Size Matters: Progressive Image Resizing for Improved ImageNet-Sketch Classification

Anonymous CVPR submission

Paper ID

Abstract

We address the classification of the ImageNet-Sketch dataset using ResNets, EfficientNets, and Vision Transformers. Progressive image resizing during training improved validation accuracy to 90%, outperforming standard augmentations and optimizer adjustments. This highlights resizing as an effective strategy for sketch image classification.

1. Introduction

The ImageNet-Sketch dataset [9] challenges models due to its abstract object representations. We aim to achieve high accuracy by exploring diverse architectures and training strategies.

2. Methodology

We began by experimenting with several convolutional neural network architectures, including ResNet18 and ResNet50 [2], which achieved validation accuracies of 76% and 81%, respectively. Recognizing the potential of more advanced architectures, we evaluated EfficientNet variants [7] [8] (B3, B4, B5, and V2M), obtaining accuracies ranging from 82% to 87%. We also tested a Vision Transformer (ViT Base16) [1], which achieved 84% accuracy.

All models were trained with images resized to 224×224 pixels and normalized using standard ImageNet mean and standard deviation values. We used stochastic gradient descent (SGD) with a learning rate of 0.1, momentum of 0.5, and a batch size of 64 over 10 epochs, using pretrained weigth on ImageNet [5].

Data augmentations, including grayscale, Gaussian blur, and Canny filters [6], showed no significant improvements, with accuracies stagnating at 84%. Using Adam optimizer [4] converged faster but also plateaued at 84%.

Our most effective strategy involved progressively increasing the input image size during training as done here [3]. We first trained the model with 224×224 images for 10 epochs using SGD. We then increased the image size to 384×384 and continued training for another 10 epochs with a reduced learning rate of 0.05 and a batch size of 24. Finally, we trained with 512×512 images for an additional 10 epochs, further reducing the learning rate to 0.01 and using a batch size of 12. This progressive resizing approach improved the validation accuracy to 90%.

Model	Validation Accuracy (%)
ResNet18	76
ResNet50	81
EfficientNet B3	82
EfficientNet B4	87
EfficientNet B5	84
EfficientNet V2M	85
ViT Base16	84

Table 1. Validation accuracy of different models on the ImageNet-Sketch dataset.

3. Results

Our experiments demonstrated that EfficientNet B4 provided the highest initial accuracy (Table 1). Data augmentation and optimizer changes did not significantly improve performance. However, progressive image resizing led to a substantial increase in accuracy, highlighting its effectiveness.

All training was performed on a single NVIDIA T4 GPU provided by Kaggle. Tha batch sizes were always chosen as the biggest that fits the GPU VRAM (15Go).

4. Conclusion

Progressive resizing significantly improved classification accuracy on ImageNet-Sketch, reaching 90%. Traditional augmentations and optimizer tuning had limited impact, while resizing enabled the model to capture fine-grained details. Future work could combine resizing with advanced training techniques for further gains.

References

[1] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. 1

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 1

[3] Jeremy Howard and Sylvain Gugger. Now anyone can train imagenet in 18 minutes. *fast.ai Blog*, 2018. 1

[4] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017. 1

[5] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge, 2015. 1

[6] M.C. Shin, D. Goldgof, and K.W. Bowyer. An objective comparison methodology of edge detection algorithms using a structure from motion task. In *Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No.98CB36231)*, pages 190–195, 1998. 1

[7] Mingxing Tan and Quoc V Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, pages 6105–6114. PMLR, 2019. 1

[8] Mingxing Tan and Quoc V. Le. Efficientnetv2: Smaller models and faster training, 2021. 1

[9] Haohan Wang, Songwei Ge, Eric P. Xing, and Zachary C. Lipton. Learning robust global representations by penalizing local predictive power, 2019. 1