Petals to the Metal: Flower Classification

Shailja Kartik
Artificial Intelligence
San José State University

Xinyu Li
Software Engineering
San José State University

Paul Mello
Artificial Intelligence
San José State University

Rodrigo Teixeira
Software Engineering
San José State University

shailja.kartik, xinyu.li, paul.mello, rodrigo.colassoteixeira @sjsu.edu

***Abstract*—As we begin a new era of mass extinction brought on by human influence, it is of increasing importance to document the species of the world before they disappear. The Petals to the Metal Kaggle competition challenge provides a solution to this by allowing anyone to build models to solve this classification problem. This has led to extensive research in image classification to digitize and process this documentation. This has resulted in many interesting methods being developed to provide better generalization and higher accuracy. Methods like Gaussian noise, random erasing, and flipping have all become necessary image augmentation tools to achieve top 10% accuracy's. In order to achieve our 97% accuracy model we incorporate some of these methods along side and ensemble of transfer learning models. These transfer learning models consist of using pre-built models capable of broad generalization and further training them to be tuned to our own data set. The competition provides an extensive data set of 12,000 training images for 104 classes of flowers. Despite this, we used an additional external data set which increased our total training to 80,000 images and even attempted multiplying the data further. Overall, we achieved a 12/162 ranking in the Kaggle competition and provided a well trained model for these images.**

***Keywords*— Machine Learning, Data Augmentation, Transfer Learning, Flower Image Classification**

I. Introduction

Global warming continues to destroy ecosystems across the planet. Many animals and plants are unable to survive the changing conditions. As a result, the scientific community has begun a rapid process of collecting images of flora from the Americas and Oceania before these species go extinct. In this process of collection, many flowers which are not well documented have become digitized. There are many flowers which are not categorized into any species and thus the aim of this paper is to build a classification model to classify them. We use a Kaggle data set which contains a total of over 23,000 total images total and over 100 unique species. We also incorporate an additional data set consisting of 68,000 training images to achieve our ranking.

II. Literature Review

Image classification has exploded in recent years thanks to recent advancements in hardware with the rise of Google's TPU architecture. Thanks to this rise in research, many new image classification methods have been developed. The following paragraphs contain a literature review of new methods which have been introduced to improve generalization and training of our models. These methods include data augmentation, transfer learning and increasing training data. Data augmentation consists of random erasing [1], adding noise [2], and other prominent methods in the field. Transfer learning helps to improve training accuracy by creating a general model which can then be used as a starting point for other researchers to train their models from. The key here is that the transferred model is trained on a larger data set and then introduced to a narrower data set. In doing so, the model is able to use its already generalized understanding and reinforce minor differences between each class in the training data. Noise is a general principle which can take on many forms including random erasing, cropping, and rotation. These image transformation techniques introduce random augmentations that distort the expected input and cause the model to create inferences about the data. Data augmentation techniques to introduce noise in the training data are able to equip models with a more abstract representation of the target images. This can help to improve predictions made by the model by reducing the required information necessary to distinguish between classes. Random erasing was a method first developed in 2017 by a research group from the University of Technology, Sydney. [1] This data augmentation technique proposes the use of a system which replaces a random section of the original image with noise. A random width and height for each box helps to further augment the data in unique ways. The theory behind this paper states "...training images with various levels of occlusion are generated, which reduce the risk of overfitting and makes the model robust to occlusion." [1] This process improves model generalization capabilities. With consistently better performance than other data augmentation tools, random erasing is a strong and necessary method for image classification and can be seen in figure 1.

Fig. 1. Data Augmentation: Random Erasing

Initial research demonstrated the sparse nature of our challenge data was affecting the quality of generalization in our modeling. With this in mind we elected to find and include other data sets which are applicable to the challenge goals. In our research we found another data set that contains 68,100 images, and incorporated it into our training data to improve our models capabilities. Transfer learning is a powerful tool for machine learning at large, but an especially powerful method in image classification. In a paper published in the International Conference on Image, Vision and Computing (ICIVC), they propose the use of transfer learning using Google's Inception-v3 model on flower classification to achieve 95% accuracy. [3] In principle, transfer learning consists of a model trained on a far larger data set with a wide range of images to introduce better generalization capabilities. Researchers design and train these models before releasing them to the public to use as pre-built models in their own studies. This form of modeling creates a starting architecture which, thanks to its broad generalization, is able to be further tuned to predict images based on another data set. In our initial exploration transfer learning quickly became identified as a necessity to achieve high accuracy rates. With the materials and knowledge gathered for modeling we set out to incorporate as many of these data augmentation techniques into our models. We started with a simple model without many augmentations to our data and achieved 81% accuracy. After the incorporation of these methods and upgrading our models to a transfer learning based model, we were able to successfully achieve 97.33% accuracy when using an ensembled method the transfer learning models EfficentNet and DenseNet. This improvement is due in part to the extensive data augmentation we conducted and that of increasing the epoch count to ensure enough random walking occurs in the generalization pit.

