Research in deep learning

John Hopcroft
Cornell University
Ithaca, NY

How knowledge evolves

- It is natural in engineering to make something work before developing the theory.
  - Airplanes flew before aerodynamics
  - Internal combustion before thermodynamics
- Deep learning is having a big impact on applications.
- Little research explaining why.
- Now may be the time to develop the theory supporting deep learning that can be taught.

Fooling deep learning

Cat

Automobile

A minor change to any image can change its classification to any arbitrary classification.

cat

tree

horse

dog

Can we partition space into 100 categories so that with high probability any point in one category is very close to a point in each of the 99 other categories?

In 2-dimensions not likely.

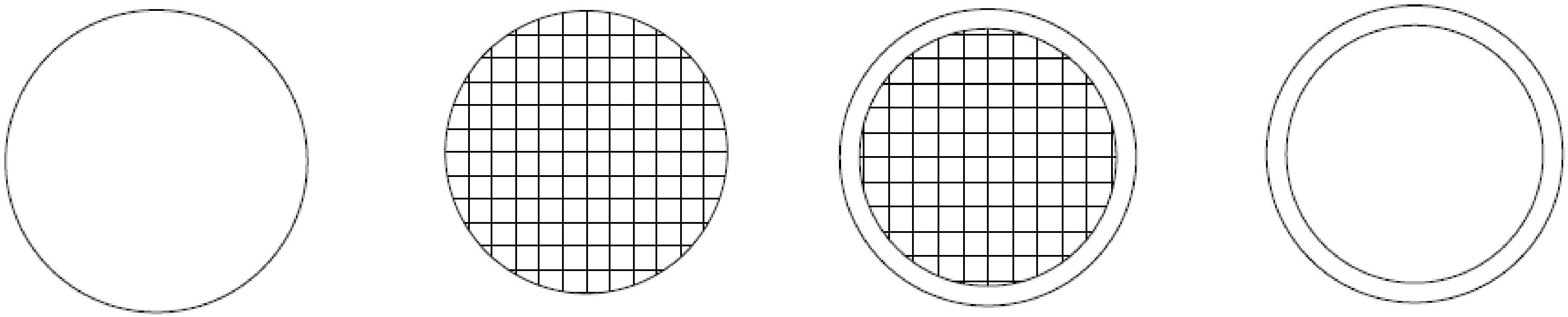
However, high dimensions is fundamentally different.

High dimensional space

- One can partition the unit sphere in high dimensions into many categories so that with high probability a point in any category is very close to a point in each of the other categories.

- Fooling may be a fundamental issue in high dimensional space and thus for deep networks.

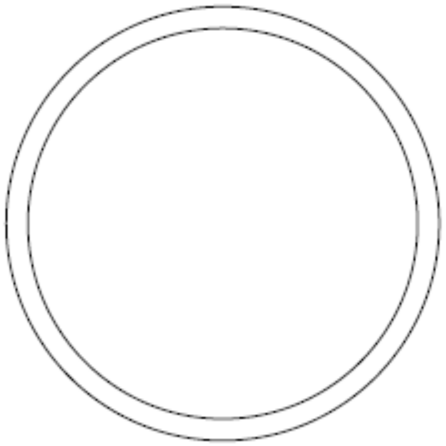
High dimensional space is fundamentally different



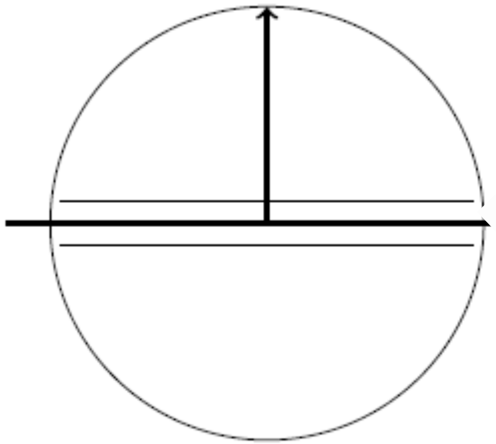
$(1-\epsilon)^d \approx 0$

Sphere in high dimension

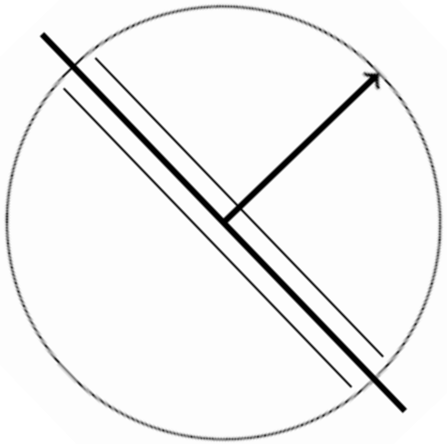
High dimensional space is fundamentally different



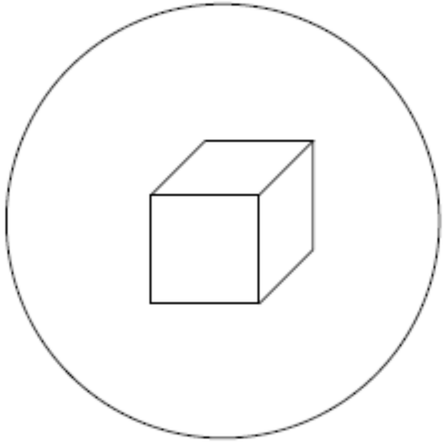
All volume
in annulus



All volume
near equator

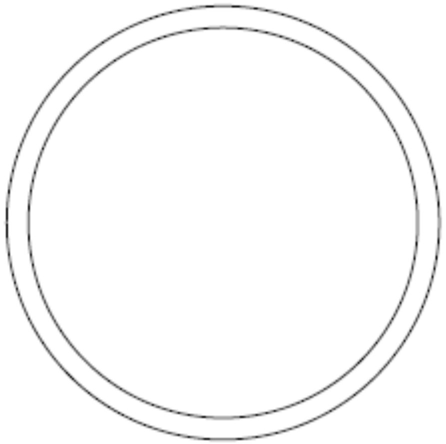


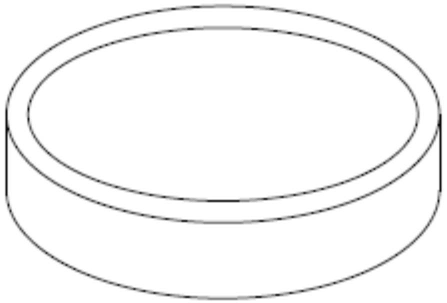
All volume near
new equator



All volume in
small box

Intersection of annulus and equator





Can now explain how to partition a sphere into 100 categories with a point in any category very close to a point in each of the other categories.

