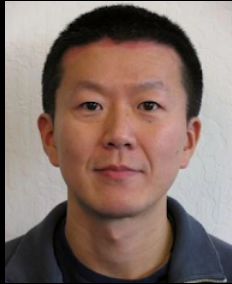


Tera-scale deep learning

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

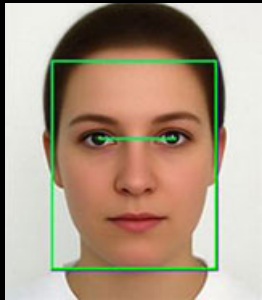


Ke Yang

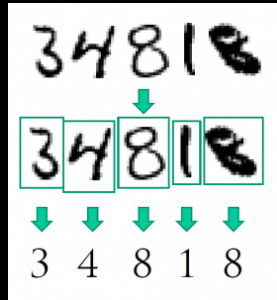
Additional Thanks:

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

Machine Learning successes



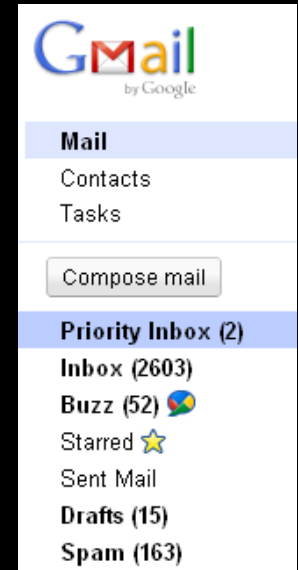
Face recognition



OCR



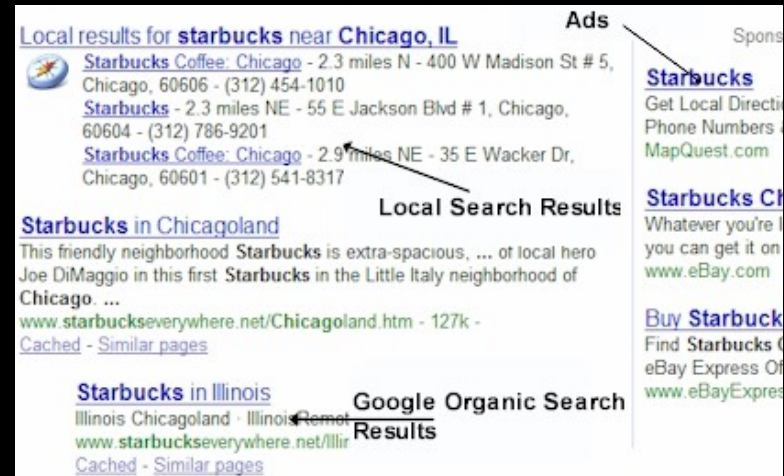
Autonomous car



Email classification

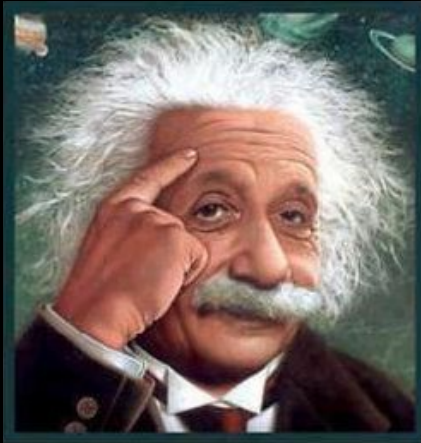


Recommendation systems



Web page ranking

Feature Extraction

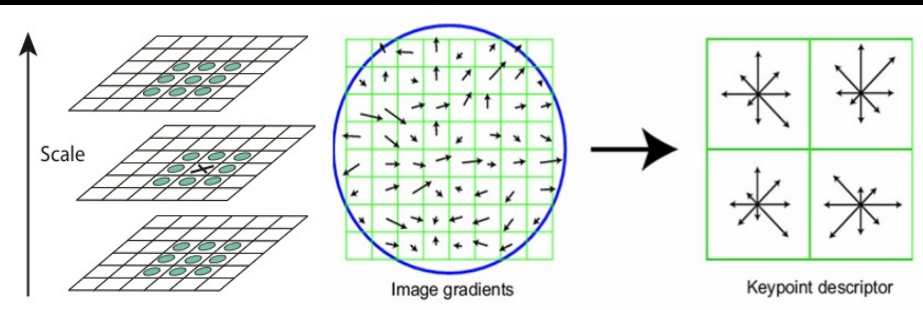


Classifier

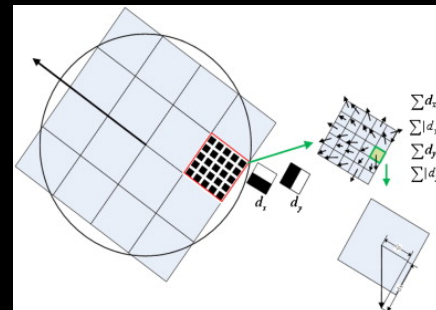
Feature extraction
(Mostly hand-crafted features)

Hand-Crafted Features

Computer vision:

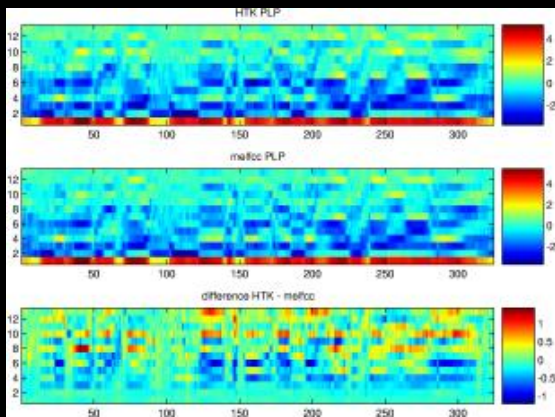


SIFT/HOG

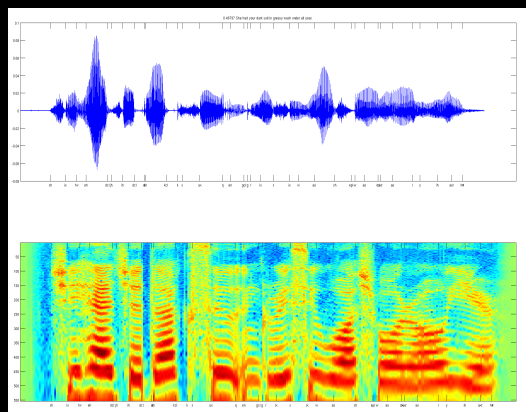


SURF

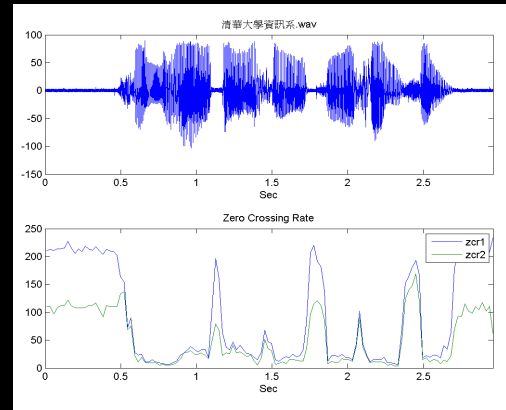
Speech Recognition:



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

A screenshot of the Facebook search interface. The search bar at the top contains the text "MGM". Below the search bar, a list of search results is displayed. The first result is "MGMT", categorized as "Musician or Band" with 518,981 fans. The second result is "MGM Grand Las Vegas", categorized as "Brand or Company" with 49,470 fans. The third result is "MGM", categorized as "Movie" with 94 fans. The fourth result is "MGM Mondo del Vino", categorized as "Brand or Company" with 341 fans. At the bottom of the search results, there is a link that says "See More Results for mgm" and a note "Displaying Top 4 Results". On the left side of the page, the user's profile information for "Jane Smith" is visible, along with navigation links for "News Feed", "Messages (5)", "Events", "Photos", "Friends (3)", "Applications", "Games", and "Trips".

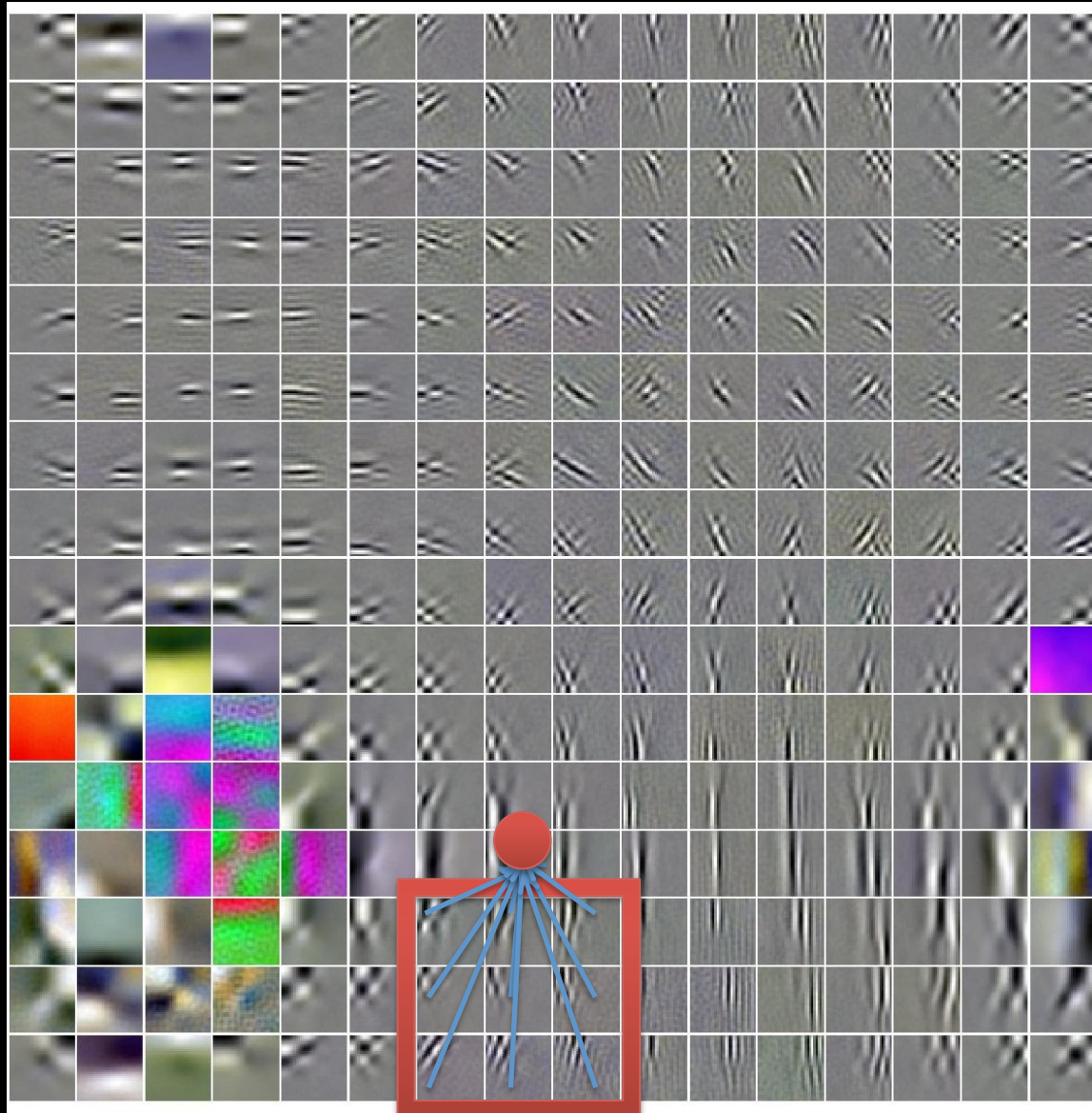
A screenshot of the Twitter homepage. At the top, there is a blue banner that says "Welcome to #NewTwitter! Read up on what's new. You can also leave the". Below the banner is the Twitter logo and a search bar. The main content area is titled "What's happening?" and features a large text input field. Below the input field, there are navigation tabs for "Timeline", "@Mentions", "Retweets", "Searches", and "Lists". At the bottom of the page, there is a purple bar that says "5 new tweets".

