

BAB I

PENDAHULUAN

1.1. Latar Belakang

Convolutional Neural Network (CNN) merupakan salah satu metode *deep learning* yang paling banyak digunakan dalam bidang *computer vision*, khususnya untuk pengolahan dan klasifikasi citra. CNN dikembangkan sebagai turunan dari *artificial neural network* yang dirancang menyerupai mekanisme kerja korteks visual manusia, sehingga mampu mengekstraksi fitur visual secara otomatis tanpa memerlukan perancangan fitur secara manual [1]. Arsitektur CNN secara umum terdiri atas beberapa lapisan utama, yaitu *convolution layer*, *pooling layer*, *fully connected layer*, dan *activation function*, yang bekerja secara hierarkis untuk mempelajari representasi fitur dari tingkat rendah hingga tingkat tinggi. Keunggulan utama CNN terletak pada kemampuannya dalam menangani data citra berdimensi tinggi serta performanya yang unggul pada berbagai tugas visual seperti klasifikasi gambar, deteksi objek, segmentasi citra, hingga pengenalan pola kompleks [2]. Oleh karena itu, CNN telah menjadi fondasi utama dalam banyak penelitian pengenalan citra modern dan terus dikembangkan melalui berbagai arsitektur dan pendekatan optimasi untuk meningkatkan akurasi dan efisiensi komputasi.

Berbagai penelitian terdahulu menunjukkan bahwa CNN telah berhasil diterapkan pada beragam tugas klasifikasi citra dengan tingkat akurasi yang tinggi. Penelitian klasifikasi citra menggunakan CNN pada *dataset CIFAR-10*, yang terdiri dari 60.000 citra berwarna berukuran 32×32 piksel dengan 10 kelas objek hewan dan kendaraan. Pemodelan dilakukan dengan menggunakan CNN sebagai model utama dengan tahapan praproses berupa *resizing* dan normalisasi citra, kemudian dilanjutkan dengan ekstraksi fitur melalui beberapa *convolution layer*, *pooling layer*, dan *fully connected layer*. Proses pelatihan dilakukan menggunakan *Stochastic Gradient Descent* dengan skema pelatihan hingga 300 *epoch*, serta evaluasi kinerja dilakukan pada data uji terpisah menggunakan metrik akurasi klasifikasi. Berdasarkan hasil pengujian, model CNN yang mampu mencapai

akurasi pengujian sebesar 93,47%, dengan waktu pelatihan sekitar 3 jam dengan menggunakan *CPU*. Hasil ini menunjukkan bahwa CNN memiliki kemampuan yang sangat baik dalam mengekstraksi fitur visual dan melakukan klasifikasi citra secara akurat, bahkan pada arsitektur yang relatif sederhana [3].

Penelitian klasifikasi citra menggunakan CNN pada klasifikasi spesies nyamuk berbasis citra digital dengan *dataset* citra nyamuk yang dikumpulkan secara mandiri sebanyak 1.800 citra yang merepresentasikan tiga genus nyamuk, yaitu *Aedes*, *Anopheles*, dan *Culex*, dengan masing-masing kelas diperluas melalui proses augmentasi data. Tahapan penelitian diawali dengan pengumpulan citra dari berbagai sumber daring, dilanjutkan dengan praproses berupa penyeragaman ukuran citra, konversi format, serta *rescaling* nilai piksel. Model CNN yang digunakan dibangun menggunakan arsitektur *custom CNN* yang terdiri dari beberapa *convolution layer*, *pooling layer*, dan *fully connected layer*, serta diimplementasikan menggunakan *framework Keras* dan *TensorFlow*. Pelatihan model dilakukan dengan pembagian data sebesar 80% untuk data latih dan 20% untuk data uji, serta menggunakan *optimizer Adam* dengan jumlah *epoch* maksimum 100. Berdasarkan hasil pengujian, model CNN setelah augmentasi mampu mencapai akurasi sebesar 84,51% dengan nilai *loss* kurang dari 0,35, yang menunjukkan bahwa CNN efektif dalam mengekstraksi fitur visual objek dengan tingkat kemiripan tinggi, meskipun performanya masih sangat dipengaruhi oleh arsitektur jaringan, strategi augmentasi, serta konfigurasi parameter pelatihan yang digunakan [4].