Random point in high dimensions is close to equator

$$\left(\frac{1}{\sqrt{d}}, \frac{1}{\sqrt{d}}, \ldots, \frac{1}{\sqrt{d}}\right)$$

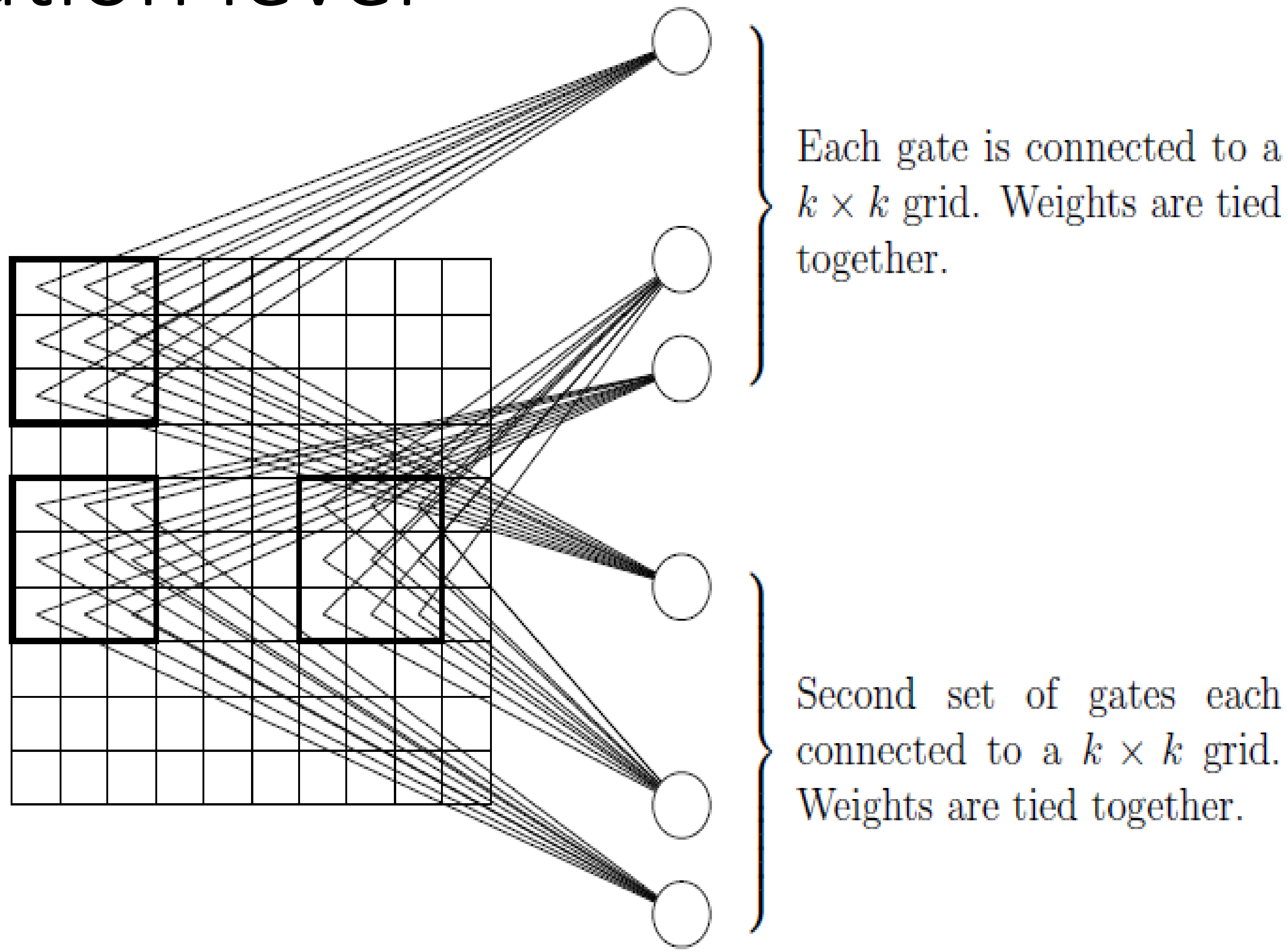
A point in one category is very close to a point in each of the other categories.

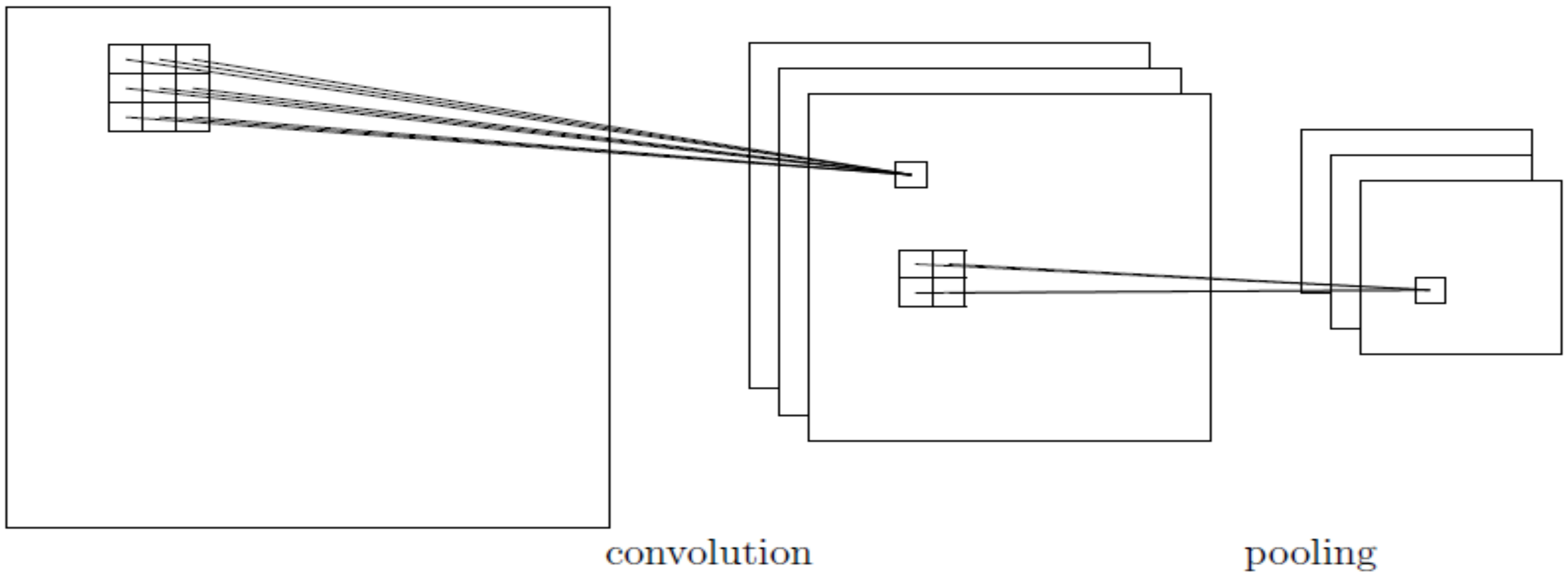
$$\left(\frac{1}{\sqrt{d}} + \epsilon, \frac{1}{\sqrt{d}}, \ldots, \frac{1}{\sqrt{d}}\right) \Longrightarrow \left(\frac{1}{\sqrt{d}}, \frac{1}{\sqrt{d}} + \epsilon, \ldots, \frac{1}{\sqrt{d}}\right)$$

Current state of deep learning research

- Most research in deep learning is experimental.
- Why experiments rather than focus on theory that can be taught?
- A possible explanation comes from some research on the cat manifold.