A screenshot of the YouTube homepage. At the top, there is the YouTube logo and the tagline "Broadcast Yourself". Below the logo, there are navigation tabs for "Videos", "Categories", "Channels", and "Community". A search bar is located in the top right corner. The main content area is divided into several sections. The "Director Videos" section features four video thumbnails with titles like "SWEET TALK", "Hoeswoe's Appeal", "Ninja Star in the Eye", and "Low Water - Strana...". The "Featured Videos" section is titled "Featured Videos selected by: DukeRighteous" and includes a video titled "Piano Lesson With A Difference" by sniper220, which has 2,725 views and a 5-star rating. Another video titled "Everyday Bravery" is also featured. On the right side of the page, there is a section titled "My: Videos - Favorites - Playlists - Inbox - Subscriptions" and a "WATCH HER DO IT BELOW:" section featuring a video of a couple.

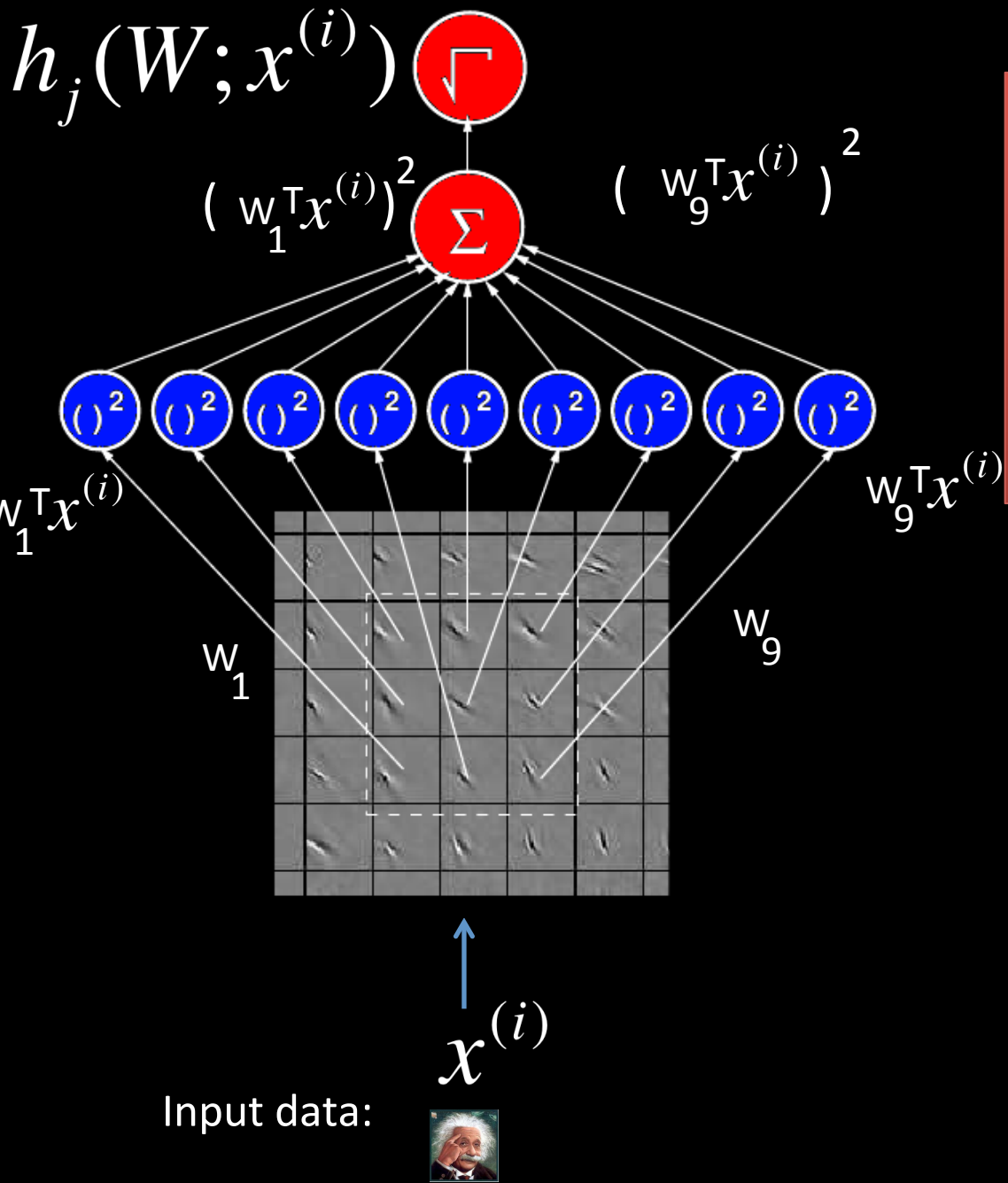
Outline

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

Topographic Independent Component Analysis (TICA)



