CS 231n and CS 230

Visualization in Fine-Grained Visual Categorization (FGVC)

Jun 9 & 12, 2018

Project team: Renato Baba (rbaba@), Klemen Cas (klemen@) and Yaya Khoja (ykhoja@)

Project Motivation

- Improve online shopping for customers and sellers
- Address a problem with critical practical applications
- Understand the particular challenges in FGVC

Data Source

Our dataset contains 1,014,544 images supplied by a Kaggle competition . Each image can have multiple ground truth labels out of 228 possible categories. The labels represent product type, color, material, etc

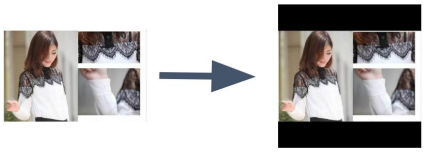
Labels: 62, 19, 14, 78, 79, 117, 131

Approach

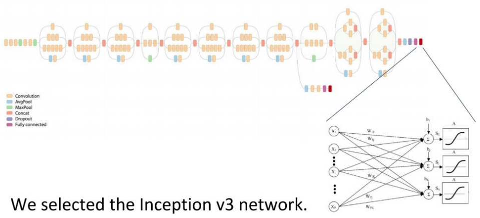
- Use transfer learning on an architecture pretrained on ImageNet and finetuned the model using the new dataset.

Data Preprocessing

- Resize images to fit pretrained model input
- Pad images so they all have uniform shape



Network Architecture



- We selected the Inception v3 network.
- Changed the final output layer to a fully connected layer with 228 outputs followed by sigmoid activation

Loss Function

We experimented with weighted and unweighted binary cross entropy

$$L = \sum_{i=1}^{C} [-\text{weight} \cdot y_i \log q_i - (1 - y_i) \log(1 - q_i)]$$

Hyperparameters

Learning Rate	Batch Size	# Training Steps
0.01 (tested with 0.005, 0.01, 0.1, 0.5, 1)	100 (train) 500 (validation)	20,000 (tested with up to 60,000)

Training Process

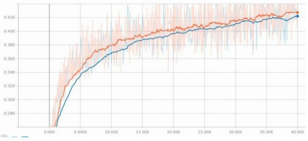
1. Started with a small subset of only 50 images to test if the model is learning and the loss function is decreasing
2. Experimented using a bigger subsets of 20,000, 75,000 and 100,000 images to tune the model hyperparameters
3. Trained the model with the best hyperparameters using the complete dataset using 2 Nvidia Tesla K80 GPUs

Results

Results on numeric evaluation metrics:

Images	Training Steps	Weight	Val acc	Test Precision	Test Recall	Test F1
50	500	1	.941	1	.167	.286
50	20k	1	.929	.500	.167	.250
100k	20k	1	.979	.846	.198	.320
100k	20k	6	.966	.370	.526	.434
100k	5k	4	.972	.458	.372	.411
1M	500	1	.976	.746	.125	.215
1M	500	4	.969	.374	.339	.355
1M	5k	4	.972	.439	.349	.389
1M	22k	4	.973	.474	.434	.453

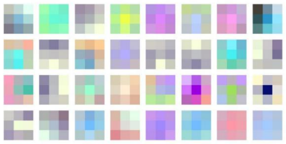
F1 score continued to improve after 22k steps (see chart for training of 75k images), hence, there is room for improvement by training the model longer.



Network Visualization

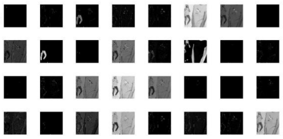
Learned Parameters:

- Represent the colors that will activate the network most
- Provide view into edge detection



Activations:

- Network clearly focuses on specific parts of the image such as seans or skin color
- Most of network activations are zero



Next Steps

- Train the model for more epochs
- Explore other architectures and compare results
- Try to implement approaches of Kaggle competition winners