Convolution level




convolution
pooling

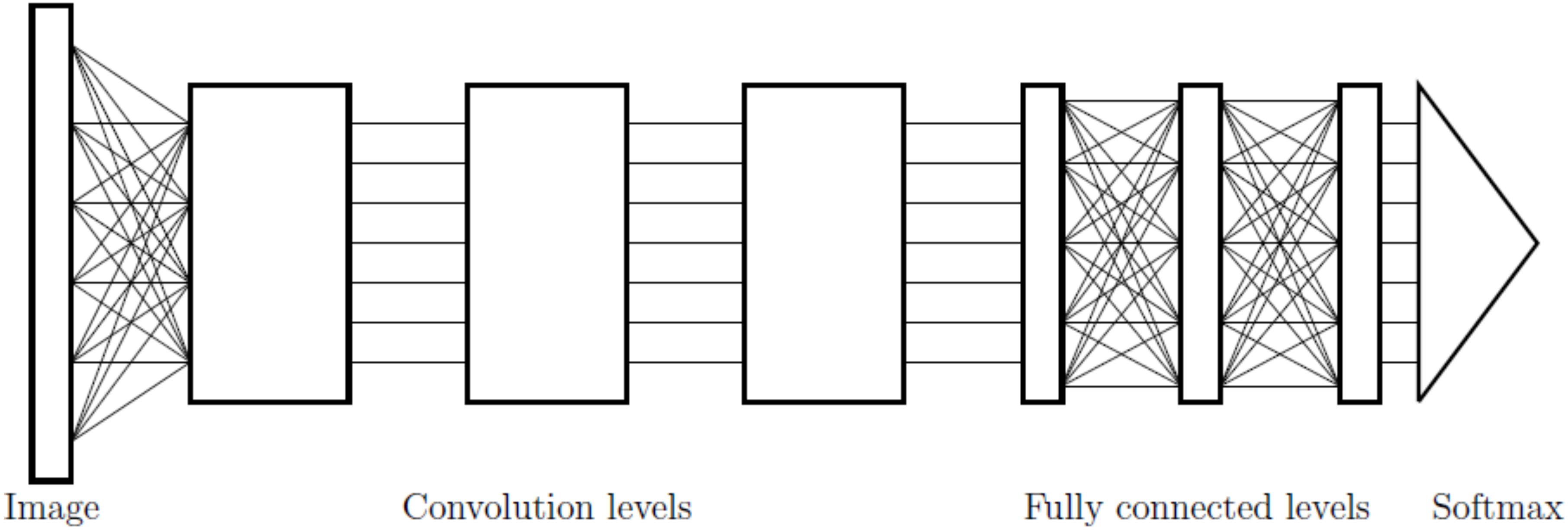
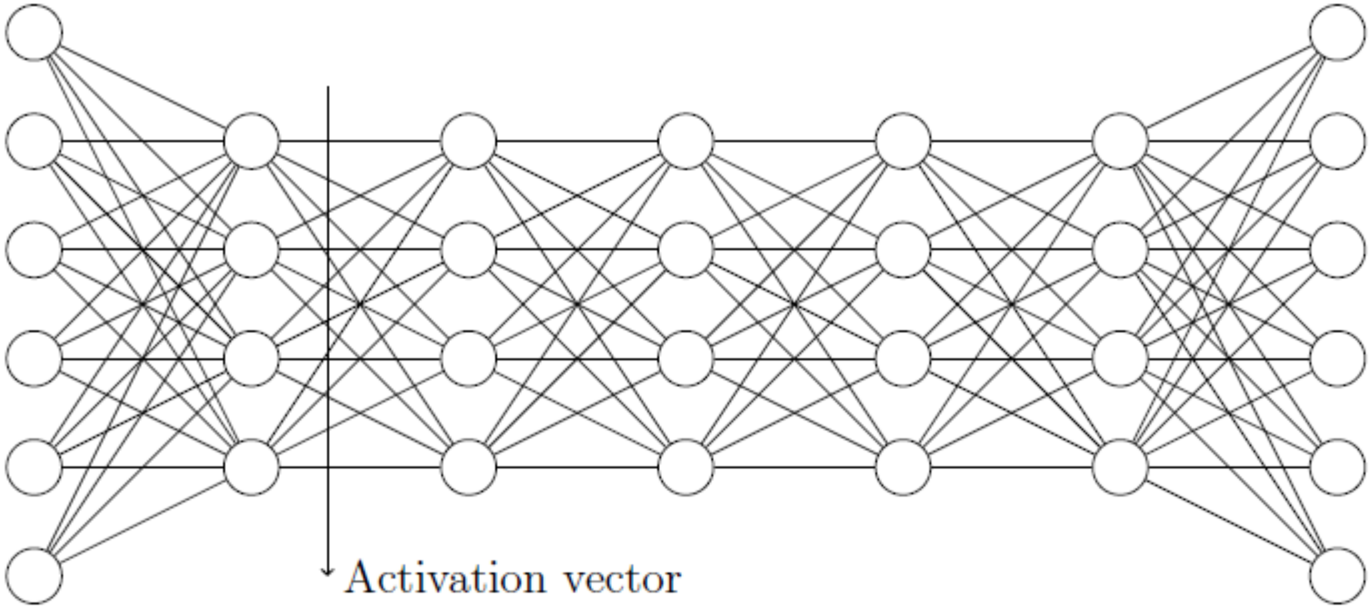


