

Tera-scale deep learning

Quoc V. Le
Stanford University and Google

Joint work with



Kai Chen



Greg Corrado



Jeff Dean



Matthieu Devin



Rajat Monga



Andrew Ng



Marc'Aurelio
Ranzato



Paul Tucker

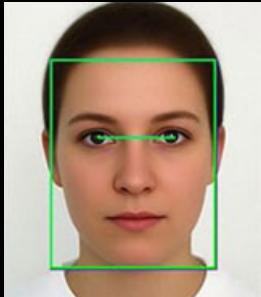


Ke Yang

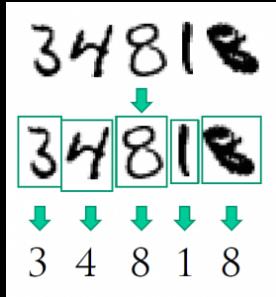
Additional
Thanks:

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

Machine Learning successes



Face recognition



OCR



Autonomous car

Gmail by Google

Mail
Contacts
Tasks
Compose mail

Priority Inbox (2)
Inbox (2603)
Buzz (52) 🇧🇷
Starred ⭐
Sent Mail
Drafts (15)
Spam (163)

Email classification

Hello, Scott Wheeler. We have recommendations for you. (Not Scott?)
Scott's Amazon.com Today's Deals Gifts & Wish Lists Gift Cards

Scott, Welcome to Your Amazon.com (If you're not Scott Wheeler, click here.)

Today's Recommendations For You
Here's a daily sample of items recommended for you. Click here to see all recommendations.

Russia Map by Michelin International David M. By Rilem Publishing Ltd
TIBETO In Search of Sunrise, Vol. 1: Asia By GU Books \$15.99
Land of the Horizons: A History of the Silk Road By Jason Goodwin \$11.95

Recommendation systems

Local results for starbucks near Chicago, IL
Starbucks Coffee: Chicago - 2.3 miles N - 400 W Madison St # 5, Chicago, 60606 - (312) 454-1010
Starbucks - 2.3 miles NE - 55 E Jackson Blvd # 1, Chicago, 60604 - (312) 786-9201
Starbucks Coffee: Chicago - 2.9 miles NE - 35 E Wacker Dr, Chicago, 60601 - (312) 541-8317

Local Search Results
Starbucks in Chicagoland
This friendly neighborhood Starbucks is extra-spacious, ... of local hero Joe DiMaggio in this first Starbucks in the Little Italy neighborhood of Chicago. ...
www.starbuckseverywhere.net/Chicagoland.htm - 127k - Cached - Similar pages

Google Organic Search Results
Starbucks in Illinois
Illinois Chicagoland - Illinois Remote
www.starbuckseverywhere.net/illir
Cached - Similar pages

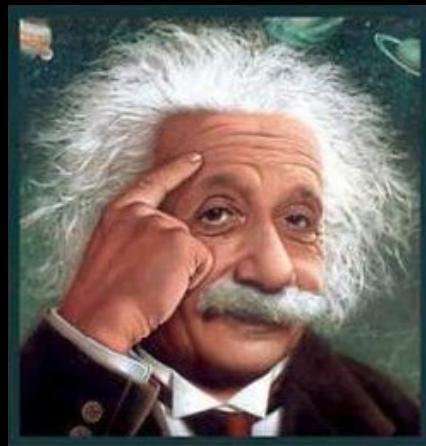
Spons
Starbucks
Get Local Direct...
Phone Numbers
MapQuest.com

Starbucks Ch
Whatever you're l...
you can get it on
www.eBay.com

Buy Starbuck
Find Starbucks O...
eBay Express Of...
www.eBayExpress

Web page ranking

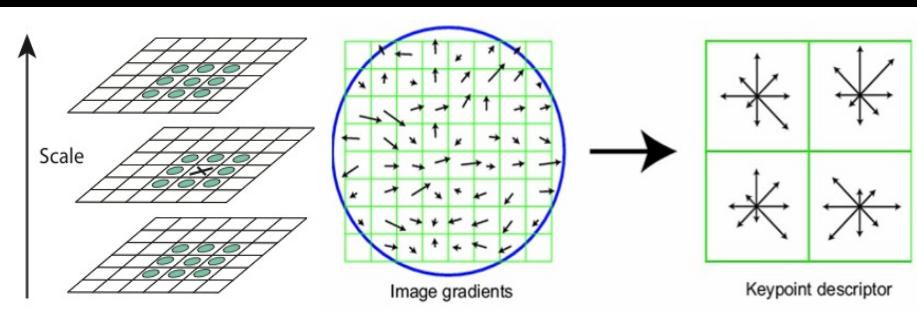
Feature Extraction



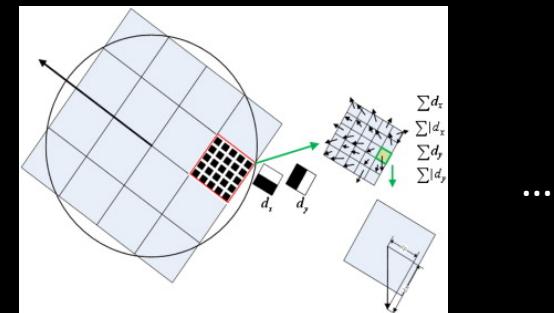
Feature extraction
(Mostly hand-crafted features)

Hand-Crafted Features

Computer vision:

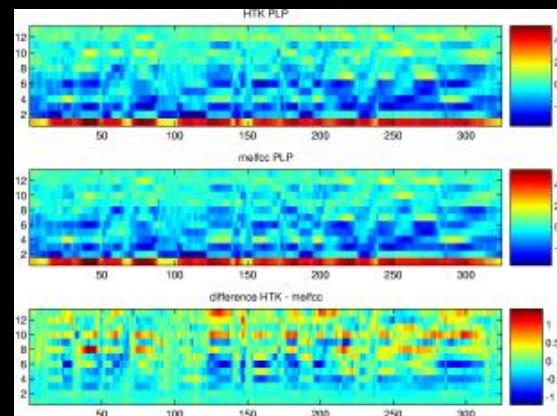


SIFT/HOG

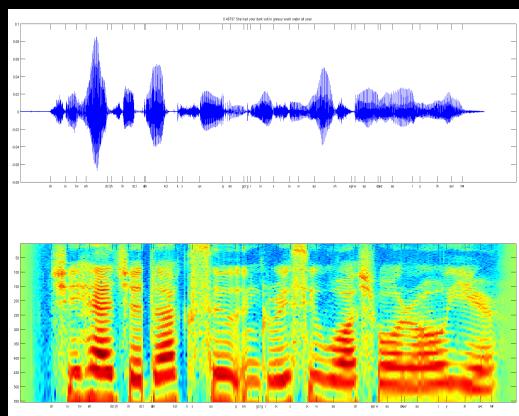


SURF

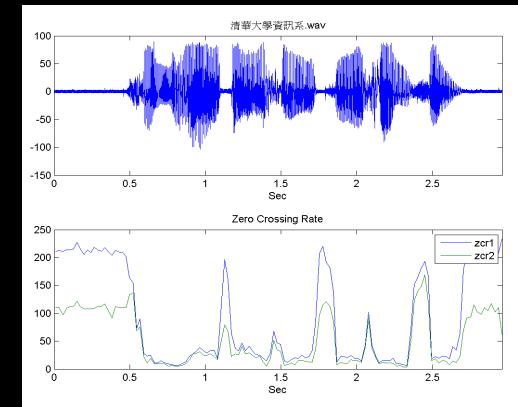
Speech Recognition:



MFCC



Spectrogram



ZCR

New feature-designing paradigm

Unsupervised Feature Learning / Deep Learning

Reconstruction ICA

Expensive and typically applied to small problems

The Trend of BigData

The image displays three separate web browser windows side-by-side, each showing search results for the query "mgm".

- Facebook Search Results:** Shows four search results:
 - MGM (Musician or Band) - 518,981 fans
 - MGM Grand Las Vegas (Brand or Company) - 49,470 fans
 - MGM (Movie) - 94 fans
 - MGM Mondo del Vino (Brand or Company) - 341 fansA link "See More Results for mgm" is visible at the bottom.
- Twitter Home Page:** Shows the Twitter header "Welcome to #NewTwitter! Read up on what's new. You can also leave the..." and the search bar. Below it, the "What's happening?" timeline section shows five new tweets. Navigation tabs include Timeline, @Mentions, Retweets, Searches, and Lists.
- YouTube Home Page:** Shows the YouTube logo and navigation menu (Videos, Categories, Channels, Community, Upload Videos). It features sections for "Director Videos" (with thumbnails for "SWEET TALK", "Moosweo's Appeal", "Ninja Star in the Eye", and "Low Water - Strange") and "Featured Videos" (with thumbnails for "Piano Lesson With A Difference" and "Everyday Bravery"). On the right, there is a video player for "30 ROCK – TINA FEY – NEW FINALE" and a sidebar for "WATCH HER DO IT BELOW."