1. Feature computation



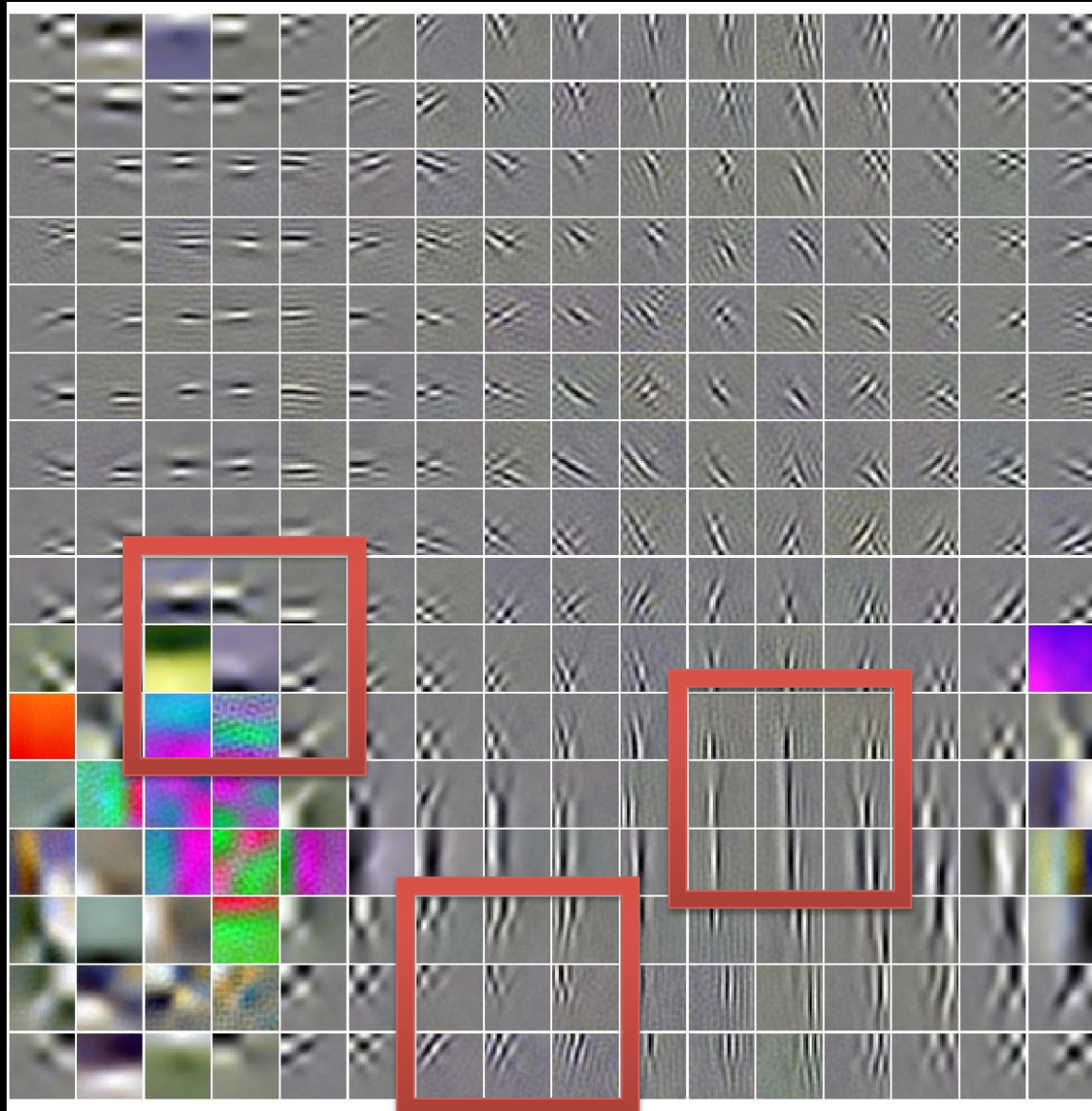
2. Learning

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

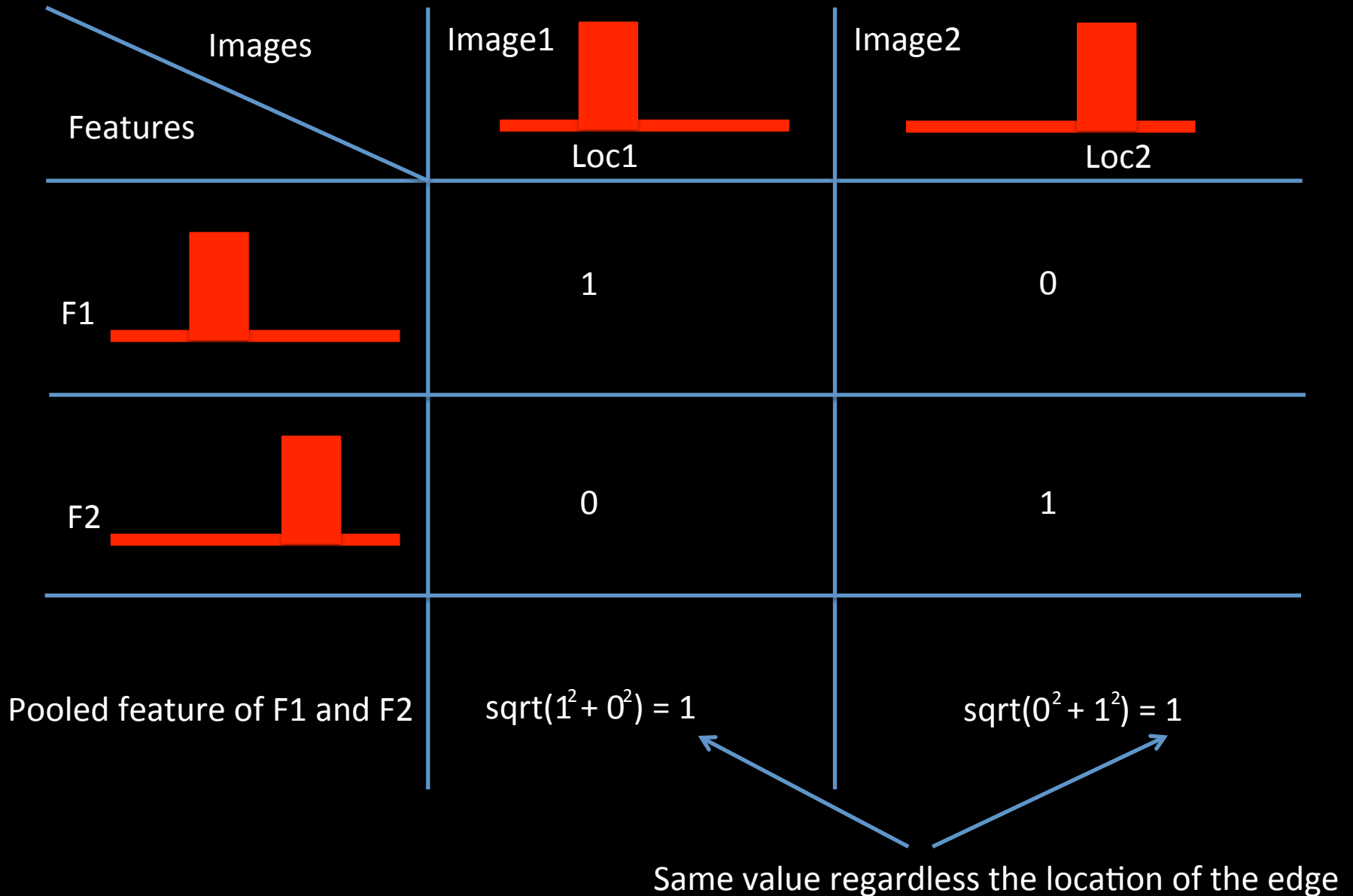
s.t. $WW^T = I$

$$W = \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_{10000} \end{bmatrix}$$

Topographic Independent Component Analysis (TICA)



Invariance explained



TICA:

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

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

Reconstruction ICA:

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

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 $\lambda \|W W^T - \mathbf{I}_k\|_{\mathcal{F}}^2$ 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:

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

Reconstruction ICA:

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

Data whitening

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

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

$$W W^T = I$$

TICA:

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

s.t. $WW^T = I$

Reconstruction ICA:

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

Data whitening

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

Why RICA?

Algorithms	Speed	Ease of training	Invariant Features
Sparse Coding	✗	✓	✗
RBM/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



