Predicting Dog Breeds

Khatanbuuvei Bold, COS '21

Abstract

Dogs are a common part of our lives. Thus, dog breed classification has potential applications in many fields. In the past, this problem was approached with methods such as face detection and part localization. However, the field of computer vision has enjoyed an unprecedented boom in recent years, thanks to developments in deep learning and convolutional neural networks (CNN's), as well as large scale datasets such as ImageNet. In this project, I attempted to build high performing dog classifiers based on transfer learning. To this end, the EfficientNetB3 pre-trained network was used in addition to the AutoAugment augmentation policies to produce models that exceed cutting edge novel approaches such as WS-DAN in accuracy.

Motivation

Classifying dog breeds is a non-trivial task that requires differentiating between very small differences among certain breeds, such as between a Kuvasz and a Great Pyrenees:

Thus, a successful model can be extended to other domains featuring fine-grained differences among subclasses, such as species of other animals or plants

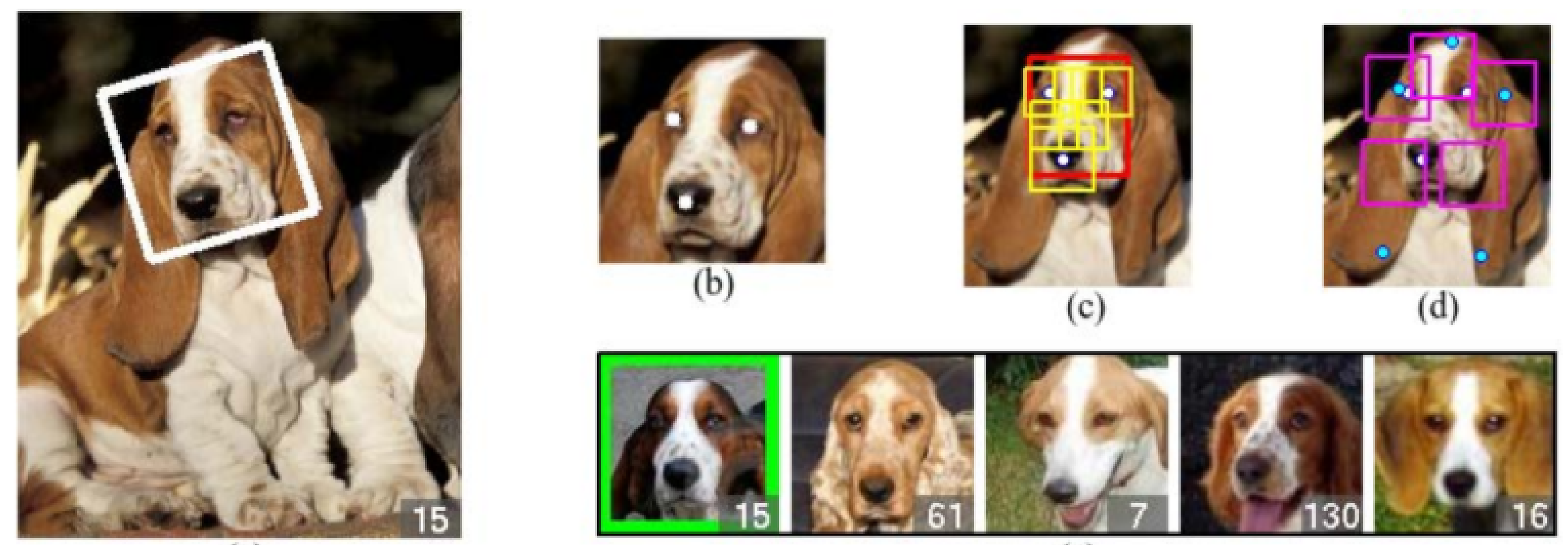
Data

Three major dog breeds datasets were used for training

- *Columbia Dogs* - 133 breeds, 8000 images [3]
- *Stanford Dogs* - 120 breeds, 20000 images [7]
- *Tsinghua Dogs* - 130 breeds, 70000 images [6]

Related Work

Pre deep learning, Liu et al. used face detection and part localization to achieve **67%** accuracy on the Columbia Dogs dataset [3].



At the time of this project, the state of the art models for fine-grained visual classification problems are weakly supervised data augmentation networks (WS-DANs), which employ data augmentation heavily. [2]

Approach

1. *EfficientNets* were used as a feature extractor for a feed-forward classifier head
2. Dropout and Gaussian Dropout were used in the classifier head to regularize the classifier, with significant accuracy gains
3. The entire *EfficientNet* convolutional neural network was fine-tuned on the respective dog breeds dataset
4. While fine tuning, AutoAugment was used to regularize the convolutional network via data augmentation.

