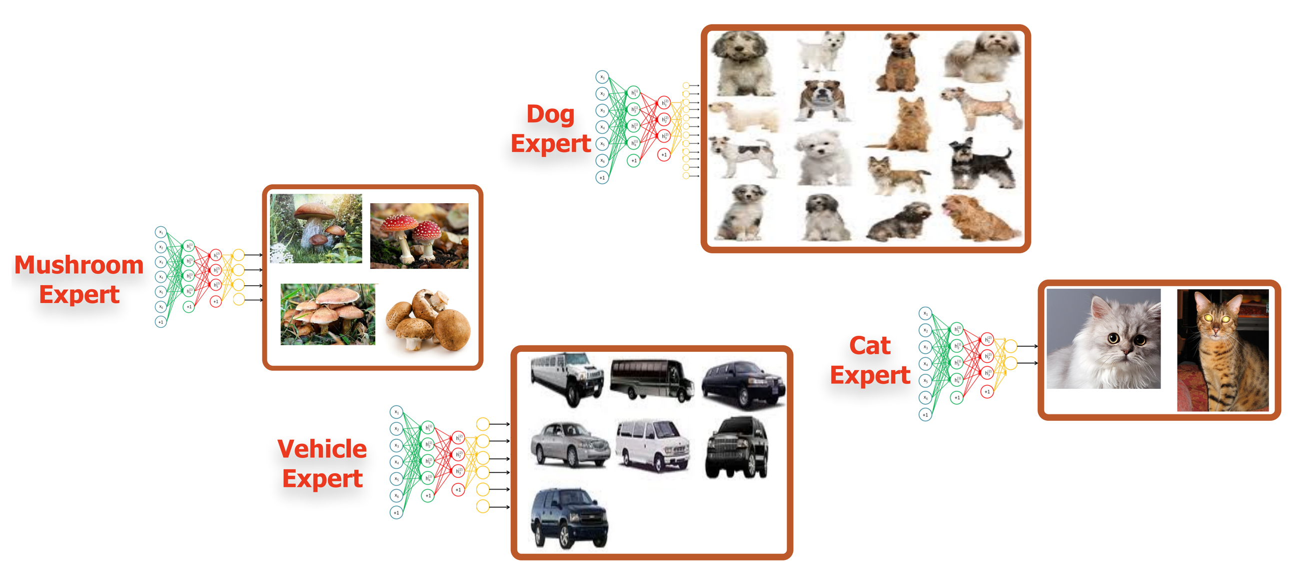
BranchConnect: Image Categorization with Learned Branch Connections

