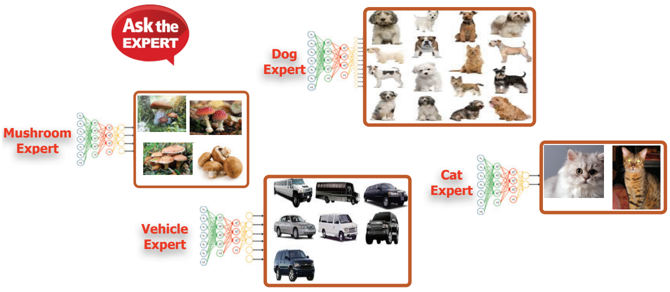
Network of Experts *for Large-Scale Image Categorization*

Karim Ahmed, Mohammad Haris Baig, Lorenzo Torresani
Department of Computer Science, Dartmouth College, USA
Code and models available at **http://vlg.cs.dartmouth.edu/projects/nofe/**

Dartmouth

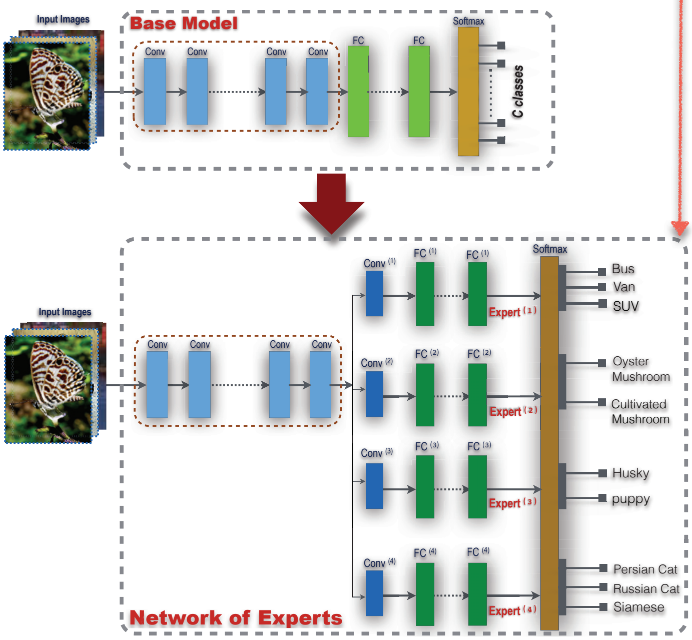
Intuition

The visual system of a layperson is a very good ***generalist*** that can accurately discriminate coarse categories but lacks the ***specialist*** eye to differentiate categories that look alike.



Contribution

Inspired by this analogy, we propose a novel tree-structured architecture *(Network of Experts)* for large-scale Image categorization. It can be built from any existing convolutional neural network (CNN), and the training is completely end-to-end.



Related Work

Our approach relates closely to methods that ***learn hierarchies of categories to train CNN experts***, such as Hinton et al. [2], Warde-Farley et al. [7], and (HD-CNN) Yan et al. [8].

But it provides the following **novel benefits** :

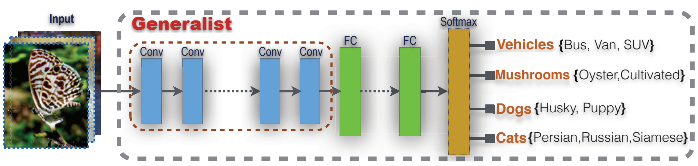
- Improved accuracy, tested for 5 different base architectures.
- Does not require training of base model.
- Does not suffer from mistakes due to routing to the wrong expert.
- End-to-end optimization over the original categorization problem.

Technical Approach

We decompose large-scale image categorization into two separate tasks:

(I).*Learning the Generalist:*

The goal of this stage is to learn K (with $K << C$) groupings of classes called *specialties*.