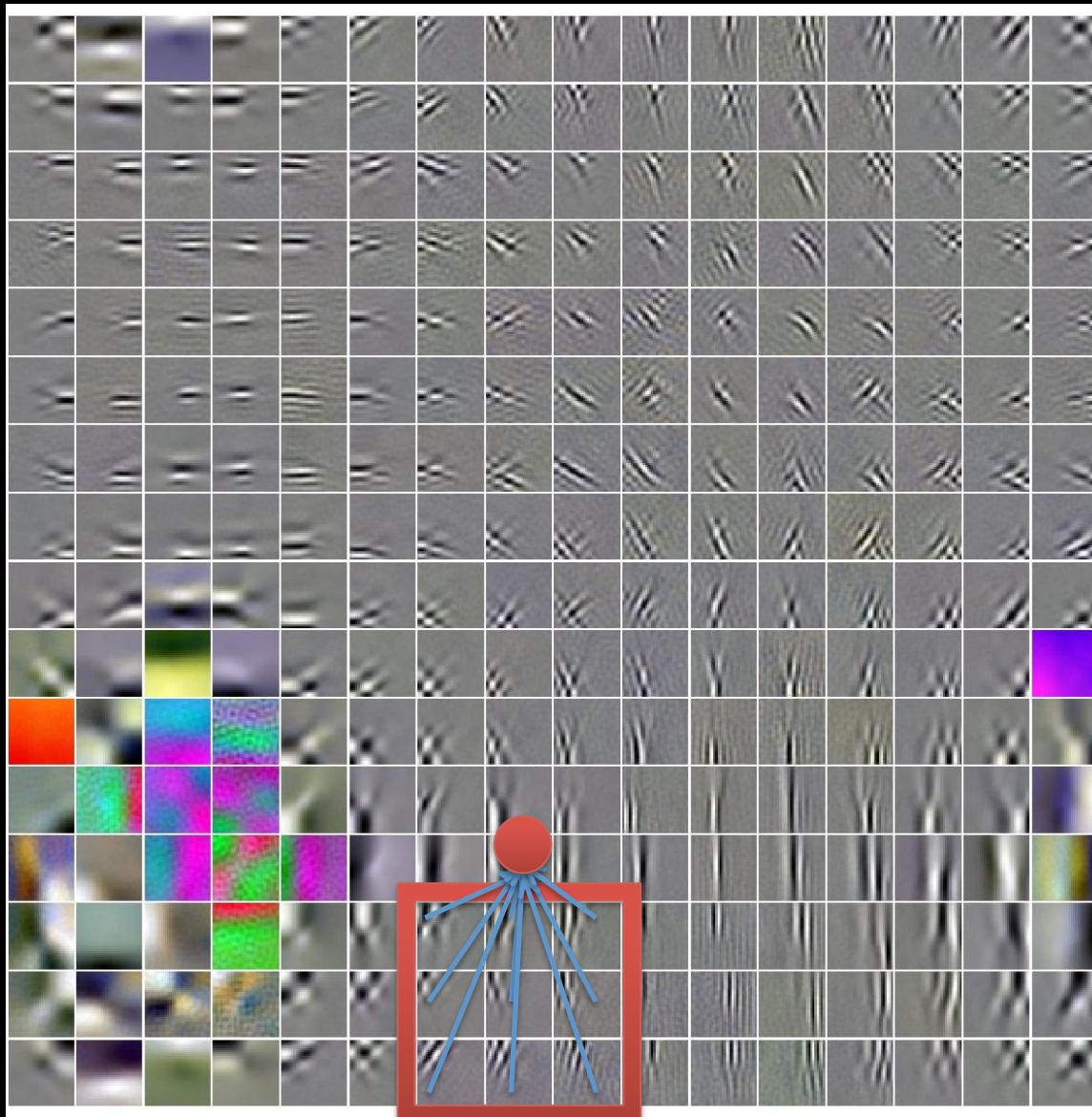
This screenshot provides a detailed view of the YouTube homepage interface.

- Header:** YouTube logo, Broadcast Yourself!, Sign Up, My Account, History, Help, Login, Search bar, and Upload Videos button.
- Top Navigation:** Director Videos, Videos, Categories, Channels, Community, and a user profile section for "My: Videos - Favorites - Playlists - Inbox - Subscriptions".
- Content Sections:**
 - Director Videos:** Displays four video thumbnails: "SWEET TALK" by LorcarrFinnegan, "Moosweo's Appeal" by tomwalkt, "Ninja Star in the Eye" by kSchwarze, and "Low Water - Strange" by lowwater.
 - Featured Videos:** Displays two video thumbnails: "Piano Lesson With A Difference" by DukeRighteous and "Everyday Bravery" by sniper20b.
 - Right Sidebar:** Shows a video player for "30 ROCK – TINA FEY – NEW FINALE" and a thumbnail for "WATCH HER DO IT BELOW." featuring Tina Fey and Steve Carell.
- Bottom Navigation:** Includes links for "Become a guest editor", "More in Music", and "Login to rate video".

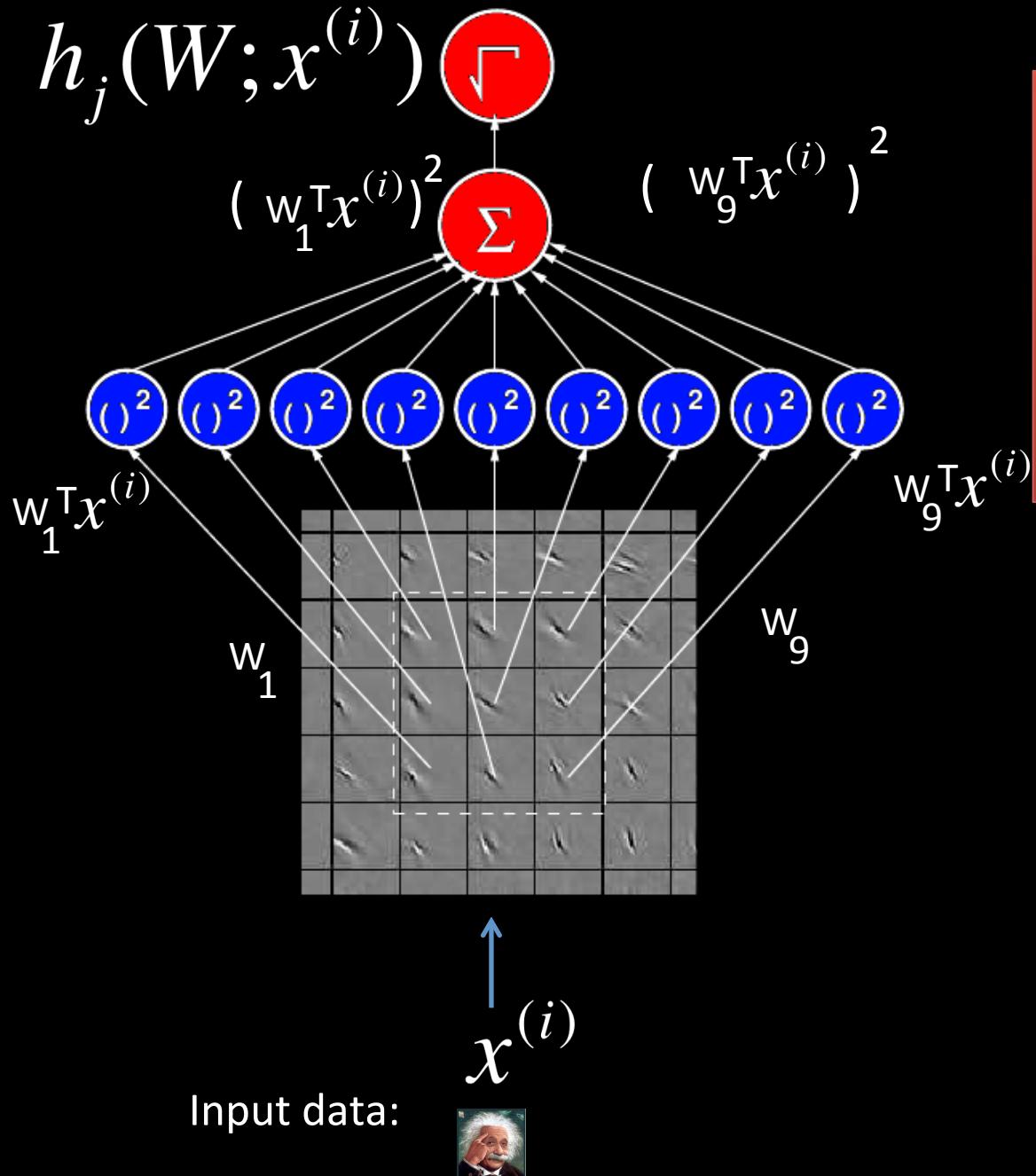
Outline

- Reconstruction ICA
- Applications to videos, cancer images
- Ideas for scaling up
- Scaling up Results

Topographic Independent Component Analysis (TICA)