Penelitian mengenai analisis waktu pelatihan CNN telah dilakukan dengan tujuan untuk mengkaji pengaruh parameter pelatihan terhadap efisiensi komputasi model. Penelitian menggunakan beberapa arsitektur CNN populer, seperti *VGG16*, *ResNet18*, *ResNet34*, *ResNet50*, dan *GoogLeNet*, yang dilatih pada *dataset CIFAR* dan *ImageNet* dengan jumlah data pelatihan mencapai 50.000 citra. Metode penelitian dilakukan dengan memvariasikan ukuran *mini-batch* pada rentang yang luas, yaitu antara 5 hingga 500 sampel, serta mengamati perubahan waktu pelatihan per *epoch* pada berbagai jenis *GPU*. Hasil penelitian menunjukkan bahwa waktu pelatihan CNN sangat sensitif terhadap perubahan ukuran *mini-batch*, di mana pada

ukuran *mini-batch* kecil waktu pelatihan meningkat secara signifikan akibat dominasi operasi *non-GPU*, sedangkan pada ukuran *mini-batch* tertentu terjadi lonjakan waktu pelatihan secara tiba-tiba akibat perubahan algoritma konvolusi pada pustaka *cuDNN*. Evaluasi kinerja model prediksi waktu pelatihan yang dikembangkan dalam penelitian tersebut menunjukkan nilai *Mean Absolute Percentage Error* (MAPE) rata-rata sebesar 1,38% pada data uji, dengan nilai maksimum mencapai 5,01%, yang menandakan bahwa estimasi waktu pelatihan CNN bersifat kompleks dan sangat dipengaruhi oleh faktor komputasi internal model. Temuan ini mengindikasikan bahwa CNN konvensional, khususnya dengan arsitektur yang dalam dan jumlah parameter besar, memiliki kelemahan utama pada aspek efisiensi komputasi dan kestabilan waktu pelatihan [5].

Keterbatasan serupa juga ditemukan pada penelitian tentang analisis penggunaan CNN untuk klasifikasi citra batik dengan *dataset* berukuran menengah. Pada klasifikasi motif batik menggunakan arsitektur CNN yang dimodifikasi dengan mengombinasikan *GoogLeNet* dan *Residual Network (IncRes)*. Penelitian tersebut menggunakan *dataset* sebanyak 7.112 citra batik yang dibagi ke dalam 11 kelas motif, dengan proses pelatihan dilakukan hingga 500 *epoch* dan *learning rate* sebesar 0,001. Hasil pengujian menunjukkan bahwa arsitektur terbaik *IncRes-2* mampu mencapai akurasi sebesar 70,84% dengan waktu eksekusi sekitar 733 ms per citra, sementara arsitektur CNN konvensional seperti *GoogLeNet* dan *ResNet-50* memerlukan waktu komputasi yang lebih lama, masing-masing sekitar 877 ms dan lebih dari 2.200 ms per citra, dengan akurasi yang lebih rendah. Hasil ini menunjukkan bahwa meskipun CNN mampu mengekstraksi fitur visual batik secara otomatis, model dengan arsitektur dalam dan jumlah parameter besar tetap menghadapi kendala pada efisiensi komputasi, waktu inferensi, serta kebutuhan sumber daya yang tinggi, khususnya ketika diterapkan pada *dataset* batik dengan variasi motif yang kompleks dan jumlah kelas yang relatif banyak [6].

Berangkat dari berbagai keterbatasan CNN konvensional terkait kebutuhan komputasi yang tinggi, jumlah parameter yang besar, serta ketergantungan pada sumber daya perangkat keras yang kuat maka dikembangkan arsitektur CNN yang lebih efisien, salah satunya adalah *MobileNetV2*. *MobileNetV2* diperkenalkan

sebagai arsitektur *deep learning* yang dirancang khusus untuk lingkungan dengan keterbatasan sumber daya, seperti perangkat *mobile* dan *embedded system*, dengan tetap mempertahankan tingkat akurasi yang kompetitif. Arsitektur ini mengusung konsep *inverted residuals* dan *linear bottleneck*, di mana koneksi residual ditempatkan pada lapisan berdimensi rendah, sementara proses ekstraksi fitur dilakukan menggunakan *depthwise separable convolution* yang secara signifikan mengurangi jumlah operasi komputasi [7].

