

Analisis Metode Augmentasi Data untuk Klasifikasi Objek pada Dataset CIFAR-10

Nanda Tri Hakiki¹, Febryanti Sthevanie², Kurniawan Nur Ramadhani³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

⁴Divisi Digital Service PT Telekomunikasi Indonesia

¹nandahakiki@students.telkomuniversity.ac.id, ²sthevanie@telkomuniversity.ac.id,

³kurniawanr@telkomuniversity.ac.id

Abstrak

Convolutional Neural Network (CNN) merupakan salah satu metode klasifikasi objek yang memiliki akurasi yang tinggi, namun akurasi tinggi ini bergantung pada kualitas dan banyaknya data latih sehingga penggunaan *big data* menjadi esensial. *Big data* juga umumnya memerlukan biaya yang besar, alternatif lainnya adalah menggunakan dataset biasa namun akurasi yang dihasilkan tidak sebaik ketika menggunakan *big data*. Augmentasi menggunakan parameter dan teknik yang tepat pada dataset dapat meningkatkan akurasi dari model yang dibangun. Salah satu dataset yang populer digunakan untuk mengembangkan model pembelajaran mesin dan CNN adalah CIFAR-10. Pada penelitian ini dilakukan eksplorasi metode augmentasi data untuk meningkatkan performa model klasifikasi menggunakan dataset CIFAR-10. Metode peningkatan data seperti *rotation*, *flip*, *shift*, dan *zoom* digunakan untuk meningkatkan jumlah dan variasi dataset. Hasil percobaan menunjukkan bahwa metode augmentasi data dapat meningkatkan akurasi model klasifikasi objek sebesar 7.73%.

Kata kunci: CNN, klasifikasi objek, augmentasi data, CIFAR-10.

Abstract

Convolutional neural networks (CNNs) are one of the most accurate object classification methods, but their accuracy depends on the quality and quantity of training data. Therefore, the use of big data is essential. However, big data also generally requires a large cost. An alternative is to use a regular dataset, but the accuracy produced is not as good as when using big data. Augmentation using the right parameters and techniques on the dataset can improve the accuracy of the model that is built. One of the most popular datasets used to develop machine learning models and CNNs is CIFAR-10. In this research, an exploration of data augmentation methods was conducted to improve the performance of the classification model using the CIFAR-10 dataset. Data enhancement methods such as rotation, flip, shift, and zoom were used to increase the number and variation of the dataset. The results of the experiment showed that data augmentation methods can improve the object classification model's accuracy by 7.73%.

Keywords: CNN, object classification, data augmentation, CIFAR-10.

1. Pendahuluan

Latar Belakang

Dalam bidang visi komputer dan pembelajaran mesin, klasifikasi objek adalah tugas fundamental yang melibatkan identifikasi kelas objek dalam sebuah gambar. Convolutional Neural Networks (CNN) telah terbukti sangat efektif dalam banyak tugas Visi Komputer, seperti pengenalan objek, deteksi objek, dan klasifikasi gambar. Namun, jaringan ini sangat bergantung pada *big data* untuk menghindari overfitting. Overfitting merujuk pada fenomena ketika sebuah jaringan mempelajari fungsi dengan varians yang sangat tinggi sehingga dapat memodelkan data pelatihan dengan sempurna. Sayangnya, banyak domain aplikasi tidak memiliki akses ke big data [1].

Salah satu metode yang digunakan dalam mengatasi tantangan ini adalah augmentasi data. Augmentasi data melibatkan pembuatan variasi data pelatihan melalui transformasi seperti rotasi, flipping, scaling, dan lainnya. Dengan cara ini, model dapat 'melihat' lebih banyak contoh dan variasi dari setiap kelas, yang berpotensi meningkatkan akurasi klasifikasi. Namun, ada banyak teknik augmentasi data yang berbeda, dan tidak semua teknik mungkin sama efektifnya untuk setiap tugas atau dataset. Hasilnya pun tidak hanya berbeda dari segi akurasi, ada beberapa metode yang memiliki akurasi sedikit lebih rendah namun dapat dicapai dengan biaya

komputasi yang jauh lebih rendah [2][3]. Oleh karena itu, penelitian ini bertujuan untuk mengeksplorasi dan mengoptimalkan proses augmentasi data untuk klasifikasi objek dan dataset yang digunakan adalah CIFAR-10.

