In-Depth Analysis of Dual Convolutional Neural Networks for Flower Identification and Segmentation using the Oxford 17-Flower Category Dataset

Prakhar Prakarsh
Computer Vision ,Machine learning in Science Department,University of Nottingham
United Kingdom
ppxpp1@nottingham.ac.uk

***Abstract*—This detailed investigation sheerly focuses on the development and evaluation of two completely distinct Convolutional Neural Networks (CNNs) specifically tailored for the task of classifying and segmenting floral images using the widely recognised Oxford 17-Flower Category Dataset. The primary CNN is meticulously crafted to identify all 17 flower categories present in the dataset, while the secondary CNN is exclusively customised to the segmentation of a particular class the daffodil, as a case study.**

The research digs deep into the complexities of the architectural designs, training methodologies, and performance metrics associated with both the CNNs, offering a broad understanding of their workings. Moreover, the study goes on exploring the potential and latent applications and its possible feasible implications of these neural networks in the larger context of image recognition and classification, which presents an exceedingly rigorous account of the development, evaluation, and potential impact of these two unique Convolutional Neural Networks in the field of flower classification and segmentation. This research not only focuses on the the abilities of the subject matter, but also extends the very idea to its future advancements making it more robust.

KEYWORDS

Convolutional Neural Networks, flower classification, segmentation, Oxford 17-Flower Category Dataset, daffodil case study, architectural designs, training methodologies, performance metrics, image recognition, data augmentation, CNN1, CNN2, U-Net architecture, evaluation metrics, real-world applications, deployment considerations, generalizability, transferability.

I. INTRODUCTION

For evaluating image classification and segmentation algorithms, the Oxford 17-Flower Category Dataset[1] consists of 1,360 images, each representing one of 17 unique floral species(refer to Fig. 1). This extensive and rigorous study solely aims to develop and assess the performance of the two distinct Convolutional Neural Networks (CNNs)[2] with separate purposes: the first , sheerly focuses on classifying all 17 flower categories, while the second targets the segmentation of daffodils in particular[3]. This research looks into, whether the strategic selection of image datasets can enhance object classification, especially for categories that exhibit significant visual similarities while maintaining noticeable variation inside the class.

This research looks into, whether the strategic selection of image datasets can enhance object classification, especially for categories that exhibit significant visual similarities while maintaining noticeable variation inside the class. To this end, a new dataset has been compiled, featuring a diverse range of flower species. The task of flower classification poses a significant challenge, as it is markedly more difficult than differentiating between a car, boat, or human[4]. Typical flower images exhibit extensive variations in viewpoint, scale, illumination, partial occlusions, multiple instances, and other factors. Furthermore, cluttered backgrounds amplify the challenge, as there is a risk of mistakenly classifying the background content instead of the flower itself.

The most significant obstacle stems from the intra-class versus inter-class variability—meaning that the variation within a class is smaller than that between different classes, yet subtle differences between flowers determine their classification. Factors that distinguish flowers can range from shape, size, and colour to unique texture patterns[5], with a combination of these aspects often playing a crucial role. Hence, the crux of this challenge lies in discovering an appropriate representation for these aspects, as well as a relatively robust method of combining them that retains the discreteness of each aspect rather than averaging over them. Nevertheless, flower species frequently display multiple values for each aspect, indicating that any class representation must adopt a "multi-modal" approach.

