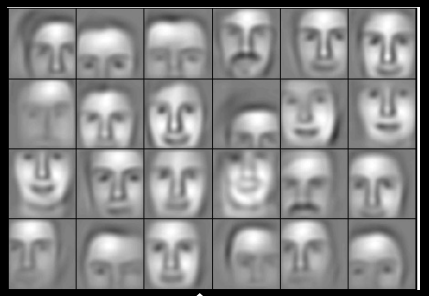
Building high-level features using large scale unsupervised learning

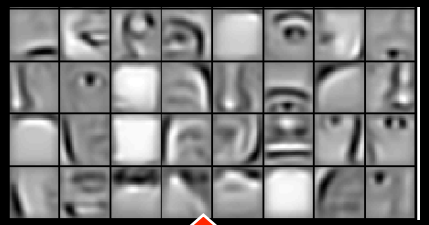
Quoc V. Le
Stanford University and Google

Joint work with: Marc'Aurelio Ranzato, Rajat Monga, Matthieu Devin, Kai Chen, Greg Corrado, Jeff Dean, Andrew Y. Ng

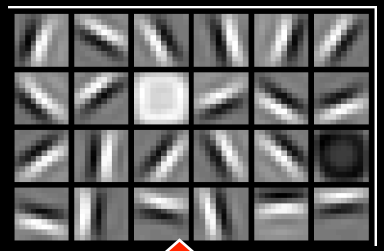
Hierarchy of feature representations



Face detectors



Face parts
(combination
of edges)



edges

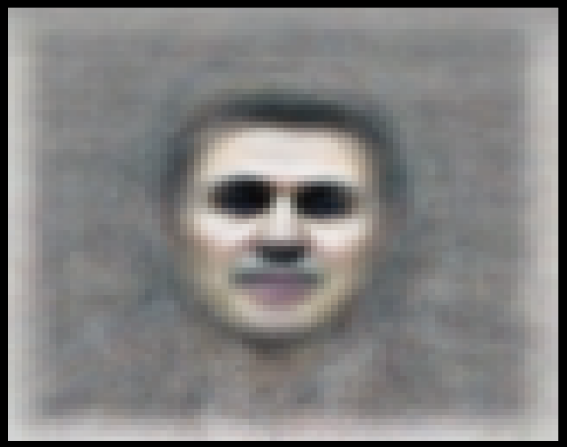
pixels

Lee et al, 2009. Sparse DBNs.