Ada banyak penelitian yang telah dilakukan tentang augmentasi data untuk klasifikasi objek. Salah satu penelitian yang paling terkenal adalah penelitian yang dilakukan oleh Krizhevsky et al. (2012). Penelitian ini menunjukkan bahwa augmentasi data dapat meningkatkan akurasi klasifikasi CNN pada dataset ImageNet sebesar 10%. Penelitian lain yang juga menunjukkan efektivitas augmentasi data untuk klasifikasi objek adalah penelitian yang dilakukan oleh Cubuk, Zoph, Mane, Vasudevan, Le (2019) dan Mueller, Hutter (2021) [4][5].

Topik dan Batasannya

Pada penelitian ini kami menggunakan dataset dari Canadian Institute for Advanced Research (CIFAR) dari penelitian sebelumnya [2]. Spesifikasi komputasi yang digunakan pada penelitian ini menggunakan T4 Tensor Processing Unit (TPU).

Tujuan

Penelitian ini bertujuan untuk mengeksplorasi dan analisis metode augmentasi sehingga dapat menghasilkan teknik augmentasi dan menentukan parameter mana yang dapat digunakan untuk menghindari *overfitting* dan meningkatkan akurasi klasifikasi kelas pada dataset CIFAR-10 menggunakan metode CNN.

Organisasi Tulisan

Susunan untuk penelitian ini terdiri dari lima bagian yaitu:

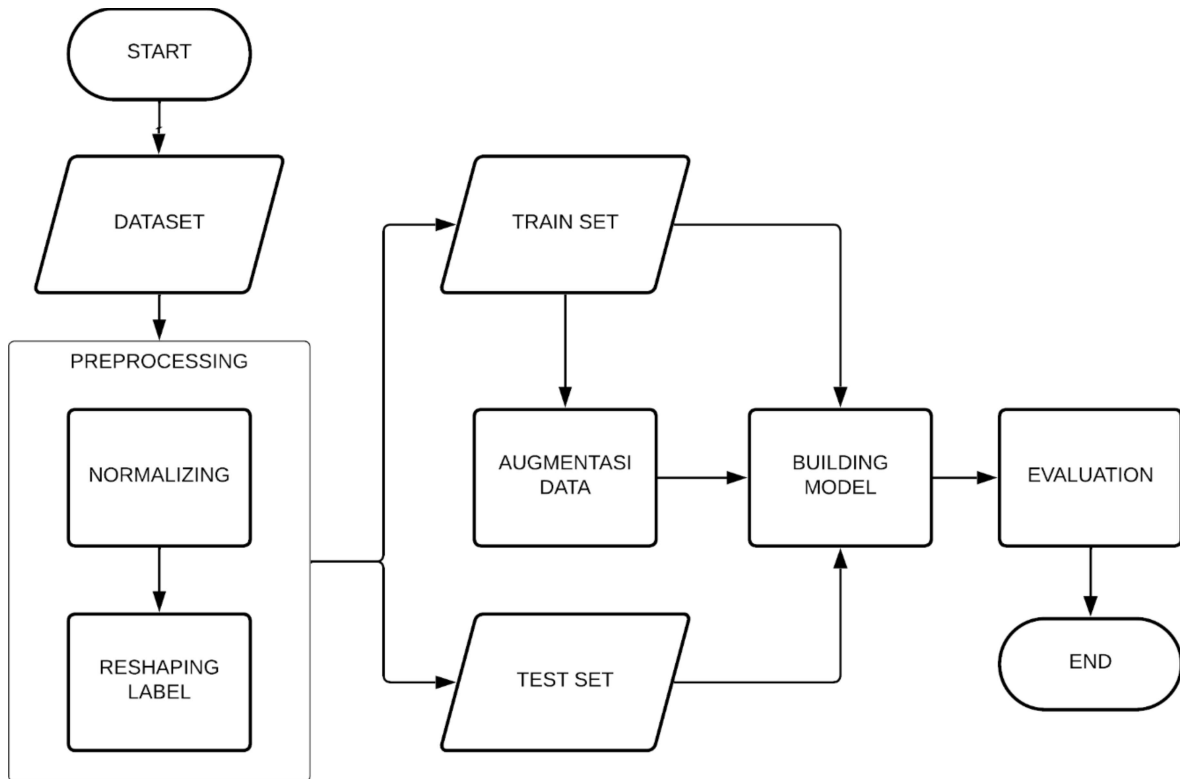
1. Bagian pertama membahas pendahuluan yang berisi latar belakang, topik dan batasannya, tujuan, dan organisasi tulisan.
2. Bagian kedua membahas studi terkait.
3. Bagian ketiga membahas sistem yang dibangun dari mulai deskripsi sistem, preprocessing dataset, augmentasi data, sampai model arsitektur yang digunakan.
4. Bagian ketiga membahas evaluasi, berisi penjelasan dan analisis dari hasil sistem yang dibangun.
5. Bagian kelima membahas kesimpulan dari penelitian ini, menjelaskan apakah penelitian yang dilakukan mencapai hasil yang diinginkan dan apa saja yang dapat dikembangkan untuk penelitian selanjutnya.