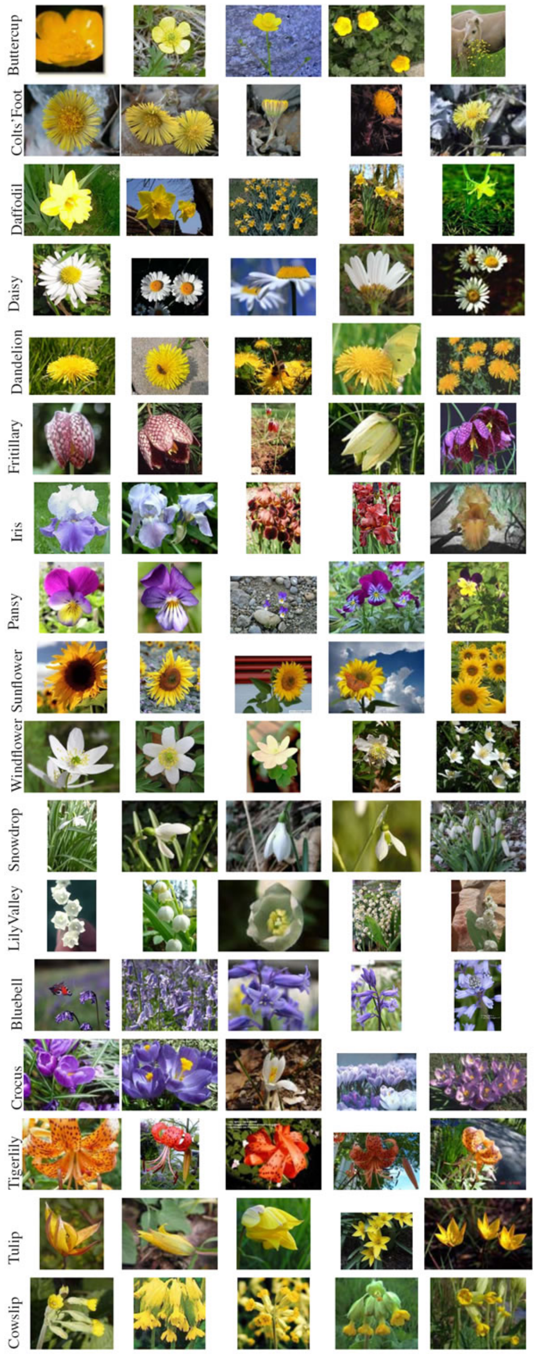

Fig. 1. Dataset image[1]

II. Method

A. CNN Architectures

This undertaking encompasses two distinct Convolutional Neural Network (CNN) designs , both of which have been sculpted from scratch , each scrupulously crafted to address the challenges associated with classification and segmentation. Both the networks are developed from scratch and are composed of several fundamental building blocks, which include convolutional layers responsible for feature extraction, activation functions that introduce non-linearity, pooling layers that reduce spatial dimensions while preserving essential features, and fully connected layers dedicated to high level reasoning and final output generation. By embodying these necessary components, the two CNN architectures demonstrate proficiency in their respective domains of image classification and segmentation.

B. CNN1: Classification

Examining the primary Convolutional Neural Network (CNN) in this study, denoted as CNN1

$$(I * K)[i, j] = \sum_m \sum_n I[i - m, j - n] * K[m, n] \qquad (1)$$

where I is input image/feature map K is the kernel , the element-wise products of the input image and kernel values for all positions covered by the kernel results in a feature map that highlights specific features or patterns in the input image as defined by the filter/kernel. So the CNN1 reveals a carefully made design which is aimed at categorising 17 distinct flower species. The architecture of this network features an assembly of numerous convolutional layers, arranged sequentially to optimise performance[2]. Following each of these layers, a Rectified Linear Unit (ReLU) activation function as shown below in the equation,

$$f(x) = \max(0, x) \qquad (2)$$

is employed which significantly enhances the network's non-linearity and, as a result, boosts its capacity to identify hard to recognise patterns within the input data[13].

Besides convolutional layers and activation functions, the CNN1 architecture also includes max-pooling layers, which function to efficiently reduce the spatial dimensions of the input while preserving the most essential features[13]. This approach enables the network to maintain an effective and accurate recognition process throughout its operation.

Reaching the pinnacle of the CNN1 architecture, a series of fully connected layers can be found, all of which are directly linked to a softmax activation function,as expressed below .

$$\sigma(z)_i = \frac{\exp(z_i)}{\sum_j \exp(z_j)} \qquad (3)$$

This critical component is responsible for generating the probability distribution across the various flower categories present in the dataset. By adopting this systematic methodology, the network is well-equipped to detect subtle patterns and features within the input images. As a direct consequence, this trait significantly contributes to the network's improved performance in image classification tasks, enhancing it the capacity to accurately distinguish between diverse range of flower types[8].