Keunggulan *MobileNetV2* dalam aspek efisiensi komputasi dibuktikan melalui penelitian yang membandingkan performa inferensi berbagai arsitektur CNN pada perangkat *edge* dengan sumber daya terbatas. Penelitian tersebut mengevaluasi beberapa keluarga model *deep learning*, antara lain *MobileNetV1–V3*, *EfficientNet V1* dan *V2*, *ResNet*, *VGG*, dan *InceptionV3*, yang diuji pada empat platform *edge* berbeda yaitu *NVIDIA Jetson Nano*, *Google Coral USB*, *Google Coral PCIe*, dan *Intel Neural Compute Stick 2*. Evaluasi dilakukan dengan mengukur waktu inferensi dan *frames per second* pada berbagai konfigurasi ukuran *input* dan skala model. Hasil pengujian menunjukkan bahwa *MobileNetV2* secara konsisten menghasilkan waktu inferensi yang jauh lebih rendah dibandingkan arsitektur CNN konvensional. Pada *input* berukuran 224×224 , *MobileNetV2* mencapai kecepatan hingga 365 FPS pada *Google Coral USB* dengan waktu inferensi rata-rata sekitar 2,73 ms, sementara model CNN yang lebih berat seperti *ResNet50* dan *VGG16* hanya mencapai kecepatan di bawah 40 FPS dengan waktu inferensi di atas 25 ms. Selain itu, pada perangkat *Jetson Nano*, *MobileNetV2* tetap menunjukkan stabilitas waktu inferensi dengan latensi yang lebih rendah dibandingkan *EfficientNet* dan *ResNet*, yang cenderung mengalami lonjakan latensi akibat keterbatasan memori. Hasil ini menegaskan bahwa desain arsitektur *MobileNetV2* yang menggunakan *depthwise separable convolution* dan *inverted residuals* mampu secara signifikan menekan beban komputasi tanpa mengorbankan kemampuan ekstraksi fitur [8].

Meskipun *MobileNetV2* dirancang sebagai arsitektur CNN yang efisien dari sisi komputasi, performa model tidak hanya ditentukan oleh arsitektur jaringan, tetapi juga sangat dipengaruhi oleh pengaturan *hyperparameter* yang digunakan selama proses pelatihan. *Hyperparameter* seperti *learning rate*, *batch size*, dan

dropout rate berperan dalam mengendalikan proses pembaruan bobot, stabilitas gradien, serta kemampuan model dalam menghindari *overfitting*. Pengaturan *hyperparameter* yang kurang tepat dapat menyebabkan model gagal mencapai konvergensi optimal, mengalami pelatihan yang tidak stabil, atau menghasilkan performa yang tidak konsisten meskipun menggunakan arsitektur yang dirancang secara efisien [9]. Selain pengaturan *hyperparameter*, pemilihan *optimizer* merupakan komponen penting lain dalam proses optimasi model CNN. *Optimizer* berfungsi mengarahkan proses pencarian solusi terbaik pada ruang parameter dengan mengatur bagaimana gradien digunakan untuk memperbarui bobot jaringan. Setiap *optimizer* memiliki karakteristik yang berbeda dalam menangani laju pembelajaran, adaptasi gradien, serta kecepatan konvergensi, sehingga dapat menghasilkan perilaku pelatihan yang berbeda pada *dataset* dan arsitektur yang sama. Oleh karena itu, pemilihan *optimizer* yang sesuai menjadi faktor penentu dalam mencapai keseimbangan antara kecepatan pelatihan, stabilitas konvergensi, dan kemampuan generalisasi model. Kombinasi antara strategi *hyperparameter tuning* yang sistematis dan pemilihan *optimizer* yang tepat menjadi langkah krusial untuk memastikan bahwa arsitektur CNN yang efisien seperti *MobileNetV2* dapat menghasilkan performa klasifikasi yang optimal dan stabil [10].