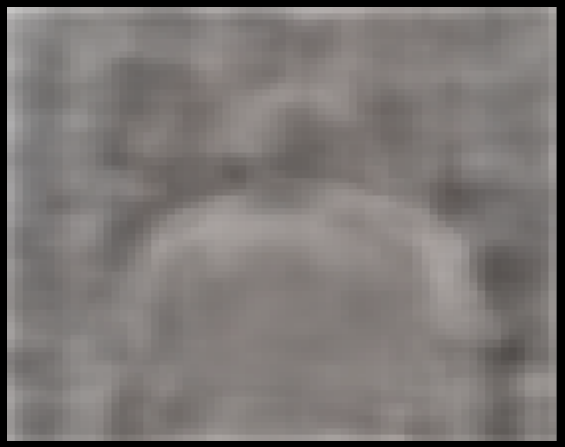
Faces

Random images from the Internet

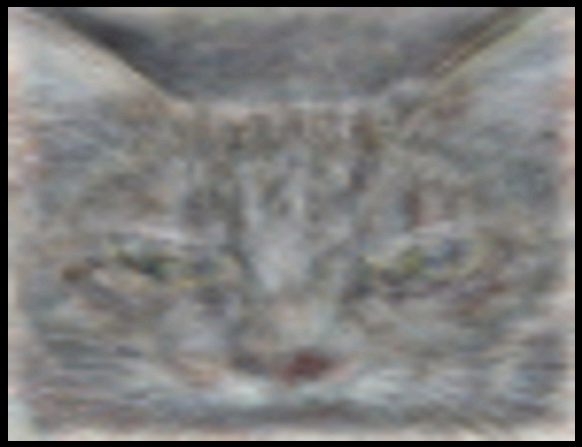
Key results



Face detector

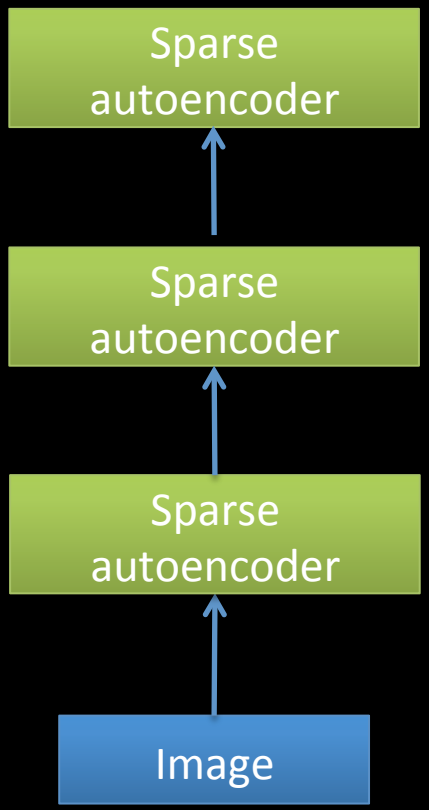


Human body detector



Cat detector

Algorithm



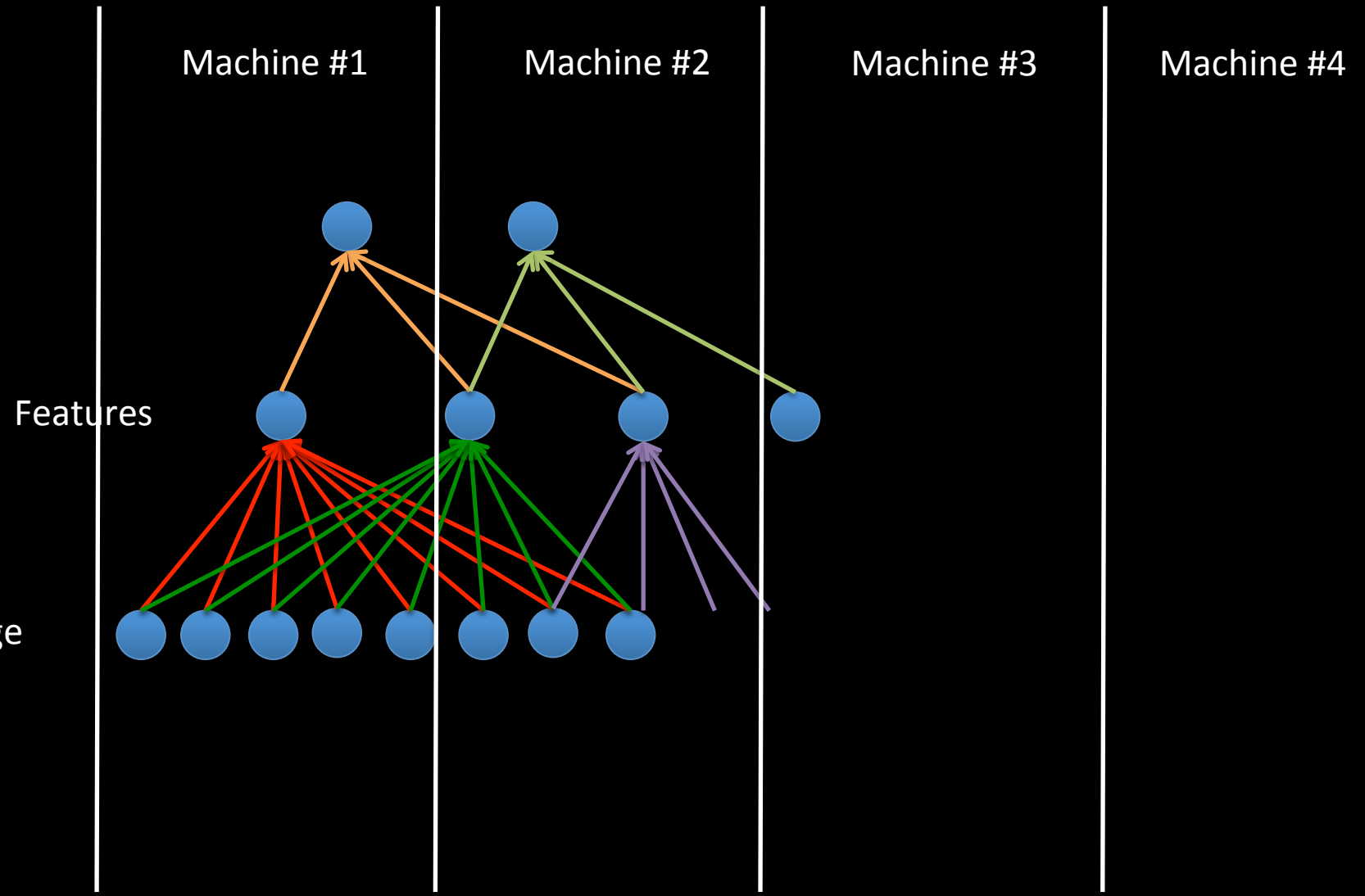
Each RICA layer = 1 filtering layer + pooling layer + local contrast normalization layer

See Le et al, NIPS 11 and Le et al, CVPR 11 for applications on action recognition, object recognition, biomedical imaging