C. CNN2: Segmentation

The second CNN, referred to as CNN2, is specifically made to segment daffodils within images. This network utilises an encoder-decoder framework that features skip connections which enables the capture of both high-level attributes and convoluted details. The architecture consists of convolutional

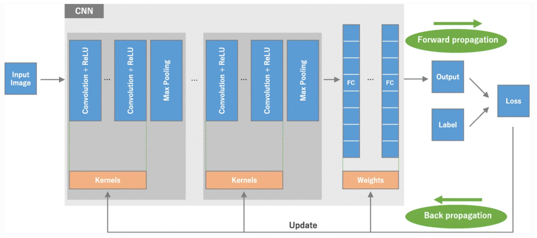

Fig. 2. A diagram illustrating CNN classification architecture[19]

layers, ReLU activation functions, max-pooling layers, upsampling layers, and a final sigmoid activation function responsible for producing a binary segmentation mask. These layers are eventually merged and returned as a layerGraph object.

The U-Net architecture[6] is favoured for semantic segmentation task purposes in computer vision due to its efficiency in retaining spatial information and preserving spatial resolution throughout the encoding and decoding stages. The inclusion of skip connections allows the model to simultaneously utilise low-level and high-level feature maps, thereby enhancing the segmentation outcomes. Also, compared to other fully convolutional networks, the U-Net architecture(example shown in the fig.3) possesses fewer parameters, resulting in faster training times and increased memory efficiency.

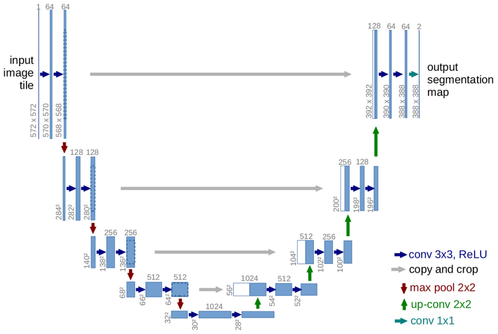

Fig. 3. An example of CNN segmentation architecture[6]

III. Training

During the training process for classification CNNs, the SGDM (stochastic gradient descent with momentum) optimizer[2] is employed, boasting a 0.9 momentum that surpasses Adam in terms of accuracy. This 0.9 momentum value has been selected based on its factual potency as a default setting in a multitude of deep learning applications. By accelerating SGD optimization convergence and attenuating oscillations, it increases the overall procedural results. The 0.9 value accumulates past gradients, balancing their impact with more recent ones, and in turn, smoothing the optimisation trajectory. While not a rigid guideline, this value has proven effective for a variety of tasks and models,specially in this case it has shown better results as compared to other values, but the optimal values can vary depending upon the specific problems and datasets .

To promote versatility of training and extend the prediction capacity, shuffle is set to occur every epoch. With a learning rate of 0.001 and a batch size of 32, the model undergoes training for 34 epochs. For segmentation CNN training, the Adam optimizer[7] is utilised with a 1e-4 learning rate and a batch size of 16. Similar to the classification CNN, this model is trained for 50 epochs, incorporating early stopping based on validation loss and shuffling every epoch to heighten complexity and improve training quality.

The fine-tuned hyperparameters for example learning rate, batch size, and training epochs using grid search and cross-validation methods. To improve generalisation and prevent overfitting, the employed data augmentation techniques (see Fig. 4), including random rotation, scaling, and flipping. These changes ultimately boosted the performance of the CNNs in both classification and segmentation tasks substantially.

