Technical Report: Multi-Level Image Classification

Project Title: Oxford Flowers-102 Fine-Grained Classification

Task: Implementation of a systematic multi-level framework for high-accuracy image recognition.

---

1. Dataset Selection and Technical Justification

Dataset Characteristics

The Oxford Flowers-102 dataset was selected for this project due to its specific focus on fine-grained classification. The dataset contains 102 flower categories, with each class consisting of between 40 and 258 images. Unlike standard object detection datasets, fine-grained tasks require models to distinguish between species that share high visual similarity, such as subtle variations in petal texture or stamen structure.

Practical Justification

- **Feature Discrimination:** The dataset demands strong feature extraction capabilities, making it an ideal testbed for attention mechanisms and ensemble learning.
- **Feasibility:** With 8,189 images, the dataset is large enough to demonstrate deep learning principles while remaining manageable for training under limited compute constraints.
- **Standardization:** It is a recognized benchmark in computer vision research and is natively supported by the torchvision library, ensuring experimental reproducibility.

---

2. Level 1: Baseline Model (Transfer Learning)

Objective

To establish a performance baseline using a robust, pre-trained convolutional neural network to leverage features learned from large-scale data.

Approach

- **Architecture:** ResNet-50 (Pre-trained on ImageNet).
- **Methodology:** The backbone layers were frozen to retain general feature extractors (edges, shapes), while a new classification head tailored to 102 classes was trained.
- **Configurations:** Input resolution was set to 224 x 224, using the Adam optimizer and Cross-Entropy loss for efficient convergence.

```
Epoch [1/7] | Train Acc: 0.1853 | Val Acc: 0.4637
Epoch [2/7] | Train Acc: 0.7618 | Val Acc: 0.6588
Epoch [3/7] | Train Acc: 0.9314 | Val Acc: 0.7735
Epoch [4/7] | Train Acc: 0.9686 | Val Acc: 0.8608
Epoch [5/7] | Train Acc: 0.9922 | Val Acc: 0.8873
Epoch [6/7] | Train Acc: 1.0000 | Val Acc: 0.9000
Epoch [7/7] | Train Acc: 1.0000 | Val Acc: 0.9098
```

```
Test Accuracy: 0.8853472109286062
```

---

3. Level 2: Intermediate Techniques (Data Augmentation and Fine-Tuning)

Objective

To improve the model's ability to generalize to unseen data and reduce the risk of overfitting caused by the limited sample size per class.

Techniques Applied

- **Data Augmentation:** Implemented random resized crops, horizontal flips, rotation, and color jitter to artificially increase dataset diversity and force the model to learn rotation-invariant features.
- **Partial Fine-Tuning:** The deeper residual blocks (specifically layer 3 and layer 4) were unfrozen. This allowed the model to adapt high-level semantic features specifically to floral patterns while keeping early-layer general feature detectors intact.

```
Epoch [1/7] | Train Acc: 0.1480 | Val Acc: 0.3882
Epoch [2/7] | Train Acc: 0.5941 | Val Acc: 0.6618
Epoch [3/7] | Train Acc: 0.7824 | Val Acc: 0.7510
Epoch [4/7] | Train Acc: 0.8745 | Val Acc: 0.8284
Epoch [5/7] | Train Acc: 0.9461 | Val Acc: 0.8843
Epoch [6/7] | Train Acc: 0.9755 | Val Acc: 0.9049
Epoch [7/7] | Train Acc: 0.9882 | Val Acc: 0.9137
Test Accuracy: 0.8812815091884859
```

4. Level 3: Advanced Architecture Design (Attention Mechanism)

Objective

To enhance feature selectivity by allowing the network to focus on the most discriminative parts of the image, which is critical for fine-grained tasks where the background or leaves might otherwise distract the model.

Approach

- **Architecture:** ResNet-50 integrated with a Convolutional Block Attention Module (CBAM).
- **Mechanism:** CBAM was added after each residual stage to provide both channel and spatial attention.
- **Functionality:** Channel attention helps the model identify "what" features are important (e.g., specific color gradients), while spatial attention helps the model identify "where" the most relevant information is located (e.g., the center of the flower bloom).

```
Epoch [1/3] | Train Acc: 0.9990 | Val Acc: 0.9118
Epoch [2/3] | Train Acc: 0.9951 | Val Acc: 0.9147
Epoch [3/3] | Train Acc: 0.9980 | Val Acc: 0.9147
Test Accuracy: 0.8882745161814929
```

5. Level 4: Expert Techniques (Ensemble Learning)

Objective

To maximize accuracy and robustness by combining the predictive strengths of multiple models with different inductive biases.

Approach

- **Ensemble Method:** Soft Voting (Probability Averaging).
- **Models Combined:** The Level 2 Fine-tuned ResNet-50 and the Level 3 CBAM-enhanced ResNet-50.
- **Logic:** By averaging the class probabilities from both models, the system reduces the variance of individual predictions and improves overall generalization.