Image
Convolution levels
Fully connected levels
Softmax

Activation space



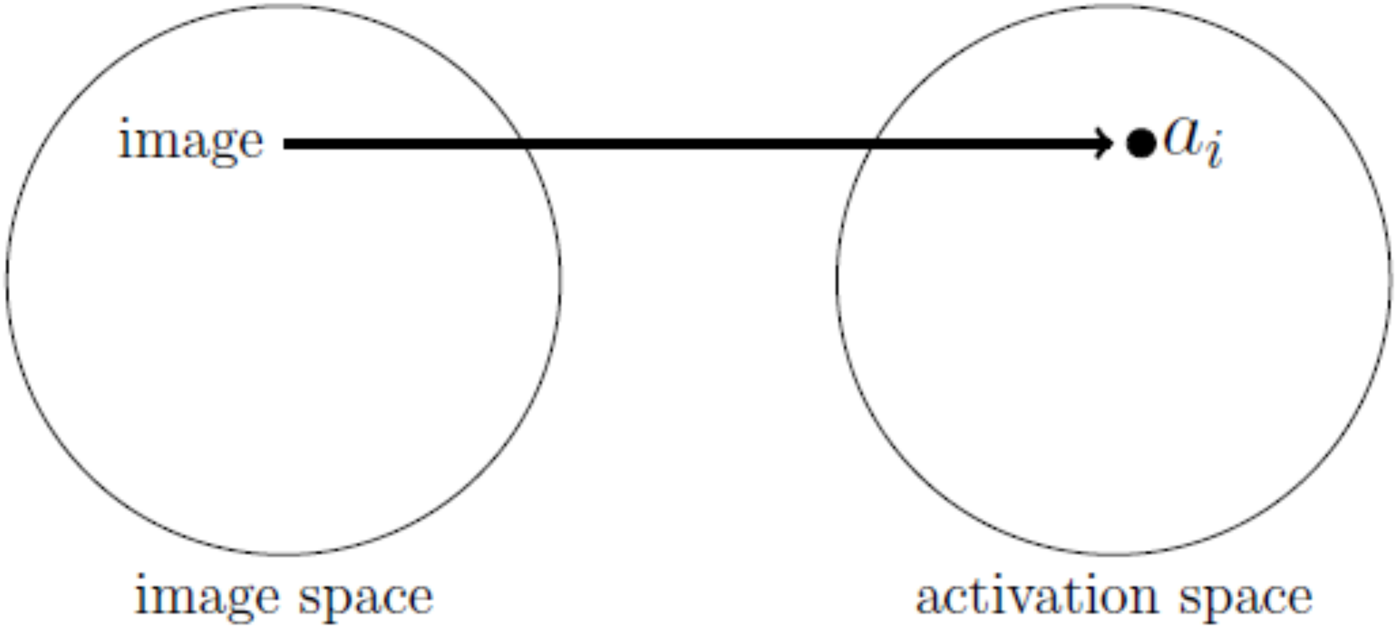


image
a_i
image space
activation space