Very large model -> Cannot fit in a single machine

-> Model parallelism, Data parallelism

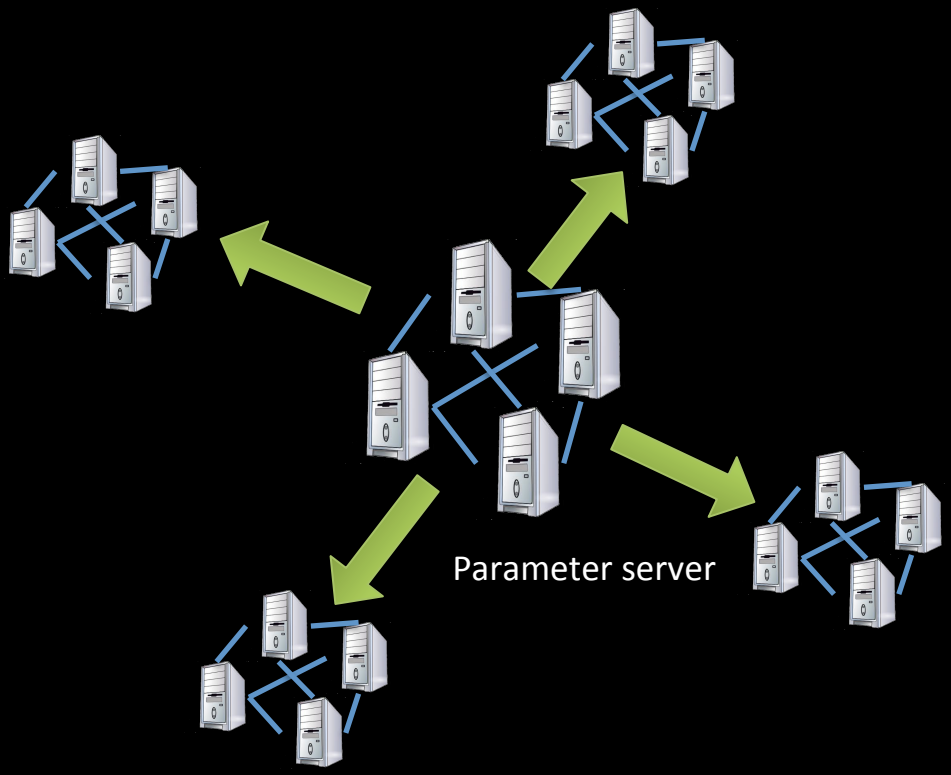
Local receptive field networks



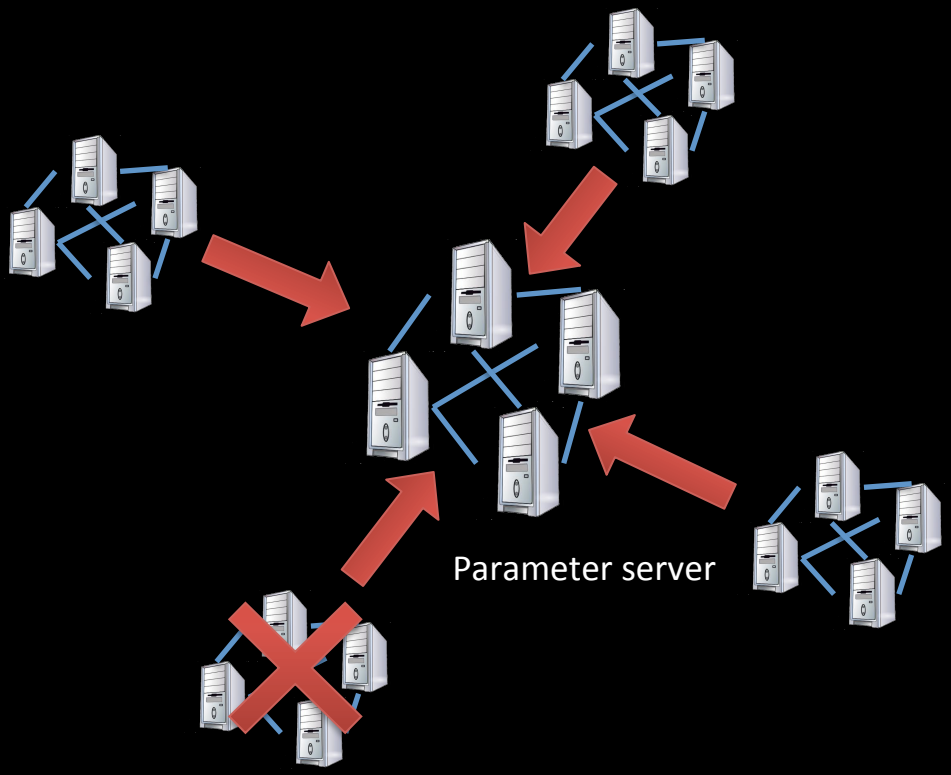
Le, et al., *Tiled Convolutional Neural Networks*. NIPS 2010