Final Outcome: This approach resulted in the highest test accuracy of 90.55%.

```
ensemble_acc = evaluate_ensemble(model_l2, model_l3, test_loader)
print("Level-4 Ensemble Test Accuracy:", ensemble_acc)
```

```
Level-4 Ensemble Test Accuracy: 0.9055130915596031
```

---

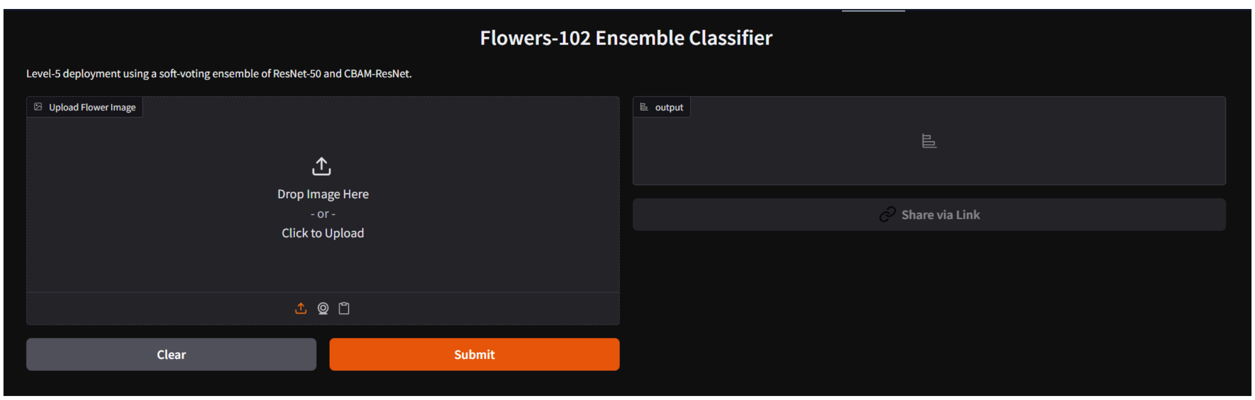
6. Level 5: Production-Ready Deployment

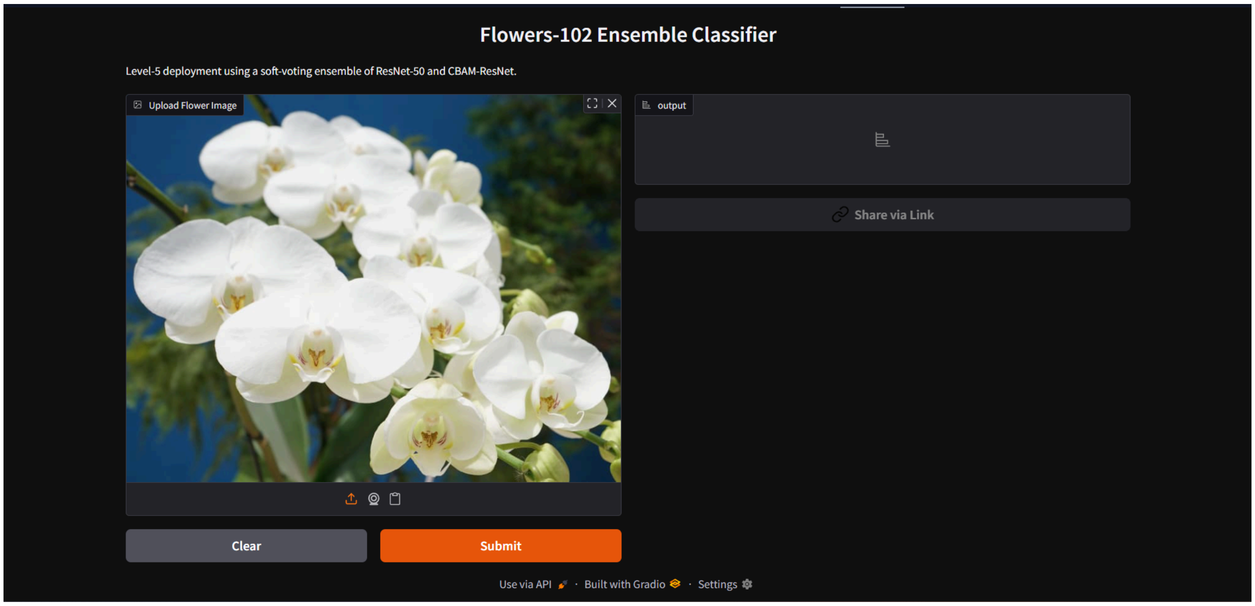
Objective

To transition the trained model into a functional, user-facing application, demonstrating end-to-end machine learning engineering.