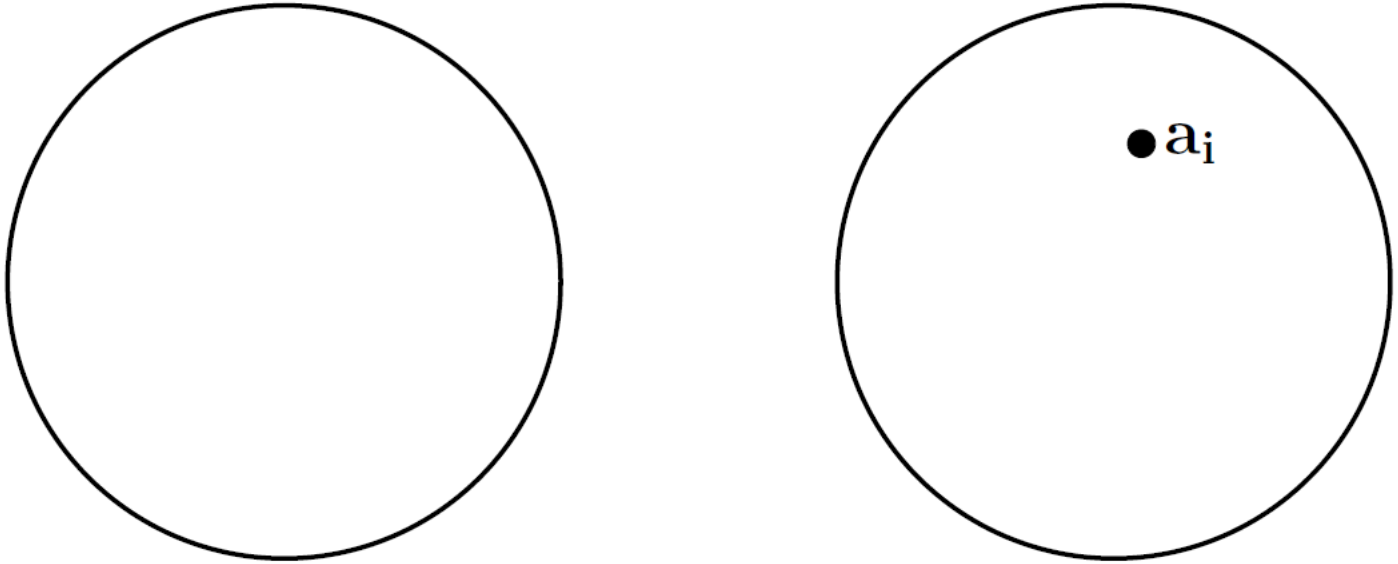


image space

activation space

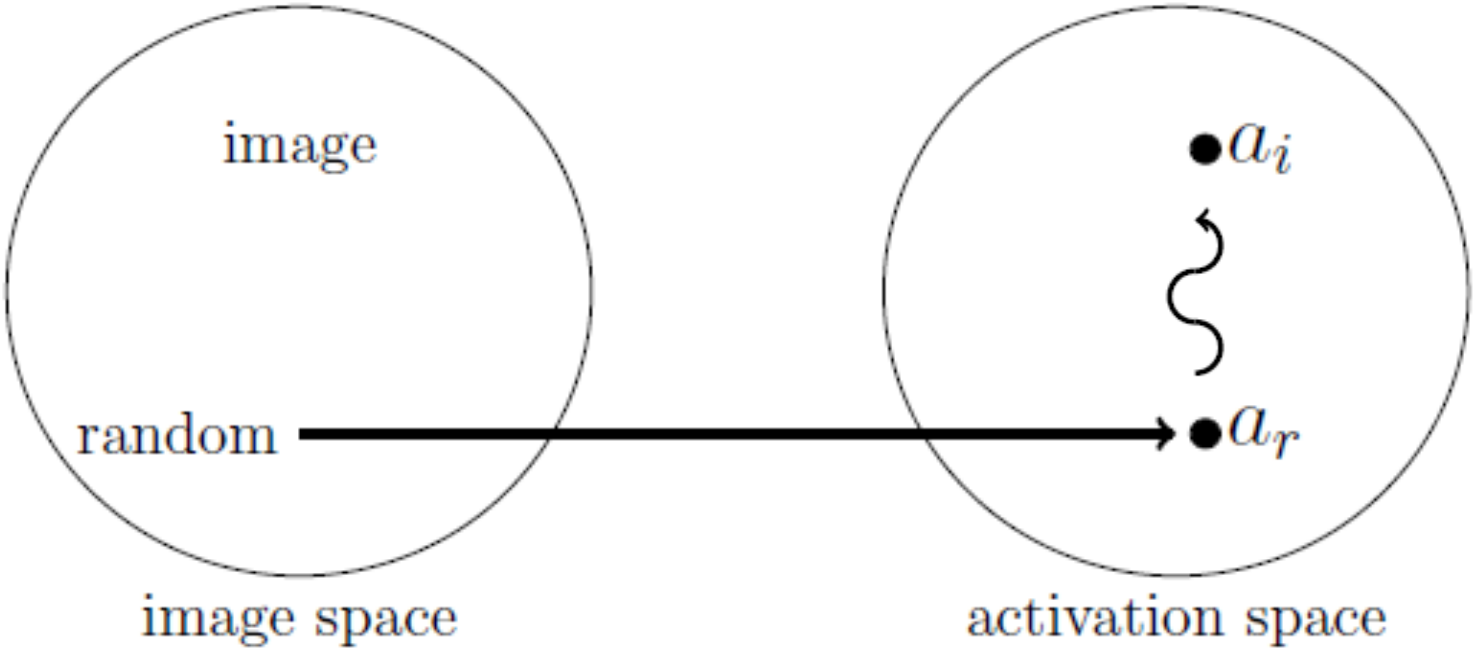
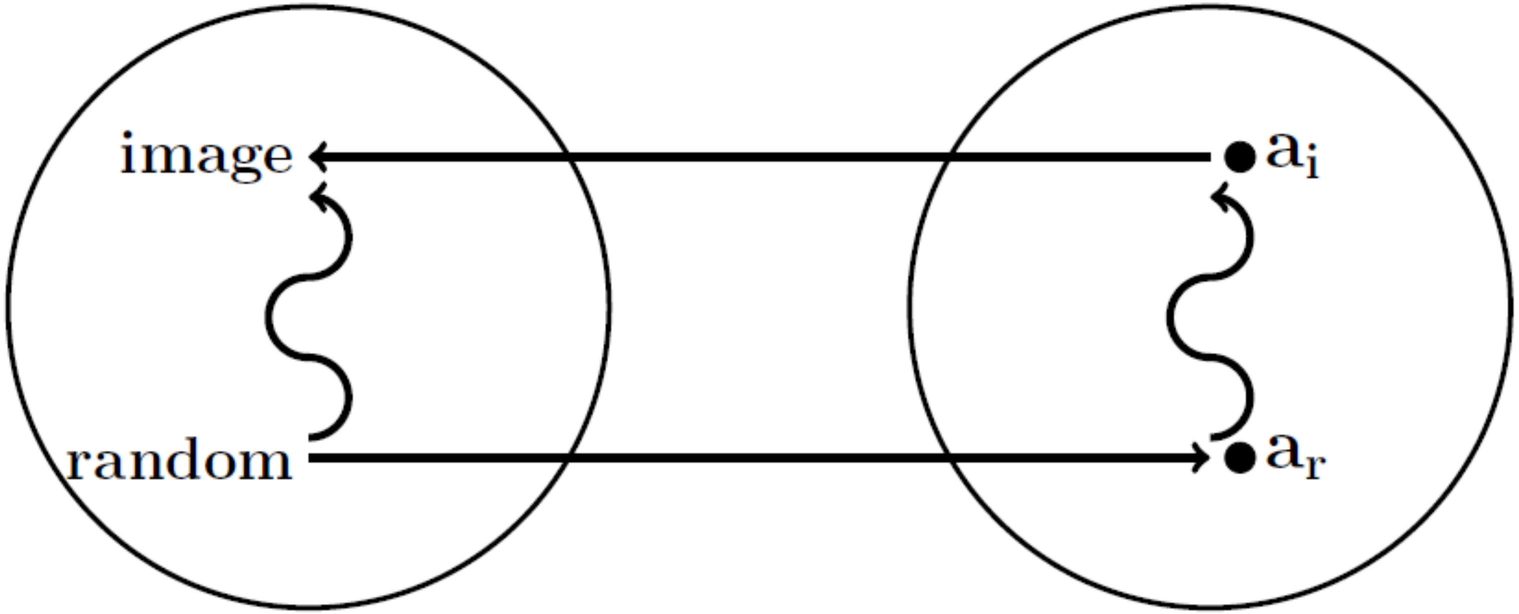
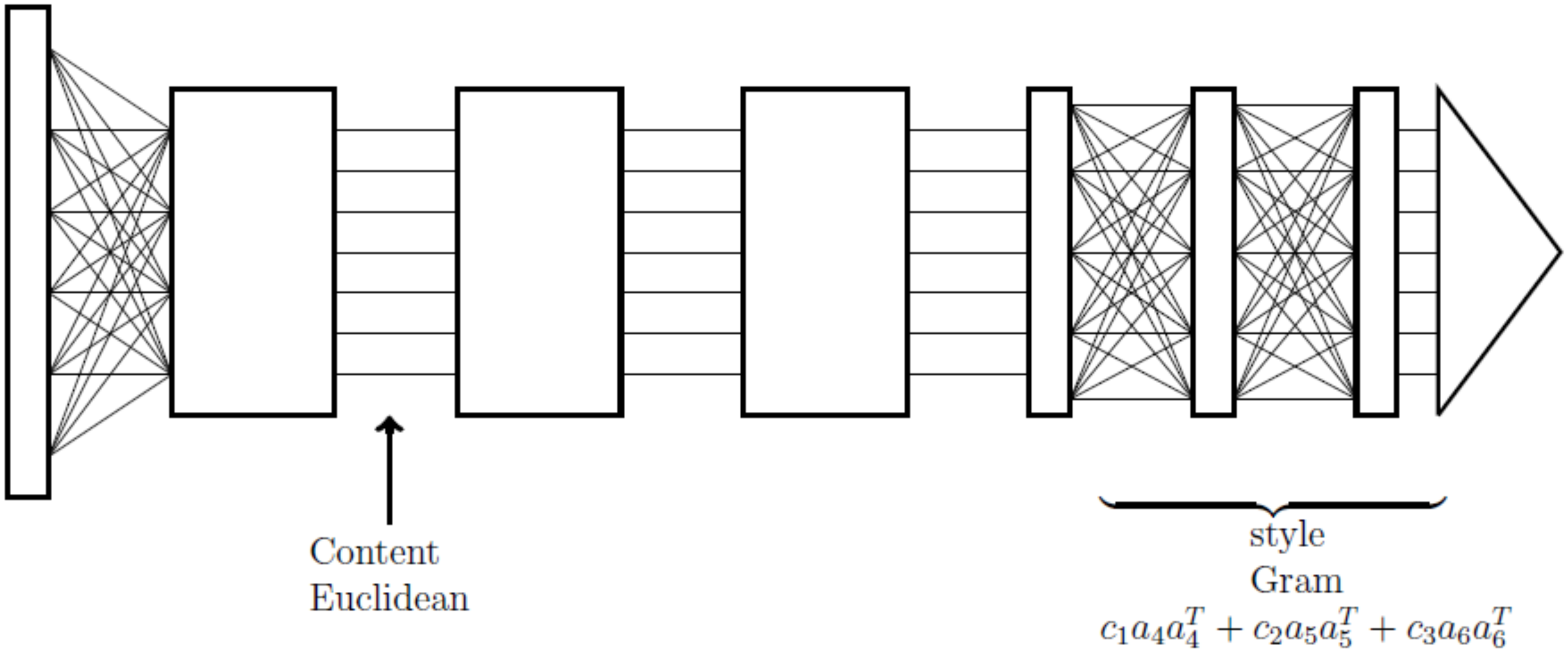