Penelitian terdahulu menunjukkan bahwa penerapan *hyperparameter tuning* memiliki peran yang sangat signifikan dalam meningkatkan performa model CNN pada arsitektur *MobileNet*. Model hibrida yang mengintegrasikan *Standard U-Net* dan *Improved MobileNet-V3* untuk diagnosis kanker kulit berbasis citra dermoskopi, di mana proses optimasi *hyperparameter* dilakukan menggunakan *Bayesian Optimization* dan *Grid Search*. Penelitian tersebut memfokuskan optimasi pada parameter penting seperti *learning rate*, *batch size*, *dropout rate*, serta konfigurasi arsitektur, dengan tujuan meningkatkan akurasi sekaligus efisiensi pembelajaran model. Evaluasi dilakukan pada *dataset HAM-10000*, dan hasil eksperimen menunjukkan bahwa model dengan *hyperparameter* teroptimasi mampu mencapai akurasi sebesar 98,86%, *precision* 97,84%, *recall* 96,35%, serta *F1-score* 97,32%, yang secara konsisten lebih tinggi dibandingkan model dengan parameter *default*. Temuan ini menegaskan bahwa *hyperparameter tuning* bukan

sekadar proses tambahan, melainkan komponen krusial yang secara langsung memengaruhi stabilitas konvergensi, kemampuan generalisasi, dan kinerja akhir model CNN [11].

Penelitian mengenai dampak pemilihan *optimizer* terhadap performa CNN mengevaluasi sejauh mana perbedaan *optimizer* memengaruhi hasil pelatihan CNN dibandingkan faktor lain seperti arsitektur jaringan. Eksperimen dilakukan menggunakan 3 *dataset* yang berbeda, yaitu *Cats and Dogs*, *Natural Images*, dan *Fashion MNIST*, dengan total data sekitar 100.000 data. Pelatihan menggunakan 4 arsitektur CNN dengan kedalaman yang berbeda. Proses pelatihan juga mengkonfigurasikan *optimizer* yang bervariasi dengan jumlah 9 *optimizer* untuk melihat pengaruh *optimizer* terhadap akurasi, *loss*, dan konvergensi. Pada *dataset Cats and Dogs*, *optimizer Adam* menunjukkan performa paling unggul dengan akurasi klasifikasi tertinggi yaitu 91,1% dibandingkan *optimizer* lain, meskipun *Adam* memiliki waktu konvergensi yang sedikit lebih lama daripada *Nadam*. Sementara itu, pada *dataset Natural Images*, *optimizer Adam* kembali menghasilkan performa terbaik dengan akurasi 98,1%. Untuk *dataset Fashion MNIST*, *Nadam* menunjukkan performa akurasi tertinggi yaitu 83,1%, dengan waktu konvergensi yang lebih besar dibandingkan *Adam*. Secara keseluruhan, hasil tersebut menyimpulkan bahwa *Adam* dan *Nadam* merupakan pilihan *optimizer* yang paling konsisten untuk tugas klasifikasi citra. *Optimizer Adadelta* tercatat sebagai yang terburuk pada seluruh *dataset* dan arsitektur yang diuji, dengan akurasi yang rendah serta konvergensi yang tidak stabil. Penelitian ini juga menemukan bahwa peningkatan kedalaman CNN memang dapat meningkatkan akurasi [12].

Sebagai objek penelitian, batik memiliki karakteristik visual yang kompleks dan menantang untuk diklasifikasikan, sehingga relevan untuk dikaji menggunakan pendekatan *deep learning*. Batik merupakan warisan budaya Indonesia yang telah diakui secara global oleh UNESCO sebagai *Masterpiece of Oral and Intangible Heritage of Humanity* pada tanggal 2 Oktober 2009, yang menegaskan nilai historis, filosofis, dan artistik batik sebagai lebih dari sekadar kain bermotif [13]. Setiap motif batik mengandung pola visual khas yang tersusun dari kombinasi bentuk geometris, kurva organik, repetisi, serta variasi warna yang halus, sehingga

menuntut model klasifikasi mampu mengekstraksi fitur tekstur dan pola secara presisi. Dalam penelitian ini, beberapa motif batik dipilih sebagai objek kajian, di antaranya Kawung, Mega Mendung, dan Parang, yang merupakan motif populer dan sering dijadikan representasi batik klasik dalam berbagai konteks akademik maupun komersial [14]. Motif Kawung memiliki pola geometris simetris berbentuk elips yang tersusun berulang dan melambangkan kesucian serta keseimbangan hidup, sementara motif Mega Mendung dikenal dengan pola awan berlapis dan gradasi warna yang mencerminkan nilai filosofis ketenangan dan kebijaksanaan masyarakat Cirebon [15][16]. Motif Parang, sebagai salah satu motif tertua, memiliki pola garis diagonal berulang yang melambangkan kekuatan, keberanian, dan kesinambungan, serta sering digunakan dalam berbagai karya seni dan *fashion modern* [17]. Selain ketiga motif utama tersebut, penelitian ini juga melibatkan beberapa motif lain sebagai kelas tambahan, seperti Sekar Jagad, Sido Mukti, Sido Luhur, dan Singo Barong.