1. Feature computation



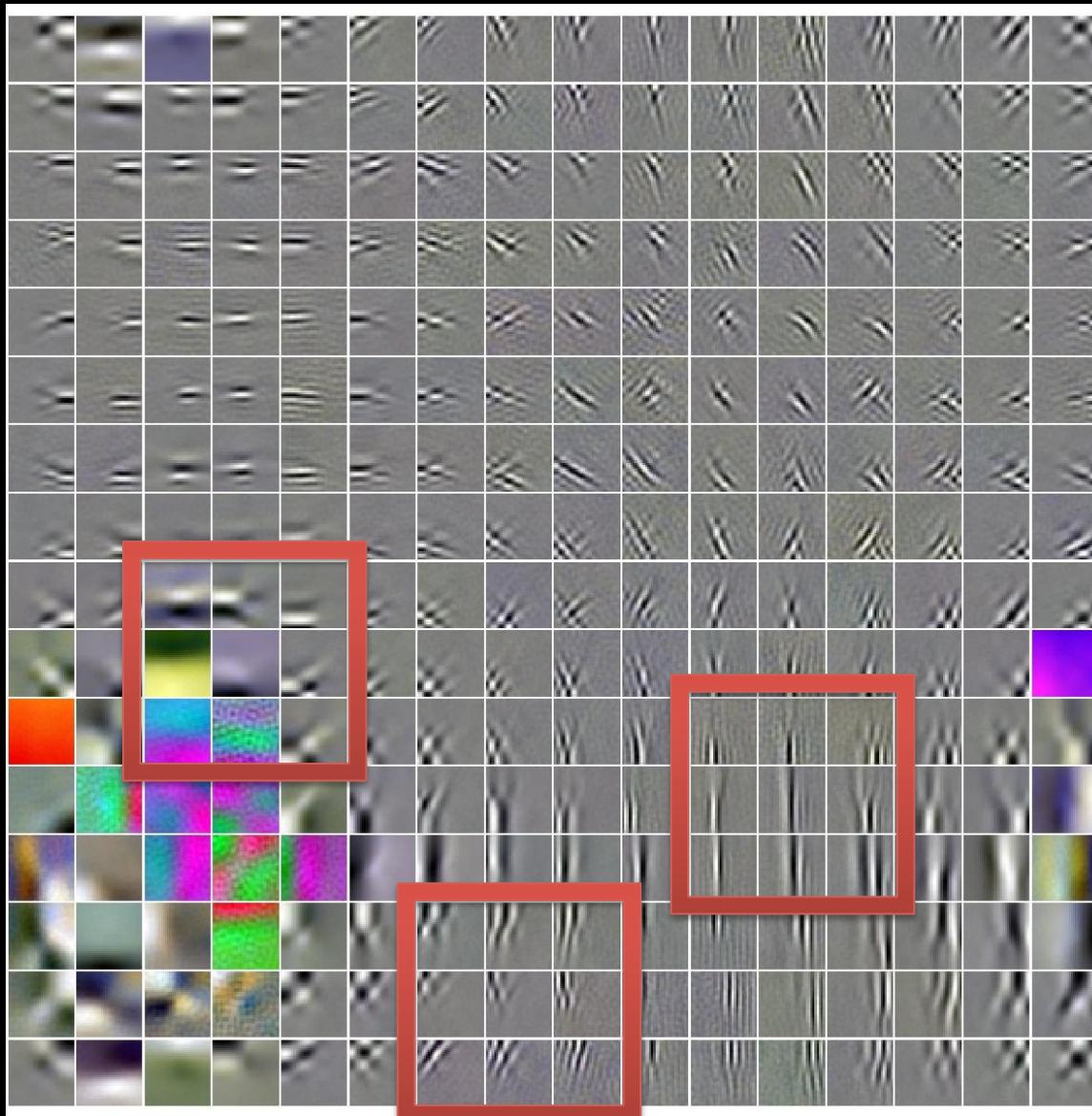
2. Learning

$$\min_W \sum_j \sum_i h_j(W; x^{(i)})$$

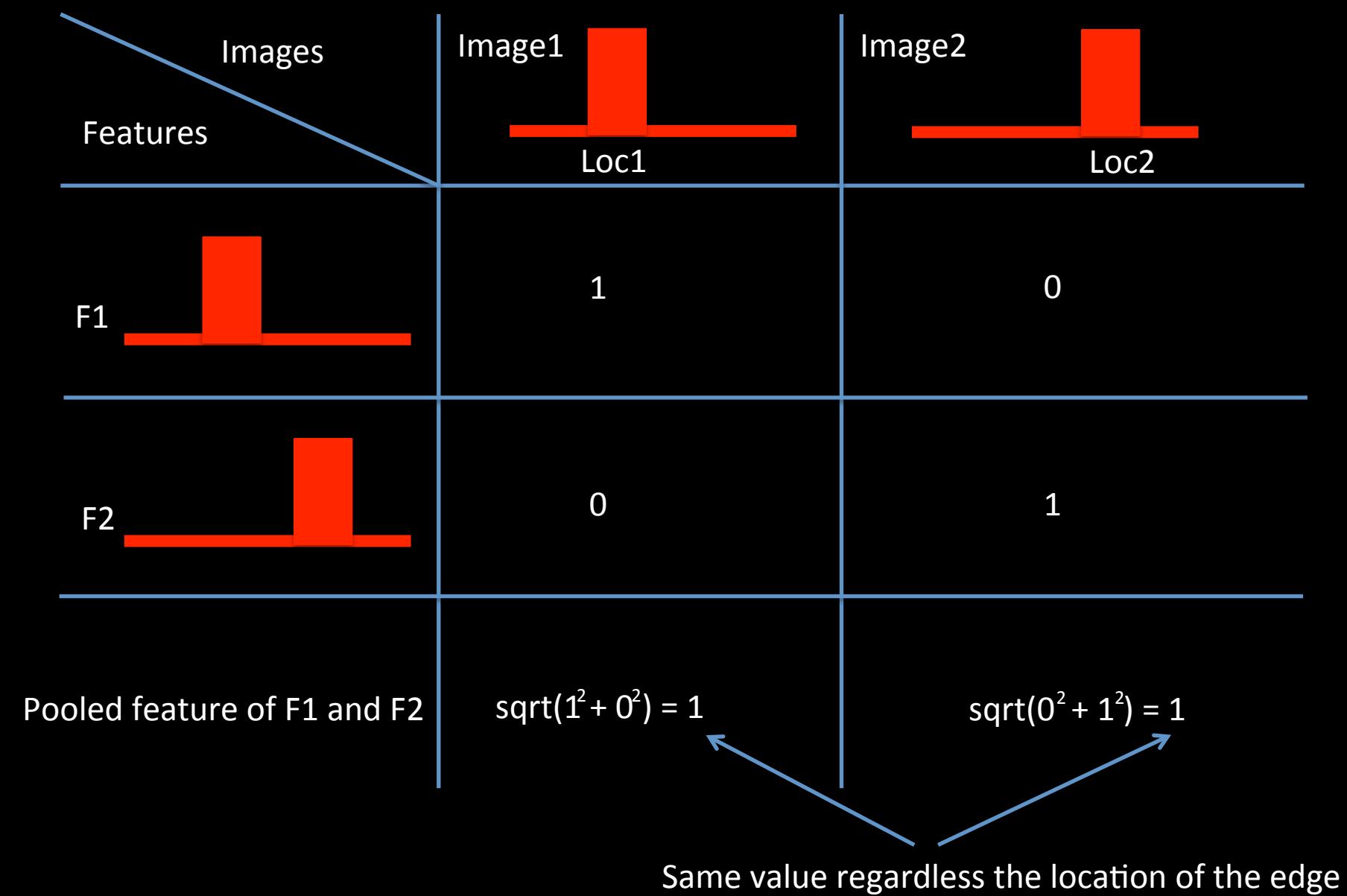
$$s.t. \quad WW^T = I$$

$$W = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{10000} \end{bmatrix}$$

Topographic Independent Component Analysis (TICA)



Invariance explained



TICA:

$$\min_W \sum_j \sum_i h_j(W; x^{(i)})$$

$$s.t. \quad \boxed{WW^T = I}$$

Reconstruction ICA:

$$\longrightarrow \min_W \left[\frac{\lambda}{m} \sum_{i=1}^m \|W^T W x^{(i)} - x^{(i)}\|_2^2 + \sum_j \sum_i h_j(W; x^{(i)}) \right]$$

Lemma 3.1 When the input data $\{x^{(i)}\}_{i=1}^m$ is whitened, the reconstruction cost $\frac{\lambda}{m} \sum_{i=1}^m \|W^T W x^{(i)} - x^{(i)}\|_2^2$ is equivalent to the orthonormality cost $\lambda \|W^T W - \mathbf{I}\|_{\mathcal{F}}^2$.

Lemma 3.2 The column orthonormality cost $\lambda \|W^T W - \mathbf{I}_n\|_{\mathcal{F}}^2$ is equivalent to the row orthonormality cost $\lambda \|WW^T - \mathbf{I}_k\|_{\mathcal{F}}^2$ up to an additive constant.

- Equivalence between Sparse Coding, Autoencoders, RBMs and ICA
- Build deep architecture by treating the output of one layer as input to another layer

Reconstruction ICA:

$$\min_W \left[\frac{\lambda}{m} \sum_{i=1}^m \|W^T W x^{(i)} - x^{(i)}\|_2^2 + \sum_j \sum_i h_j(W; x^{(i)}) \right]$$

Reconstruction ICA:

$$\frac{\lambda}{m} \sum_{i=1}^m \|W^T W x^{(i)} - x^{(i)}\|_2^2$$

$$WW^T = I$$

$$\lambda \|WW^T - I\|_F^2 \quad \leftarrow \quad \lambda \|W^T W - I\|_F^2$$

↓ Data whitening

TICA:

$$\min_W \sum_{i=1}^m \sum_{j=1}^N p_j^i(W; x^{(i)})$$

s.t.

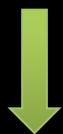
$$WW^T = I$$



$$\lambda \|WW^T - I\|_F^2$$

Reconstruction ICA:

$$\min_W \left[\frac{\lambda}{m} \sum_{i=1}^m \|W^T W x^{(i)} - x^{(i)}\|_2^2 + \sum_j \sum_i h_j(W; x^{(i)}) \right]$$



Data whitening

$$\lambda \|W^T W - I\|_F^2$$

Why RICA?

Algorithms	Speed	Ease of training	Invariant Features
Sparse Coding	✗	✓	✗
RBMs/Autoencoders	✓	✗	✗
TICA	✗	✓	✓
Reconstruction ICA	✓	✓	✓

Summary of RICA

- Two-layered network
- Reconstruction cost instead of orthogonality constraints
- Learns invariant features