2. Studi Terkait

Augmentasi data mencakup serangkaian teknik yang meningkatkan ukuran dan kualitas data latih sehingga model deep learning yang lebih baik dapat dibuat menggunakan data tersebut. Algoritma peningkatan gambar yang tercakup meliputi transformasi geometris, peningkatan ruang warna, filter kernel, pencampuran gambar, penghapusan acak, peningkatan ruang fitur, adversarial training, generative adversarial networks, neural style transfer, dan meta learning [1]. Dua metode augmentasi data yang paling populer dan efektif dalam pelatihan CNN adalah random flipping dan random cropping [6]. Sementara untuk menghindari overfitting pada arsitektur lain seperti ResNet, Wide ResNet, PyramidNet, and ResNeXt dapat menggunakan metode ShakeDrop. Metode tersebut diusung Yamada dkk (2018) [7]. Meskipun untuk ImageNet hanya mencapai akurasi sekitar 79% tetapi pada penelitiannya didemonstrasikan beberapa kondisi dimana ShakeDrop dapat bekerja dengan baik dan bisa lebih efisien dibanding Shake-Shake. Kabir (2023) [3]. mengusulkan reduksi kelas latar belakang untuk mencapai generalisasi yang lebih baik dengan komputasi yang lebih rendah dibandingkan dengan multitask learning untuk membantu peneliti dan organisasi dengan daya komputasi terbatas. Adapun selain menggunakan CNN dan variasinya. Dosovitskiy dkk (2021) [2]. mengeksplorasi arsitektur vision transformers (ViT) untuk pengenalan citra. Ketika dilakukan pra-latih pada sejumlah besar data dan dipindahkan ke beberapa benchmark pengenalan citra berukuran menengah atau kecil (ImageNet, CIFAR-100, VTAB, dll.), ViT mencapai hasil yang mendekati CNN dan hanya membutuhkan sumber daya komputasi yang jauh lebih sedikit untuk dilatih. Zhong dkk (2020) [6]. mengenalkan metode augmentasi data Random Erasing. Penghapusan wilayah persegi secara acak ini menghasilkan data latih yang rentan dari penutupan objek dan juga overfitting. Meskipun sederhana, random erasing melengkapi teknik augmentasi data yang umum digunakan seperti cropping dan random flipping, dan telah terbukti meningkatkan kinerja CNN pada tugas klasifikasi gambar, deteksi objek, dan mengidentifikasi ulang orang. Selain itu Liu dkk (2022) [8]. juga mengusulkan RandomMix, metode yang dapat meningkatkan keandalan model, tidak memerlukan banyak perhitungan tambahan, dan mudah dimasukkan ke dalam pipeline pelatihan. Eksperimen pada dataset CIFAR-10/100, Tiny-ImageNet, ImageNet, dan Google Speech Commands menunjukkan bahwa RandomMix mencapai kinerja lebih baik daripada metode data augmentation campuran state-of-the-art lainnya [9].

3. Sistem yang Dibangun

Model yang dibuat pada tugas akhir ini adalah model augmentasi dataset agar dapat memperkaya data latih sehingga dapat meningkatkan akurasi dan menghindari overfitting. Skema Pembangunan model dapat dilihat pada Gambar 1.