Karim Ahmed, Lorenzo Torresani

Dartmouth

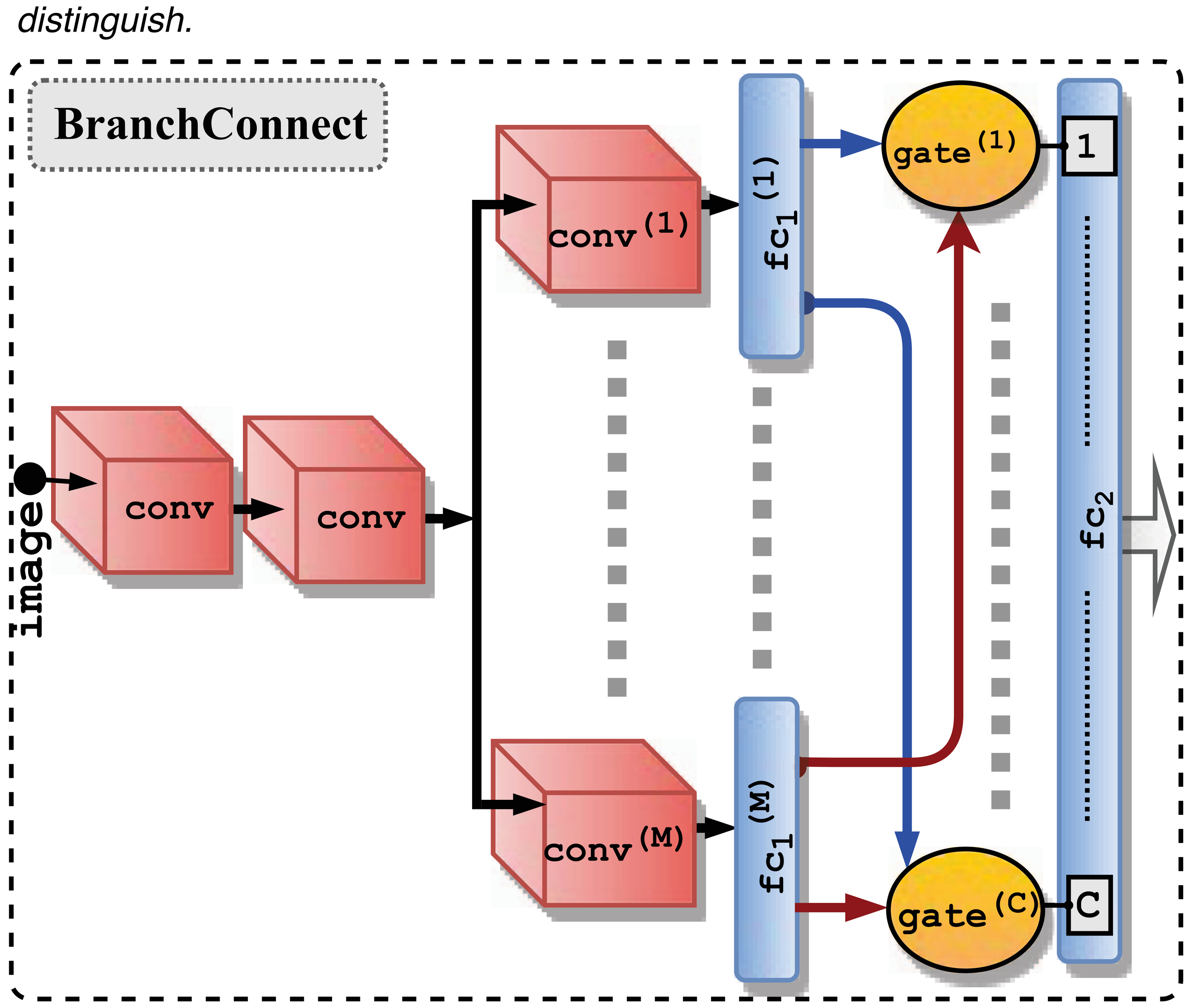
Intuition

*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

End-to-end learning of separate visual features for the different classes to distinguish.



- *The architecture of **BranchConnect** for classification of C classes. The branches implement M < C feature extractors. Class-specific gates connect the M branches to the C classes in the last fully-connected layer.*
- *The learned gates determines for each class the subset of features to use.*

Technical Approach

Goal: learn branch connectivity to classes from data by optimizing training objective

- Minimize $\ell(\theta, \mathbf{g})$ using backpropagation over $\mathbf{g}$ and θ
- During training, we update auxiliary real-valued gates $\mathbf{g}_c^r \in [0,1]^M$
- Constrain the number of active branch connections per class to be a constant, ***K*** (a hyperparameter)

Forward propagation:

1. Stochastically binarize $\mathbf{g}_c^r \in [0,1]^M$ into $\mathbf{g}_c^b \in \{0,1\}^M$ s.t. $\sum_{m=1}^{M} g_{c,m}^b = K$
2. Perform forward pass using binary gates $\mathbf{g}_c^b \in \{0,1\}^M$

$$F_c = \sum_{m=1}^{M} g_{c,m}^b \cdot E_m$$

F_c : the c[th] neuron in the last fully connected layer

E_m : m[th] branch activations

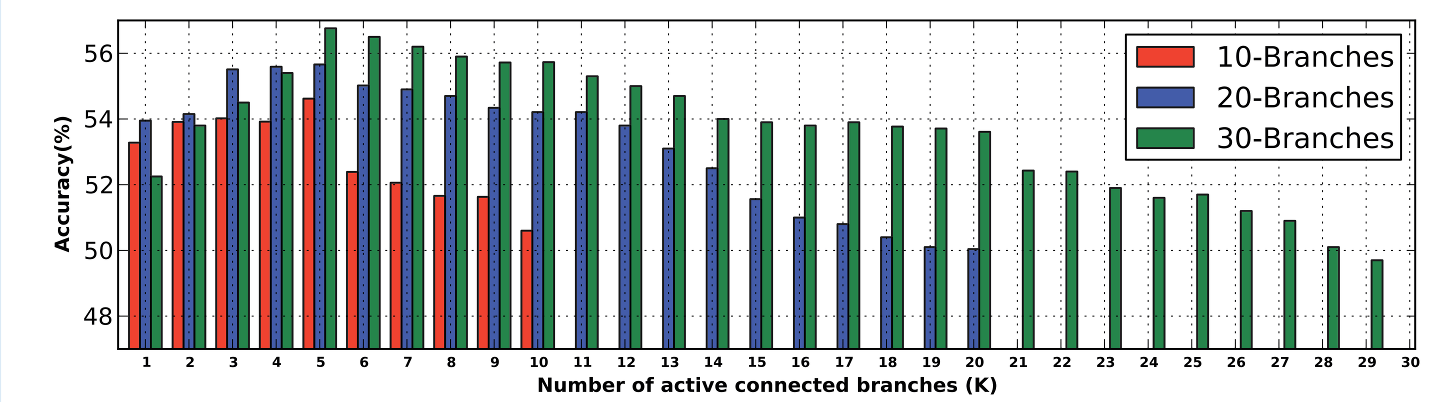
Parameter update:

Compute $\frac{\partial \ell}{\partial g_{c,m}^b}$, and update real-valued masks $g_{c,m}^r$

$$g_{c,m}^r \leftarrow \text{clip}(g_{c,m}^r - \eta \cdot \frac{\partial \ell}{\partial g_{c,m}^b})$$