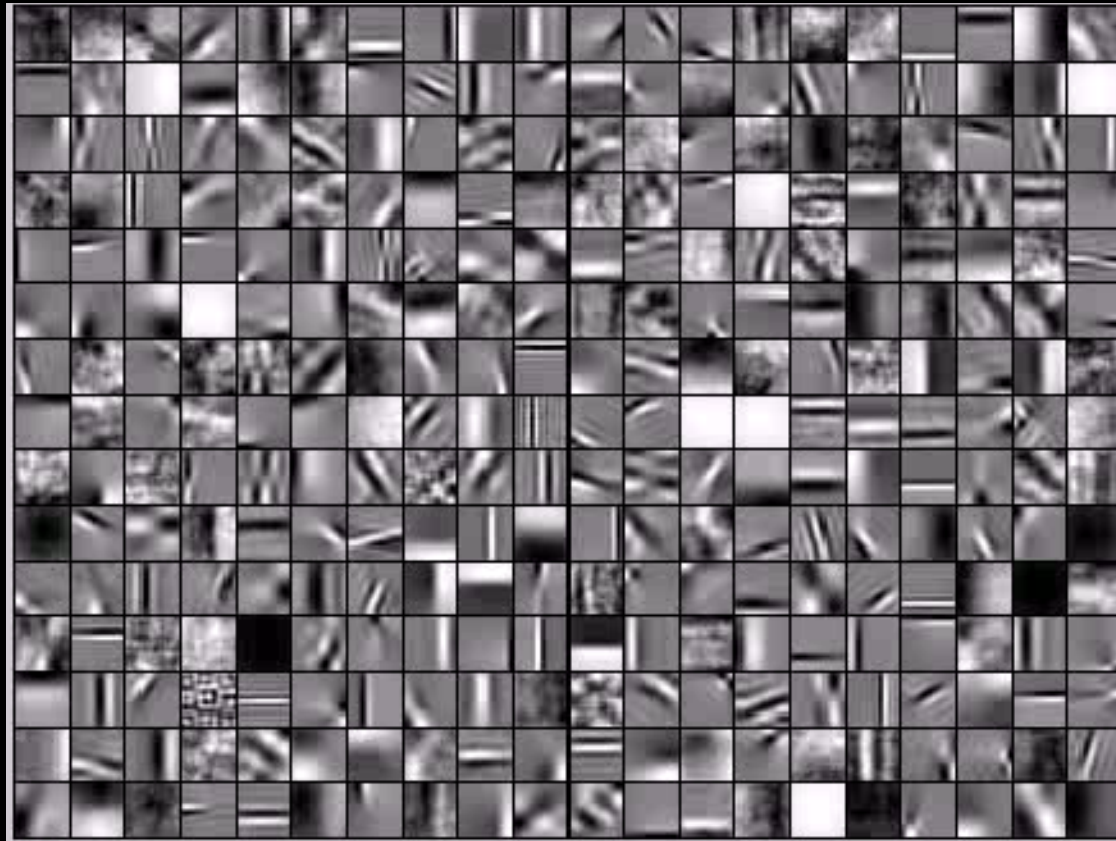
Run



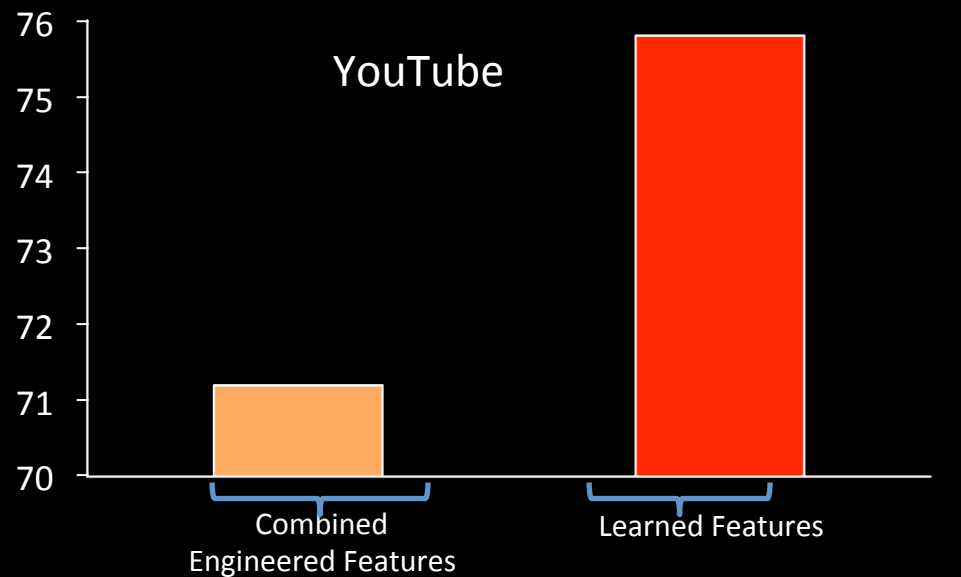
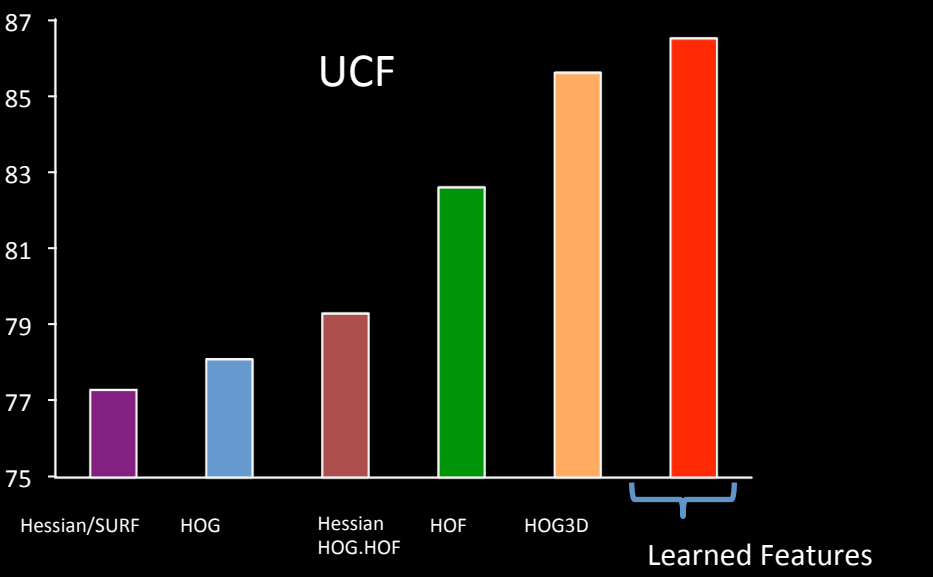
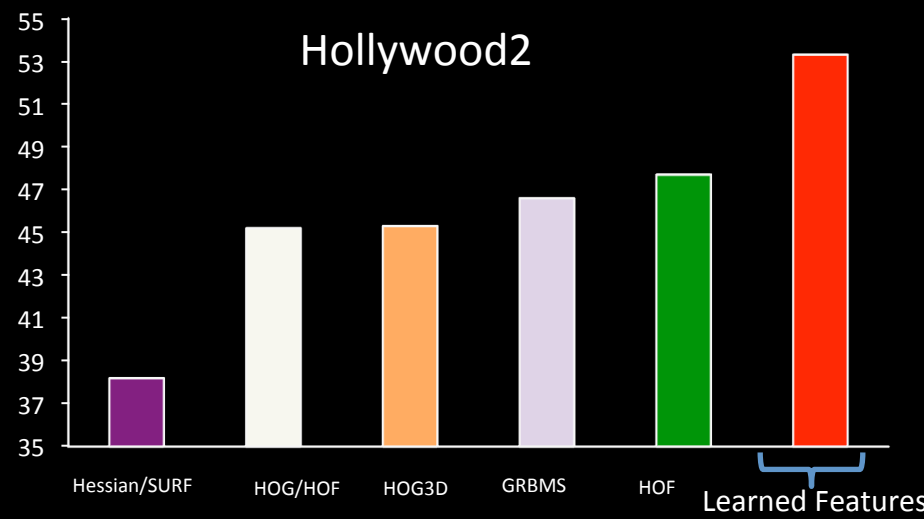
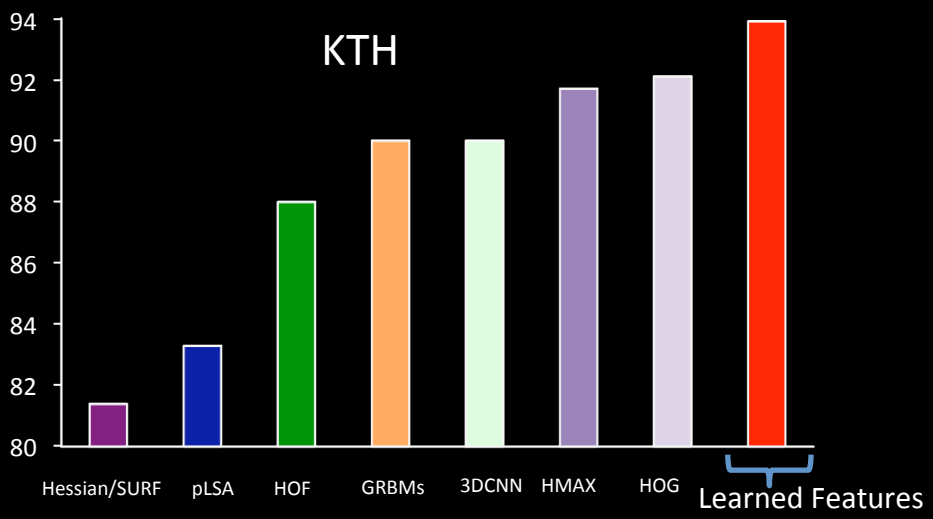
Stand up



Shake hands

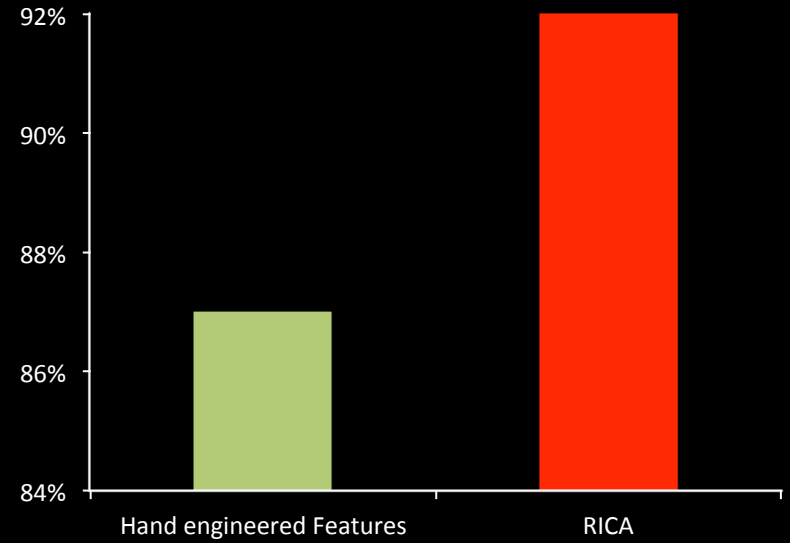
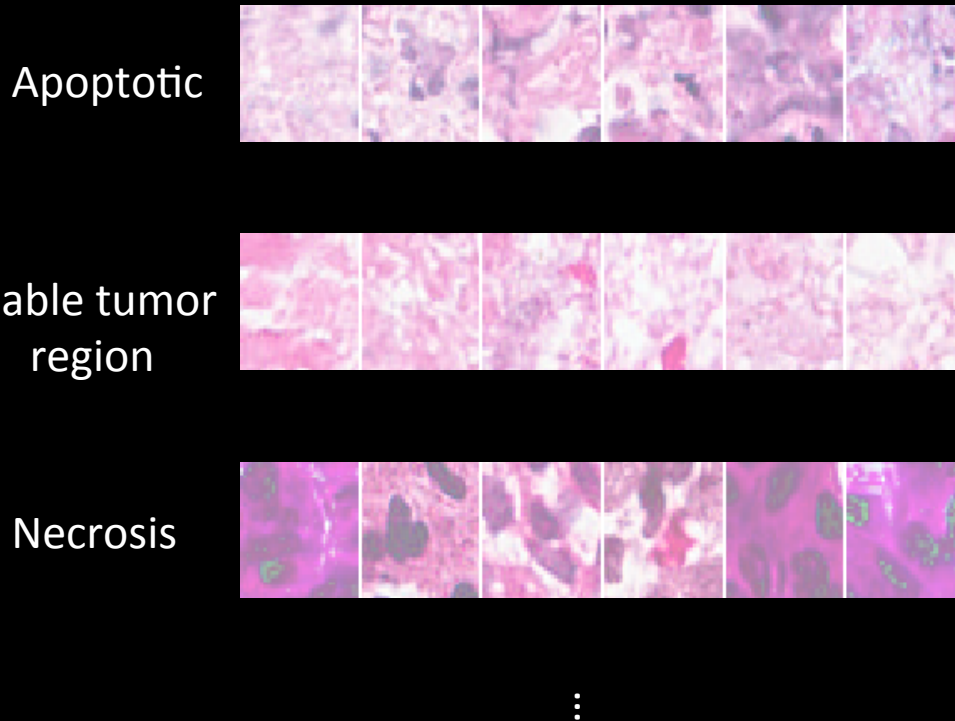


Le, et al., *Learning hierarchical spatio-temporal features for action recognition with independent subspace analysis*. CVPR 2011