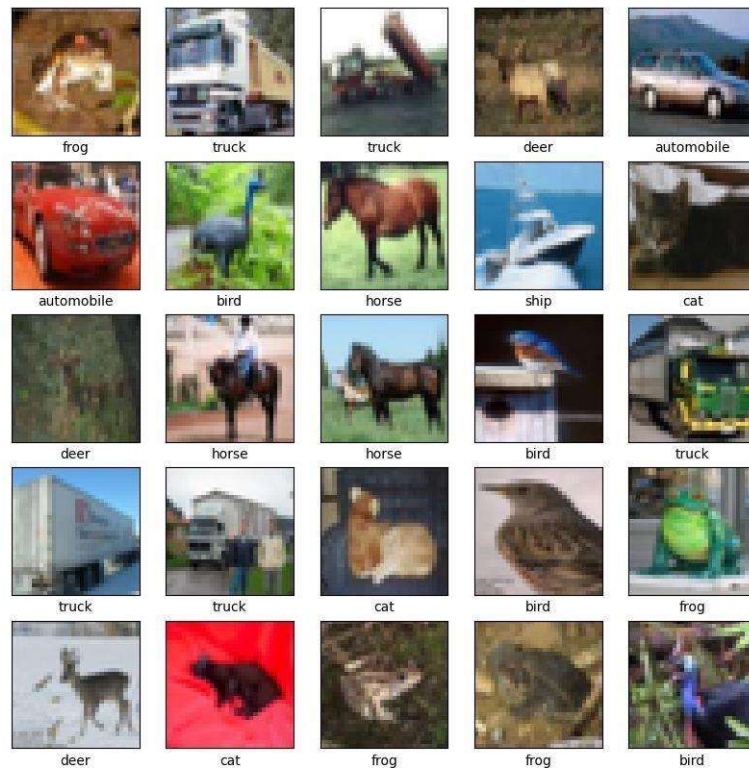


Gambar 1. Flowchart Model

Langkah pertama adalah membagi dataset menjadi data latih dan data uji. Kemudian dilakukan *preprocessing* pada dataset yang melibatkan normalisasi dan pembentukan ulang label. Data latih dan uji masing-masing dibagi menjadi 5 fold untuk kemudian digunakan sebagai validasi silang. Selanjutnya data latih diaugmentasi dengan menerapkan berbagai transformasi pada gambar untuk meningkatkan ukuran dan variasi dataset. Langkah terakhir adalah membangun model dengan CNN dan divalidasi menggunakan *K-fold Cross Validation*.

Preprocessing Dataset

Dataset yang digunakan adalah CIFAR-10 yang terdiri dari 60.000 gambar berwarna dengan ukuran 32x32 piksel yang dikategorikan ke dalam 10 kelas. Terdapat 50.000 gambar untuk proses pelatihan dan 10.000 gambar untuk proses pengujian. Ke-10 kelas tersebut mencakup kategori hewan dan kendaraan, yakni airplane, automobile, bird, cat, deer, dog, frog, horse, ship, dan truck. CIFAR-10 dipilih karena ukurannya yang kecil dan mudah diakses namun memiliki variasi gambar yang tinggi, dataset ini juga memiliki jumlah gambar yang sama untuk tiap kelasnya sehingga model tidak bias pada kelas tertentu. Sampel dari dataset dapat dilihat pada Gambar 2. Setelah memisahkan dataset menjadi data latih dan data uji selanjutnya dilakukan normalisasi data, karena dataset yang digunakan berisi citra berwarna maka operasi normalisasinya akan membagi tiap elemen dalam array dengan 255.0 yang mana nilai tersebut nilai maksimal citra warna. Setelahnya dataset dibentuk ulang menjadi array satu dimensi dengan mereduksi dimensi label.