Quoc V. Le

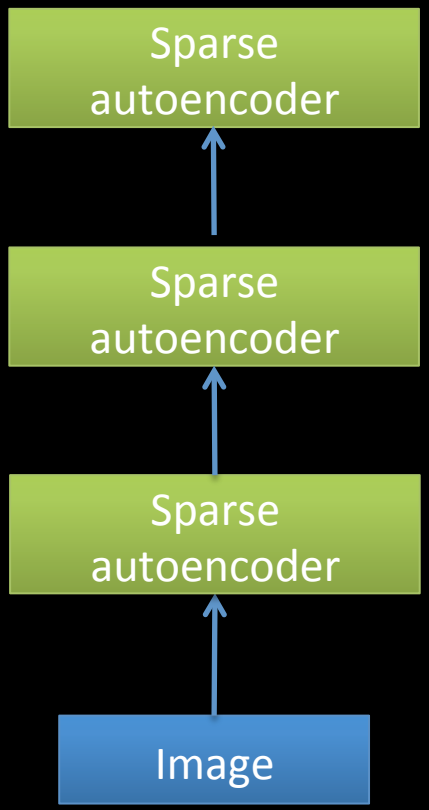
Asynchronous Parallel SGDs



Asynchronous Parallel SGDs



Training



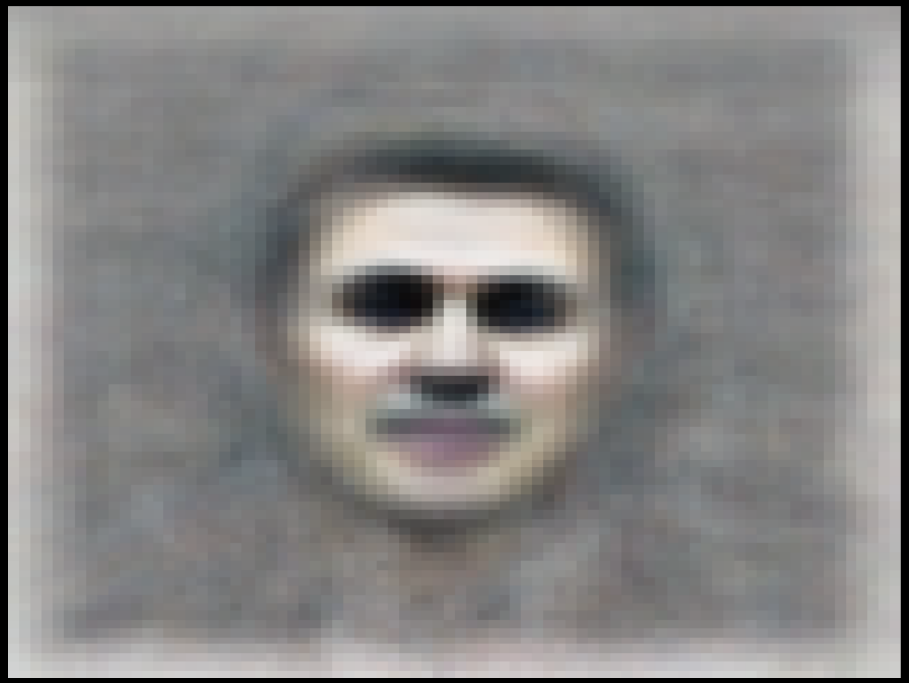
Dataset: 10 million 200x200 unlabeled images from YouTube/Web

Train on 1000 machines (16000 cores) for 1 week

1.15 billion parameters
- 100x larger than previously reported
- Small compared to visual cortex