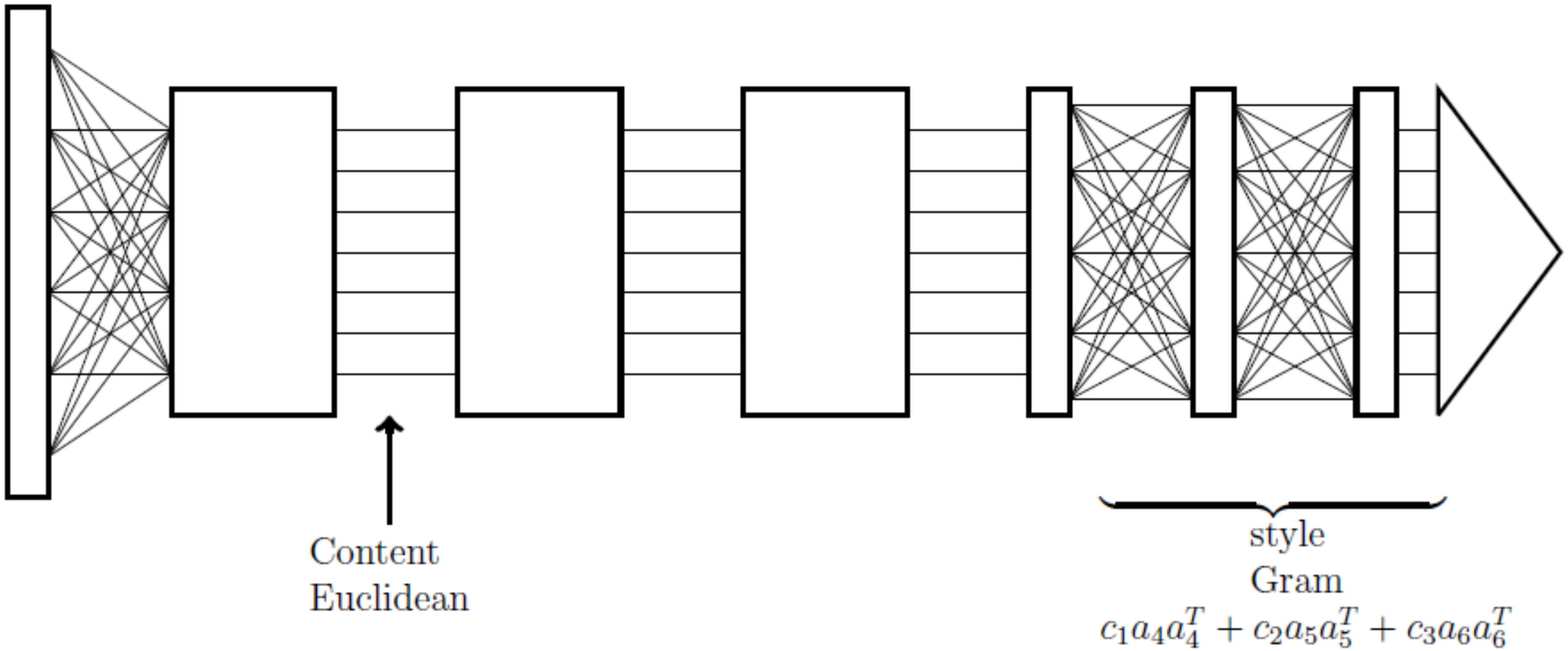
image
random
image space
a_i
a_r
activation space

Image from activation vector





$\lambda_1 \times \text{content difference} + \lambda_2 \times \text{style difference}$



$\lambda_1 \times$ content difference $+ \lambda_2 \times$ style difference

Changing young to old

Jacob R. Gardner*, Paul Upchurch*, Matt J. Kusner, Yixuan Li, Kilian Q. Weinberger, Kavita Bala, John E. Hopcroft

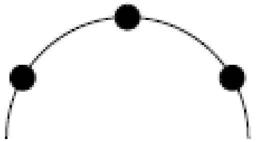
Chinese Painting Meets Cornell Campus

* Top left: Elegant photo taken by Parvez Sukheswalla.

* Bottom left: Guanzhong Wu, *Bridges*, 1985.

What dimension is the cat manifold?

- Many researchers mention that the cat manifold is low dimensional.
- We asked what is the dimension?
- With only a thousand cat images we had a set of activation vectors that were far apart.



Dimension of the cat manifold

- We created a thousand closes activations by modifying a cat image with random noise.
- Determined the dimension of the hyperplane tangent to these activation vectors.
- Realized that determining the dimension was complicated by the fact that we did not have a mathematical definition of the set of cat images.
- Why not create some mathematical categories of images and try to prove some theorems?