Gambar 2. Sampel CIFAR-10

Augmentasi Data

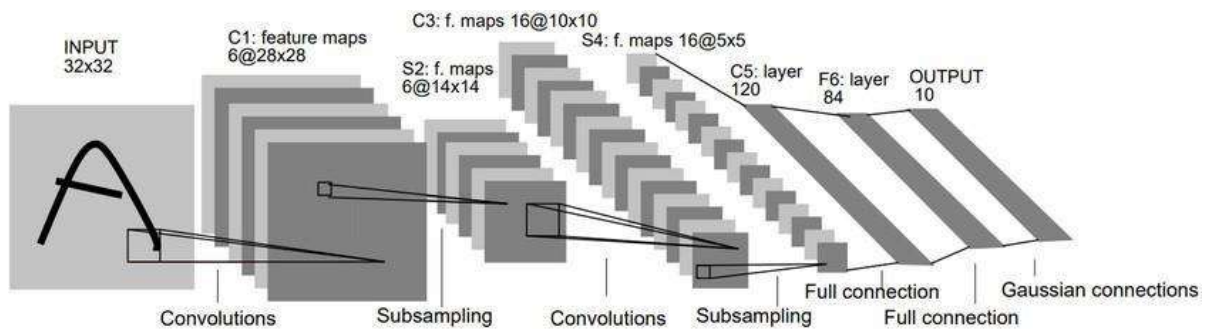
Augmentasi data pada pengenalan citra diimplementasikan agar tidak terjadi overfitting pada saat pelatihan model. Teknik augmentasi hanya dilakukan pada data latih. Implementasi augmentasi data dilakukan menggunakan data generator Keras, dari banyaknya parameter yang dicoba hanya beberapa parameter saja yang digunakan. Nilai dan parameter augmentasi yang digunakan pada penelitian ini dapat dilihat pada Tabel 1.

Tabel 1 Parameter Augmentasi

Parameter	Nilai
Rotation range	10-80
Horizontal flip	True
Vertical flip	True
Width shift	0.1-0.3
Height shift	0.1-0.3
Zoom range	0.2-0.3

Convolutional Neural Network (CNN)

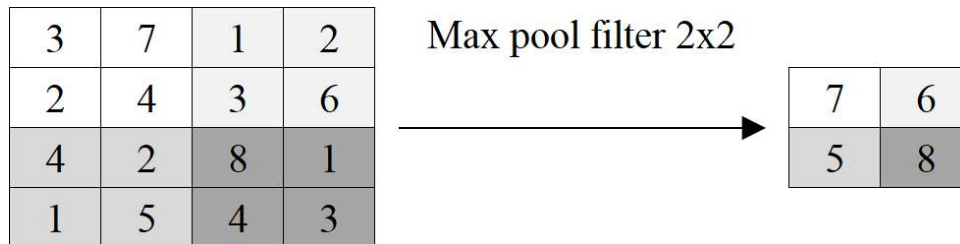
Model yang digunakan pada penelitian ini adalah CNN sederhana yaitu dengan arsitektur LeNet-5[10] yang dikenalkan oleh Yann LeCun, arsitekturnya dapat dilihat pada gambar 3. Lapisan utama yang digunakan pada model ini adalah *convolutional*, *pooling*, dan *fully connected layer*. [11]



Gambar 3. Arsitektur Convolutional Neural Network (CNN) LeNet-5 (Adaptasi dari [10])

Ekstraksi ciri dari citra input dilakukan menggunakan convolutional layer dengan cara menjumlahkan produk antara filter konvolusi dan citra input. Filter konvolusi adalah matriks koefisien yang dapat dipelajari dari data. Keluaran lapisan konvolusi adalah peta fitur, yang merupakan representasi gambar input yang menyoroti fitur-fitur spesifik yang relevan dengan tugas yang sedang dilakukan [12]. Filter size yang digunakan adalah 32, 64, 128 yang meningkat berurut dengan channel. Dan untuk kernel size yang digunakan adalah 3x3 dengan fungsi aktivasi ReLU.

Pooling layer digunakan untuk mengurangi dimensi representasi fitur untuk mengurangi ukuran model dan meningkatkan performa generalisasinya. Model yang dibangun menggunakan Max pool dengan filter 2x2 dan tidak menggunakan stride. Contoh dari penggunaan pooling layer dapat dilihat pada Gambar 4.