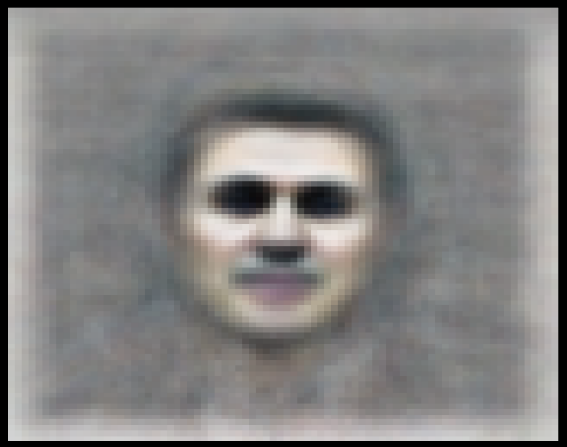
Top stimuli from the test set

Quoc V. Le



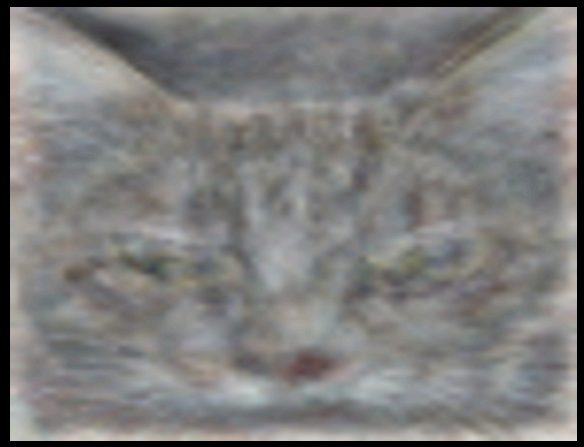
Optimal stimulus via optimization

Quoc V. Le



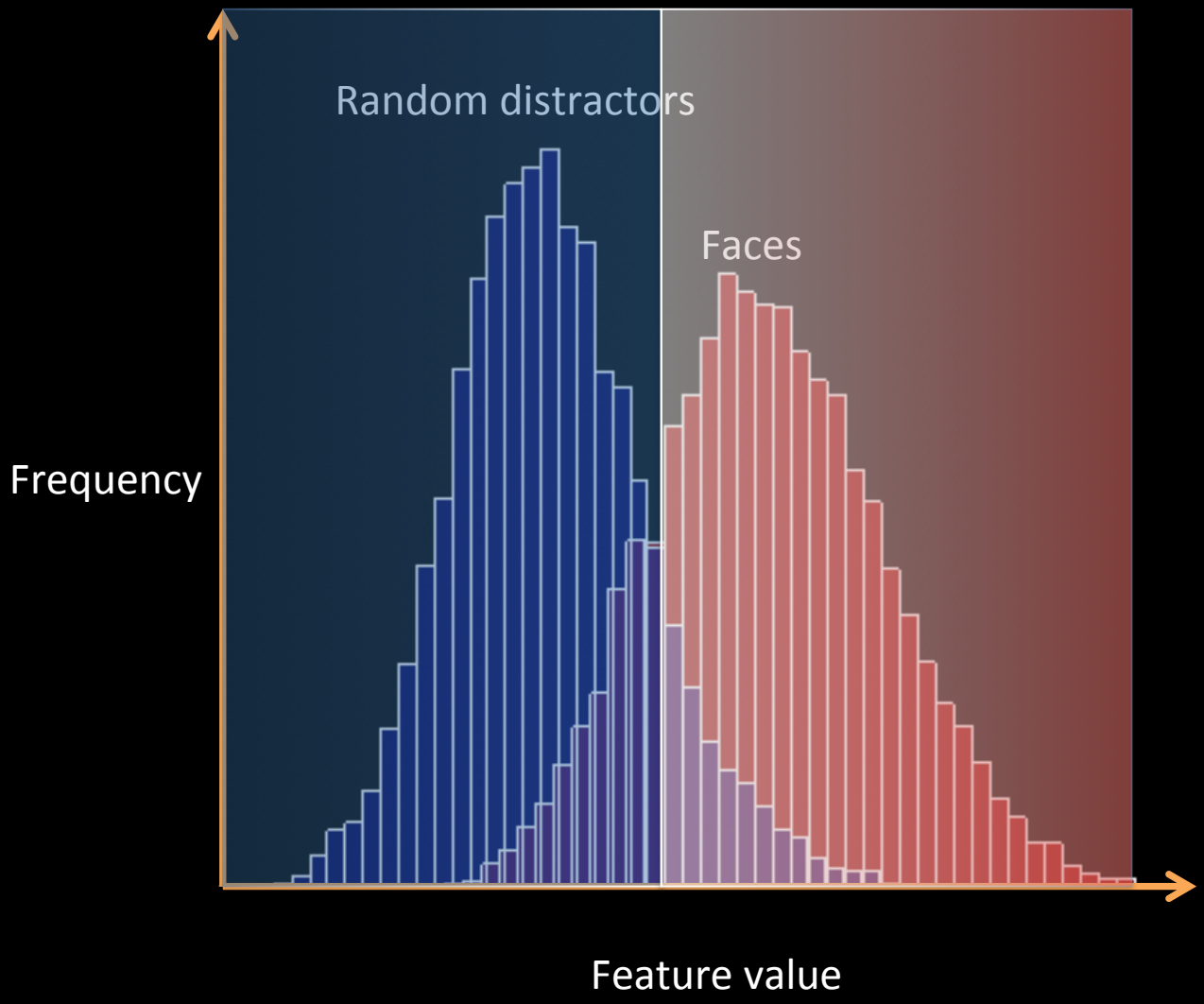
Face detector

Human body detector



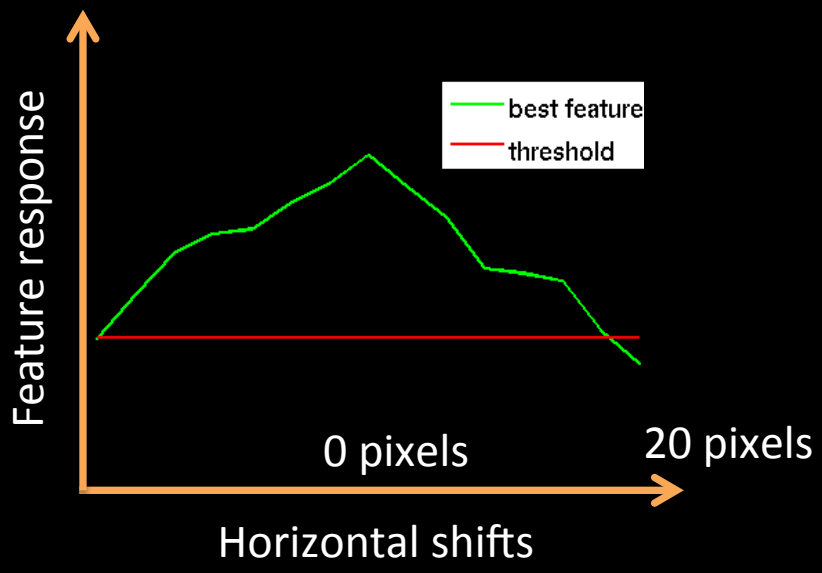
Cat detector

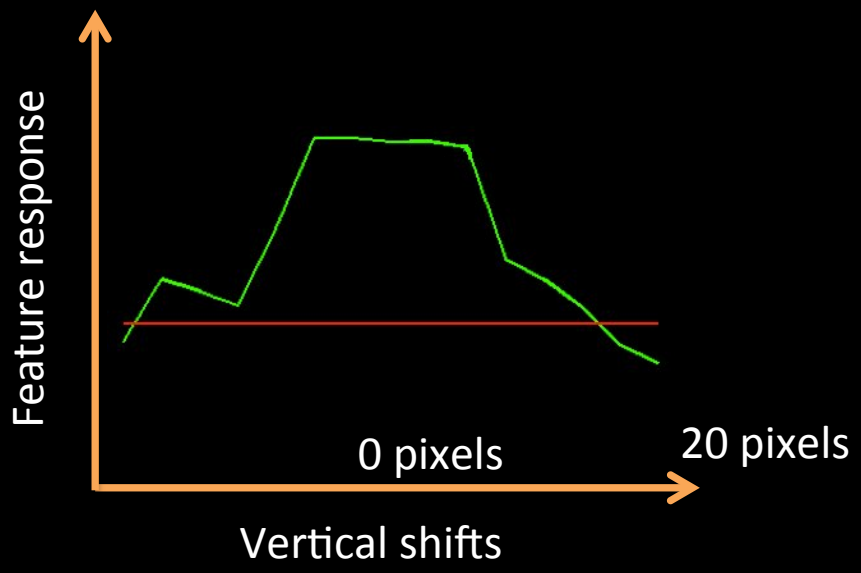