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

Cancer classification

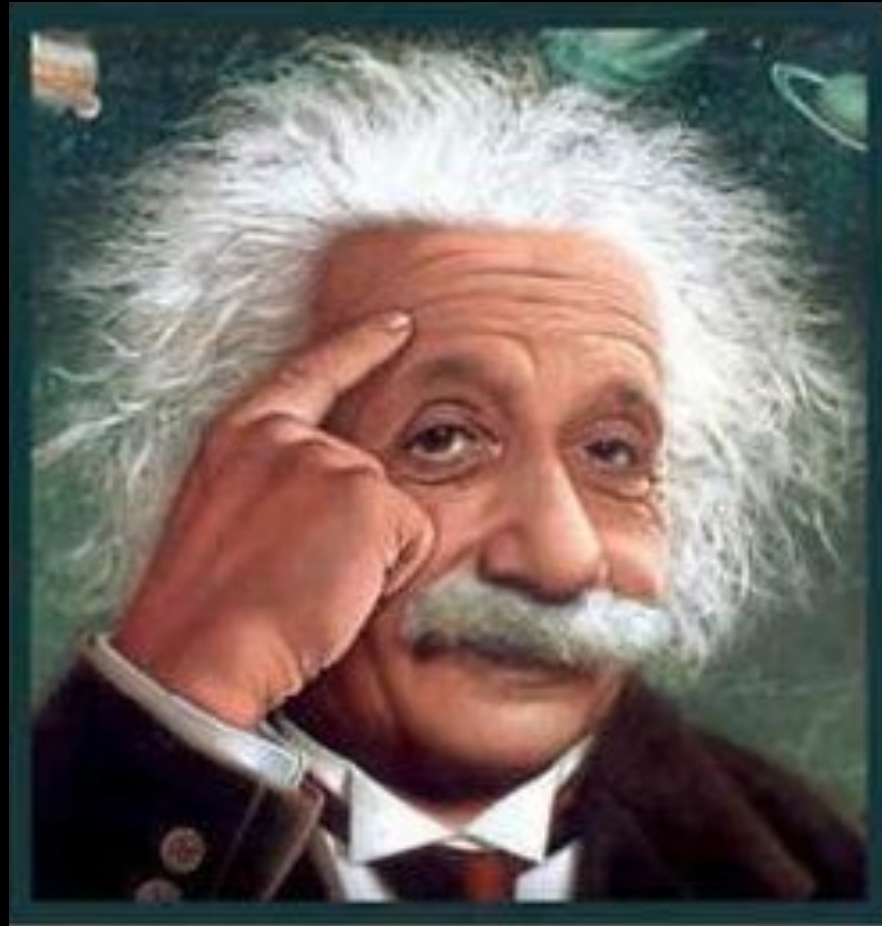


Scaling up deep RICA networks

Scaling up Deep Learning

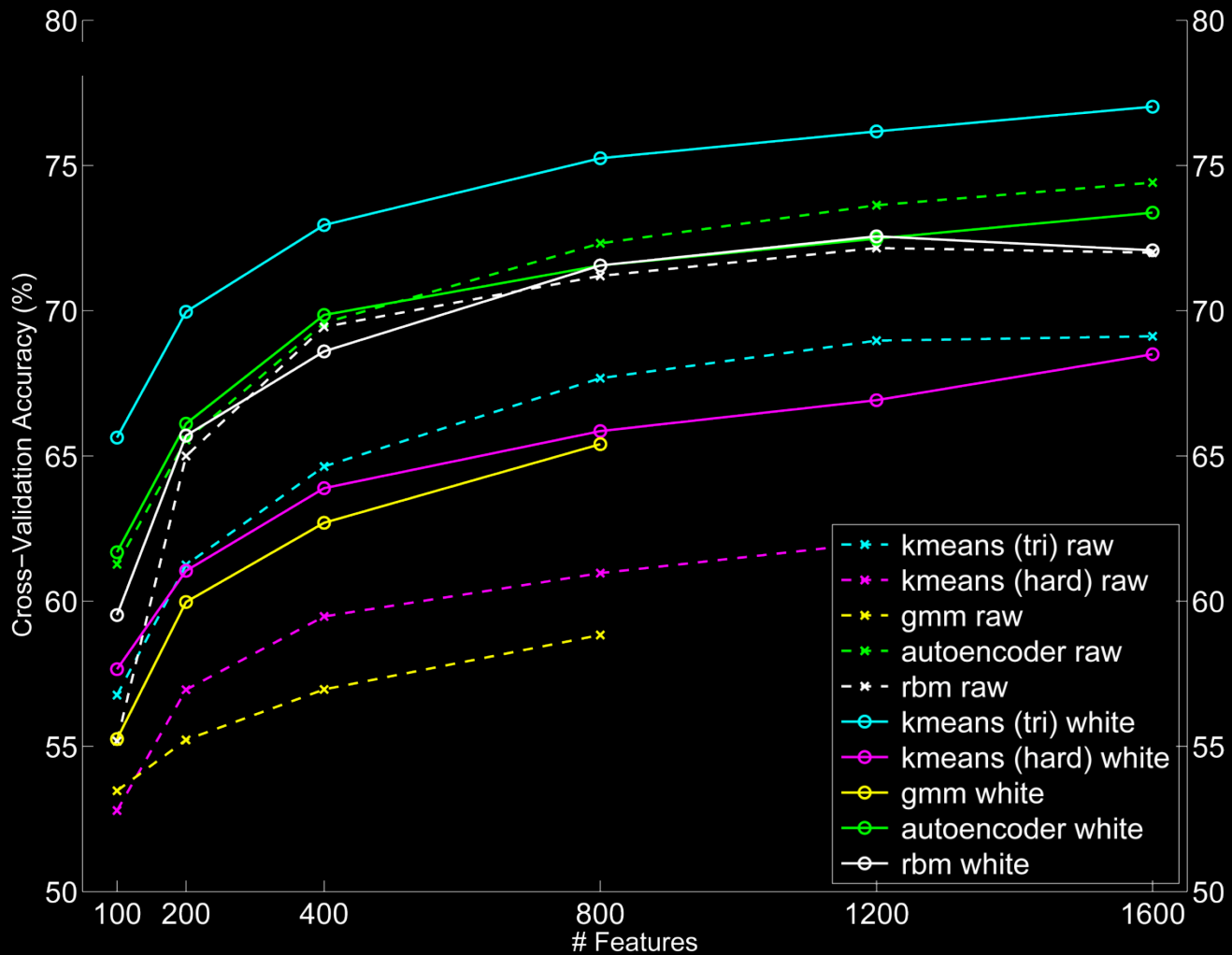


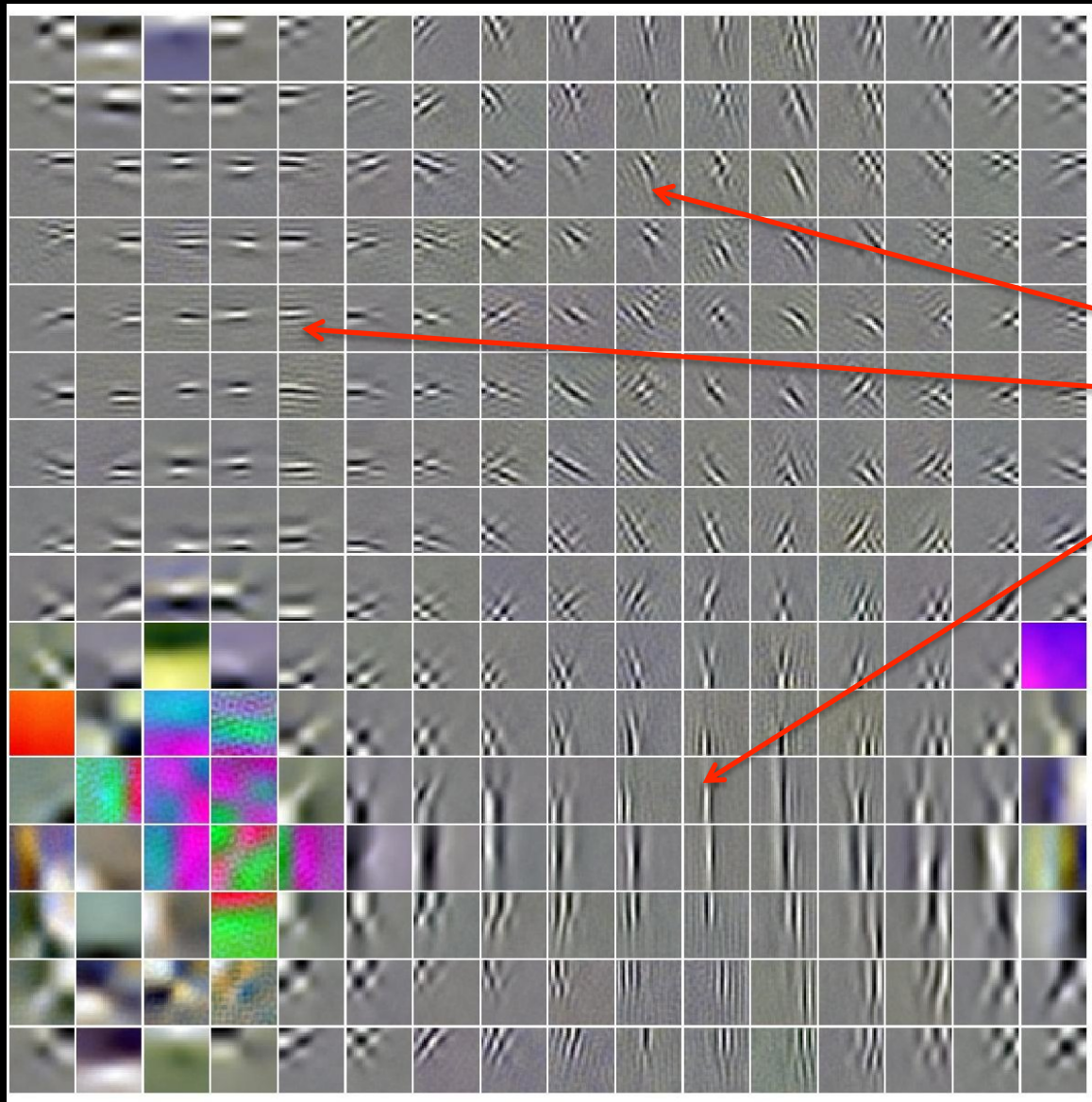
Deep learning data



Real data

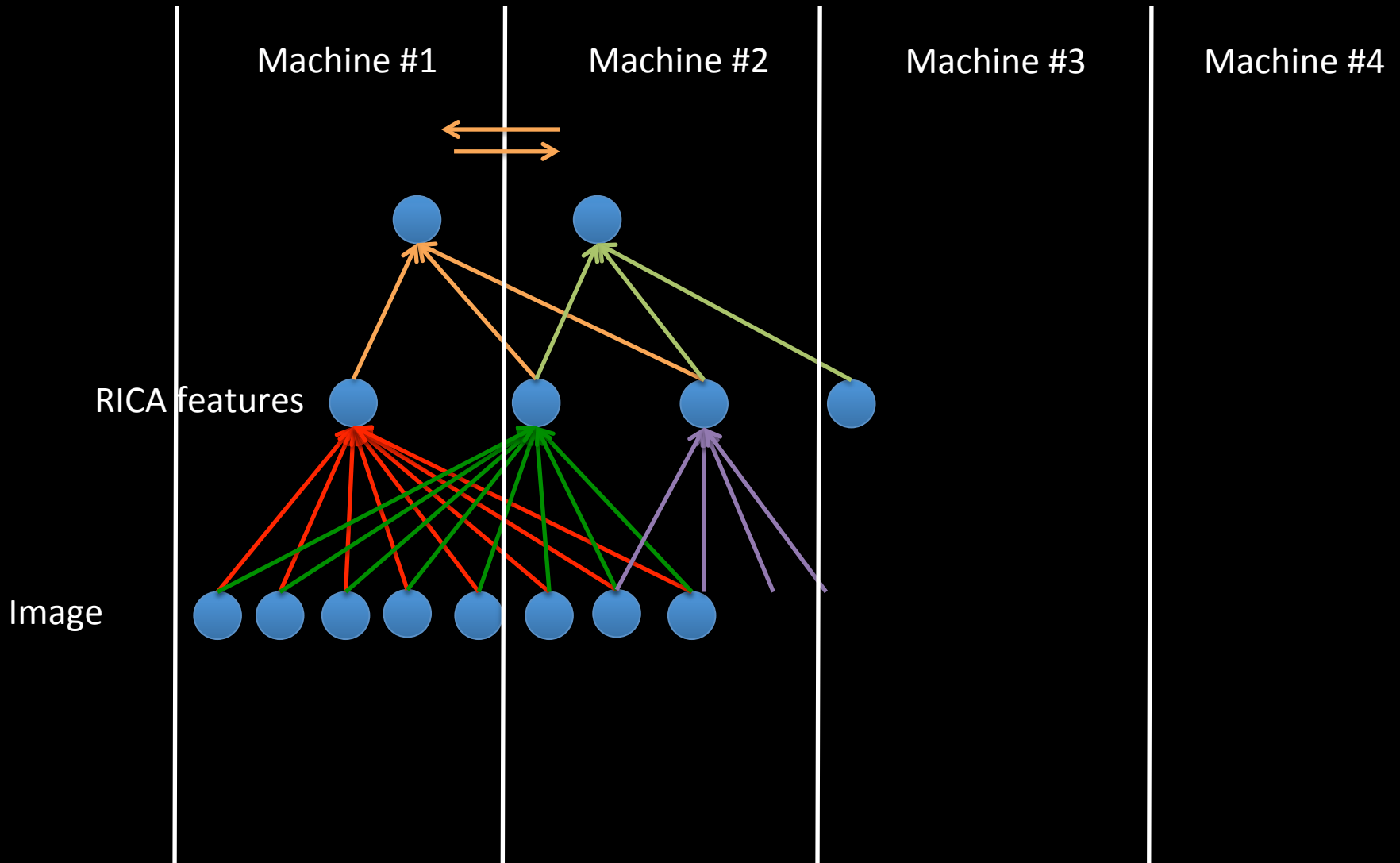
It's better to have more features!





