Computer Vision Based Flower Classification

Jitesh Parapoil, Rushikesh Kankar, Saicharan Reddy Banda, Avinash Compani

December 06, 2022

Abstract

In the world of flower classification, it's quite tough yet interesting to correctly identify and sort a huge range of flower types just by looking at them. Our project tackles this issue by using deep learning, focusing on two tools called Vision Transformer (ViT) and DenseNet201. We worked with a collection of over 12,753 pictures of 104 different kinds of flowers. Using TensorFlow and TensorFlow Hub, we processed and augmented this dataset to enhance the diversity and robustness of the training system. The augmentation techniques, including random adjustments in brightness, contrast, hue, and saturation, were helpful in preparing the model to handle real-world variations in flower images. The main approach is the use of DenseNet201 integrated within TensorFlow's Sequential API, enabling us to utilize its deep convolutional network capabilities. This choice was based on DenseNet's proficiency in managing vanishing gradient problems and its efficiency in feature propagation which is perfect for figuring out the flower types. Apart from this we also trained the Vision Transformer model, known for its ability to handle sequential data and capture global context effectively it is considered an important aspect when dealing with the subtle differences and tiny details that makes each flower species unique. Through thorough training, validation, and testing, we achieved good accuracy, demonstrating the effectiveness of these advanced models in classifying complex image data. The project contributes to the field of botanical research by providing a tool for accurate flower species classification and also showcases how machine learning can help us make appreciate nature and make sense of it in a better way.

1. Introduction

1.1 Background and Motivation

Identifying flower species is as interesting as it is complex, involving a rich variety of shapes and minute differences. Traditionally, expert botanists have tackled this task with thorough study and deep knowledge. But now, techniques in machine learning and computer vision are paving new paths for automating this task, providing scalable and efficient methods. Our project is driven by the goal to use these to accurately sort more than 100 types of flowers, making the most of their power to pick out the tiny details in complex image data.

1.2 The Challenge of Fine-Grained Classification

Flower species classification falls under the domain of fine-grained visual classification (FGVC), where the differences between classes are subtle, yet critically important. This task is particularly challenging due to the high intra-class variation and low inter-class variation, meaning that individuals of the same species can look quite different, while different species can appear very similar. The Challenge is with training a model that can not only recognize these subtle differences but also generalize well across a diverse set of images.

1.3 Leveraging Deep Learning Models

Recent advancements in deep learning, especially in models like the Vision Transformer (ViT) and DenseNet, have shown great potential in undertaking FGVC tasks. The ViT model, known for its effectiveness in processing sequential data, captures global context and complex details within images. On the other hand, DenseNet201, a convolutional network, is known for its efficiency in feature propagation and managing vanishing gradients. These characteristics are important when dealing with the fine patterns and features of flower images.

1.4 Objective and Scope

Our aim is to create a reliable and precise system that can recognize over 100 flower species from pictures. With a large dataset of 12,753 images, we plan to show how well-advanced deep learning models can perform on this intricate task. Our project covers preparing the data, training the models, evaluating their work, and analyzing the results to achieve high accuracy and a deeper understanding of how the models do their job.

1.5 Project Structure and Approach

We prepared the dataset with TensorFlow, adjusting it to fit the model’s needs. To make the dataset more versatile and to prevent the model from being too narrow in its learning, we applied various data augmentation techniques. We then concentrated on training the DenseNet201 and Vision Transformer models, assessed their performance, and compared how well they classified flower species.

2. Method

Our approach to classifying over 100 flower species utilized a combination of advanced machine learning techniques, focusing on the Vision Transformer (ViT) and DenseNet201 models. This section details the methodology, from dataset preprocessing to model architecture and training.

2.1 DenseNet201

For our initial architecture, we utilized TensorFlow's implementation of DenseNet201, a deep convolutional network known for its efficiency, ability in feature propagation and gradient flow. The model was setup to include Global Average Pooling and a Dense output layer with 104 units, corresponding to the number of flower species, with softmax activation for multi-class classification.

2.2 Vision Transformer (ViT)

The Vision Transformer (ViT) represents a transformative approach to image processing, departing from the conventional convolutional neural networks (CNNs). Introduced by Google Research, ViT adopts the foundational Transformer architecture originally designed for natural language processing tasks. In ViT, images are treated as sequences of fixed-size non-overlapping patches, resembling tokens in language models. Each patch undergoes linear embedding, forming a sequence input for the Transformer model. Implemented and trained the ViT model, processing images as sequences of patches. For this current implementation, we made use of **vit.vit_l32** from **vit_keras**, configured for the image size, with sigmoid activation and a pre-trained top layer for 104 flower classes.