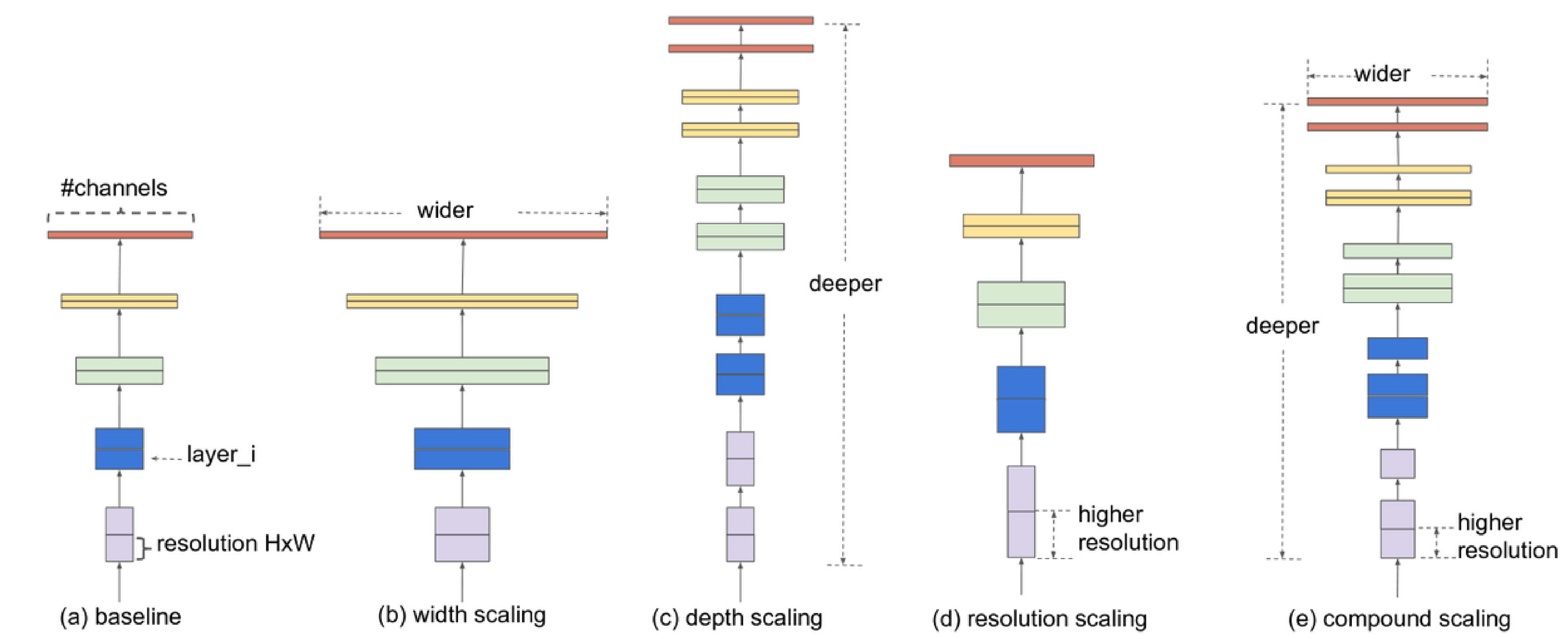
Results and Conclusions

Model	Validation Loss	Validation Accuracy	Test Loss	Test Accuracy	Epochs
Columbia Dogs	0.2338	0.9341	**0.2550**	**0.9175**	15
Stanford Dogs	**0.2362**	**0.9289**	-	-	15
Tsinghua Dogs	**0.4085**	**0.8800**	-	-	10

The table above shows the final validation and test set performance. "Epochs" only indicate the epochs iterated during the fine tuning phase. The model outperforms WS-DANs on the Stanford and Tsinghua dogs datasets by **0.69%** and **1.4%**, respectively. In future works, using higher EfficientNet baselines with an identical approach, or using EfficientNets as a backbone for WS-DANs will likely result in even greater performance.

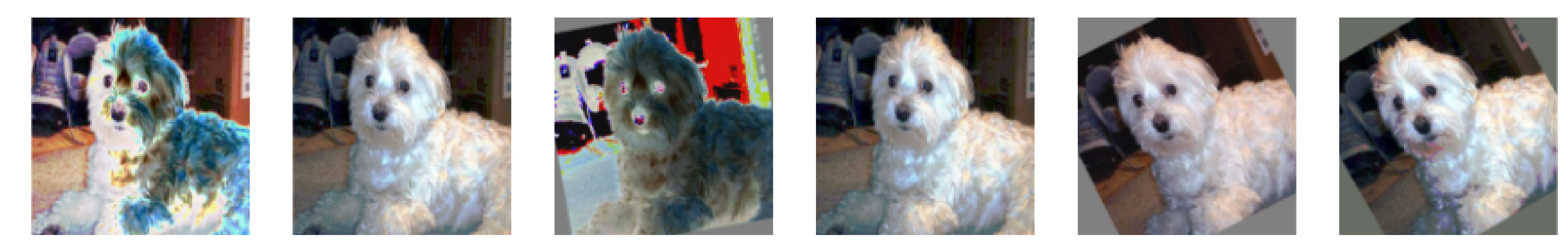
EfficientNets

EfficientNets[5] are a family of networks that simultaneously scale the width, depth and input resolution of CNN's by constant ratios, as opposed to scaling one dimension. This produces high performing, computationally inexpensive networks. The EfficientNetB3 baseline was used as a pre-trained feature extractor for this project.



AutoAugment

AutoAugment[1] randomly applies transformations to images while training, according to a policy learned by reinforcement learning. The policies learned on ImageNet[4] were used for this project. Examples:



References

[1] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "Autoaugment: Learning augmentation policies from data," 2019.
[2] T. Hu, H. Qi, Q. Huang, and Y. Lu, "See better before looking closer: Weakly supervised data augmentation network for fine-grained visual classification," 2019.
[3] J. Liu, A. Kanazawa, D. Jacobs, and P. Belhumeur, "Dog breed classification using part localization," in European conference on computer vision. Springer, 2012, pp. 172–185.
[4] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," International Journal of Computer Vision (IJCV), vol. 115, no. 3, pp. 211–252, 2015.
[5] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," 2020.
[6] D.-N. Zou, S.-H. Zhang, T.-J. Mu, and M. Zhang, "A new dataset of dog breed images and a benchmark for fine-grained classification," Computational Visual Media, 2020. [Online].
[7] A. Khosla, N. Jayadevaprakash, B. Yao, and L. Fei-Fei, "Novel dataset for fine-grained image categorization," in First Workshop on Fine-Grained Visual Categorization, IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, June 2011.

Advised by Prof. Olga Russakovsky