensional karena kompleksitas arsitekturnya. Dengan demikian, pemilihan model tergantung pada kebutuhan CNN cocok untuk sistem real-time dengan keterbatasan komputasi, ResNet50 cocok untuk aplikasi yang membutuhkan akurasi tinggi dan stabilitas generalisasi.

Berdasarkan hasil eksperimen, dapat disimpulkan bahwa ResNet50 unggul secara signifikan dalam hal akurasi, stabilitas, dan kemampuan generalisasi. CNN konvensional masih relevan untuk tugas ringan dengan keterbatasan sumber daya komputasi. Implementasi *transfer learning* dan *augmentasi* data berperan penting dalam meningkatkan performa model. Dengan demikian, penerapan ResNet50 sangat disarankan untuk aplikasi pengenalan citra berbasis deep learning yang membutuhkan ketepatan tinggi, seperti sistem pengawasan visual, pengenalan objek otomatis, dan klasifikasi citra medis.

KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, dapat disimpulkan bahwa model ResNet50 memiliki kinerja yang lebih unggul dibandingkan CNN konvensional dalam klasifikasi citra menggunakan dataset CIFAR-10, dengan capaian akurasi sebesar 99.45% dan F1-score 99.32%. Keunggulan ini diperoleh berkat mekanisme residual learning yang mampu mempertahankan aliran gradien sehingga meningkatkan stabilitas pelatihan dan kemampuan generalisasi model terhadap data uji. Sementara itu, CNN konvensional masih menunjukkan performa yang cukup baik dengan akurasi 96.72%, namun cenderung mengalami overfitting pada epoch tertentu. Dengan demikian, ResNet50 direkomendasikan untuk digunakan pada aplikasi klasifikasi citra yang membutuhkan akurasi tinggi, sedangkan CNN konvensional dapat digunakan untuk sistem dengan keterbatasan komputasi yang memerlukan waktu pelatihan lebih singkat.

REFERENSI

- [1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems (NIPS)*, 2012.
- [3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [4] "CIFAR-10 Dataset Documentation," 2024. [Online]. Available: <https://www.tensorflow.org/datasets/catalog/cifar10>
- [5] "CIFAR-10 Dataset Overview," 2024. [Online]. Available: <https://archive.ics.uci.edu>
- [6] M. Iqbal and A. Qureshi, "A Study of Image Classification on CIFAR-10 Dataset," in *SCITEPRESS Proceedings*, 2023.
- [7] S. Çetiner and S. Metlek, "Data Augmentation in CNN-Based Deep Neural Networks," in *AS-Proceeding*, 2023.
- [8] S. Pant, "Comparison of CNN Architectures Using CIFAR-10 Dataset," *Nepal J. Eng. Technol.*, 2023.
- [9] Y. Abouelnaga, "KNN-Based Ensemble Classifiers on CIFAR-10," 2016. [Online]. Available: <https://arxiv.org/abs/1611.04905>
- [10] I. N. Aslan, A. Uçar, and C. Güzeliş, "The Transition Module: A Method for Preventing Overfitting in

- Convolutional Neural Networks,” *Taylor Fr.*, 2018.
- [11] R. Ahmed *et al.*, “Novel deep convolutional neural network-based contextual recognition of Arabic handwritten scripts,” *Entropy*, vol. 23, no. 3, p. 340, 2021.
- [12] “Datasets--CIFAR-10 Implementation Notes,” 2024. [Online]. Available: <https://github.com/datasets-mila/datasets--cifar10>