PERBANDINGAN HRNet, ConvNeXT, SWIN TRANSFORMER UNTUK KLASIFIKASI CITRA KUPU-KUPU

M. Raditya Adhirajasa 2157051004

Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung

KATA KUNCI

Computer Vision

ABSTRAK

Eksperimen ini membandingkan kinerja tiga model deep learning, yaitu HRNet, ConvNeXt, dan Swin Transformer, dalam klasifikasi citra spesies kupu-kupu. Hasil eksperimen menunjukkan bahwa HRNet mencapai akurasi 100%, dengan presisi, recall, dan F1-Score masing-masing juga 100%, menjadikannya model paling efektif untuk tugas ini. ConvNeXt, sebaliknya, menunjukkan performa sangat rendah dengan akurasi 8,12% dan F1-Score 0,02%, mengindikasikan ketidaksesuaian arsitektur model dengan dataset yang digunakan. Sementara itu, Swin Transformer menghasilkan akurasi 91,88%, presisi 92,91%, recall 91,88%, dan F1-Score 91,87%, memberikan alternatif yang cukup baik meskipun tidak sebaik HRNet. Kesimpulannya, HRNet menunjukkan keunggulan dalam klasifikasi citra kupu-kupu, sementara ConvNeXt memerlukan perbaikan signifikan dan Swin Transformer memberikan hasil yang kompetitif.

1. Pendahuluan

Kupu-kupu merupakan bagian dari ordo Lepidoptera yang memiliki ciri khas pada pola sayapnya, menjadikannya salah satu kelompok serangga dengan keragaman spesies yang sangat tinggi. Diperkirakan terdapat sekitar 15.000 hingga 20.000 spesies kupu-kupu di dunia, dengan 1.600 spesies di antaranya ditemukan di Indonesia. Kupu-kupu memainkan peran penting dalam ekosistem, terutama dalam membantu proses penyerbukan tanaman, serta menjadi indikator bioekologi untuk memantau perubahan kualitas lingkungan [1].

Pengenalan dan klasifikasi spesies kupu-kupu menghadapi tantangan besar karena kemiripan karakteristik antarspesies. Kebutuhan akan teknologi pengenalan berbasis citra semakin meningkat untuk mendukung konservasi dan penelitian biodiversitas. Dalam beberapa dekade terakhir, metode berbasis *deep learning*, khususnya *Convolutional Neural Network* (CNN), telah menunjukkan kinerja unggul dalam pengenalan pola citra. Model seperti EfficientNet-B0 dan VGG-16 telah berhasil digunakan untuk mengklasifikasi spesies

kupu-kupu dengan akurasi hingga 97,91% [2].

Eksperimen ini bertujuan untuk membandingkan kinerja tiga model, yaitu HRNet, ConvNeXt, dan Swin Transformer, dalam tugas klasifikasi gambar kupu-kupu yang terdiri dari delapan kelas: batik cap, harimau kuning hijau, hijau biru, jarak, jojo, pantat merah, raja helena, dan raja limau. Pemilihan model ini didasarkan pada kemampuan mereka dalam menangkap fitur spasial, lokal, dan global secara efisien pada berbagai resolusi gambar, sehingga diharapkan dapat memberikan hasil yang lebih akurat dan konsisten dibandingkan model konvensional. Studi ini juga menggunakan teknik augmentasi data untuk meningkatkan keragaman dataset, yang menjadi langkah penting dalam mencegah overfitting serta meningkatkan performa model.

2. Tinjauan Pustaka

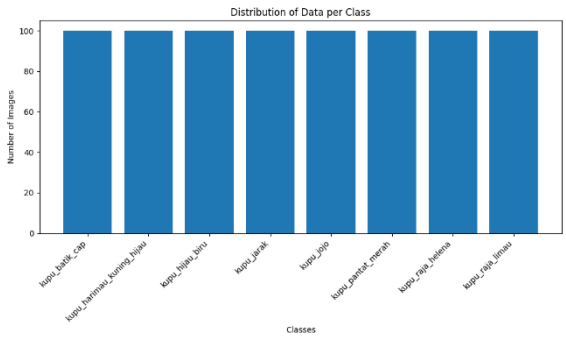
Penelitian sebelumnya telah menunjukkan bahwa penggunaan metode deep learning, khususnya Convolutional Neural Networks (CNN), memberikan hasil yang sangat baik dalam tugas klasifikasi citra kupu-kupu. Micheal dan Hartati (2022) mengembangkan sistem klasifikasi spesies kupu-kupu menggunakan arsitektur CNN VGG-16 dan LeNet. Studi mereka menunjukkan bahwa VGG-16, dengan lapisan konvolusi yang lebih mendalam, mampu memberikan akurasi tertinggi sebesar 93% menggunakan optimasi Adam dibandingkan LeNet yang hanya mencapai akurasi 67%. Mereka juga menyoroti pentingnya augmentasi data untuk meningkatkan keragaman dataset, yang membantu model dalam belajar fitur visual yang kompleks pada citra.

