The Power of Inception: Tackling The Tiny ImageNet Challenge

Pedro Milani

CS-231N Spring 2017

Introduction

The ImageNet Challenge [1] is an important tool to develop and benchmark visual recognition algorithms. The dataset contains almost 1.5 million images, labeled into 1000 classes. The top performing models are based on deep Convolutional Neural Networks (CNNs) since 2012. Currently, state-of-the-art top-1 validation accuracies are around 80% [4].

I have worked on the Tiny ImageNet Challenge, a smaller version of that challenge. The dataset contains 100k training images (with 10k validation and 10k test images), split into 200 classes. Each image is 64x64 pixels. Examples are shown below:

bullfrog

backpack

potter's wheel

Choosing an Algorithm

- **Goal:** maximize top-1 accuracy in the test set.
- **Constraints:** limited time and computational resources.

Inception-v3 has the best mix of accuracy and cost!

Inception layers were introduced in GoogleNet [2] to apply convolutional filters efficiently. Improvements were made in [3].

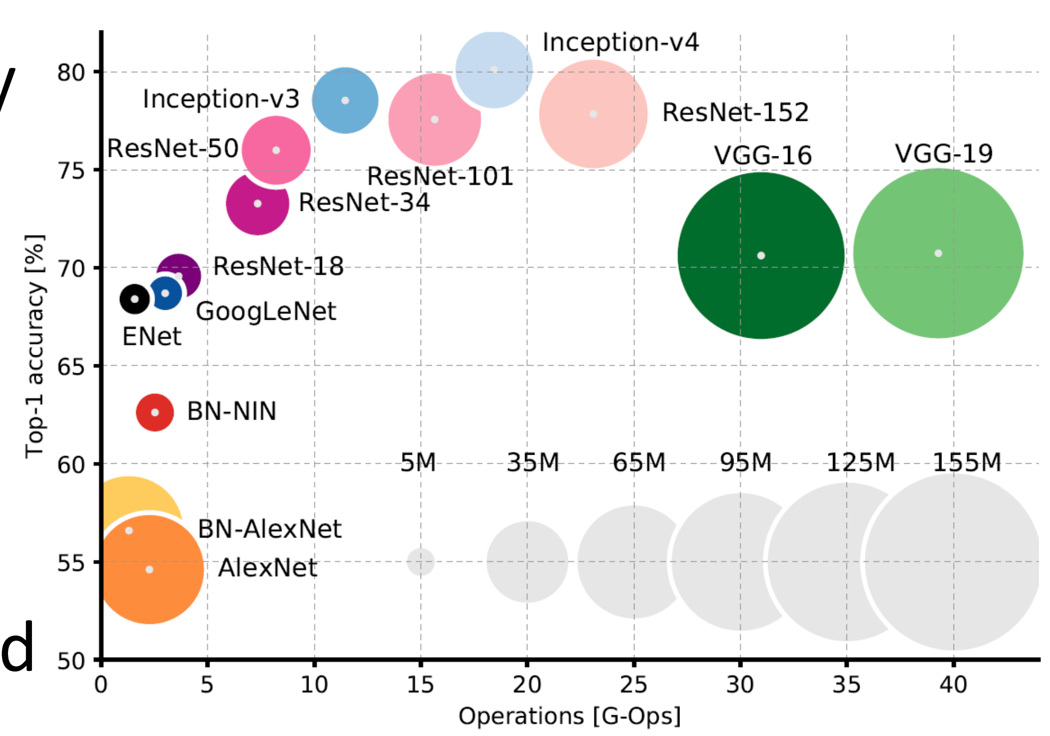


Figure 1: Chart from [4] analyzing different CNN models used in ImageNet.

http://knowyourmeme.com/photos/531557-we-need-to-go-deeper. [2] actually cites this.