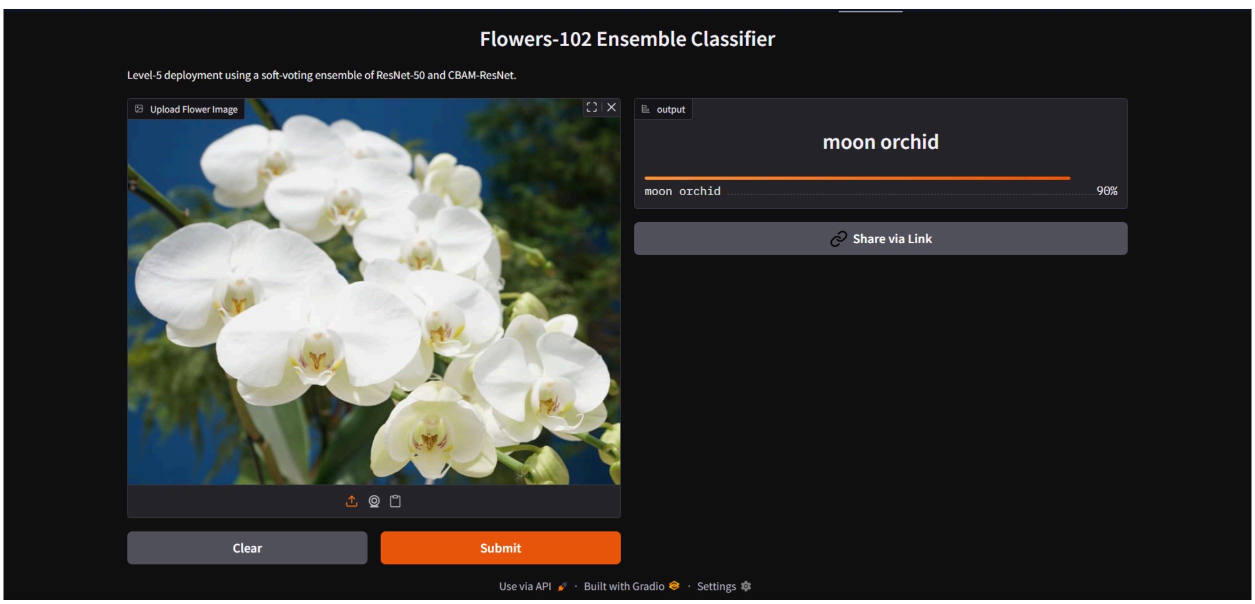
Implementation

- **Framework:** An interactive Gradio application was developed.
- **Hosting:** Deployed via Hugging Face Spaces for public accessibility.
- **Features:** Supports image uploads, processes inference through the ensemble pipeline, and returns human-readable flower names.





Instruction to use :
-Download any from image the given 102 categories of flowers
-Visit the [LINK] and upload the picture
-Click on the 'Submit' button to see the prediction.



---

7. Results and Comparative Analysis

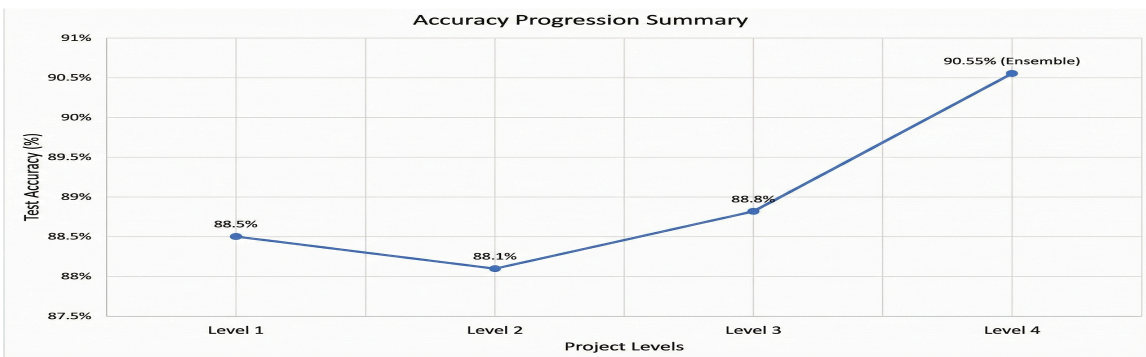
Accuracy Progression Summary

Level	Description	Test Accuracy

Level 1	Baseline ResNet-50	88.5%
Level 2	Augmented + Fine-Tuned	88.1%
Level 3	CBAM Attention Model	88.8%
Level 4	**Soft-Voting Ensemble**	**90.55%**

Accuracy Progression Graph

The following graph visualizes the impact of each architectural and training improvement throughout the project lifecycle.



---

8. Constraints and Conclusion

Challenges and Mitigations

The project was executed under strict compute and time constraints. Limited GPU access necessitated the use of transfer learning and selective fine-tuning rather than training large models from scratch. Time constraints led to a focus on methodological improvements (Attention and Ensembling) over exhaustive hyperparameter tuning.

Conclusion

This project demonstrates a rigorous progression from a baseline transfer learning approach to a production-ready ensemble system. By integrating attention mechanisms and ensemble methods, the final model achieved a 90.55% accuracy on a challenging fine-grained dataset, proving that strategic architectural choices can overcome limitations in data size and compute power.

Name : Abhishek Choudhary

USAR [AIML]