Gambar 4. Teknik max pool

Setelah layer terhubung secara penuh, model ini menggunakan proses konvolusi untuk memprediksi kelas citra. Dropout diaktifkan dengan nilai 0.2 sebagai regulasi untuk mengurangi overfitting. Setelahnya 10 kelas citra diprediksi menggunakan fungsi aktivasi softmax sebagai peubah vector agar dapat menjadi distribusi probabilitas.

4. Evaluasi

Hasil Pengujian

Sebelum menentukan teknik yang akhirnya dipilih, kami mencoba beberapa teknik lain sebagai perbandingan untuk mencari parameter apa saja yang berpengaruh untuk meningkatkan performa dan akurasi. Pada penelitian ini terdapat tiga model yang berhasil dibangun menggunakan CNN yaitu 1) *Baseline* CNN tanpa augmentasi, 2) CNN menggunakan 4 teknik augmentasi, 3) CNN menggunakan 6 teknik augmentasi. Parameter yang digunakan pada model 2 dan 3 dapat dilihat pada Tabel 2.

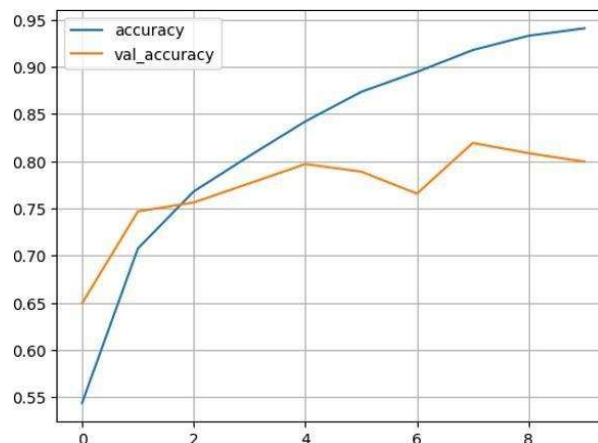
Tabel 2. Penggunaan parameter

Model	Parameter	Nilai Parameter
2	Rotation range	20
	Horizontal flip	True
	Width shift range	0.1
	Height shift range	0.1
3	Rotation range	15
	Horizontal flip	True
	Vertical flip	True
	Width shift range	0.1
	Height shift range	0.1
	Zoom range	0.3

Seluruh model dilatih pada runtime Python 3 menggunakan hardware accelerator T4 GPU sebanyak 50 epoch, evaluasi yang digunakan adalah *K-fold Cross Validation* dengan K=5. Gambar 5 menunjukkan model 1 yaitu CNN tanpa augmentasi yang dijadikan *baseline*. Gambar 6 menunjukkan hasil pengujian dari model 2. Gambar 7 menunjukkan hasil pengujian model 3.

Baseline

- 15s 10ms/step
- loss: 0.1695
- accuracy: 0.9412
- val_loss: 0.7341
- val_accuracy: 0.8080
- lr: 0.0010



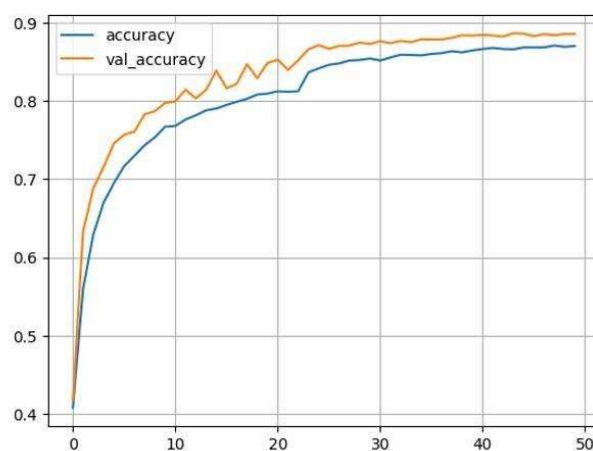
Gambar 5. Model 1 (tanpa augmentasi)

Model 2

- 32s 10ms/step
- loss: 0.3780
- accuracy: 0.8699
- val_loss: 0.3920
- val_accuracy: 0.8853
- lr: 1.0000e-05

Technique used:

- rotation_range = 20
- horizontal_flip = True
- width_shift_range = 0.1
- height_shift_range = 0.1