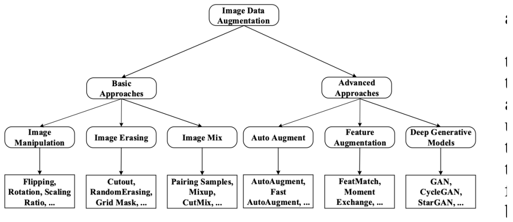
Image Data Augmentation
Basic Approaches
Advanced Approaches
Image Manipulation
Image Erasing
Image Mix
Auto Augment
Feature Augmentation
Deep Generative Models
Flipping, Rotation, Scaling Ratio, ...
Cutout, RandomErasing, Grid Mask, ...
Pairing Samples, Mixup, CutMix, ...
AutoAugment, Fast AutoAugment, ...
FeatMatch, Moment Exchange, ...
GAN, CycleGAN, StarGAN, ...

Fig. 4. Data augmentation techniques[18]

IV. Evaluation Metrics

Enhancing dataset diversity through data augmentation helps restrain overfitting while promoting generalisation. Utilising standardised 256x256 images strengthens performance, hinting at promising results for additional classification tasks using varied evaluation metrics[9].

The process of semantic segmentation consists of data preparation, training, and evaluation. Data is loaded and further divided, while U-Net architecture and training options are established.Gradually, the network undergoes training, with a validation set directed to assess its performance. The labeloverlay function facilitates result visualisation by superimposing predicted segmentation labels on input images, granting a lucid depiction of classified sections, which supports the evaluation of model performance.

The EvaluateSemanticSegmentation function calculates metrics, via comparison of predictions against ground truth data. These metrics include global accuracy, mean accuracy, mean IoU, weighted IoU, and mean BF score.

Favoured for semantic segmentation tasks, the U-Net architecture expertly captures both context and localisation. The Adam optimizer refines parameters while tracking progress. The selected techniques boast precision,and computational efficiency. U-Net enjoys widespread recognition in segmentation tasks, while Adam subtly handles sizable datasets and high-dimensional spaces. Data augmentation, early stopping[8], and evaluation metrics collectively guarantees an exhaustive performance assessment .

V. Results and Discussions

This thorough study conducts a comprehensive analysis of the proposed Convolutional Neural Networks' (CNNs) efficacy in classifying and segmenting flowers from the Oxford 17-Flower Category Dataset, (refer to fig. 5)it illustrates the CNNs' performance.

Two distinct network structures, designed and optimised for robust performance, challenge existing state-of-the-art techniques in image recognition and classification. Data augmentation methods during training substantially enhance model effectiveness, improving their adaptability and feasibility.

So in a nutshell, the suggested CNNs accurately categorise flowers in the Oxford Dataset, demonstrating increased robustness through data augmentation. This promising performance competes with advanced approaches, paving the way for future progress in broader contexts.

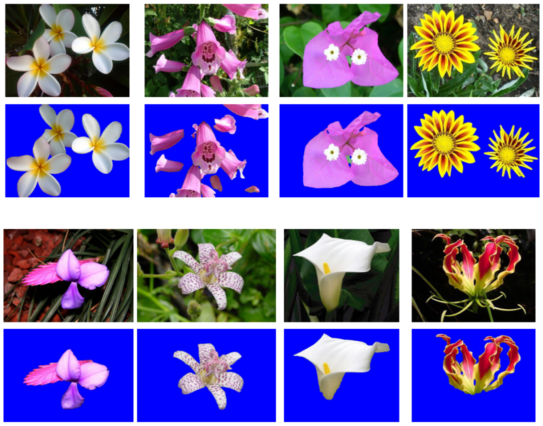

Fig. 5. Image segmentation results[1]

VI. Performance Comparison with Popular Architectures

This section compares the proposed models (CNN1 and CNN2) performances in contrast with popular architectures like VGG[8], ResNet[9], and MobileNet[14], discussing their relative performance, advantages, and disadvantages.

VGG[8] is known for capturing intricate image details and its simple architecture. However, it has a large number of parameters, making it computationally expensive and memory-intensive.

ResNet[9] addresses the vanishing gradient problem through skip connections, allowing deep architecture and faster convergence. Nevertheless, it requires substantial resources and can be complex to implement.

MobileNet[14] is designed for efficiency on resource-constrained devices, achieving competitive performance despite its smaller size. Its drawbacks include slightly lower performance compared to VGG or ResNet and the challenging implementation of depth-wise separable convolutions.