Results

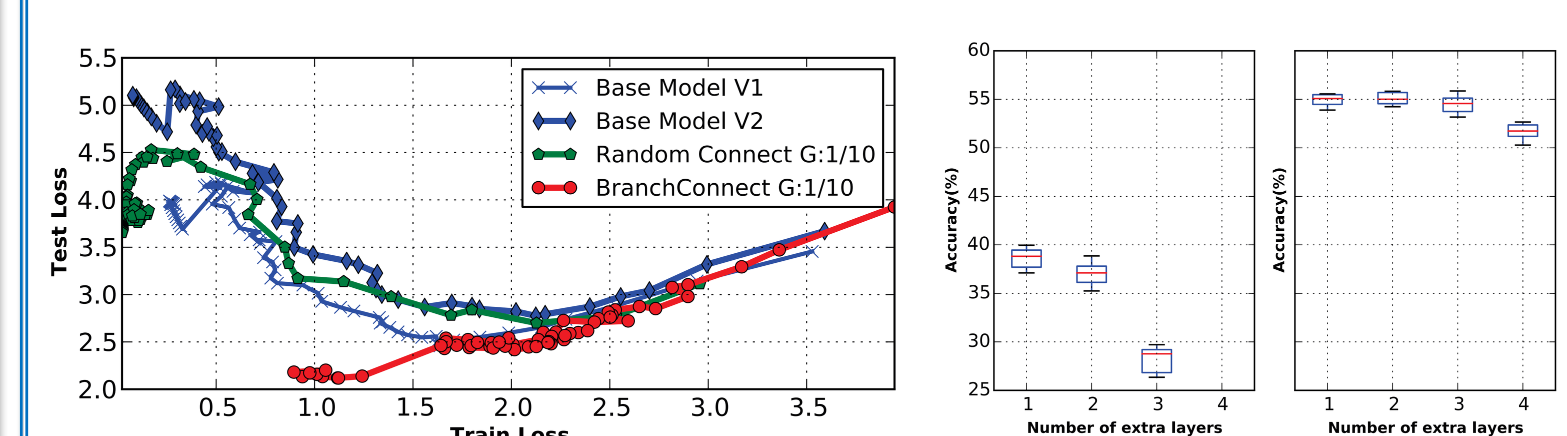
CIFAR-100 (100 classes, 50K training examples)



	Method	depth	#params	Accuracy
AlexNet-Quick	Base Model V1	5	0.15M	44.3
	Base Model V2	5	1.20M	40.26
	NofE [1]	6	1.27M	49.09
	BranchConnect G:1/10	5	1.20M	**53.28**
	BranchConnect G:5/10	5	1.20M	**54.62**
Alexnet-Full	Base Model V1	4	0.18M	54.04
	Base Model V2	4	0.64M	50.42
	NofE [1]	5	1.12M	56.24
	BranchConnect G:1/10	4	0.64M	**57.34**
	BranchConnect G:6/10	4	0.64M	**60.27**
NIN [3]	Base Model V1	9	1.38M	64.73
	Base Model V2	9	1.61M	65.24
	HD-CNN [2]	n/a	n/a	65.64
	NofE [1]	11	4.66M	65.91
	BranchConnect G:1/10	9	1.61M	**66.10**
	BranchConnect G:5/10	9	1.61M	**66.45**

	Method	depth	#params	Accuracy
ResNet5[4]	Base Model V1	56	0.86M	69.66
	Base Model V2	56	1.47M	70.72
	BranchConnect G:1/10	56	1.47M	**71.24**
	BranchConnect G:5/10	56	1.47M	**71.98**
ResNet56-4X [1]	Base Model V1 [1]	56	13.6M	72.23
	Base Model V2	56	25.4M	73.12
	NofE [1]	58	25.5M	74.71
	BranchConnect G:1/10	56	25.4M	**75.55**
	BranchConnect G:5/10	56	25.4M	**75.72**

BranchConnect acts as regularizer



CIFAR-10
(10 classes, 50K training examples)

Architecture	Method	Accuracy
AlexNet-Quick	Base Model	76.86
	BranchConnect G:3/5	**82.84**
AlexNet-Full	Base Model	82.78
	BranchConnect G:3/5	**85.00**
ResNet-56 [4]	Base Model	92.04
	BranchConnect G:3/5	**92.46**

ImageNet
(1000 classes, 1.28M training examples)

Architecture	Method	Accuracy
AlexNet	Base Model	58.71
	NofE [1]	61.29
	BranchConnect G:5/10	**63.49**
ResNet50 [4]	Base Model [7]	76.15
	BranchConnect G:5/10	**77.39**
	BranchConnect G:8/15	**77.68**
ResNet101 [4]	Base Model [7]	77.37
	BranchConnect G:5/10	**78.19**

References

[1] K. Ahmed et al. "*Network of Experts for large scale image categorization.*", ECCV 2016.

[2] Yan et al. "*HD-CNN: hierarchical deep convolutional neural networks for large scale visual recognition.*", ICCV 2015

[3] Lin et al. "Network in Network", ICLR 2014.

[4] He et al. "*Deep residual learning for image recognition.*" CVPR. 2016.