Applications of RICA

Action recognition



Sit up



Drive Car



Get Out of Car



Eat



Answer phone



Kiss



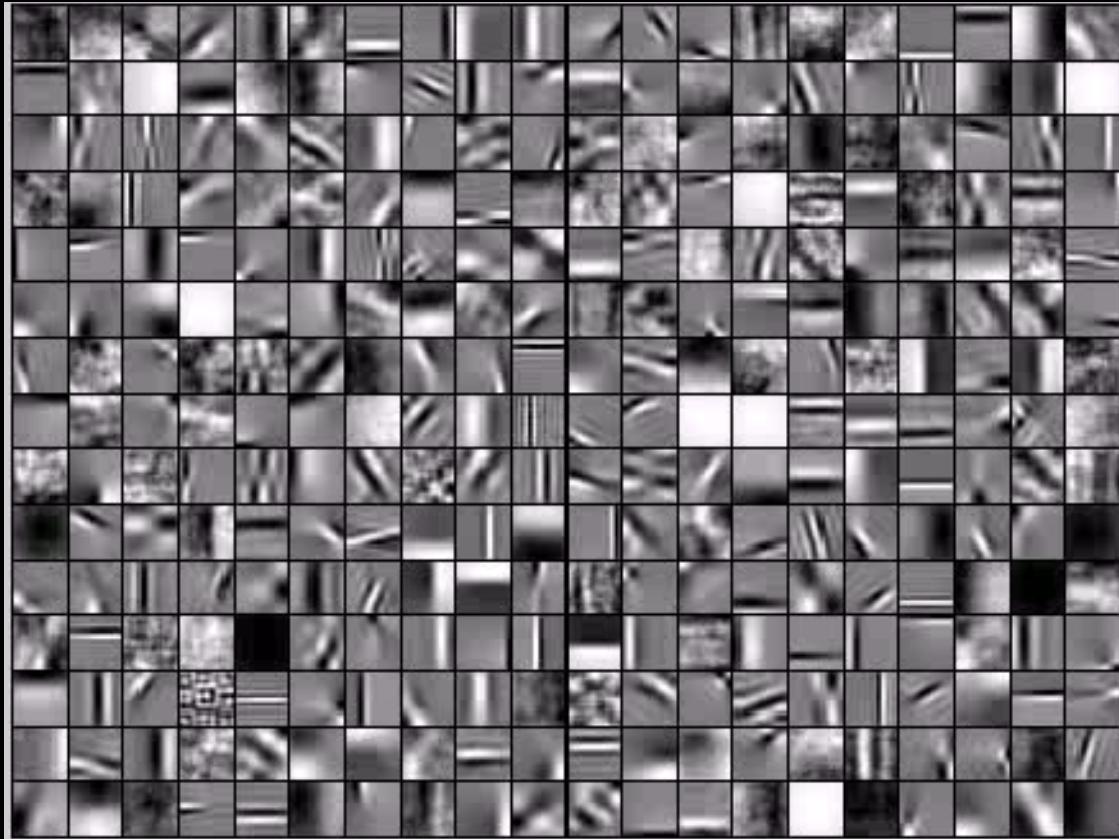
Run



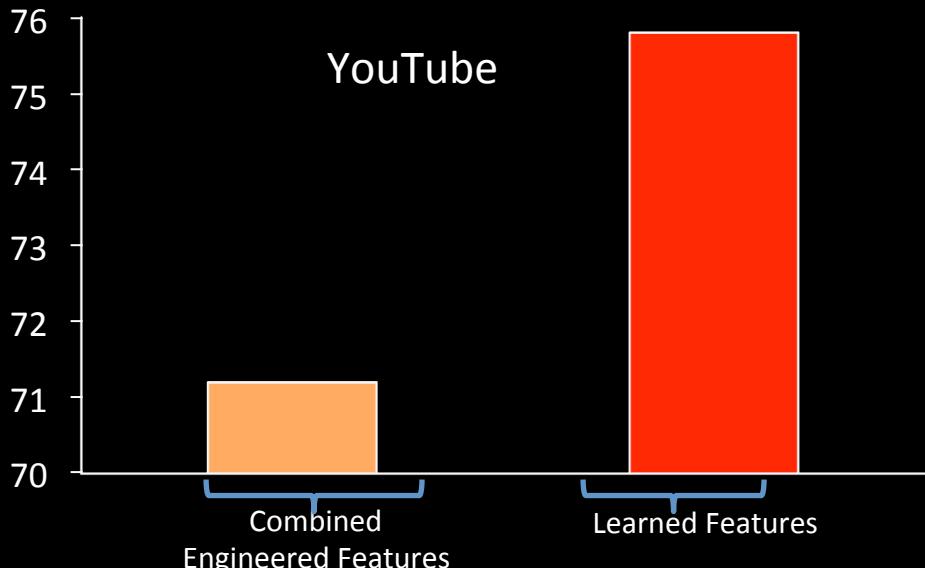
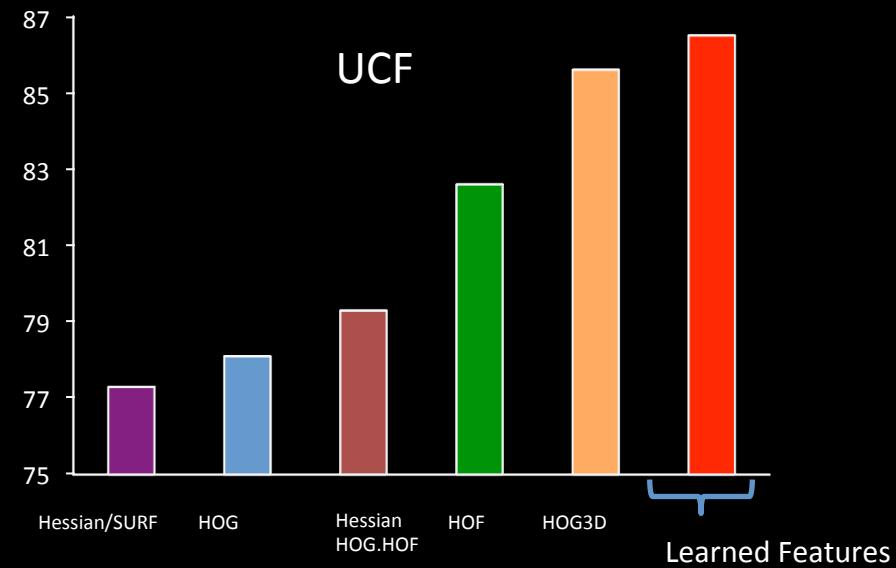
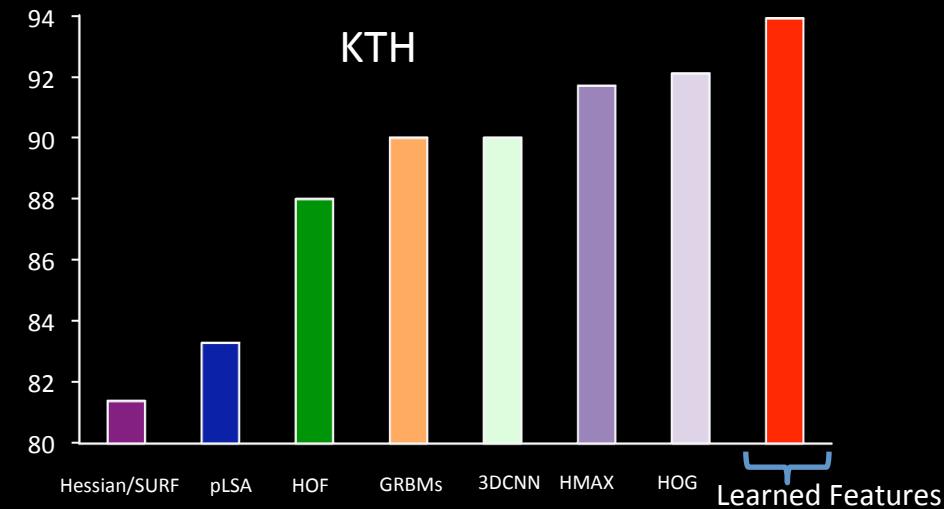
Stand up



Shake hands



Le, et al., *Learning hierarchical spatio-temporal features for action recognition with independent subspace analysis*. CVPR 2011



Cancer classification

Apoptotic



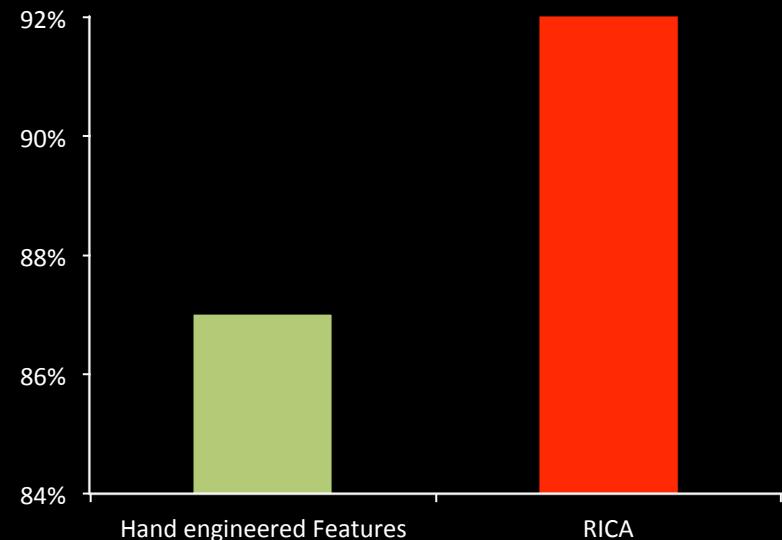
Viable tumor
region



Necrosis



:

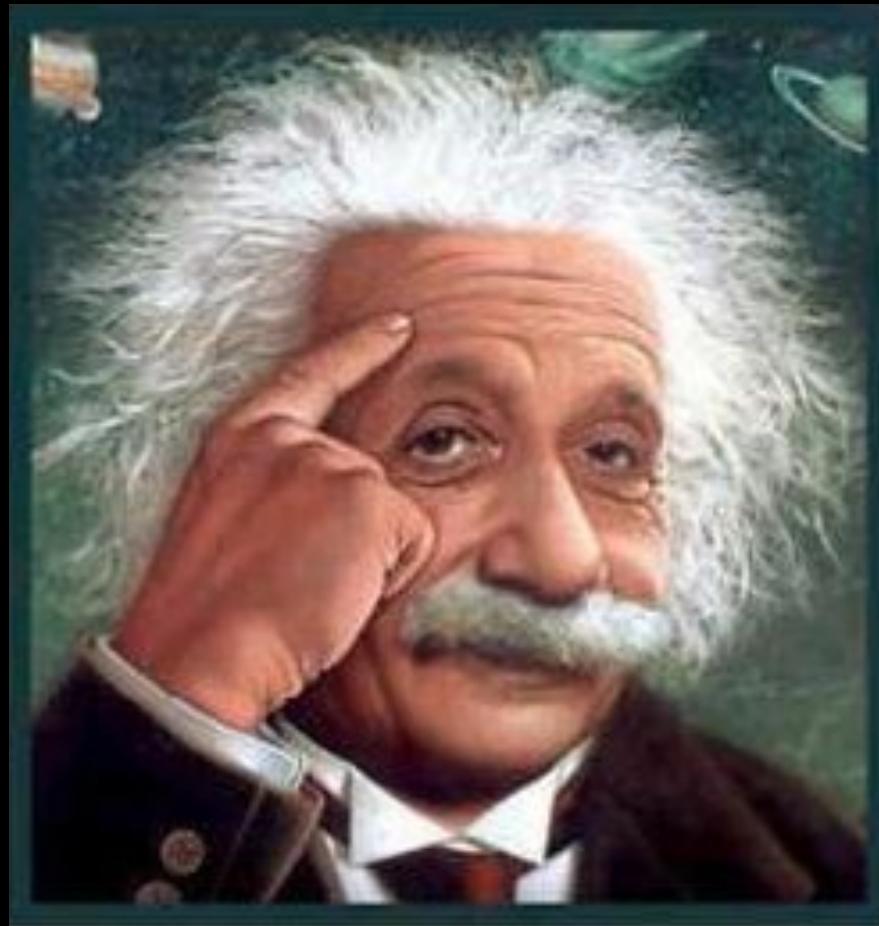


Scaling up deep RICA networks

Scaling up Deep Learning

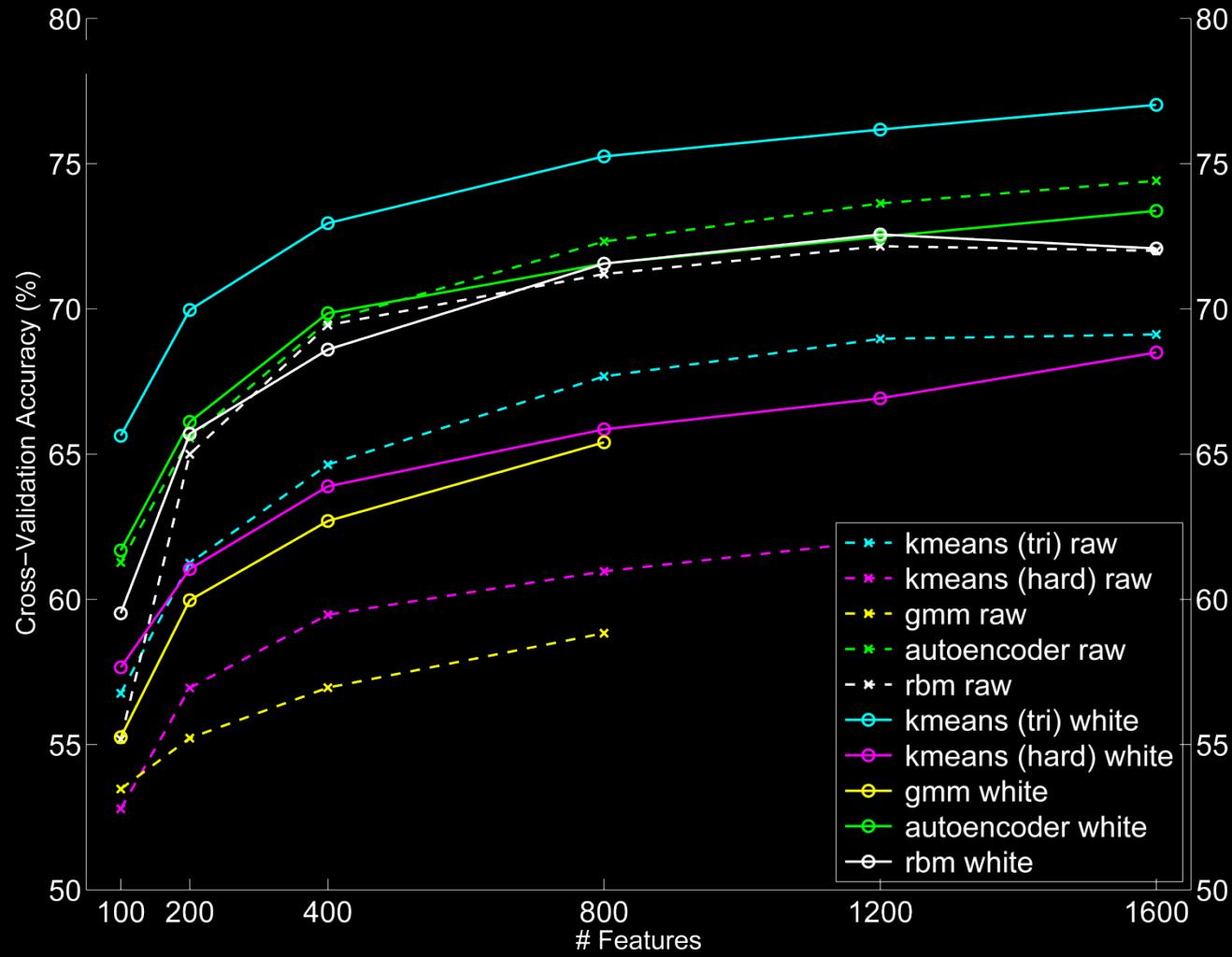


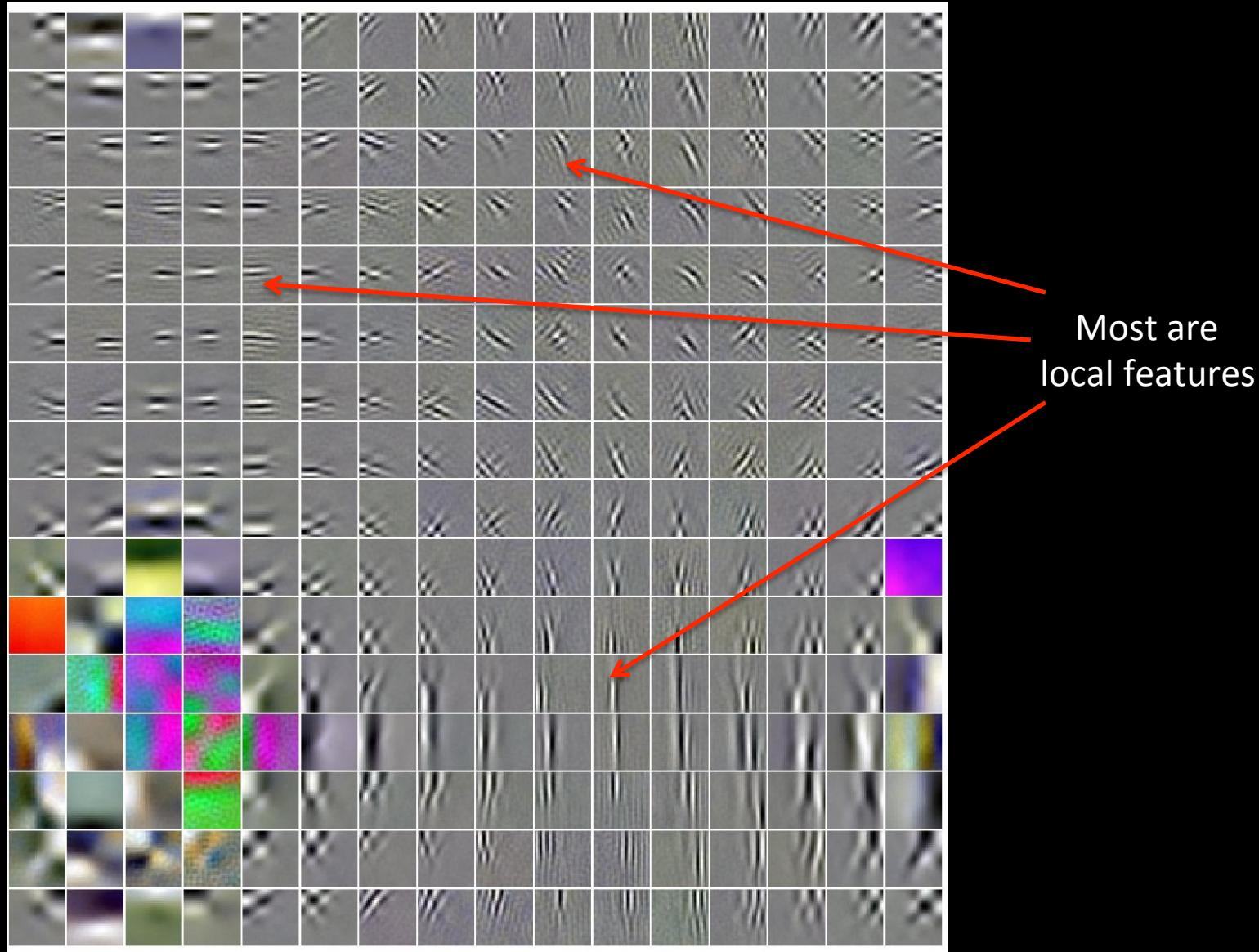
Deep learning data



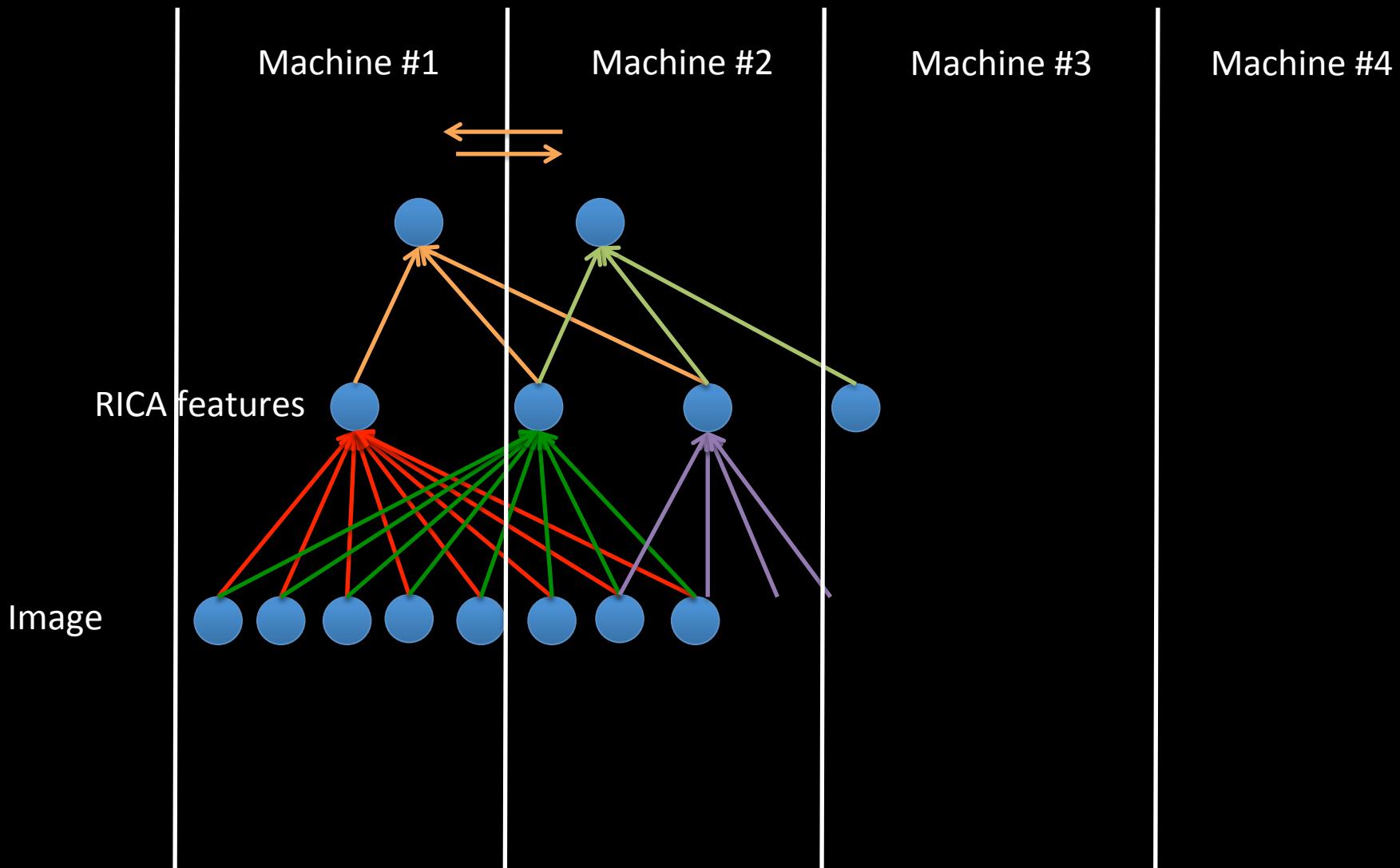
Real data

It's better to have more features!