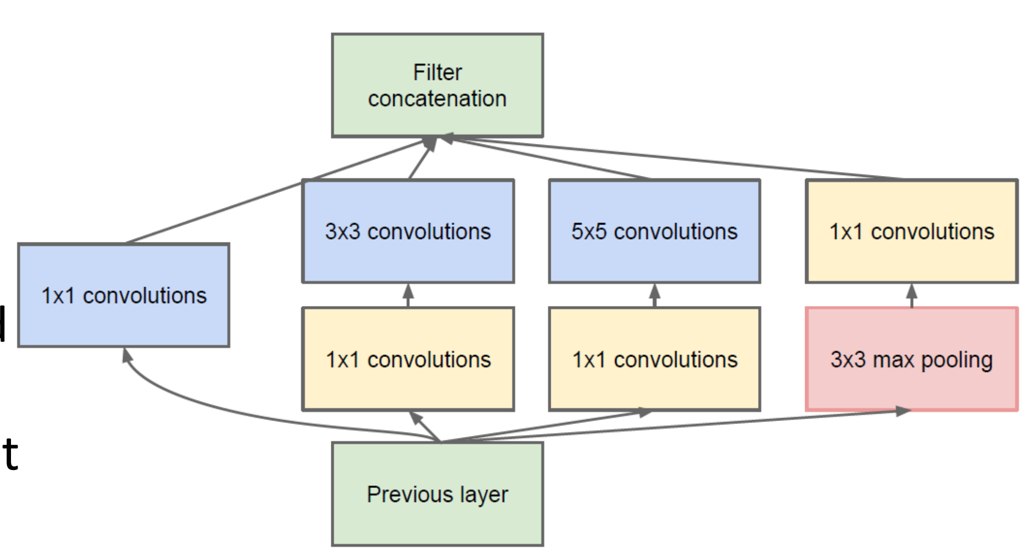


Figure 2: Inception layer used in GoogleNet [2]

Architecture and Results

My Own:

- I designed and trained from scratch a simplified architecture inspired on figure 3.
- Fewer layers/filters, no side classifier, no dropout, heavy use of batch normalization.
- ~30min/epoch on Tesla K80 GPU.

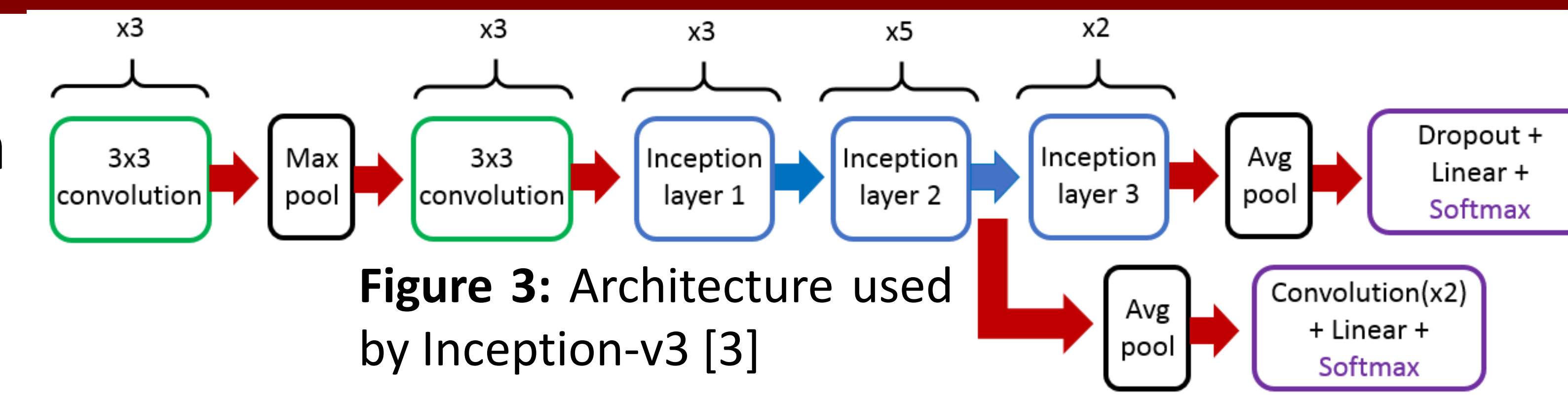
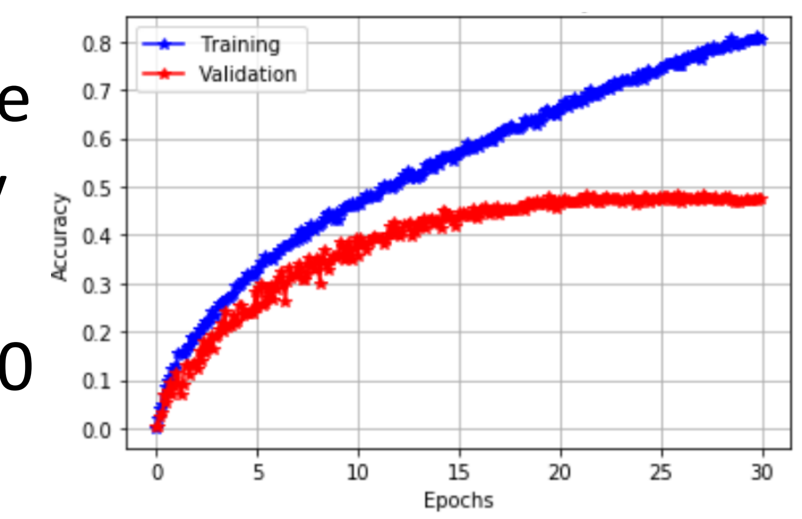


Figure 3: Architecture used by Inception-v3 [3]

Transfer Learning:

- I used the 2048 activations after the last layer of the pre-trained Inception-v3 model to train a linear map to the 200 classes.
- Online tutorial/code by [5].