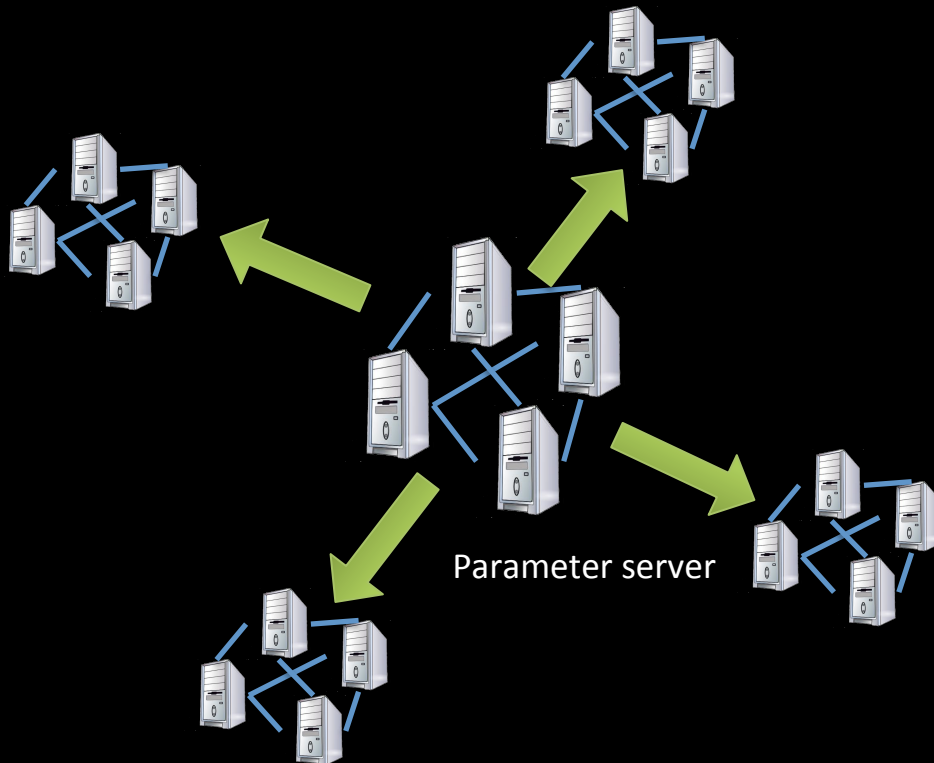
Most are
local features

Local receptive field networks

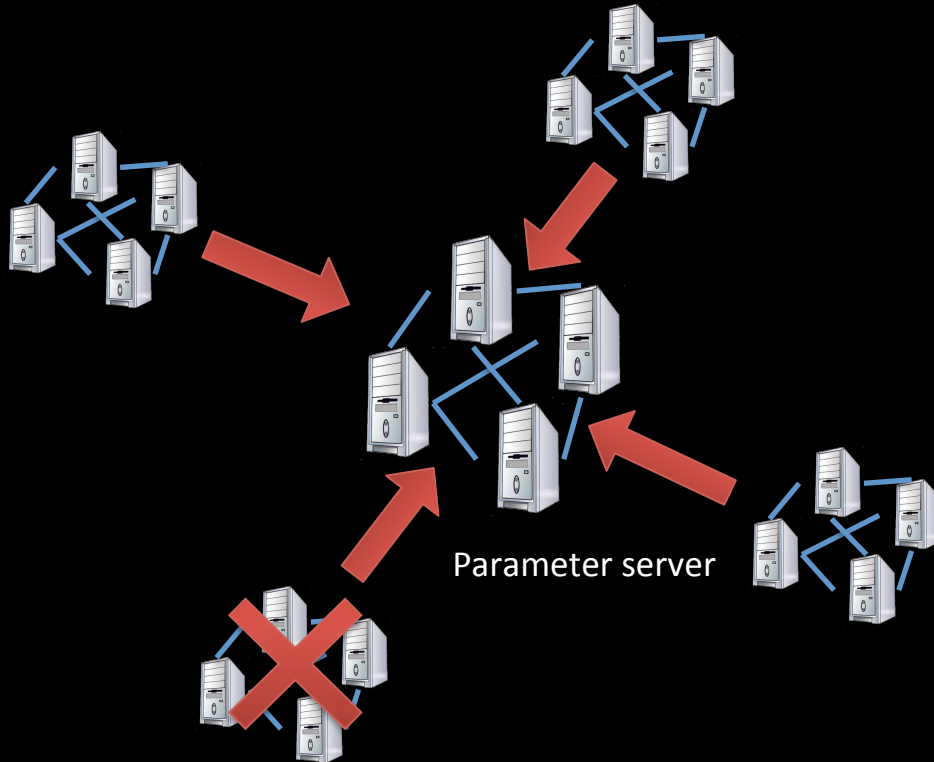


Challenges with 1000s of machines

Asynchronous Parallel SGDs



Asynchronous Parallel SGDs



Summary of Scaling up

- Local connectivity
- Asynchronous SGDs

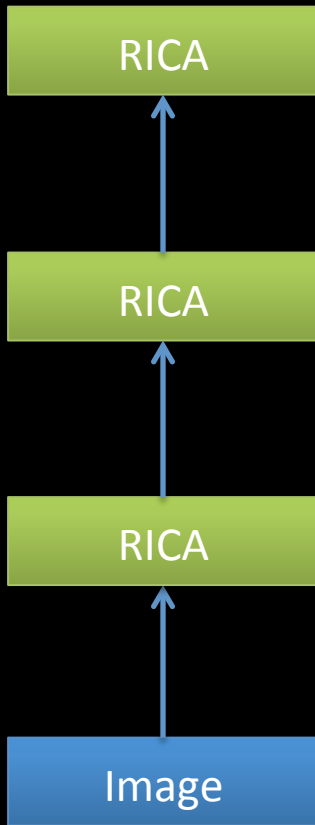
