Large-Scale Image Classification On A Shoestring

Olivier Morère[1,2], Hanlin Goh[1], Antoine Veillard[2], Vijay Chandrasekhar[1]

1. Institute for Infocomm Research, A*STAR, Singapore
2. Université Pierre et Marie Curie, Paris, France

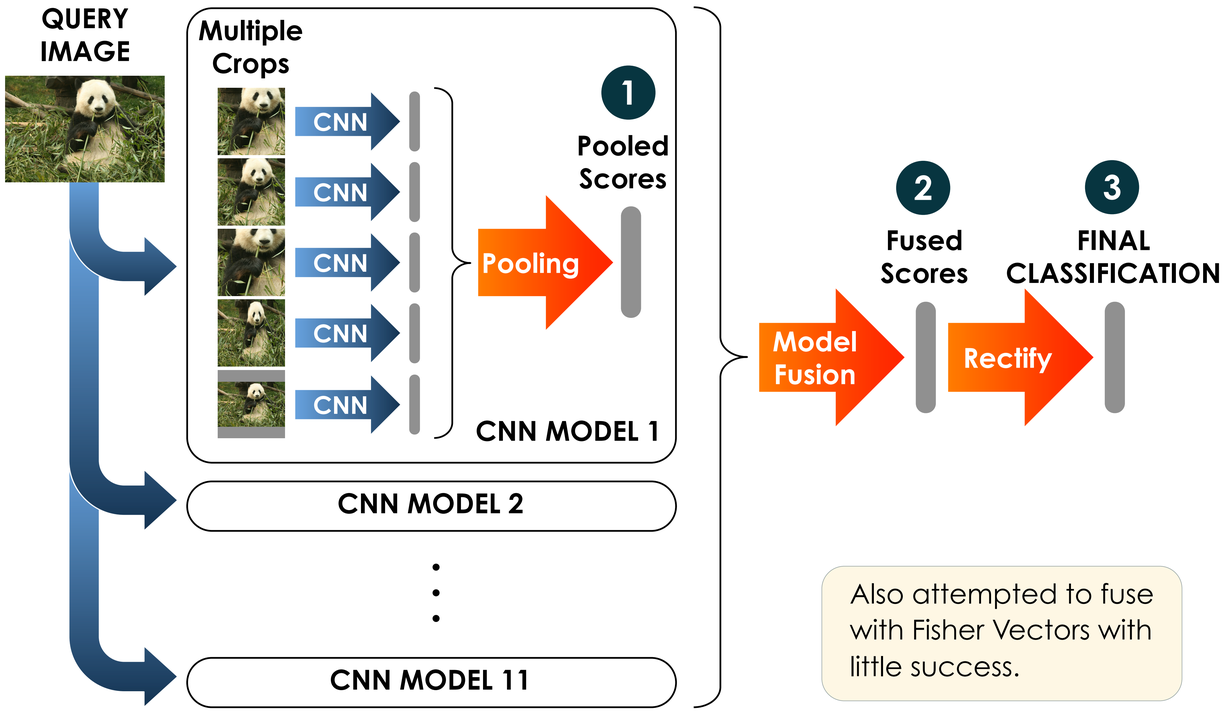
RESOURCE CONSTRAINTS

- Low Memory GPU Cluster
  - NVIDIA GTX580 (1.5GB)
- Small Hadoop Cluster
- Team formed in June 2014

OUR APPROACH

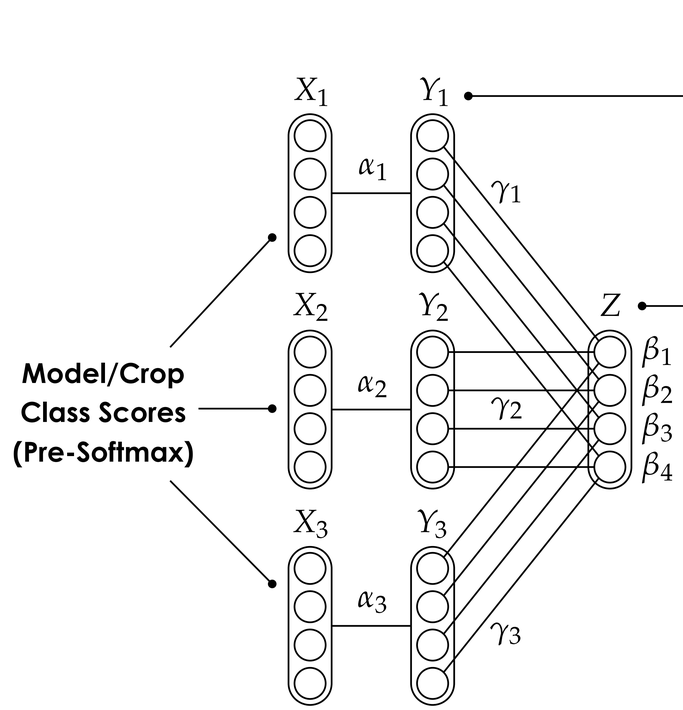
- Train Small CNN models
- Study Existing Pre-Trained Models
- Multi-Crop Pooling
- Model Fusion
- Rectify Outputs

ADAPTIVE FUSION OF MULTIPLE CNN



Also attempted to fuse with Fisher Vectors with little success.

MULTI-CROP POOLING & MODEL FUSION



Final Layer Softmax

$$y_{ni} = \frac{\exp(\alpha_n x_{ni})}{\sum_{j=1}^{I} \exp(\alpha_n x_{nj})}$$

Softmax Averaging

$$z_i = \frac{\sum_{n=1}^{N} y_{ni}\gamma_n \exp(\beta_i y_{ni})}{\sum_{n=1}^{N} \gamma_n \exp(\beta_i y_{ni})}$$

Optimization

Gradient Descent & Grid Search

α : Shared parameter
β : Shared parameter
γ_n : Grouped regularization

IMAGENET CLASSIFICATION RESULTS

Models Submitted	Top-1 Error Rate	Top-5 Error Rate
1 Single CNN model	32.491	12.128
2 Adaptive fusion of multiple CNN models	31.715	11.403
3 Adaptive fusion of multiple CNN models with output rectification	31.733	11.326