Penelitian terdahulu mengenai klasifikasi motif batik juga menunjukkan bahwa pendekatan *deep learning* berbasis arsitektur ringan mampu memberikan kinerja yang sangat tinggi ketika dikombinasikan dengan strategi pelatihan yang tepat. Salah satu penelitian yang relevan adalah klasifikasi motif Batik Semarang menggunakan arsitektur *MobileNetV2* yang dipadukan dengan teknik data *augmentation*. Penelitian tersebut menggunakan *dataset* sebanyak 3.020 citra batik yang terdiri dari 10 kelas motif Batik Semarang, dengan pembagian data ke dalam data latih dan validasi. Tahapan praproses dilakukan melalui penyeragaman ukuran citra menjadi 224×224 piksel serta normalisasi nilai piksel, diikuti dengan augmentasi data berupa rotasi, *flipping*, *zoom*, dan translasi untuk meningkatkan keragaman data latih. Model *MobileNetV2* dilatih menggunakan pendekatan *transfer learning* berbasis bobot pra-latih *ImageNet* dan dimodifikasi dengan penambahan beberapa lapisan *fully connected* serta *dropout* untuk menyesuaikan karakteristik motif batik. Hasil evaluasi menunjukkan bahwa model mampu mencapai akurasi sebesar 100% pada validasi, dengan nilai *precision*, *recall*, dan *F1-score* rata-rata di atas 0,98 pada seluruh kelas motif. Analisis *confusion matrix* menunjukkan tingkat kesalahan klasifikasi yang sangat kecil, terutama pada motif

yang memiliki kemiripan visual, sehingga menegaskan efektivitas *MobileNetV2* dalam mengekstraksi pola geometris dan tekstur kompleks batik sekaligus mempertahankan efisiensi komputasi yang tinggi [18].

Penelitian tentang kombinasi *MobileNetV2* dengan *transfer learning* untuk mengklasifikasikan motif Kawung, Mega Mendung, dan Parang menggunakan dataset berjumlah 3.000 citra batik. Penelitian tersebut melaporkan akurasi pengujian yang sangat tinggi, mencapai 89,16%, dengan waktu pelatihan relatif singkat sekitar 18 menit, sehingga menegaskan keunggulan *MobileNetV2* dari sisi efisiensi komputasi dibandingkan arsitektur CNN yang lebih besar seperti *DenseNet* atau *ResNet*. Penelitian ini juga membandingkan beberapa arsitektur CNN, seperti *DenseNet*, *Inception*, dan *ResNet*, pada tugas klasifikasi batik dan menunjukkan bahwa model berarsitektur ringan cenderung memberikan kompromi terbaik antara akurasi dan waktu komputasi, terutama ketika *dataset* memiliki variasi motif yang kompleks namun jumlah data terbatas [19].

Meskipun demikian, sebagian besar penelitian terdahulu masih berfokus pada perbandingan arsitektur model atau penerapan *transfer learning*, tanpa mengeksplorasi secara mendalam strategi optimasi pelatihan. Aspek seperti perbandingan beberapa *optimizer* secara sistematis, penggunaan berbagai metode *hyperparameter tuning*, serta pencatatan waktu pencarian *hyperparameter* sebagai indikator biaya komputasi sering kali belum dibahas secara komprehensif. Selain itu, evaluasi performa umumnya hanya dilaporkan berdasarkan satu kali proses pelatihan, sehingga stabilitas model terhadap variasi data dan konfigurasi pelatihan belum sepenuhnya teruji. Cela inilah yang menjadi dasar penelitian ini, yaitu tidak hanya mengevaluasi efektivitas *MobileNetV2* dalam klasifikasi motif batik, tetapi juga menganalisis pengaruh kombinasi *optimizer* dan metode *hyperparameter tuning* secara menyeluruh, termasuk dari sisi akurasi, *loss*, jumlah *epoch*, dan waktu pencarian parameter, untuk memperoleh konfigurasi yang paling stabil dan efisien.