... And more

- 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

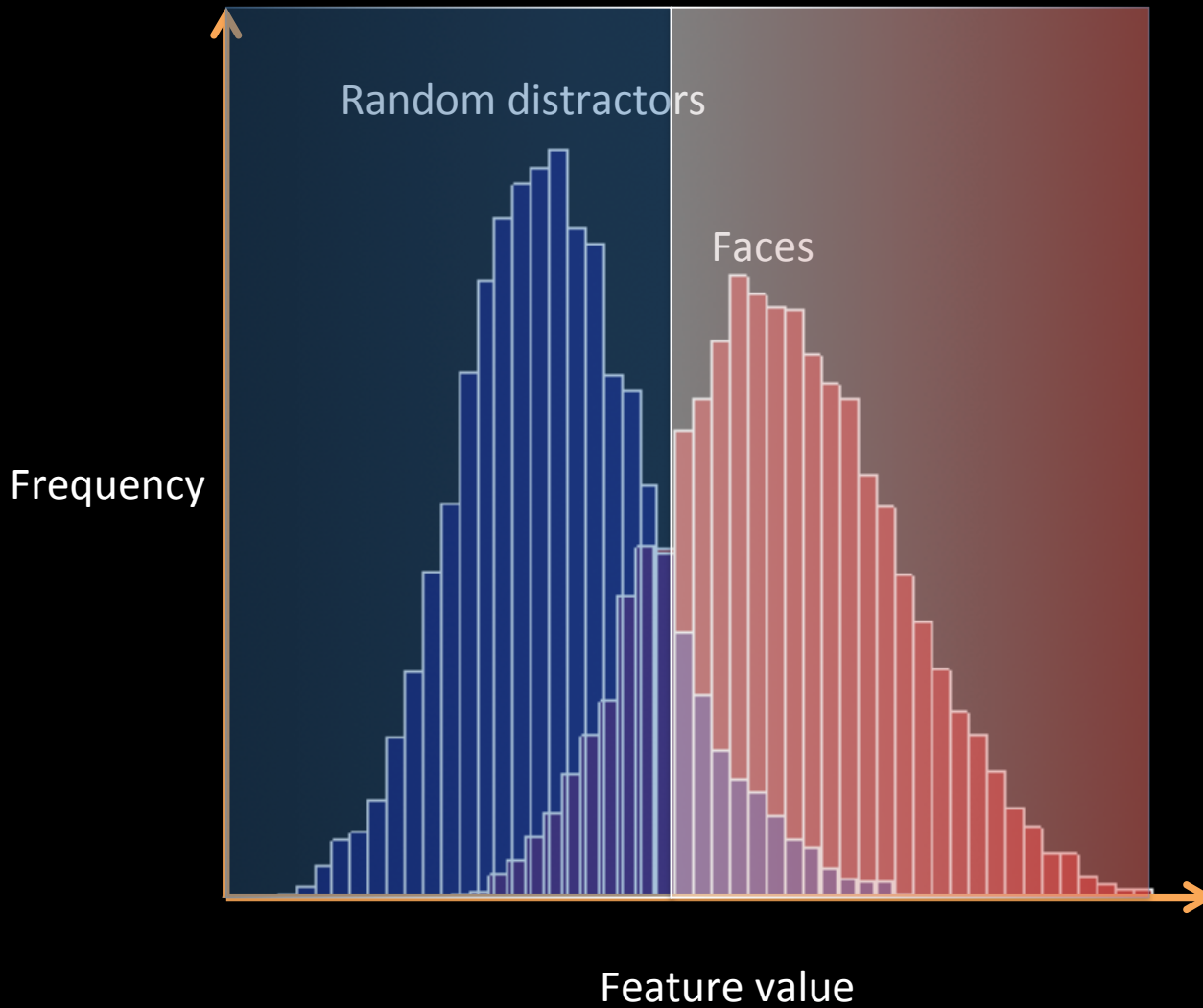
The face neuron



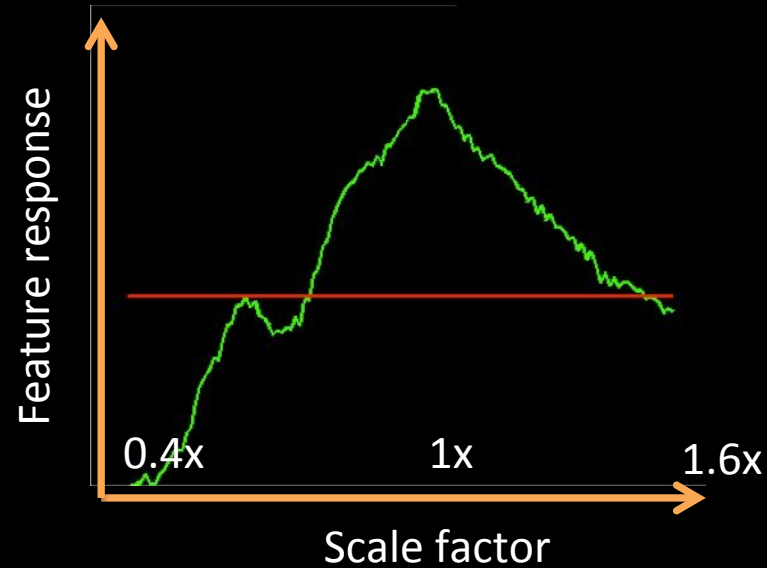
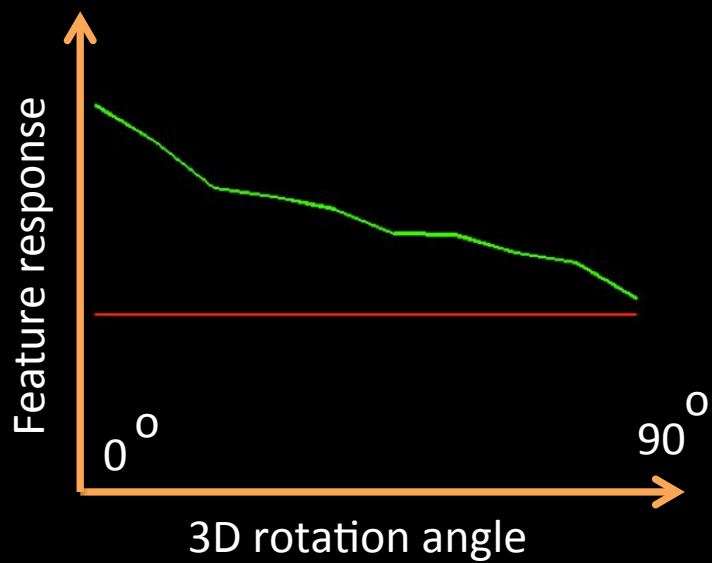
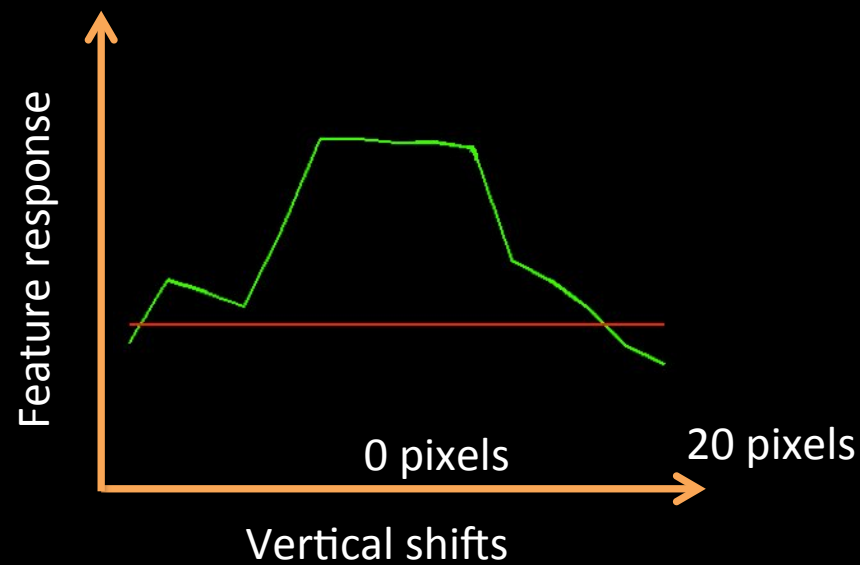
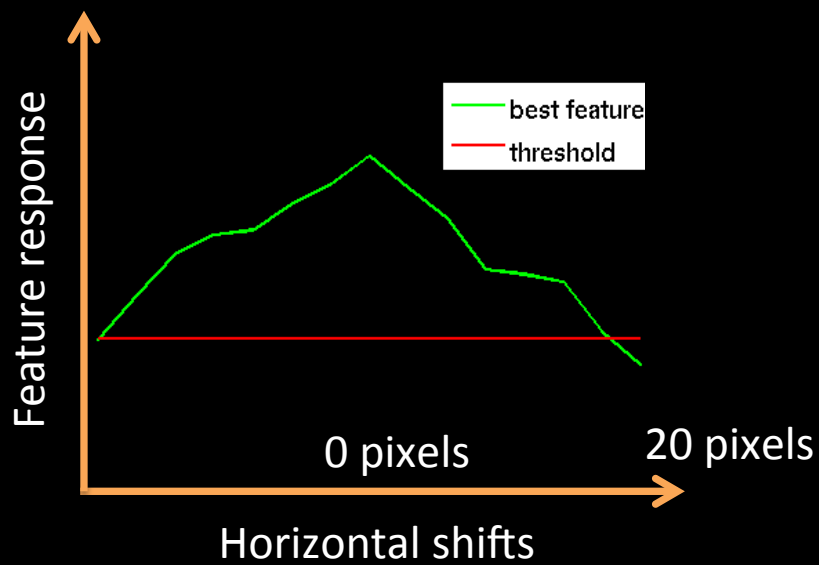
Top stimuli from the test set