Quoc V. Le

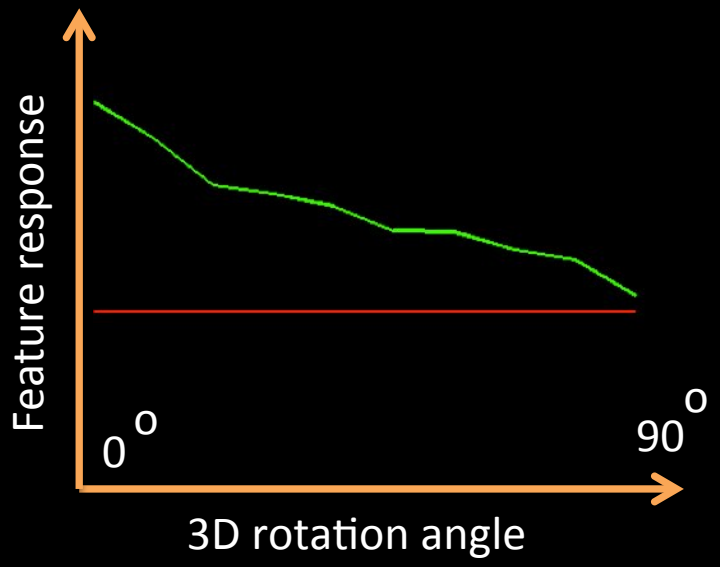


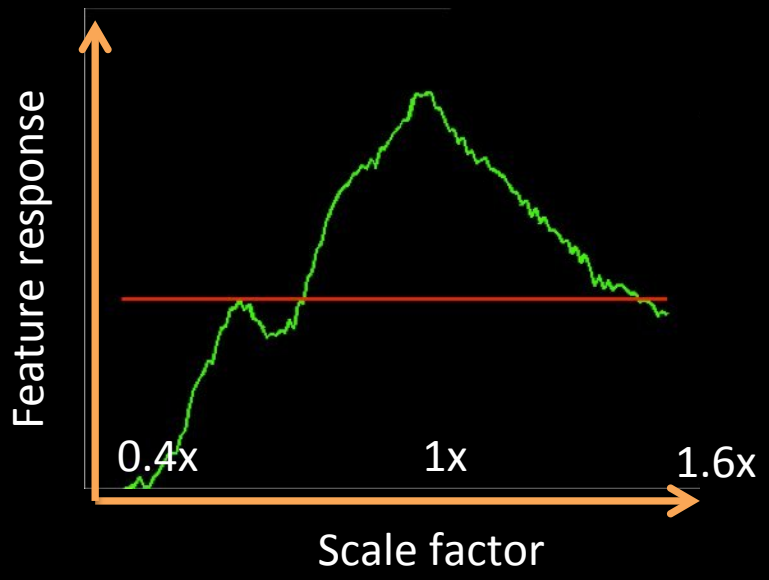
Quoc V. Le

Invariance properties









Quoc V. Le

ImageNet classification

20,000 categories, 16,000,000 images

Hand-engineered features (SIFT, HOG, LBP), Spatial pyramid, SparseCoding/Compression, Kernel SVMs