III. Data Exploration and Preprocessing

We elected to use a data set consisting of the competition data as well as external data. Each contains a wide range of image data with dimensions of either 192, 224, 331, or 512 pixels in both width and height. To retain the most information in the system we elected to use the 512x512 data set. Despite the increased computational cost, initial tests demonstrate the model performs better at generalizing with the additional information. The challenge data set contains a total of over 23,000 images and is broken down into the following split consisting of 12,750 training images, 3,700 validation images, 7,350 unlabelled images, and over 100 unique species. Our external data set contains an additional 68,100 training images which will directly help in improving training and validation accuracy.

Fig. 2. Data Set

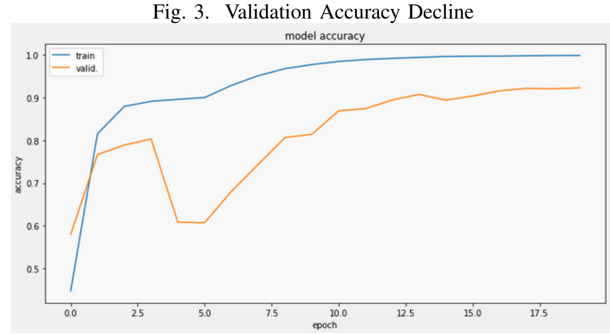
Figure 2 demonstrates the cleanliness of the images and a small portion of the classes which are present in the data set. Particularly, there is good clarity, lighting, and focus on each flower. Despite this, some images are slightly distorted such that colors of the flower are lost to a gray-scale image. Despite this, our validation data also contains a few gray-scaled images, so we decided to leave it as it was. Our external data also has the same properties which was good for validation accuracy. Preliminary testing demonstrated the importance of data augmentation techniques and transfer learning. Initial models were trained with and without these transformations. We found that any models without image augmentation resulted in a 20% decrease in accuracy when compared to the same model with data augmentation. Expanding the quantity of training data also resulted in a profound improvement on generalized accuracy by increasing the number of images in each class. This ratio of training data to classes directly improves generalization in the model.

IV. Experiments

1) Problem Space: We conducted a lot of preliminary testing to ensure that our models were capable of meeting our desired accuracy goals. Our literature review provided us with significant insight into the tools which would be necessary to increase our chances of success. We found techniques such as transfer learning models, data augmentation, ensembled models, and hyperparameter tuning. Subsequently, we incorporated any methods which helped to improve our chances of success. We found that with the right learning rate schedulers and proper data augmentation combinations we could significantly raise the performance of our transfer learning models.

2) Model Selection: Preliminary testing demonstrated that a high performing model would need to use transfer learning. For this we tested wide but shallow networks and compared them to deep but narrow networks. We found that, on average, deeper networks tended to perform better, but a mixture of these two types of models would be ideal. Models like EfficentNet and DenseNet provided a good balance between these two properties. We tested other models like Xception and ResNet, but they did not provide similar validation accuracy's. In most cases we found validation accuracy significantly dropped before slowly rising. Given the training times incurred by such a massive data set it was infeasible to wait for validation to improve once it significantly declined for some models. In figure 3 we demonstrate one of these declines in validation accuracy.

Fig. 3. Validation Accuracy Decline

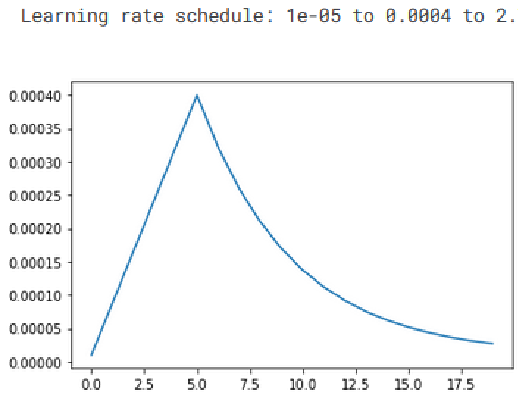


Despite this initial transfer modeling, we noticed that the incorporation of different variations of data augmentations made huge differences in the generalization capabilities of these models. Meaning that, models were sensitive to the type of data augmentation techniques we added. This resulted in some extensive research to find what combination of augmentations would provide the best accuracy's. Over time we found that flipping, random erasing, and saturation were necessary techniques while sharpness hurt model performance. This is likely due to the cleanliness of the validation data set.