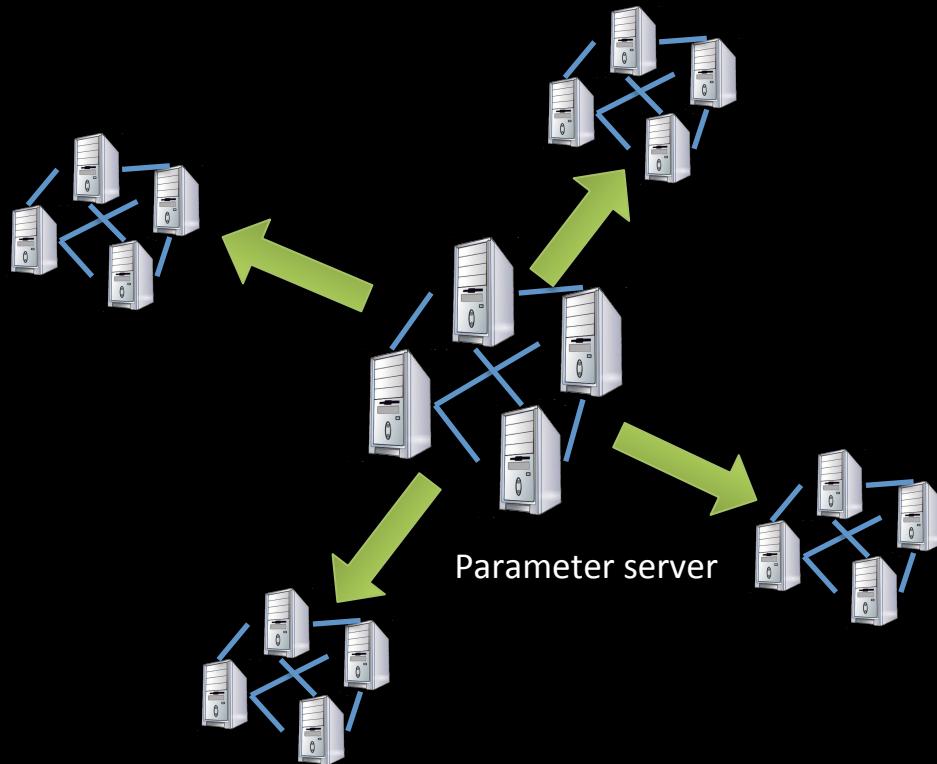


Local receptive field networks

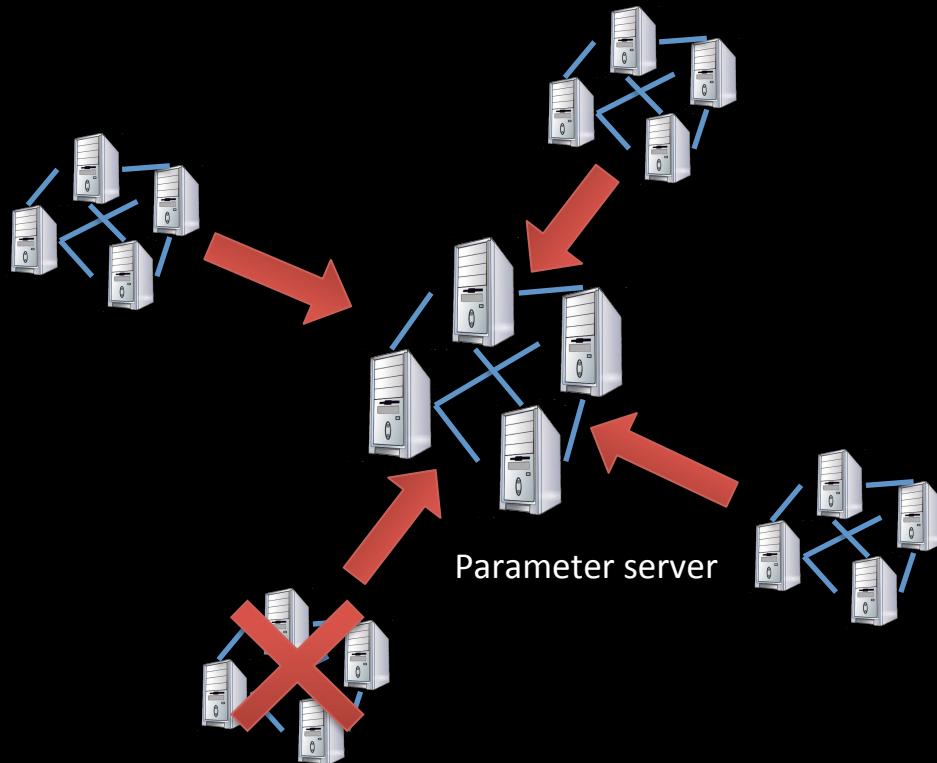


Challenges with 1000s of machines

Asynchronous Parallel SGDs



Asynchronous Parallel SGDs



Summary of Scaling up

- Local connectivity
- Asynchronous SGDs

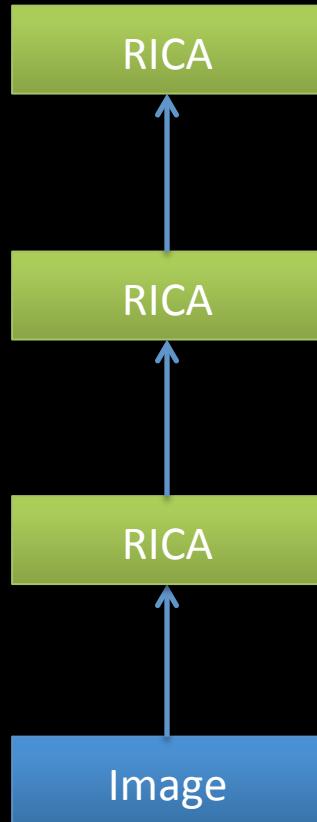
... And more

- RPC vs MapReduce
- Prefetching
- Single vs Double
- Removing slow machines
- Optimized Softmax
- ...