20,000 is a lot of categories…

…
smoothhound, smoothhound shark, Mustelus mustelus
American smooth dogfish, Mustelus canis
Florida smoothhound, Mustelus norrisi
whitetip shark, reef whitetip shark, Triaenodon obseus
Atlantic spiny dogfish, Squalus acanthias
Pacific spiny dogfish, Squalus suckleyi
hammerhead, hammerhead shark
smooth hammerhead, Sphyrna zygaena
smalleye hammerhead, Sphyrna tudes
shovelhead, bonnethead, bonnet shark, Sphyrna tiburo
angel shark, angelfish, Squatina squatina, monkfish
electric ray, crampfish, numbfish, torpedo
smalltooth sawfish, Pristis pectinatus
guitarfish
roughtail stingray, Dasyatis centroura
butterfly ray
eagle ray
spotted eagle ray, spotted ray, Aetobatus narinari
cownose ray, cow-nosed ray, Rhinoptera bonasus
manta, manta ray, devilfish
Atlantic manta, Manta birostris
devil ray, Mobula hypostoma
grey skate, gray skate, Raja batis
little skate, Raja erinacea
…

Stingray

Mantaray

Quoc V. Le

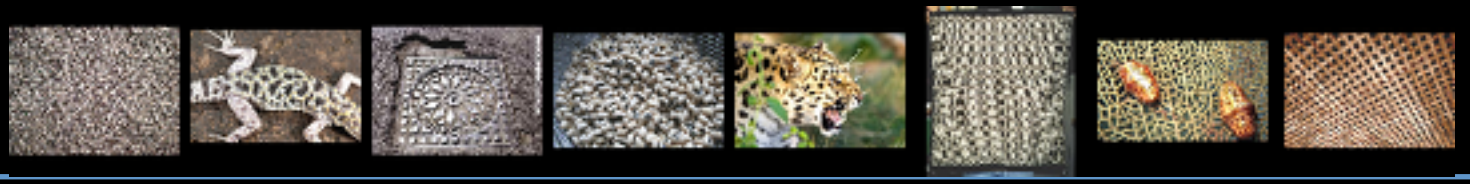
0.005%	9.5%	?
Random guess	State-of-the-art (Weston, Bengio ‘11)	Feature learning From raw pixels

0.005%	9.5%	15.8%
Random guess	State-of-the-art (Weston, Bengio ‘11)	Feature learning From raw pixels

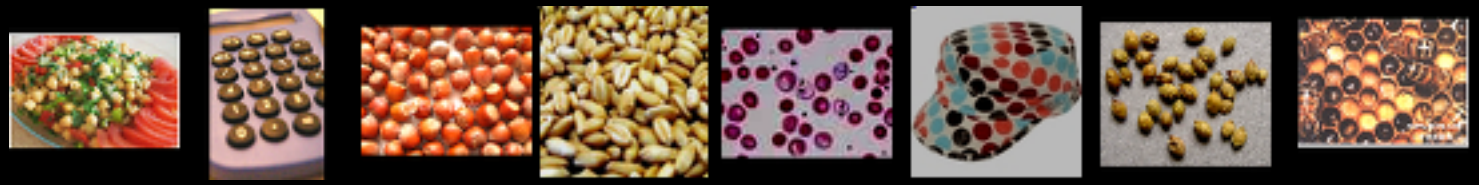
ImageNet 2009 (10k categories): Best published result: 17% (Sanchez & Perronnin ‘11), Our method: 19%

Using only 1000 categories, our method > 50%

Feature 1



Feature 2



Feature 3

Feature 4

Feature 5

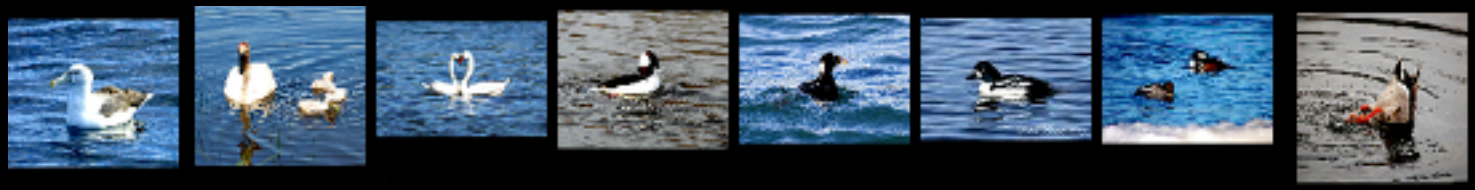
Quoc V. Le

Feature 6

Feature 7

Feature 8

Feature 9



Quoc V. Le

Feature 10

Feature 11

Feature 12

Feature 13

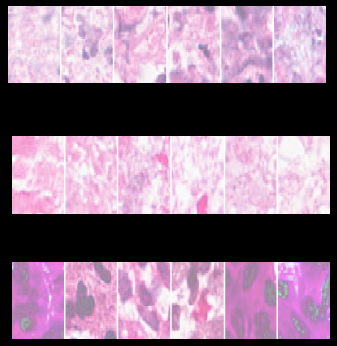
Quoc V. Le

Conclusions

- RICA learns invariant features
- Face neuron with totally unlabeled data with enough training and data
- State-of-the-art performances on
  - Action Recognition
  - Cancer image classification
  - ImageNet