Using a linear rate scheduler, such as the one demonstrated in figure 4, we were able to improve the tractability of our model. This is due to ensuring that the learning rate is low enough that the massive batch size can help learn at a reasonable rate. Interestingly, these principles did not apply to our tested model optimizers. Where a constant learning rate of .00001 proved to be the most effective when compared to other magnitudes of learning rates.

After extensively testing many models and training for many hours we found that EfficentNet and DenseNet, in an ensembled prediction method, would provide the strongest accuracy. The learning rate scheduler demonstrated in figure 4 and Adam optimization was used

Fig. 4. LR Scheduler for Model Fitting

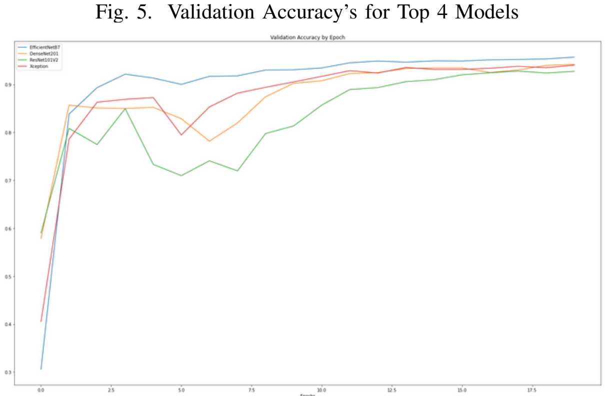


for both models. We trained each on the 80,000 total image training images from the combined Kaggle and external data set. These 80,000 images were filtered through our data augmentation method and provided an increase in generalization abilities. In this culmination of work we present a model which is 97.33% accurate. Despite this the model requires training on TPU's due to the size of both the data's dimensions and its quantity.

V. Results

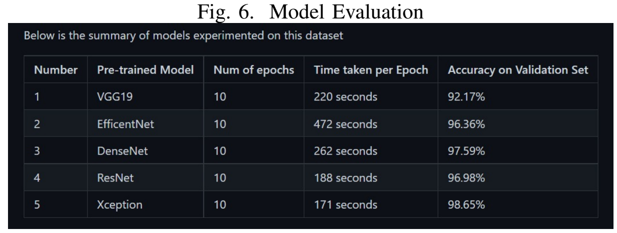
1) Evaluation: We evaluated our models on a few key metrics. Necessary training time and validation accuracy. For a majority of our models, they required a minimum of 3 minutes of computation per epoch. Using 10 epochs as a training run we found that models like Xception provided the highest training accuracy at 98.65% but failed at validation accuracy's. DenseNet simultaneously provided a good mixture of both metrics taking 4 minutes and 30 seconds per epoch and achieved 96% training accuracy. Despite these longer results, our model was able to generalize better and reached significantly better results for the validation data. In figure 5 we demonstrate the overall validation accuracy we achieved when running for 20 epochs. While in figure 6 we demonstrate the final results of the model evaluations from our preliminary testing.

Fig. 5. Validation Accuracy's for Top 4 Models



2) Results Analysis: The Kaggle competition uses a macro F1 score to determine the success of our models prediction abilities. We achieve 97.33% validation accuracy and demonstrate that success using a confusion matrix consisting of each species of flower. Figure 7 demonstrates this accuracy matrix. We can see that a majority of our predictions are accurate leading to an F1 score of 95%.