The two dimensional T category

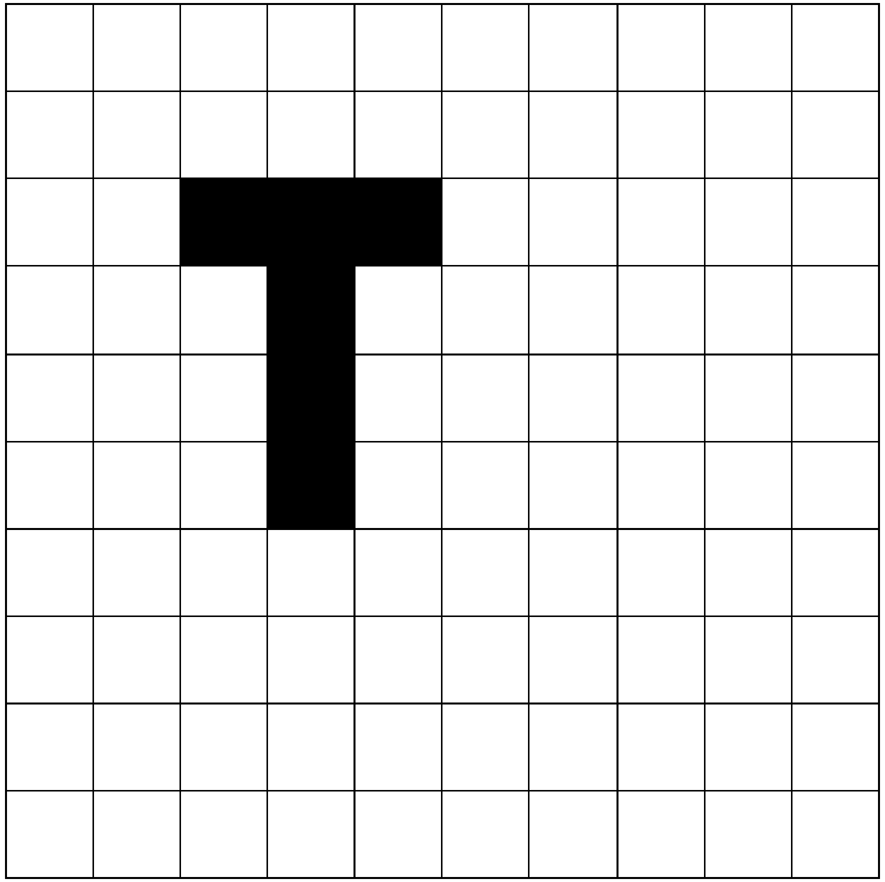


Image categories

- The T category of images is a two dimensional category in 100 dimensions.
- It is actually a vector space when images are converted to vectors.
- Would like a manifold rather than a vector space.
- Find categories similar to cat and dog in CIFAR10.
- Train and see if behavior is similar to CIFAR data
- If behavior similar then prove some theorems using the mathematical categories.

Mathematical category of images

- Having mathematically defined categories of images gives you a way to prove theorems that can be taught.
- Would like the categories to be manifolds rather than vector spaces.
- Would like categories to be 200 dimensional in 1024 dimensional space.
- Prefer that the manifolds not be symmetric.

Training a deep network

- Create an error function indicating how far from correct classification each image is.
- The error function has a term for each of possibly 100,000 images.
- Network may have a million weights.
- Gradient descent finds a local minimum of the error function by taking the derivative of the 100,000 terms with respect to each of the million weights.