10 million 200x200 images

1 billion parameters

Training



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

Train on 2000 machines (16000 cores) for 1 week

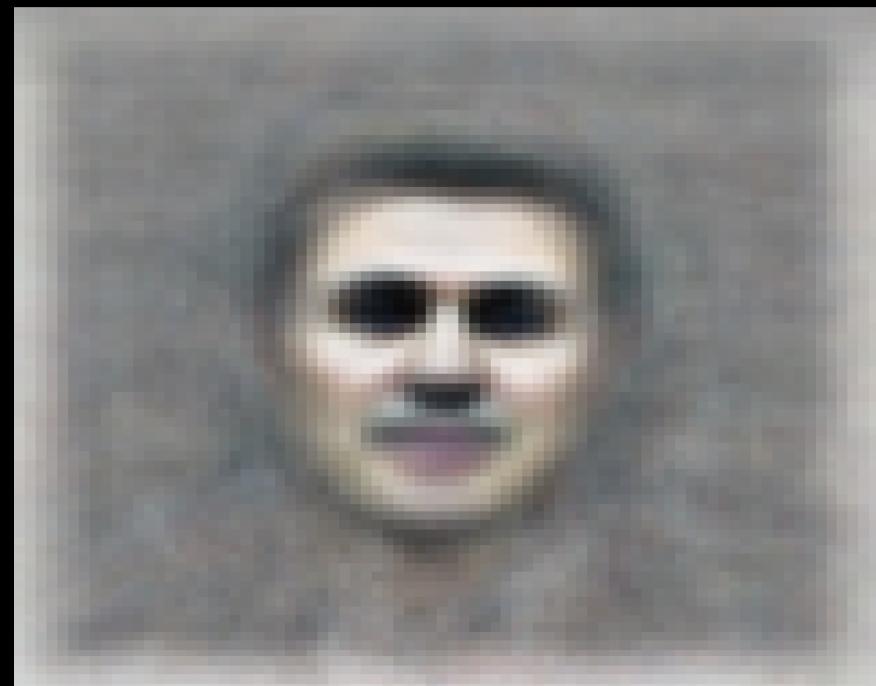
1.15 billion parameters

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

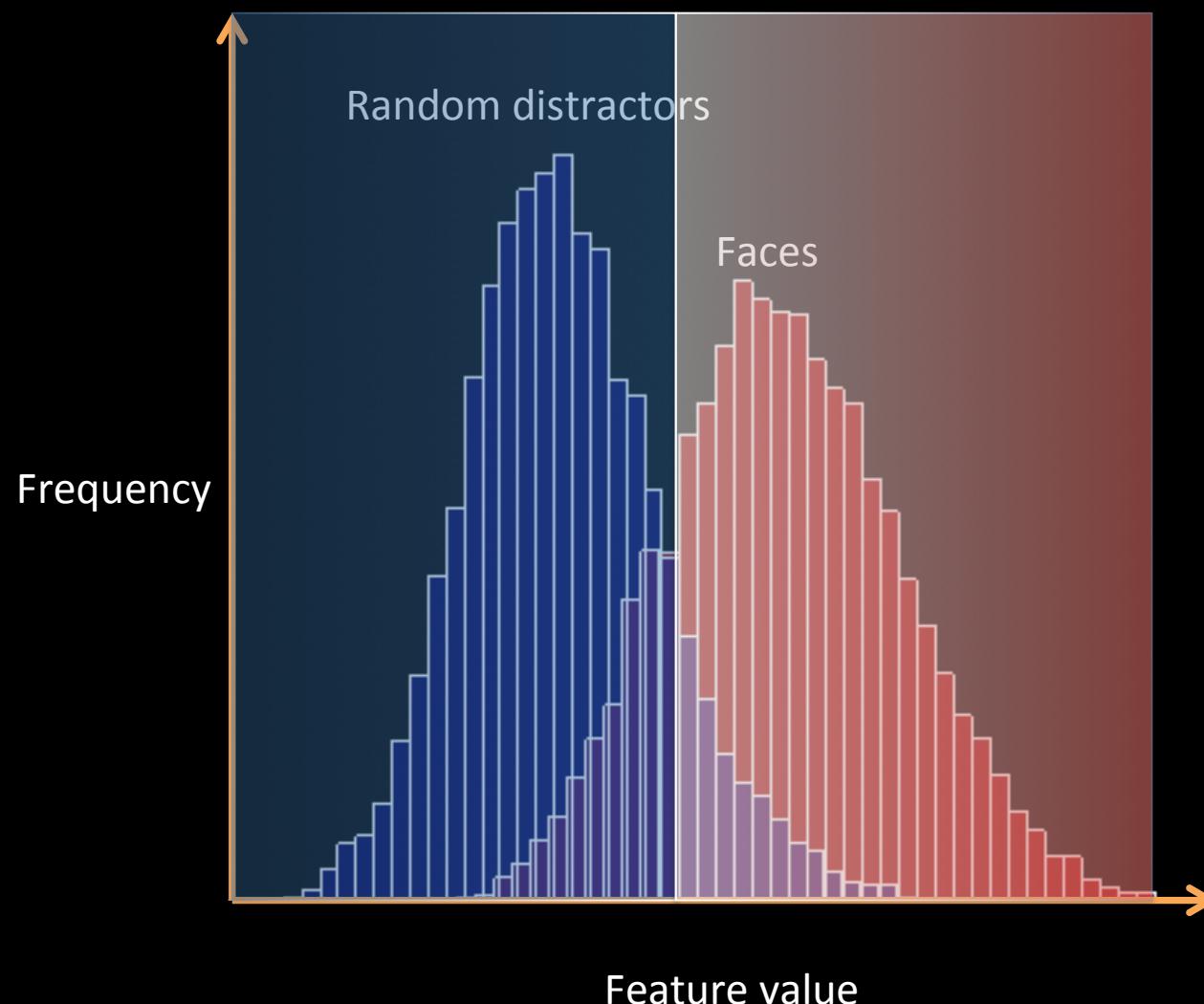
The face neuron



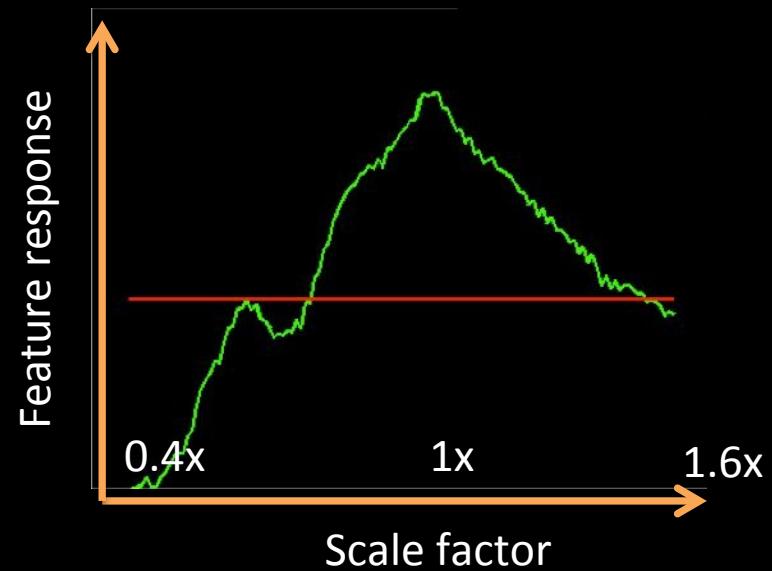
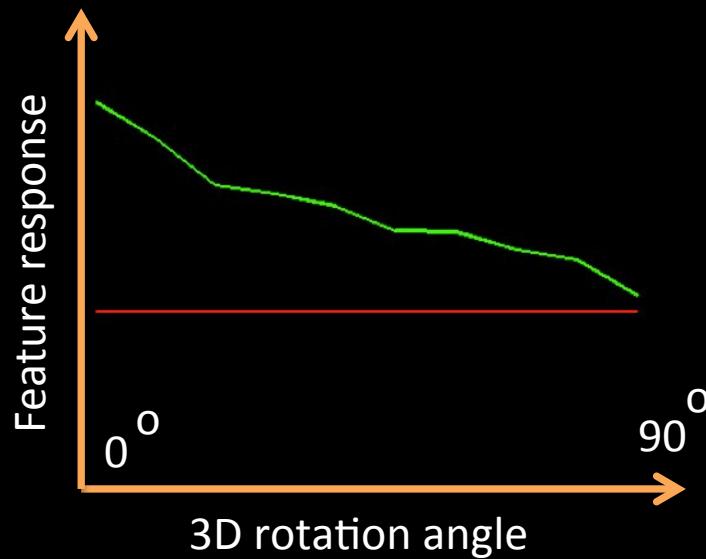
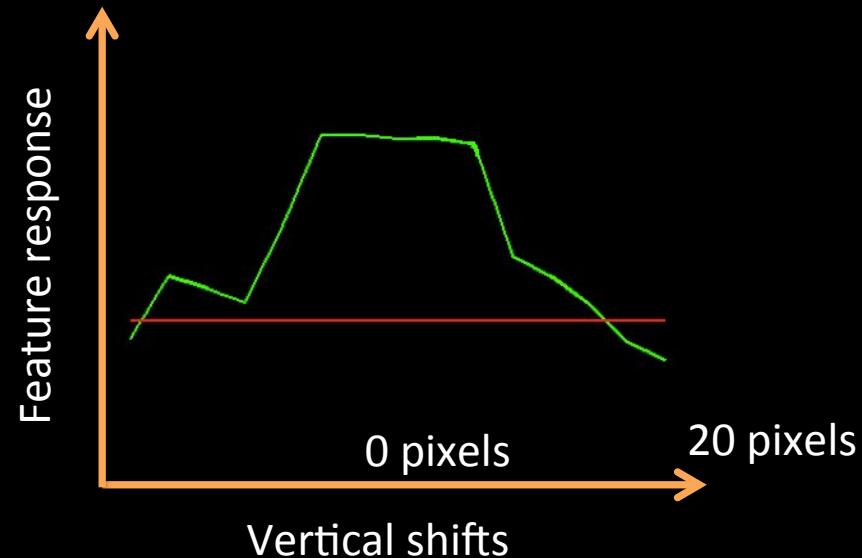
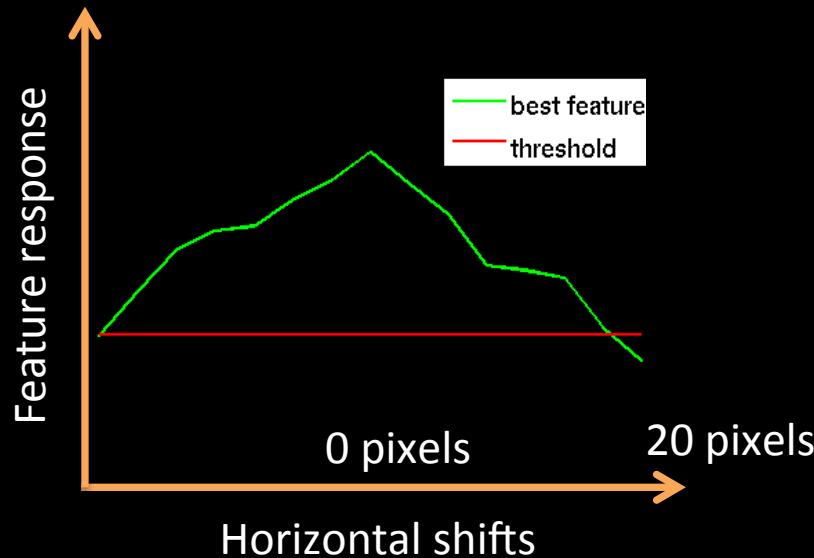
Top stimuli from the test set