Given *base model* objective:

$$E_{\mathrm{b}}(\theta;\mathcal{D}) = \underbrace{R(\theta)}_{Regularization} + \frac{1}{N}\sum_{i=1}^{N} \underbrace{L(\theta;\ \overbrace{x^i}^{Image},\ \overbrace{y^i}^{Class\ Label})}_{Classification\ Loss} \quad (1)$$

We propose the following ***generalist*** objective:

$$E_{\mathrm{g}}(\theta^G,\ell;\mathcal{D}) = R(\theta^G) + \frac{1}{N}\sum_{i=1}^{N} L(\theta^G; x^i,\ \underbrace{\ell(y^i)}_{Assigned\ Specialty}) \quad (2)$$

where $\ell(y^i)$ maps classes to specialties

i.e. $\forall y^i \in \{1,\ldots,C\}$ assign $\ell(y^i) \in \{1,\ldots,K\}$ where $\underbrace{K}_{\#Specialties} << \underbrace{C}_{\#Classes}$

Minimized via alternation between:

1. Optimizing parameters θ^G while keeping specialty labels ℓ fixed (traditional SGD).
2. Updating specialty labels ℓ given the current estimate of weights θ^G.

(II).*Training the Network of Experts:*

The *Network of Experts* is a tree-structured architecture:

- The trunk of the tree is finetuned from the *Generalist* and contains shared features.
- The trunk splits into K branches corresponding to the K learned specialties.
- Each branch is an *expert* optimized to distinguish the classes within its specialty.
- Final softmax layer over all C classes calibrates the outputs of the K experts.

Results

CIFAR100

Model	K=2	K=5	K=10	K=20	K=50
NoFE	53.3	55.0	**56.2**	55.7	55.33
Base: AlexNet-C100			54.0		

Table 1: Top-1 accuracy (%) on CIFAR100 for the base model (AlexNet-C100) and Network of Experts (NoFE) using varying number of experts (K)

maple_tree, oak_tree, pine_tree, willow_tree, palm_tree
apple, cloud, poppy, rose, tulip
dolphin, seal, shark, turtle, whale
baby, boy, girl, man, woman

Table 2: Example of specialties learned from CIFAR100

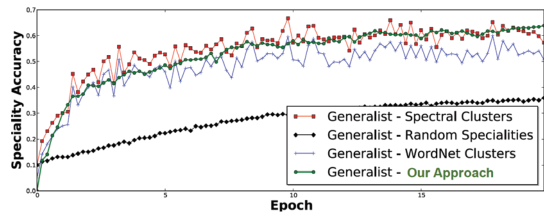
Architecture	Base Model	NoFE
AlexNet-C100 [4]	54.04	**56.24**
AlexNet-Quick-C100 [3]	37.94	**45.58**
VGG11-C100 [6]	68.48	**69.27**
NIN-C100 [5]	64.73	**67.96**
ResNet56-C100 [1]	73.52	**76.24** *Best Published Result*

Table 3: CIFAR100 top-1 accuracy (%) for 5 different CNN base architectures and corresponding NoFE models.

Approach	Top-1	# params	Avg. Inference time
NoFE using NIN	**67.96**	$4.7M$	**0.0071 secs**
HD-CNN [8] using NIN	67.38	$9.2M$	0.0147 secs

Table 4: Our NoFE compared to HD-CNN [8], using NIN[5] as a base model on CIFAR100.

ImageNet



Approach	Top-1	# params
Base: AlexNet-Caffe [4]	58.71	$60.9M$
NoFE, K=10	**61.29**	$40.4M$
NoFE, K=40	60.85	$151.4M$
NoFE, K=10 (spectral clustering)	56.10	$40.4M$

Table 5: Top-1 accuracy on the ImageNet validation set using AlexNet and our NoFE.

References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016.
[2] Geoffrey E. Hinton, Oriol Vinyals, and Jeffrey Dean. Distilling the knowledge in a neural network. *CoRR*, abs/1503.02531, 2015.
[3] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.
[4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012.
[5] Lin, Min, Qiang Chen, and Shuicheng Yan. Network in network. In *International Conference on Learning Representations, 2014 (arXiv:1409.1556).*, 2014.
[6] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.
[7] David Warde-Farley, Andrew Rabinovich, and Dragomir Anguelov. Self-informed neural network structure learning. *CoRR*, abs/1412.6563, 2014.
[8] Zhicheng Yan, Hao Zhang, Robinson Piramuthu, Vignesh Jagadeesh, Dennis DeCoste, Wei Di, and Yizhou Yu. HD-CNN: hierarchical deep convolutional neural networks for large scale visual recognition. In *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015*, pages 2740–2748, 2015.