The proposed models (CNN1 and CNN2) are tailored for the Oxford 17-Flower Category Dataset and leverage data augmentation techniques for robustness and adaptability. However, the dataset's limited size may cause overfitting or restrict generalizability. The models' applicability might also be limited due to the dataset's narrow scope, and the transferability of the techniques to other domains remains uncertain as of now.

VII. Real-world Applications and Deployment

The potential real-world applications and deployment considerations for the proposed CNN architectures (CNN1 and CNN2).

Potential applications include botanical research, agriculture, mobile applications, and environmental monitoring. The models can aid in identifying flower species, optimising crop yields, and tracking ecological changes.

Key deployment considerations include:

Inference speed: Optimise for fast inference(fastly and efficiently extracting information from videos and images) using techniques like model pruning and quantisation[15]. Memory constraints: Minimise memory footprint in resource-constrained environments through techniques such as weight sharing and knowledge distillation[15,16]. Adapting to different image resolutions: Ensure adaptability to varying resolutions with methods like multi-scale training[17]. Robustness and generalisation: Enhance performance by employing data augmentation[18], regularisation techniques, and transfer learning. These considerations will help maximise the proposed models' effectiveness in real-world scenarios.

VIII. Conclusion

In conclusion, this detailed report dives deep into the development, evaluation, and potential outcomes of two well-designed Convolutional Neural Networks (CNN1 and CNN2) for the purpose of classifying and segmenting floral images using the Oxford 17-Flower Category Dataset. These models illustrate exceptional performance in classification and segmentation, with data augmentation techniques contributing to their robustness and feasibility [18].

Comparing this with other popular architectures such as VGG [8], ResNet [9], and MobileNet [14] highlights the strength and weakness of the proposed models, offering valuable insights into their overall performance and areas for improvement. Furthermore, the report traverses real-world implementations and deployment considerations, including crucial factors like inference speed, memory constraints, adaptability to different image resolutions, and robustness in various scenarios.

A thorough analysis of these factors ensures optimal model performance in real-life applications, laying the groundwork for advancements in flower classification and segmentation while providing practical insights on Convolutional Neural Networks' diverse applications [10].

References

[1] Nilsback, M. E., Zisserman, A. (2008). Automated flower classification over a large number of classes. In Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing (pp. 722-729).

[2] Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems (pp. 1097-1105).

[3] Long, J., Shelhamer, E., Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3431-3440).

[4] Everingham, M., Van Gool, L., Williams, C. K., Winn, J., Zisserman, A. (2010). The Pascal Visual Object Classes (VOC) Challenge. International Journal of Computer Vision, 88(2), 303-338.

[5] Lowe, D. G. (1999). Object recognition from local scale-invariant features. In Proceedings of the International Conference on Computer Vision (Vol. 2, pp. 1150-1157).

[6] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 234-241). Springer, Cham.

[7] Kingma, D. P., Ba, J. (2014). Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980.

[8] Simonyan, K., Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556.

[9] 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 (pp. 770-778).

[10] Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep learning. MIT Press.

[11] Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., Brendel, W. (2018). ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. arXiv preprint arXiv:1811.12231.

[12] Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S. (2014). CNN features off-the-shelf: an astounding baseline for recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 806-813).

[13] Nair, V., Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning (ICML-10) (pp. 807-814).

[14] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv preprint arXiv:1704.04861.

[15] Han, S., Mao, H., Dally, W. J. (2015). Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding. arXiv preprint arXiv:1510.00149.

[16] Hinton, G., Vinyals, O., Dean, J. (2015). Distilling the Knowledge in a Neural Network. arXiv preprint arXiv:1503.02531.

[17] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S. (2017). Feature Pyramid Networks for Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2117-2125).

[18] Shorten, C., Khoshgoftaar, T. M. (2019). A Survey on Image Data Augmentation for Deep Learning. Journal of Big Data, 6(1), 60.

[19] Yamashita, R., Nishio, M., Do, R.K.G. et al. Convolutional neural networks: an overview and application in radiology. Insights Imaging 9, 611–629 (2018).