Syamsudin et al. (2024) dalam penelitian mereka menggunakan arsitektur EfficientNet-B0, yang merupakan salah satu model CNN terkini yang dirancang untuk efisiensi dalam skala parameter dan akurasi. Penelitian ini berhasil mengklasifikasi 25 spesies kupu-kupu dan ngengat dengan akurasi 97,91%, yang menunjukkan peningkatan performa dibandingkan metode sebelumnya seperti GoogLeNet dengan alurasi 97,5% dan Mask R-CNN dengan akurasi 83,62%. Mereka juga menggarisbawahi peran transfer learning dalam mengurangi kebutuhan data pelatihan yang besar tanpa mengorbankan akurasi

3. Metodologi

3.1. Dataset

Penelitian ini menggunakan dataset yang terdiri dari delapan kelas kupu-kupu, yaitu: batik cap, harimau kuning hijau, hijau biru, jarak, jojo, pantat merah, raja helena, dan raja limau. Setiap kelas awalnya memiliki jumlah gambar yang bervariasi, dengan beberapa kelas yang kekurangan data citra. Untuk memastikan distribusi data yang seimbang, dilakukan augmentasi data berupa rotasi pada kelas-kelas yang kekurangan gambar, sehingga setiap kelas memiliki 100 citra kupu-kupu. Proses augmentasi ini bertujuan untuk meningkatkan keragaman dan jumlah data pelatihan, serta membantu model mengenali pola visual dari berbagai sudut pandang [5].



Gambar 1 Distribusi data

Setiap gambar dalam dataset memiliki dimensi yang disesuaikan menjadi 224×224 piksel agar sesuai dengan masukan model deep learning. Dataset ini kemudian dibagi ke dalam subset data pelatihan dan pengujian dengan rasio 80:20, untuk memastikan evaluasi model yang adil dan representatif.

Gambar 2 Sample gambar kupu-kupu

3.2. HRNet

High-Resolution Network (HRNet) adalah arsitektur deep learning yang dirancang untuk mempertahankan representasi resolusi tinggi sepanjang proses ekstraksi fitur. Tidak seperti arsitektur konvolusi tradisional yang cenderung mengurangi resolusi gambar melalui proses downsampling bertahap, HRNet menjaga jalur resolusi tinggi secara paralel dengan jalur resolusi menengah dan rendah [7]. Pendekatan ini memungkinkan jaringan untuk mempertahankan detail spasial sambil tetap memanfaatkan fitur global yang lebih abstrak dari resolusi rendah. HRNet mengintegrasikan informasi dari berbagai jalur resolusi melalui mekanisme fusion, sehingga menghasilkan representasi fitur yang kaya dan akurat.

3.3. ConvNeXT

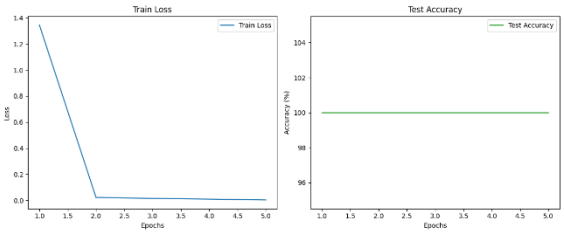
ConvNeXt adalah arsitektur deep learning berbasis Convolutional Neural Network (CNN) yang dirancang untuk memanfaatkan keunggulan desain modern dari Vision Transformers (ViT) sambil tetap mempertahankan prinsip dasar CNN. ConvNeXt dikembangkan dengan menyederhanakan struktur CNN tradisional seperti ResNet dan menambahkan fitur desain yang ditemukan dalam ViT, seperti depthwise convolution, LayerNorm, dan headless architectures. Pendekatan ini menjadikan ConvNeXt lebih efisien dalam menangkap pola visual kompleks dengan performa yang kompetitif terhadap ViT dalam berbagai tugas visi komputer, termasuk klasifikasi citra [4].

3.4. Swin Transformer

Swin Transformer adalah model deep learning berbasis Vision Transformer (ViT) yang dirancang untuk menangkap hubungan spasial lokal maupun global dalam gambar secara lebih efisien [6]. Salah satu inovasi utama dari Swin Transformer adalah penggunaan mekanisme shifted windows, di mana input gambar dibagi menjadi blok-blok kecil (windows) yang saling tumpang tindih. Pendekatan ini memungkinkan model untuk menangkap hubungan antarblok secara hierarkis sambil mempertahankan efisiensi komputasi. Selain itu, Swin Transformer menggunakan struktur pyramid yang mirip dengan CNN untuk menghasilkan representasi multi-resolusi, menjadikannya sangat cocok untuk tugas yang memerlukan analisis detail dan konteks global, seperti klasifikasi citra kupu-kupu.