Optimal stimulus
by numerical optimization



Invariance properties

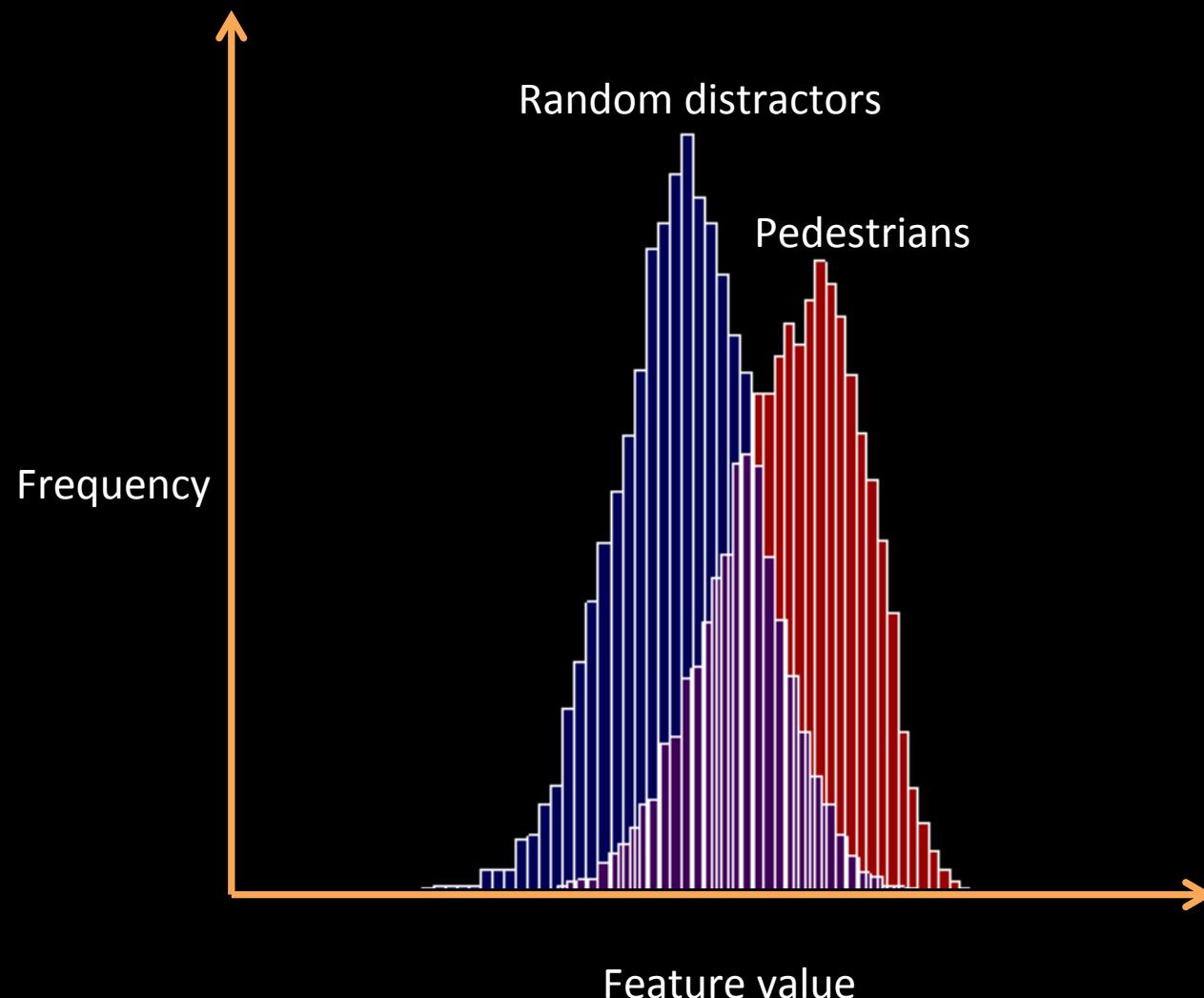




Top stimuli from the test set



Optimal stimulus
by numerical optimization

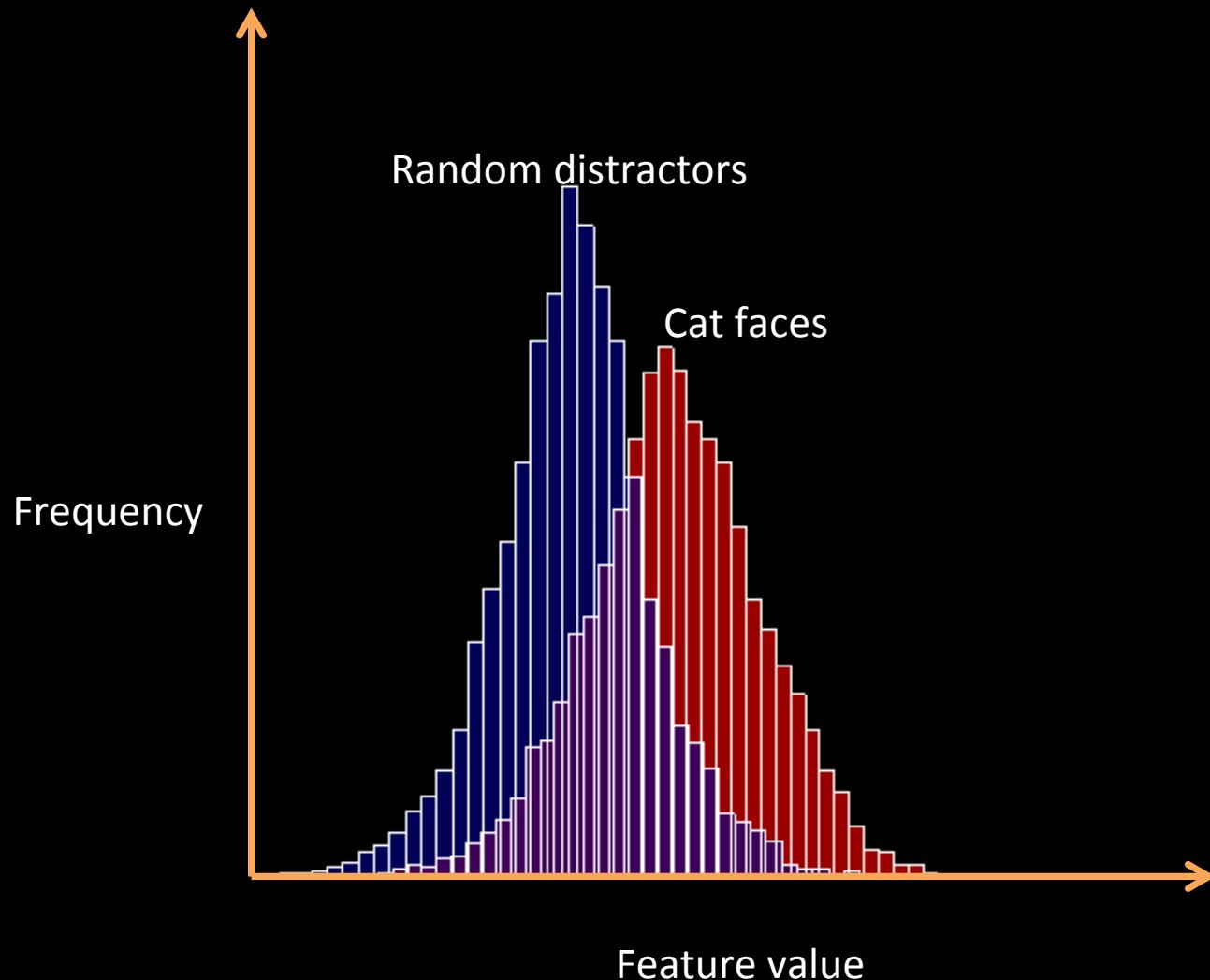




Top stimuli from the test set



Optimal stimulus
by numerical optimization



MARCJACOBS.COM

The New York Times

Tuesday, June 26, 2012 Last Update: 9:57 AM ET



TRY A TIMES DIGITAL SUBSCRIPTION: 4 WEEKS FOR 99¢.

[CLICK HERE](#)

Search

INFO DIRECT

Follow Us:



Subscribe to Home Delivery | Personal

WORLD

U.S.

POLITICS

NEW YORK

BUSINESS

DEALBOOK

TECHNOLOGY

SPORTS

SCIENCE

HEALTH

ARTS

STYLE

OPINION

Autos

Blogs

Books

Cartoons

Classifieds

Crosswords

Dining & Wine

Education

Event Guide

Fashion & Style

Home & Garden

Jobs

Magazines

Turkish Border Is Crucial to Syrian Fight as Rebels Evolve

By NEIL MacFARQUHAR

A network of activists has taken advantage of the tensions between Turkey and Syria to build a supply chain for those opposed to the government of President Bashar al-Assad.

- [Photographs](#) | [Video](#)
- [Assad Supporters Suspected in Beirut Unrest](#) **8:07 AM ET**
- [Post a Comment](#) | [Read \(15\)](#)

Turkey Warns Syrian Forces Not to Approach Border

By SEBNEM ARSU, ALAN COWELL and PAUL GEITNER **12 minutes ago**

Buoyed by support from NATO allies, the Turkish prime minister vowed on

LIVE DAILY AT 10 AM ET



Why News Corporation may divide itself in two. | The anxiety of being a digital parent. | Google's artificial brain scours the Internet in search of ... cats.

DEALBOOK

News Corporation Considers Dividing Itself

By MICHAEL J. DE LA MERCE

The embattled media company may separate its publishing arm from its larger entertainment division.

Euro Zone Leaders to Debate Joint Banking Union

By STEPHEN CASTLE **8:51 AM ET**

Details of a plan for tighter fiscal unity, including the creation

OPINION *

THE SUPREME COURT ACTS

Editorial: The basis of the harsh Arizona immigration law is rejected by the court.

Editorial: The justices allow the biggest donors to keep buying elections.

- Brooks: The Particular
- Bruni: Capt.
- Nocera: The
- Room for De Muslim Bro
- Gil Troy: Pr Politicking

MARKETS *

At 10:03 AM ET

S&P 500 Dow Nasdaq

1,317.65 12,515.03 2,851.97

+0.83 +12.37 +15.81

+0.30% +0.10% +0.56%

GET QUOTES

My Portfolio >

Stock, ETFs, Funds

Go

4 WEEK

BONOBOS
Men's Clothing

**SUMMER
SALE**

ImageNet classification

22,000 categories

14,000,000 images

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

22,000 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 obesus*

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



Best stimuli

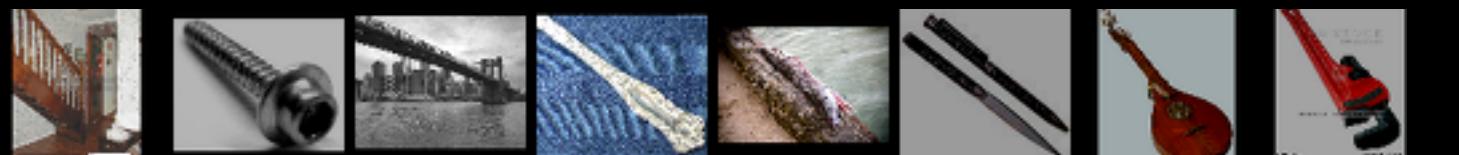
Feature 1



Feature 2



Feature 3



Feature 4



Feature 5



Best stimuli

Feature 6



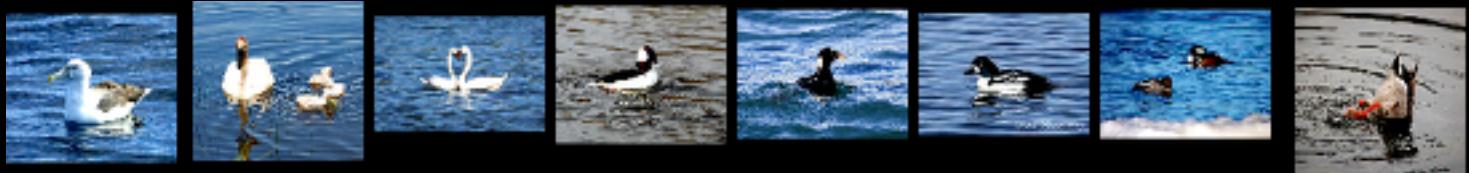
Feature 7



Feature 8



Feature 9



Best stimuli

Feature 10



Feature 11



Feature 12



Feature 13



0.005%

Random guess

9.5%

State-of-the-art
(Weston, Bengio '11)

?

Feature learning
From raw pixels

0.005%

Random guess

9.5%

State-of-the-art
(Weston, Bengio '11)

15.8%

Feature learning
From raw pixels

ImageNet 2009 (10k categories): Best published result: 17%
(Sanchez & Perronnin '11),
Our method: 20%

Using only 1000 categories, our method > 50%

Other results

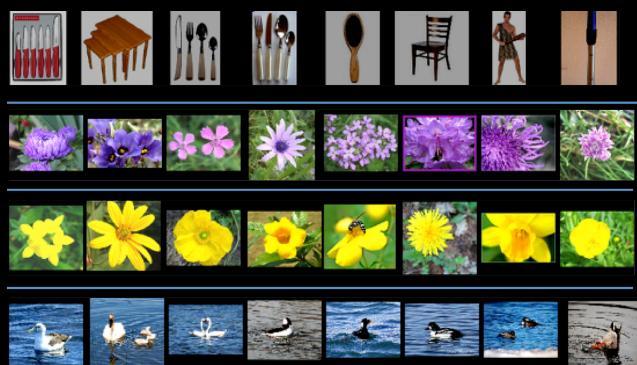
- We also have great features for
 - Speech recognition
 - Word-vector embedding for NLPs

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



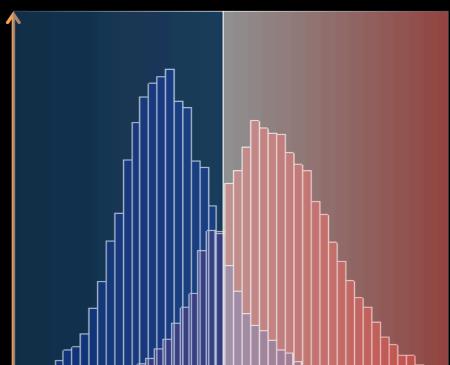
Cancer classification



Feature visualization

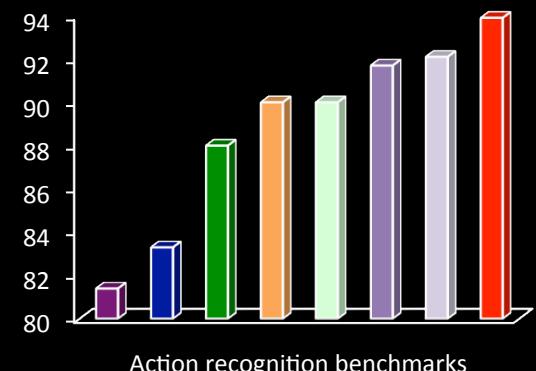


Action recognition



Face neuron

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



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.