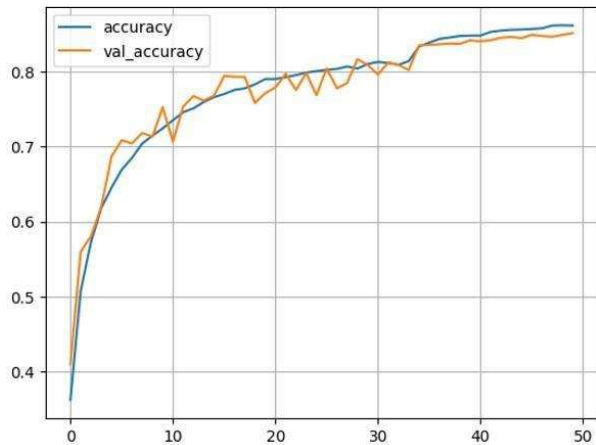
Gambar 6. Model 2

Model 3

- 26s 10ms/step
- loss: 0.3988
- accuracy: 0.8612
- val_loss: 0.4395
- val_accuracy: **0.8511**
- lr: 1.0000e-04

Technique used:

- rotation_range = 15
- horizontal_flip = True
- vertical_flip = True
- width_shift_range = 0.1
- height_shift_range = 0.1
- zoom_range = 0.3



Gambar 7. Model 3

Beberapa kelas mengalami peningkatan yang signifikan. Contohnya kelas bird dan frog yaitu sebesar, peningkatan ini terjadi pada kedua model. Namun pada model 2 terdapat penurunan akurasi sebesar 3.9% pada kelas anjing. Sementara untuk kelas kendaraan cenderung tidak memiliki peningkatan yang besar karena akurasi dari *baseline* untuk kelas seperti airplane, automobile, ship, dan truck sudah mencapai rata-rata akurasi 88.4%. Tabel 3 menunjukkan persentase kenaikan akurasi tiap kelas.

Tabel 3. Peningkatan akurasi tiap kelas

Kelas	Baseline	Model 2	Peningkatan	Model 3	Peningkatan
airplane	86.2	88.2	2	87.9	1.7
automobile	91.9	94.5	2.6	96.9	5
bird	70.9	80.2	9.3	85.6	14.7
cat	62.7	66.4	3.7	70.7	8
deer	80.7	82	1.3	86.4	5.7
dog	73	69.1	-3.9	77.2	4.2
frog	83.6	93.2	9.6	95	11.4
horse	84.8	90.7	5.9	90.4	5.6
ship	90	91.8	1.8	93.9	3.9
truck	85.5	92.5	7	94	8.5

Analisis Hasil Pengujian

Hasil pengujian dari tiga model yang dibangun berhasil menunjukkan bahwa penggunaan data augmentasi dapat meningkatkan akurasi. Gambar 5, 6, dan 7 menunjukkan model 2 yang menggunakan *rotation*, *flipping*, *width shift* dan *height shift* dapat meningkatkan akurasi sebesar 7.73%, sedangkan model 3 dengan menambah parameter *zoom* dan *vertical split* dapat meningkatkan akurasi sebesar 4.31% dari *baseline*.

Perubahan 5 poin pada parameter rotation range tidak memiliki pengaruh yang besar dan dapat diabaikan untuk kasus ini, tetapi pada penelitian ini setiap penggunaan rotation range diatas 30 poin akurasi dapat turun bahkan model menjadi lebih buruk dari *baseline*. CIFAR-10 juga tidak terlalu sensitif terhadap zoom, tetapi penggunaan parameter zoom dalam rentang 0.2-0.35 dapat meningkatkan akurasi sampai 1.2%. Sedangkan pada pengujian yang telah dilakukan *vertical flip* cenderung menurunkan akurasi sebesar 2.1 ± 0.4 %.

Model 2 yang hanya menggunakan 4 parameter menghasilkan akurasi lebih tinggi dibanding model 3 namun memiliki kecenderungan *overfitting*. Terbukti pada pengujian ini bahwa dengan lebih banyak parameter akan menambah variasi dataset sehingga dapat menurunkan *overfitting*. Perlu diperhatikan juga bahwa aktivasi parameter bersifat acak sehingga dalam pengujian yang berbeda dapat menghasilkan akurasi dan performa yang berbeda, sehingga dilakukan validasi silang dengan lima iterasi untuk tiap modelnya.