2.3 Transfer Learning Application:

Our project harnesses transfer learning to boost our models' performance. This method repurposes models pre-trained on large datasets, like ImageNet, to our specific task of flower classification. By fine-tuning these established models, we efficiently adapted them to recognize the nuances of different flower species, thus accelerating the training process and enhancing accuracy. Transfer learning proved essential, especially with our relatively smaller dataset, by providing a depth of pre-learned visual understanding, which significantly contributed to the effectiveness of the classification system.

2.4 Iterative Model Improvements

We initially focused on data preprocessing and augmentation techniques. Upon observing the models' performance, we iteratively fine-tuned hyperparameters and model configurations to improve accuracy and generalization capabilities. We leveraged transfer learning in both Vision Transformer (ViT) and DenseNet for tailored adaptation to image tasks. By pre-training on large datasets like ImageNet and fine-tuning on target data, we efficiently customized the models. This approach accelerated convergence, enhanced generalization, and optimized performance even with limited target data.

One major challenge was handling the high intra-class variation within the dataset. We addressed this by implementing a custom learning rate scheduler and employing data augmentation techniques like brightness and contrast adjustments.

3. Experiments

This section outlines the experimental setup, procedures, and methodologies applied to evaluate the performance of the models – DenseNet201 and Vision Transformer (ViT) – in classifying over 100 flower species.

3.1 Dataset Preparation and Preprocessing

The basis of our project is a diverse dataset comprising over 12,753 images of various flower species. We obtained this dataset from an online resource, Kaggle, and pre-processed each image to ensure it was ready for the models. This preparation involved key steps such as resizing images, normalization and data augmentation. Resizing involves standardizing all images to a uniform size of 224x224 pixels to match the input size requirements of the DenseNet201 and ViT models. In normalization, we scale the pixel values to a range of 0 to 1. This normalization serves in the convergence of the training process by stabilizing the gradients. Finally, we enhance the model's ability to generalize and to mitigate overfitting by implementing data augmentation techniques. These included random adjustments in brightness, contrast, flipping, hue changes etc.

3.2 Training and Validation

3.2.1 DenseNet201

The DenseNet201 architecture from TensorFlow was pre-trained on ImageNet and was further employed for fine-tuning on a flower dataset. The model underwent optimization using the 'adam' optimizer and 'sparse_categorical_crossentropy' as the loss function. Training spanned 30 epochs with a batch size of 32, striking a balance between computational efficiency and training duration. A custom learning rate scheduler was implemented, starting at a base rate of 3e-5 and dynamically adjusting based on epoch progress for improved training.

For validation, a distinct set of images was reserved from the training dataset, allowing an assessment of the model's generalization to unseen data. Key metrics, including validation loss and accuracy, were monitored after each epoch to evaluate model performance.

3.2.2 Vision Transformer (ViT)

In this phase, the ViT model, initially pre-trained, was tailored to suit the characteristics of the dataset. The model's unique approach processes images as sequences of patches, enabling the capture of global dependencies within each image. The training methodology mirrored that of DenseNet201, implementing similar optimization strategies. Additionally, specific adjustments were made to the ViT model configuration, ensuring compatibility with the flower dataset's distinct requirements, such as input size and class count. To assess performance consistently, a validation strategy similar to DenseNet201 was employed, using a separate dataset for ViT's validation phase.

This approach allowed for a comprehensive evaluation of the model's adaptability and generalization capabilities.

3.3 Results

This section presents the findings from our experiments and discusses the implications of these results. We analyze the performance of both the DenseNet201 and Vision Transformer (ViT) models and provide insights into their effectiveness in classifying flower types.

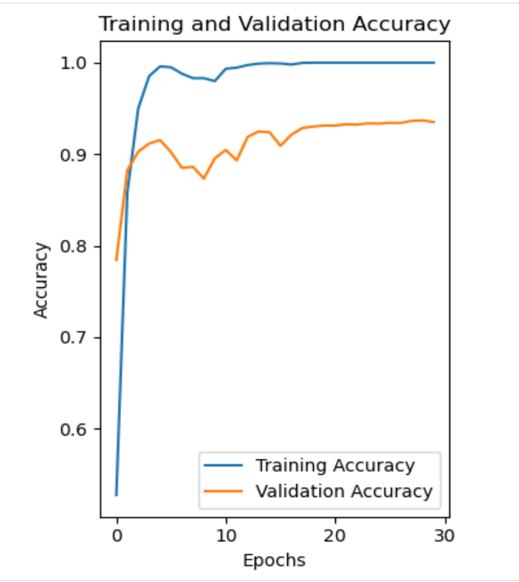
3.3.1 DenseNet201:

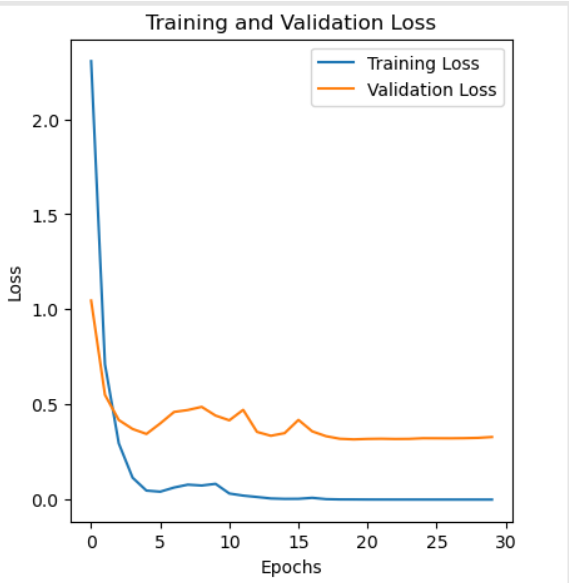
- **Training Loss**: Demonstrated a significant decrease over epochs, indicating effective learning and optimization. The final training loss was recorded at 0.00015437, showcasing the model's ability to minimize the error in predictions during the training process.

- **Validation Loss**: Showed consistent improvement with a final value of 0.32912, suggesting the model's good generalization capabilities when presented with unseen data.

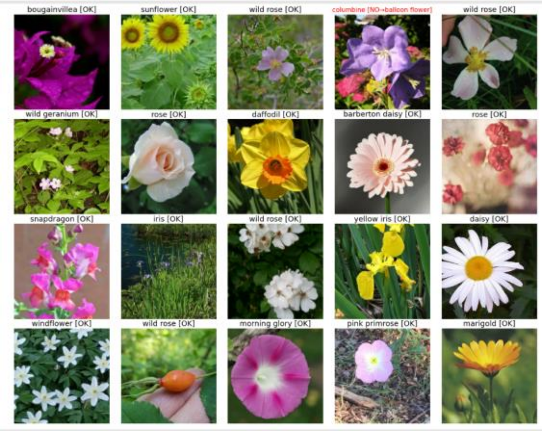
- **Training Accuracy**: Achieved a remarkable final training accuracy of 100.00%, pointing towards a perfect fit on the training dataset.

- **Validation Accuracy**: Attained a high validation accuracy of 93.51%, reflecting the model's robustness and reliability in classifying flower species accurately when validated against unseen data.





- A subset of images from the test dataset was processed through the DenseNet201 model. The model's predictions were put together with the actual labels to provide a qualitative evaluation of its classification accuracy.



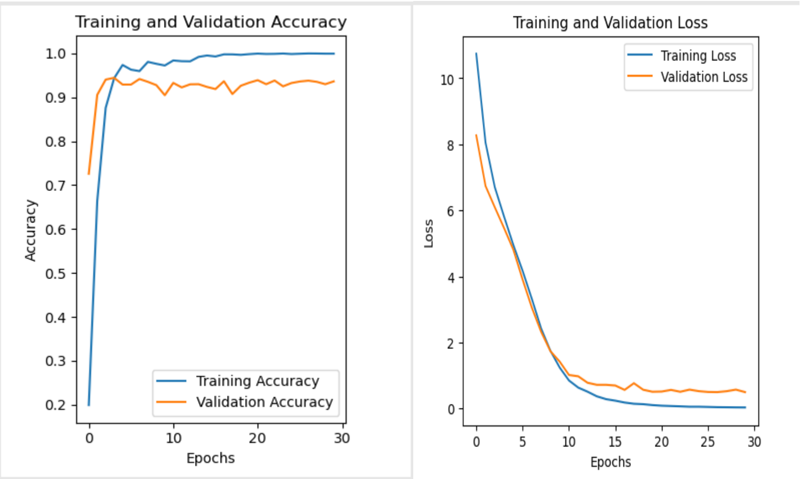
3.3.2 Vision Transformer (ViT):

- **Training Loss**: Shows progressive reduction to a final training loss of 0.03266023, confirming the model's effectiveness in learning from the training dataset.

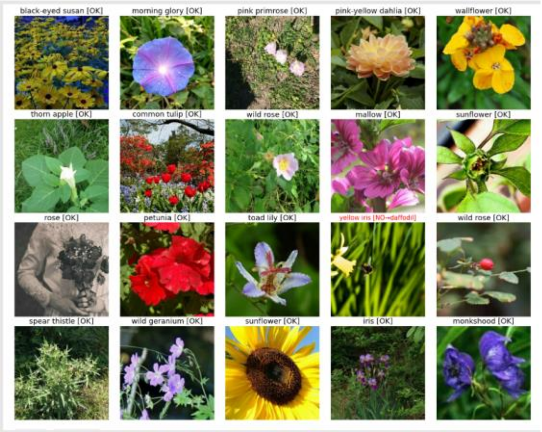
- **Validation Loss**: Recorded a final validation loss of 0.49732, which, though it is higher than the training loss, still indicates a strong performance on the validation set.