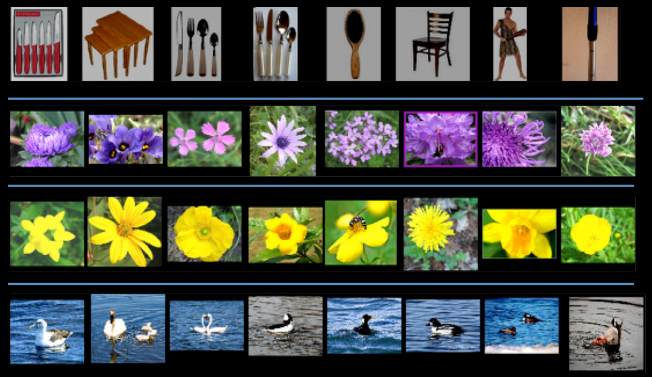
ImageNet

0.005%	9.5%	15.8%
Random guess	Best published result	Our method

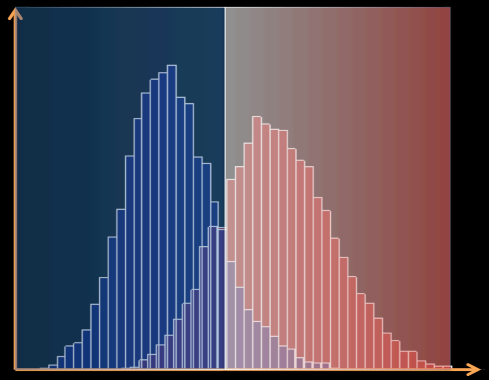


Cancer classification

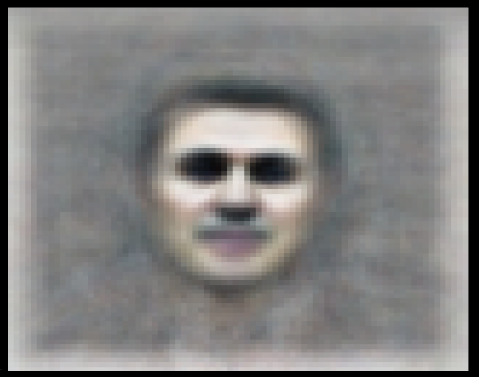
Action recognition



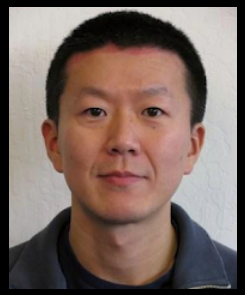
Feature visualization



Face neuron



Joint work with



Kai Chen

Greg Corrado

Jeff Dean

Matthieu Devin

Rajat Monga

Andrew Ng

Marc'Aurelio Ranzato

Paul Tucker

Ke Yang

<u>Additional Thanks:</u>

Samy Bengio, Zhenghao Chen, Tom Dean, Pangwei Koh, Mark Mao, Jiquan Ngiam, Patrick Nguyen, Andrew Saxe, Mark Segal, Jon Shlens, Vincent Vanhouke, Xiaoyun Wu, Peng Xe, Serena Yeung, Will Zou

References

- Q.V. Le, M.A. Ranzato, R. Monga, M. Devin, G. Corrado, K. Chen, J. Dean, A.Y. Ng. **Building high-level features using large-scale unsupervised learning**. *ICML*, 2012.
- Q.V. Le, J. Ngiam, Z. Chen, D. Chia, P. Koh, A.Y. Ng. **Tiled Convolutional Neural Networks**. *NIPS*, 2010.
- Q.V. Le, W.Y. Zou, S.Y. Yeung, A.Y. Ng. **Learning hierarchical spatio-temporal features for action recognition with independent subspace analysis**. *CVPR*, 2011.
- Q.V. Le, J. Ngiam, A. Coates, A. Lahiri, B. Prochnow, A.Y. Ng. **On optimization methods for deep learning**. *ICML*, 2011.
- Q.V. Le, A. Karpenko, J. Ngiam, A.Y. Ng. **ICA with Reconstruction Cost for Efficient Overcomplete Feature Learning**. *NIPS*, 2011.
- Q.V. Le, J. Han, J. Gray, P. Spellman, A. Borowsky, B. Parvin. **Learning Invariant Features for Tumor Signatures**. *ISBI*, 2012.
- I.J. Goodfellow, Q.V. Le, A.M. Saxe, H. Lee, A.Y. Ng, **Measuring invariances in deep networks**. *NIPS*, 2009.

http://ai.stanford.edu/~quocle