Pengujian ini hanya setitik kecil dari eksplorasi teknik-teknik augmentasi data. mengingat pada penelitian ini hanya menggunakan CNN dasar dan CIFAR-10. Hasil dari penggunaan parameter yang digunakan pada penelitian ini pun akan beragam jika digunakan pada arsitektur model, dataset, regulasi, dan berbagai teknik lainnya.

5. Kesimpulan

Dapat disimpulkan dari hasil pengujian bahwa model klasifikasi objek berbasis CNN menggunakan dataset CIFAR-10 berhasil dibangun dengan akurasi sebesar 80.8%. Model 2 menggunakan parameter augmentasi *rotation*, *horizontal flip*, *width shift*, *height shift* digunakan pada dataset CIFAR-10 dan model yang dibangun menghasilkan akurasi sebesar 88.53%.

Tentunya penelitian ini memiliki kekurangan dan dapat dieksplorasi lebih lanjut. Oleh karena itu kami menyarankan untuk membuat regulasi tambahan atau menggabungkan penggunaan parameter dengan teknik augmentasi citra otomatis.

Daftar Pustaka

- [1] Shorten, C. and Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. *Journal of big data*, 6(1), pp.1-48.
- [2] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S. and Uszkoreit, J., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- [3] Kabir, H.M., 2023. Reduction of Class Activation Uncertainty with Background Information. *arXiv preprint arXiv:2305.03238*.
- [4] Cubuk, E.D., Zoph, B., Mane, D., Vasudevan, V. and Le, Q.V., 2019. Autoaugment: Learning augmentation strategies from data. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 113-123).
- [5] Müller, S.G. and Hutter, F., 2021. Trivialaugment: Tuning-free yet state-of-the-art data augmentation. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 774-782).
- [6] Zhong, Z., Zheng, L., Kang, G., Li, S. and Yang, Y., 2020, April. Random erasing data augmentation. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 07, pp. 13001-13008).
- [7] Yamada, Y., Iwamura, M., Akiba, T. and Kise, K., 2019. Shakedown regularization for deep residual learning. *IEEE Access*, 7, pp.186126-186136.
- [8] Liu, X., Shen, F., Zhao, J. and Nie, C., 2022. RandomMix: A mixed sample data augmentation method with multiple mixed modes. *arXiv preprint arXiv:2205.08728*.
- [9] Perez, L. and Wang, J., 2017. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*.
- [10] LecunY, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, 86(11):2278-2324.
- [11] Phong, N.H. and Ribeiro, B., 2020. Rethinking recurrent neural networks and other improvements for image classification. *arXiv preprint arXiv:2007.15161*.
- [12] Suyanto, K., Ramadhani, N. and Mandala, S., 2019. Deep Learning Modernisasi Machine Learning Untuk Big Data. *Informatika*.
- [13] Guo, T., Dong, J., Li, H. and Gao, Y., 2017, March. Simple convolutional neural network on image classification. In *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)* (pp. 721-724). IEEE.
- [14] Fujita, K., Kobayashi, M. and Nagao, T., 2018, December. Data augmentation using evolutionary image processing. In *2018 Digital Image Computing: Techniques and Applications (DICTA)* (pp. 1-6). IEEE.
- [15] Li, Y., Kim, Y., Park, H., Geller, T. and Panda, P., 2022, October. Neuromorphic data augmentation for training spiking neural networks. In *European Conference on Computer Vision* (pp. 631-649). Cham: Springer Nature Switzerland.
- [16] Huang, Y., Cheng, Y., Bapna, A., Firat, O., Chen, D., Chen, M., Lee, H., Ngiam, J., Le, Q.V. and Wu, Y., 2019. Gpipe: Efficient training of giant neural networks using pipeline parallelism. *Advances in neural information processing systems*, 32.
- [17] Chen, P., Liu, S., Zhao, H. and Jia, J., 2020. Gridmask data augmentation. *arXiv preprint arXiv:2001.04086*.
- [18] Sultana, F., Sufian, A. and Dutta, P., 2018, November. Advancements in image classification using convolutional neural network. In *2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)* (pp. 122-129). IEEE.