4. Hasil

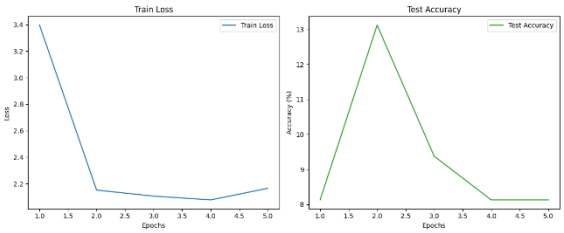
4.1. Akurasi



Gambar 3 Plot Akurasi HRNet

```
Final Test Accuracy: 100.00%
Final Precision: 1.00%
Final Recall: 1.00%
Final F1-Score: 1.00%
```

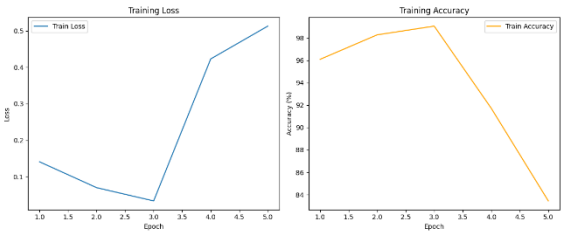
Gambar 4 Akurasi HRNet



Gambar Plot 5 Akurasi ConvNeXT

```
Final Test Accuracy: 8.12%
Final Precision: 0.01%
Final Recall: 0.12%
Final F1-Score: 0.02%
```

Gambar Akurasi 6 ConvNeXT



Gambar 7 Plot Akurasi Swin Transformer

```
Final Test Accuracy: 91.88%
Final Precision: 92.91%
Final Recall: 91.88%
Final F1-Score: 91.87%
```

Gambar 8 Akurasi Swin Transformer

HRNet mencapai akurasi, presisi, recall, dan F1-Score sebesar 100%, yang mengindikasikan bahwa model ini mampu mengklasifikasi seluruh citra dalam dataset dengan sempurna. Kemampuan HRNet untuk mempertahankan resolusi tinggi sepanjang jaringan serta mengintegrasikan informasi dari berbagai resolusi melalui mekanisme fusion menjadi faktor utama kesuksesannya dalam menangkap pola visual kompleks dari citra kupu-kupu.

Sebaliknya, ConvNeXt menunjukkan performa yang jauh lebih rendah, dengan akurasi sebesar 8,12% dan nilai F1-Score sebesar 0,02%. Rendahnya performa ConvNeXt dapat disebabkan oleh ketidaksesuaian arsitektur model dengan karakteristik dataset, misalnya pola visual kupu-kupu yang kompleks yang memerlukan pendekatan multi-resolusi atau hierarkis. Selain itu, hal ini juga dapat menunjukkan adanya kebutuhan untuk penyesuaian lebih lanjut dalam hyperparameter atau teknik augmentasi data.

Swin Transformer, meskipun tidak seoptimal HRNet, masih menunjukkan kinerja yang cukup baik dengan akurasi 91,88%, presisi 92,91%, recall 91,88%, dan F1-Score 91,87%. Meskipun tidak seakurat HRNet, Swin Transformer berhasil mengatasi sebagian besar tantangan dalam klasifikasi citra kupu-kupu, dengan hasil yang konsisten dan mendekati kinerja model terbaik.

Hasil penelitian ini menunjukkan bahwa HRNet adalah pilihan yang jauh lebih efektif dibandingkan ConvNeXt, sedangkan Swin Transformer memberikan alternatif yang kompetitif meski tidak sebaik HRNet.

Daftar Pustaka

[1] Micheal, & Hartati, E., 2022. Klasifikasi Spesies Kupu-kupu Menggunakan Metode Convolutional Neural Network. MDP Student Conference (MSC) 2022, pp. 569–572.

[2] Syamsudin, H., Khalidah, S., & Jumanto, 2024. Lepidoptera Classification Using Convolutional Neural Network EfficientNet-B0. *Indonesian Journal of Artificial Intelligence and Data Mining*, 7(1), pp. 47–56.

[3] Vishniakov, K., Shen, Z., & Liu, Z., 2024. ConvNet vs Transformer, Supervised vs CLIP: Beyond ImageNet.

[4] Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A ConvNet for the 2020s. *Facebook AI Research (FAIR) and UC Berkeley*.

[5] Mikołajczyk, A., & Grochowski, M. (2018). Data augmentation for improving deep learning in image classification problems. *Department of Control Systems Engineering, Gdańsk University of Technology*. IEEE.

[6] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows.

[7] Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., & Xiao, B. (2020). Deep High-Resolution Representation Learning for Visual Recognition. *IEEE Transactions On Pattern Analysis And Machine Intelligence.*