- **Training Accuracy**: Showed substantial improvement over training epochs, with a final training accuracy of 99.92%. This nearly perfect score demonstrates the model's capability in learning detailed features and patterns specific to the training dataset.

- **Validation Accuracy**: Achieved a final validation accuracy of 93.62%, which, similar to DenseNet201, validates the model's ability to generalize well to new, unseen images.



- A subset of images from the test dataset was processed through the ViT model. The model's predictions were put together with the actual labels to provide a qualitative evaluation of its classification accuracy.



4 Analysis and discussions

DenseNet201 demonstrated high accuracy and consistent performance, indicating its suitability for fine-grained classification tasks. The model's depth and feature propagation capabilities were instrumental in effectively distinguishing subtle differences between flower species. Similarly, the Vision Transformer (ViT), with its approach to processing images as sequences of patches, showcased potential in capturing global dependencies within images. The unique architecture of ViT suggested its effectiveness in recognizing fine variances among different flower species.

The results we’ve obtained from testing and performance evaluation confirm that both DenseNet201 and ViT not only learned the training data effectively but also generalized well to new data. The loss metrics corroborate the accuracy scores, providing a holistic view of the models' performance dynamics.

The success of both DenseNet201 and ViT offers valuable insights for future applications in botanical research and other fields requiring detailed image analysis. These findings highlight the potential of advanced deep learning models in managing tasks that require a experts.

5. Conclusion

Both DenseNet201 and ViT have exhibited remarkable effectiveness in fine-grained classification of flower species. DenseNet201 achieved notable training and validation accuracies of 100.00% and 93.51%, respectively, while ViT showcased impressive learning and generalization capabilities. The consistent high accuracy maintained by both models on the test dataset underscores their robustness and reliability in accurately classifying a diverse array of flower species. Notably, data augmentation emerged as a vital factor contributing to the enhanced performance of both models, effectively mitigating overfitting and ensuring robustness against diverse image variations.

In summary, this project stands as a robust demonstration of the power of deeplearning models in transforming the approach to complex classification tasks. The integration of deep learning in botanical research not only facilitates more efficient research methodologies but also broadens our understanding of the natural world.

6. Contribution

Jitesh Parapoil:

- Led the creation and improvement of DenseNet201model.
- Performed detailed testing on DenseNet201model.
- Produced visual data graphics for analysis and for the project report.
- Played a key role in writing and compiling the full project documentation.

Rushikesh Harikishan Kankar:

- Led the creation, improvement, and testing of Vision Transformer
- Directed the initial data analysis, influencing the choice and training of models.
- Tuned model settings for Vision Transformer model and monitored ongoing results.
- Created visuals of the models' results for easy understanding and reporting.
- Helped in writing detailed documentation throughout the project.

Saicharan Reddy Banda:

- Defined the project goals and planned how to achieve them.
- Tuned model settings for DenseNet201model model and monitored ongoing results.
- Wrote scripts to prepare data for use in model training.
- Watched over model training results to ensure accuracy.
- Managed the merging of different parts of the project report into a unified document.

Avinash Compani:

- Conducted in-depth research on model design and application.
- Oversaw the training of models, using the latest techniques.
- Explored the detailed design of both DenseNet201 and Vision Transformer models.
- Made substantial contributions to the writing of the project report, ensuring detailed and accurate methodology and results were presented.

7. References

[1]. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://arxiv.org/abs/1608.06993

[2]. Junyu Chen, Yufan He, Eric C. Frey, Ye Li, Yong Du(2021). ViT-V-Net: Vision Transformer for Unsupervised Volumetric Medical Image Registration. https://arxiv.org/abs/2104.06468

[3]. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Uszkoreit, J. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In International Conference on Learning Representations (ICLR). https://arxiv.org/abs/2010.11929

[4]. TensorFlow. (n.d.). TensorFlow Hub: A repository of trained machine learning models. Retrieved from https://www.tensorflow.org/hub

[5]. Kaggle. (n.d.). Flower Classification on TPU. Kaggle competition dataset. Retrieved from https://www.kaggle.com/c/flower-classification-with-tpus

[6]. Chollet, F. et al. (2015). Keras. https://keras.io/

[7]. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Kudlur, M. (2016). TensorFlow: A system for large-scale machine learning. In 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16). https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi

[8]. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://arxiv.org/abs/1512.03385

[9]. Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. In International Conference on Learning Representations (ICLR). https://arxiv.org/abs/1412.6980