Oleh karena itu, penelitian ini dilakukan untuk mengeksplorasi strategi optimasi pelatihan pada klasifikasi motif batik berbasis *deep learning*. Penelitian ini berfokus pada analisis efektivitas arsitektur *MobileNetV2* dalam mengklasifikasikan motif batik dengan mengombinasikan berbagai metode

hyperparameter tuning dan pilihan *optimizer*. Tidak hanya mengevaluasi performa model dari sisi akurasi dan *loss*, penelitian ini juga menitikberatkan pada aspek efisiensi, seperti kecepatan konvergensi, jumlah *epoch* pelatihan, serta waktu pencarian *hyperparameter* sebagai indikator biaya komputasi. Diharapkan penelitian ini dapat memberikan gambaran yang lebih menyeluruh mengenai strategi optimasi yang paling stabil, efisien, dan efektif untuk klasifikasi motif batik, sekaligus menjadi acuan bagi pengembangan sistem klasifikasi citra batik yang lebih praktis dan adaptif terhadap keterbatasan sumber daya komputasi di masa mendatang.

1.2. Rumusan Masalah

Berlandaskan latar belakang yang sudah dibuat, kemudian dapat diambil rumusan masalah yang menjadi permasalahan prioritas yang akan diselesaikan, antara lain:

1. Bagaimana melakukan *hyperparameter tuning* dan pemilihan *optimizer* yang optimal untuk meningkatkan performa, efisiensi sumber daya komputasi, dan konvergensi model *MobileNetV2* dalam klasifikasi motif batik?
2. Bagaimana hasil performa model, kecepatan komputasi, dan konvergensi *MobileNetV2* yang telah dioptimasi dalam mengidentifikasi ketiga motif batik tersebut?
3. Bagaimana membuat tampilan antarmuka untuk mempermudah deteksi motif batik?

1.3. Batasan Masalah

Berlandaskan uraian dalam latar belakang serta identifikasi masalah yang cukup luas seperti yang telah diuraikan sebelumnya. Berikut merupakan beberapa batasan-batasan yang diterapkan dalam penelitian:

1. Penelitian ini berfokus pada klasifikasi tiga jenis motif batik tradisional: Kawung, Mega Mendung, dan Parang.

2. Metode yang digunakan adalah CNN dengan pendekatan *Transfer Learning* menggunakan arsitektur *MobileNetV2*.
3. Data yang digunakan berupa citra digital motif batik dalam format RGB dengan ukuran seragam. Penelitian ini tidak mencakup segmentasi motif batik atau deteksi objek motif dalam gambar yang lebih kompleks.
4. Luaran penelitian ini adalah model klasifikasi yang telah dilatih dan dievaluasi, serta rancangan antarmuka dasar untuk demonstrasi klasifikasi.

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dirumuskan, penelitian ini bertujuan untuk mencapai sasaran sebagai berikut:

1. Mengetahui cara melakukan *hyperparameter tuning* dan pemilihan *optimizer* yang optimal untuk meningkatkan performa, efisiensi sumber daya komputasi, dan konvergensi model *MobileNetV2* dalam klasifikasi motif batik.
2. Mengetahui hasil performa model, kecepatan komputasi, dan konvergensi *MobileNetV2* dalam mengidentifikasi ketiga motif batik tersebut.
3. Mengetahui cara membuat tampilan antarmuka untuk mempermudah deteksi motif batik.

1.5. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi bagi pihak-pihak yang membutuhkan baik teoritis maupun praktis, diantaranya:

1. Menambah wawasan dalam bidang *Deep Learning*, khususnya dalam penggunaan model *MobileNetV2* untuk klasifikasi citra motif batik dengan efisien.
2. Menghasilkan model *MobileNetV2* yang lebih optimal dan efisien dalam mengklasifikasikan motif batik melalui penerapan *hyperparameter tuning* dan pemilihan *optimizer*.
3. Mendukung upaya pelestarian budaya batik dengan menyediakan alat identifikasi motif yang cepat dan akurat.