Optimal stimulus
by numerical optimization



Invariance properties

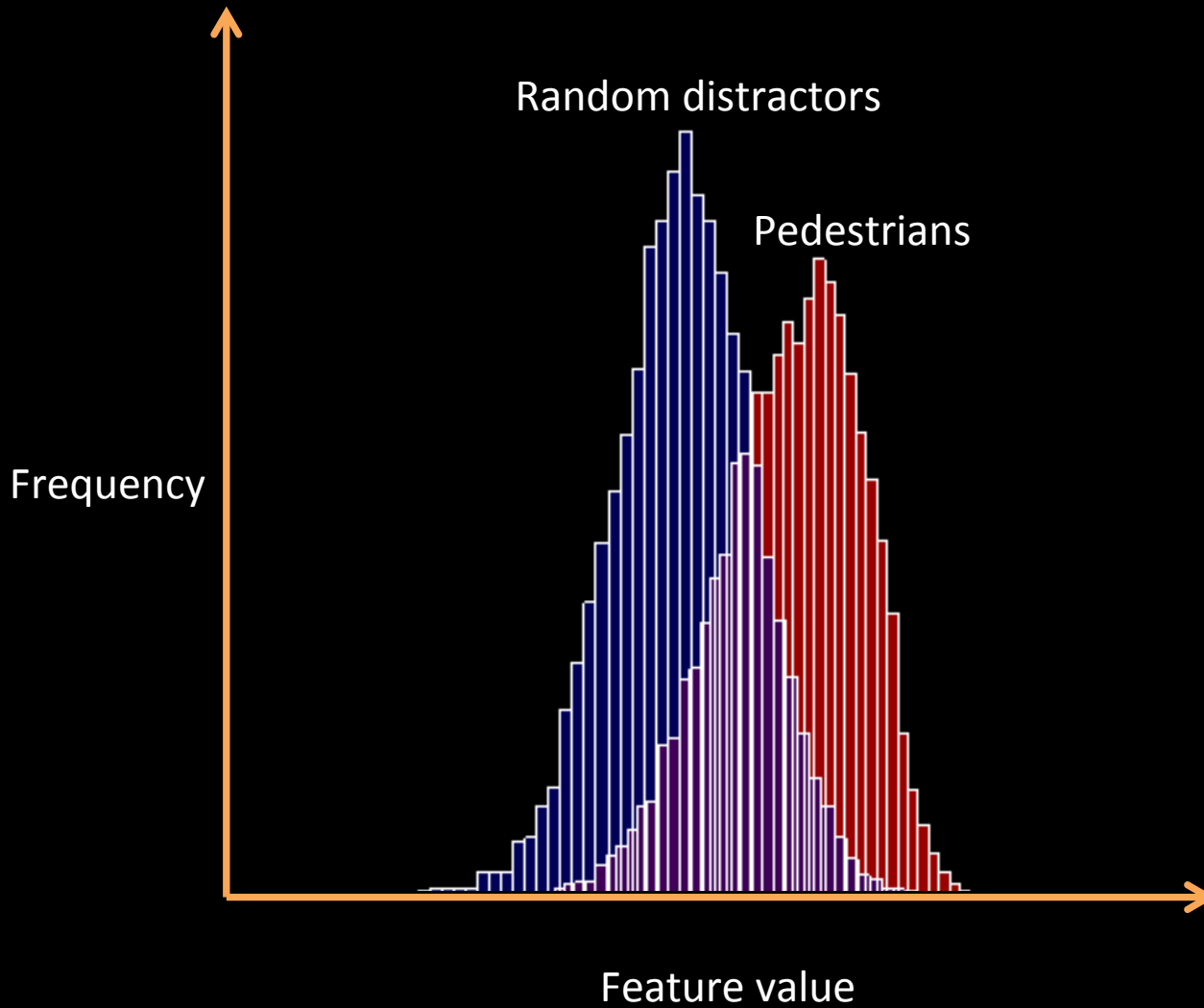




Top stimuli from the test set



Optimal stimulus
by numerical optimization

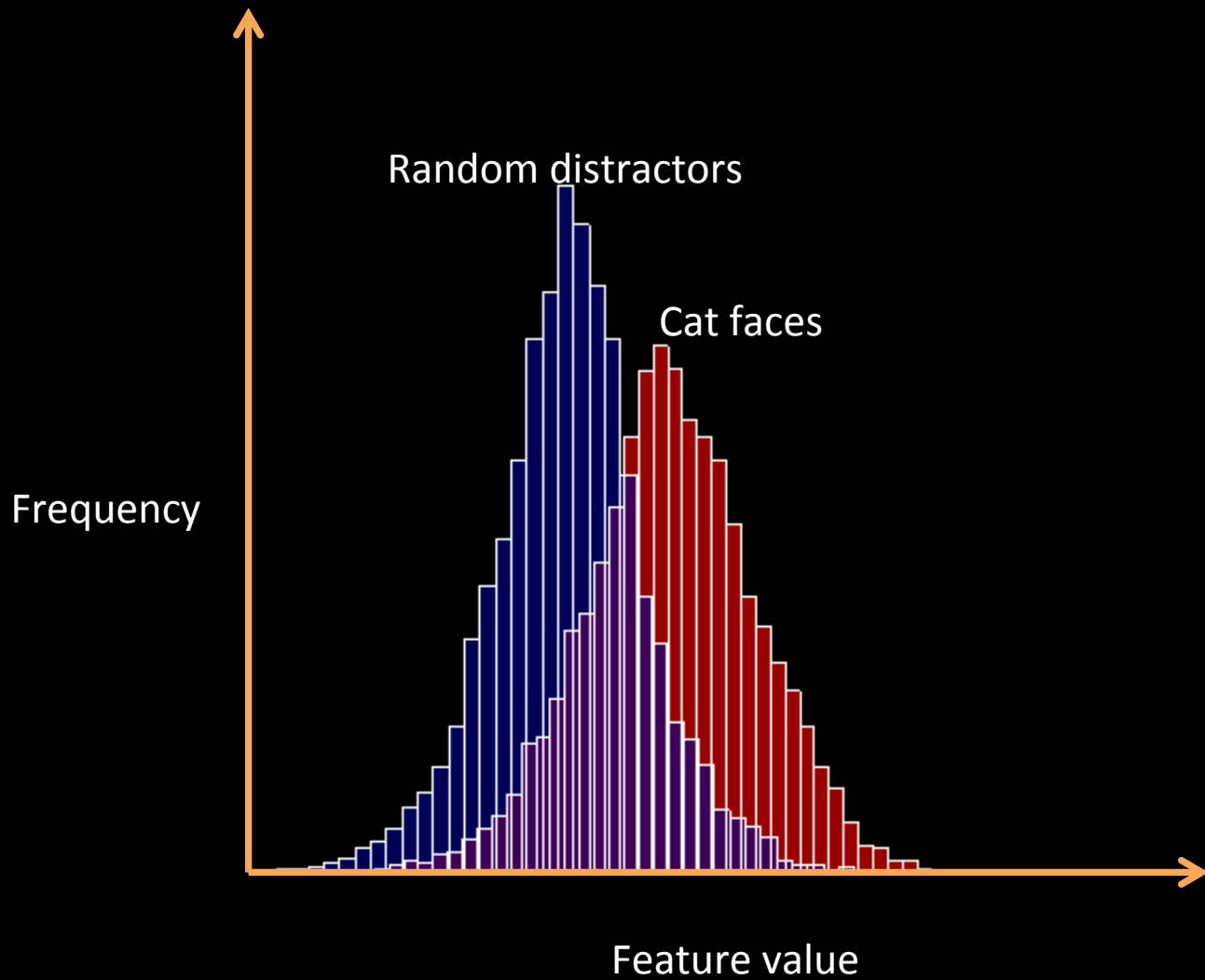




Top stimuli from the test set



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

INDIRECT

Follow Us



Subscribe to Home Delivery | Person

WORLD
U.S.
POLITICS
NEW YORK
BUSINESS
DEALBOOK
TECHNOLOGY
SPORTS
SCIENCE
HEALTH
ARTS
STYLE
OPINION

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 SEBNEH ARSU, ALAN COWELL and PAUL GEITNER
12 minutes ago

Baoyed by support from NATO allies, the Turkish foreign minister accused of a

LIVE DAILY AT 10 AM ET



Jim Wilson/The New York Times

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 MERCEO

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
- Ibrani: Capt
- Noceca: The
- Room for D
- Muslim Bro
- Gil Troy: P
- Politicking

MARKETS » At 10:03 AM ET

S&P 500	Dow	Nasdaq
1,317.85	12,515.03	2,851.97
+3.83	+12.37	+15.81
+0.30%	+0.10%	+0.56%

GET QUOTES My Portfolios »

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 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

rougtail 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

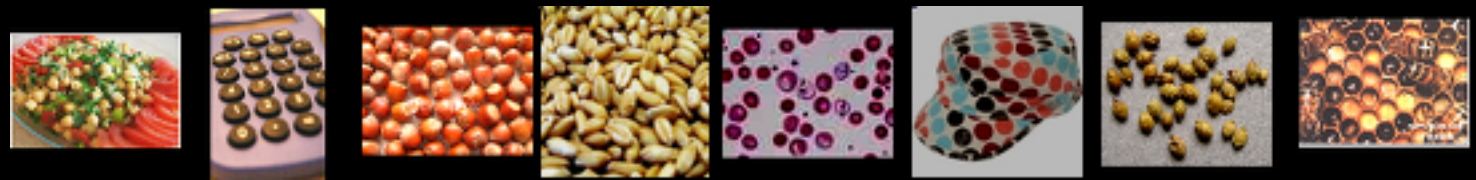


Best stimuli

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

ImageNet

0.005%

Random guess

9.5%

Best published result

15.8%

Our method



Cancer classification



Sit up



Drive Car



Get Out of Car



Eat



Answer phone



Meet



Run

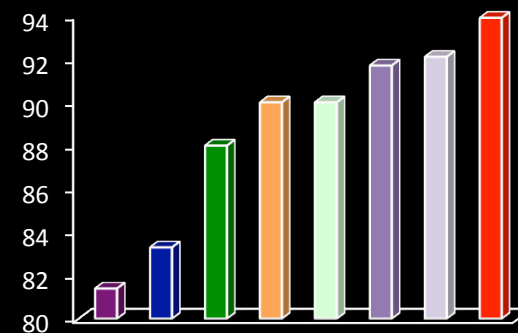


Stand up

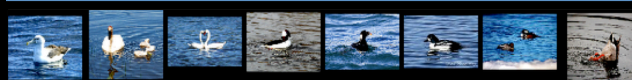


Shake hands

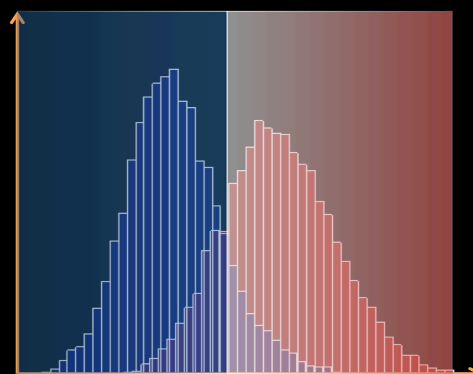
Action recognition



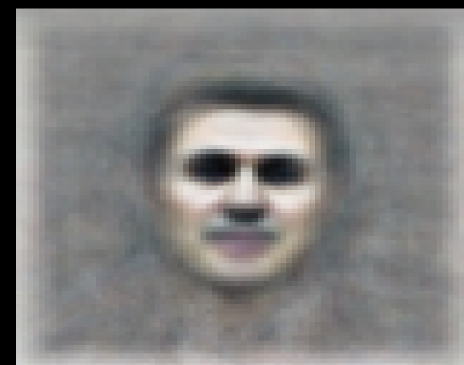
Action recognition benchmarks



Feature visualization



Face neuron



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>