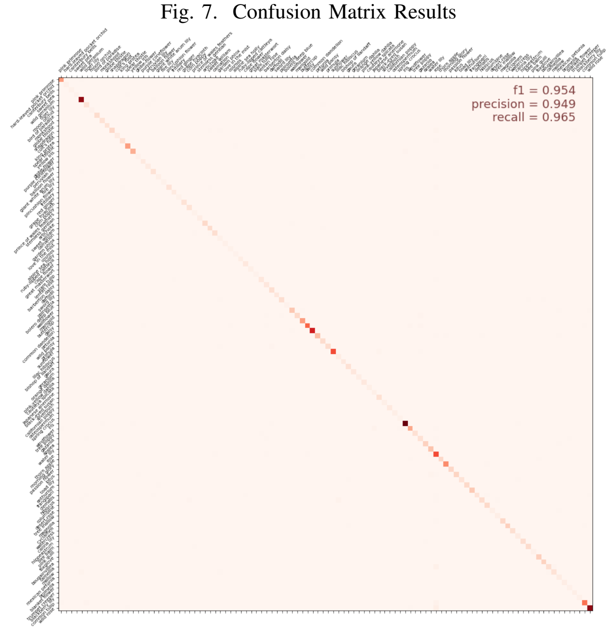
Fig. 6. Model Evaluation



Below is the summary of models experimented on this dataset

Number	Pre-trained Model	Num of epochs	Time taken per Epoch	Accuracy on Validation Set
1	VGG19	10	220 seconds	92.17%
2	EfficentNet	10	472 seconds	96.36%
3	DenseNet	10	262 seconds	97.59%
4	ResNet	10	188 seconds	96.98%
5	Xception	10	171 seconds	98.65%

Fig. 7. Confusion Matrix Results



3) Visualization: In figure 8 we demonstrate our models capabilities for proper image classification by using a common method proposed within the Kaggle competition. As we can see the model can accurately identify each flower in the validation data. Although only a small subsection, it demonstrates the models capabilities.

VI. Conclusion

We experimented with several pre-trained models to determine which one would succeed in generalizing on unseen data the best. We also used a range of data augmentations to provide variation so that our model could learn new features and get more comfortable with them. Finally, by combining an ensemble of transfer learning models, data augmentation approaches, and hyperparameter tweaking, we were able to attain a top 10% rating. All of the models we tested had their own advantages and performed well on the Validation set. However, the top model that produced this result was an ensemble of EfficientNet and DenseNet trained on both the original and external data sets and provided a score of 97.334% for the Kaggle testing set. This resulted in our models finishing 12th out of 162 other entries.

VII. Future Work

Although there is still space for development, the Kaggle competition's constraints restrict lengthy training. Another limitation was the size of the original data set. The external data set was the only reason we could attain such high accuracy with this classification model. The model classification process will undoubtedly improve if we gather more accurate training data sets with diverse orientations and angles over time. As a result, we anticipate that using additional techniques like data set multiplication and extending training duration can increase validation accuracy. With technological improvements,

Fig. 8. Validation Image Grid

new variables and metrics for evaluating the fitness and usefulness of a model for flower categorization are likely to emerge. Overall, we feel we have contributed significantly to the challenge of flora picture categorization in a collapsing environment. We believe that, with increased research, we will be able to document these species more quickly and precisely before they become extinct.

References

[1] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," *CoRR*, vol. abs/1708.04896, 2017. [Online]. Available: http://arxiv.org/abs/1708.04896

[2] C.-H. Wu and H.-H. Chang, "Gaussian noise estimation with superpixel classification in digital images," pp. 373–377, 2012.

[3] X. Xia, C. Xu, and B. Nan, "Inception-v3 for flower classification," pp. 783–787, 2017.

[4] N. R. Gavai, Y. A. Jakhade, S. A. Tribhuvan, and R. Bhattad, "Mobilenets for flower classification using tensorflow," pp. 154–158, 2017.

[5] N. Alipour, O. Tarkhaneh, M. Awrangjeb, and H. Tian, "Flower image classification using deep convolutional neural network," pp. 1–4, 2021.

[6] J. Ball, "Few-shot learning for image classification of common flora," *CoRR*, vol. abs/2105.03056, 2021. [Online]. Available: https://arxiv.org/abs/2105.03056

[7] Y. Rao, W. Zhao, Z. Zhu, J. Lu, and J. Zhou, "Global filter networks for image classification," *CoRR*, vol. abs/2107.00645, 2021. [Online]. Available: https://arxiv.org/abs/2107.00645

[8] H. Touvron, M. Cord, A. Sablayrolles, G. Synnaeve, and H. Jégou, "Going deeper with image transformers," *CoRR*, vol. abs/2103.17239, 2021. [Online]. Available: https://arxiv.org/abs/2103.17239