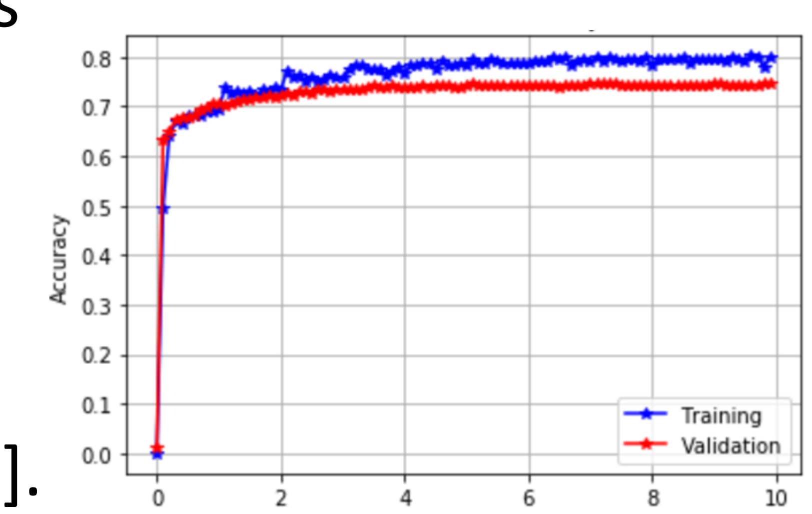


Figure 4/5: Accuracy curves for my model (top) and transfer learning (bottom).

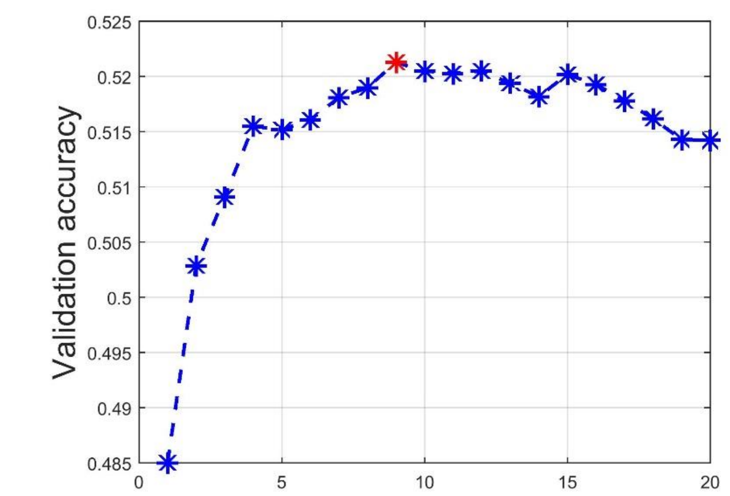


Figure 6: Accuracy from ensembles of my model

	Top-1 Accuracy	
	Validation	**Test**
Mine	48.5%	n/a
Mine ens	52.2%	47.7%
T.L.	74.56%	n/a
T.L. ens	74.60%	66.5%

Table 1: Best results to date.

Conclusions and Future Work

Conclusions:

- I obtained good accuracy with my model, but much better using transfer learning.
- Ensembling is a cheap way to boost accuracy; not effective with transfer learning.
- Big discrepancy between validation and test data; I need to investigate.

Future Work:

- Improve my model: regularization (dropout and data augmentation)
- Interpretation: filter visualization for my model, dimensionality reduction for transfer cache.

References

[1] O. Russakovsky et al. ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV), 115(3):211–252, 2015.
[2] C. Szegedy et al. Going deeper with convolutions. CoRR, abs/1409.4842, 2014.
[3] C. Szegedy et al. Rethinking the inception architecture for computer vision. CoRR, abs/1512.00567, 2015.
[4] A. Canziani et al. An analysis of deep neural network models for practical applications. CoRR, abs/1605.07678, 2016.
[5] Hvass Laboratories Tutorials. http://www.hvass-labs.org/ Acessed on 06/04/2017.

Flow Physics and Computational Engineering Laboratory at Stanford University

pmmilani@stanford.edu