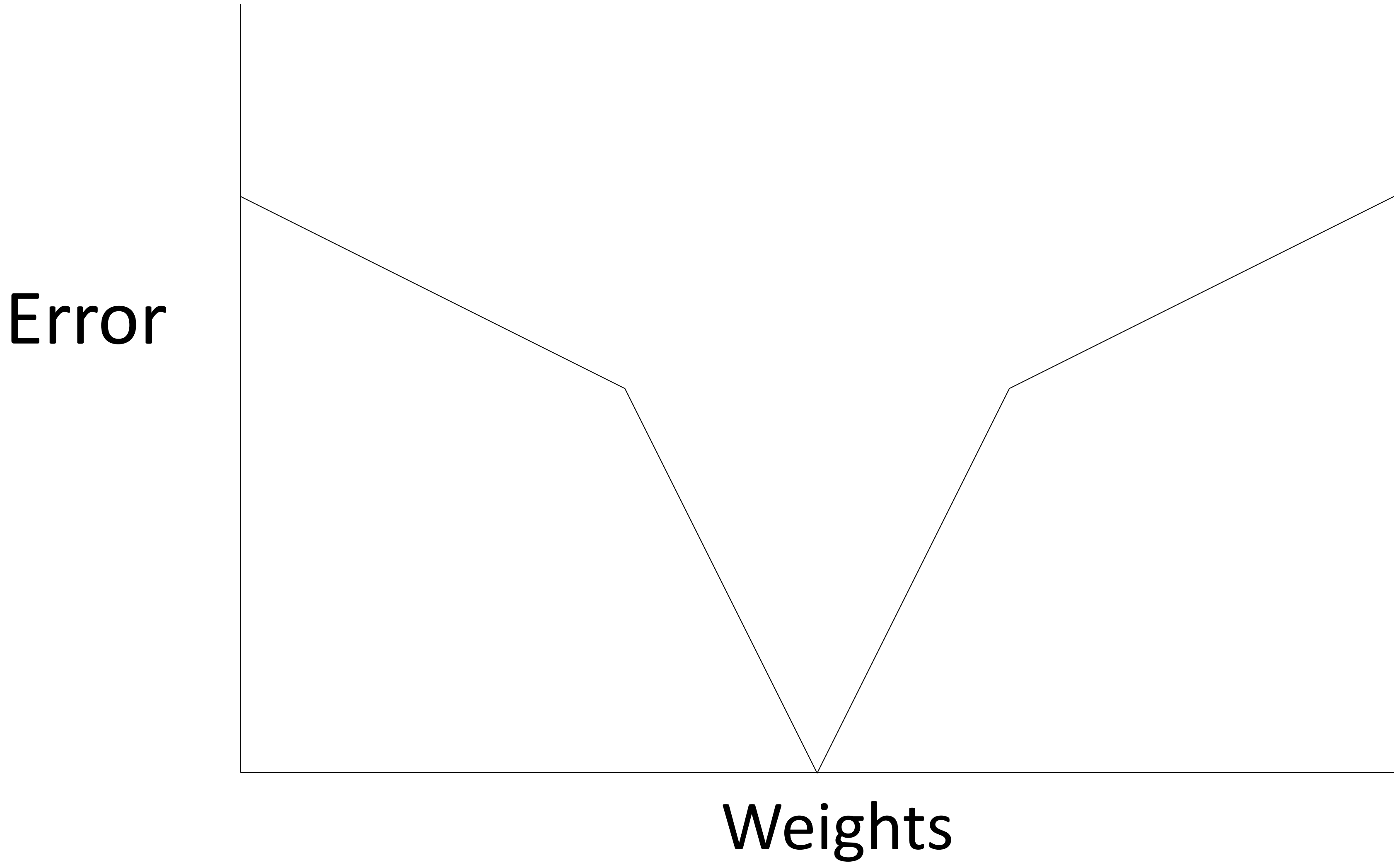
Stochastic gradient descent

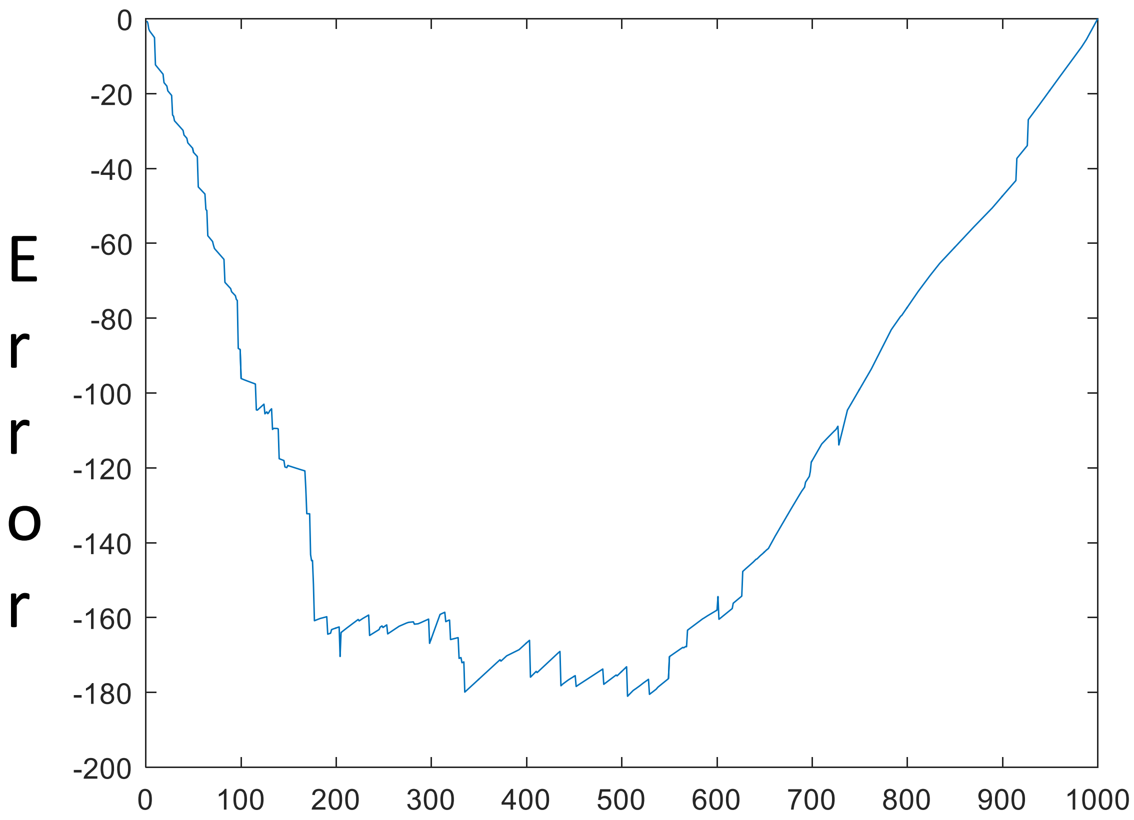
- Stochastic gradient descent randomly samples an image at each iteration and takes the derivative of only one term in the error function.

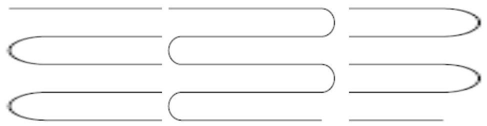
- Stochastic gradient descent is 10,000 times faster but it also usually finds a much better local minimum.

- WHY?

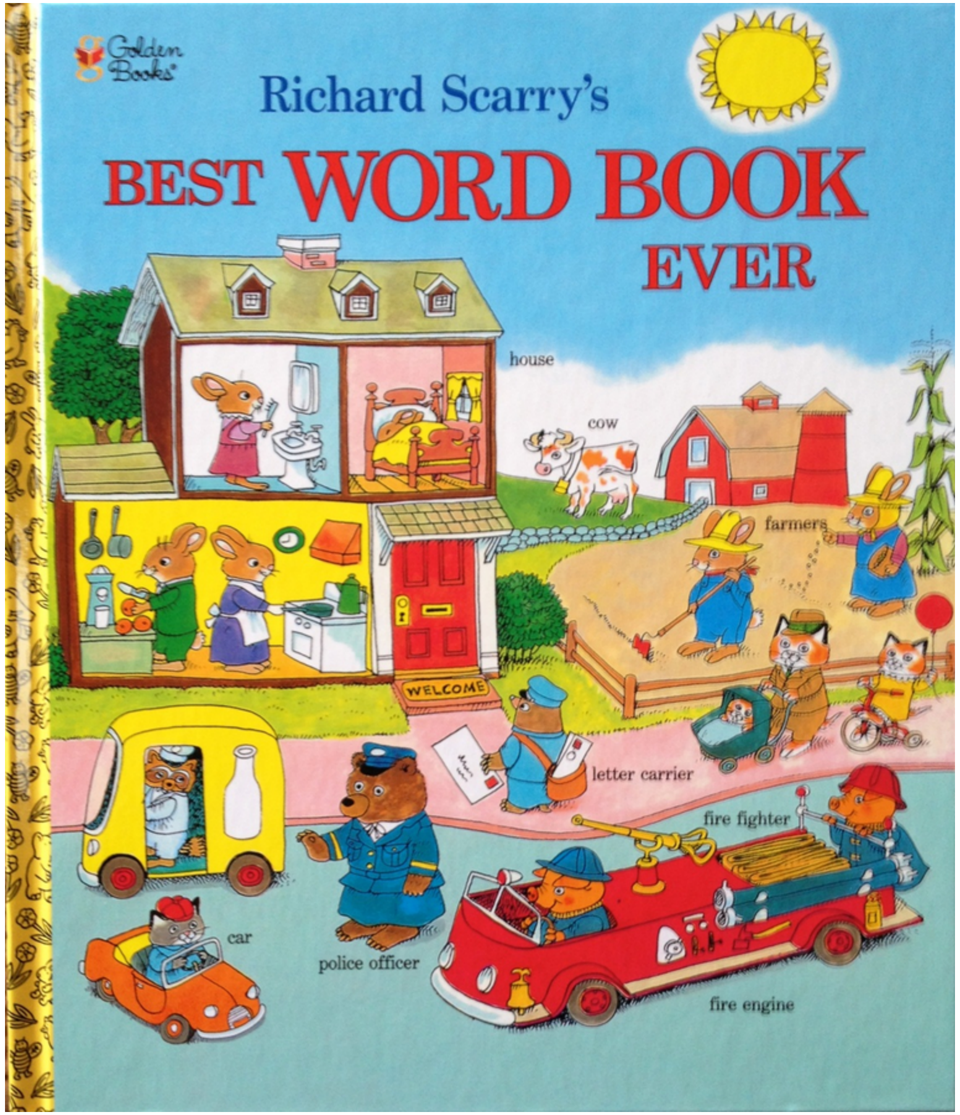
Error function for a single image




E
r
r
o
r
0
-20
-40
-60
-80
-100
-120
-140
-160
-180
-200
0
100
200
300
400
500
600
700
800
900
1000



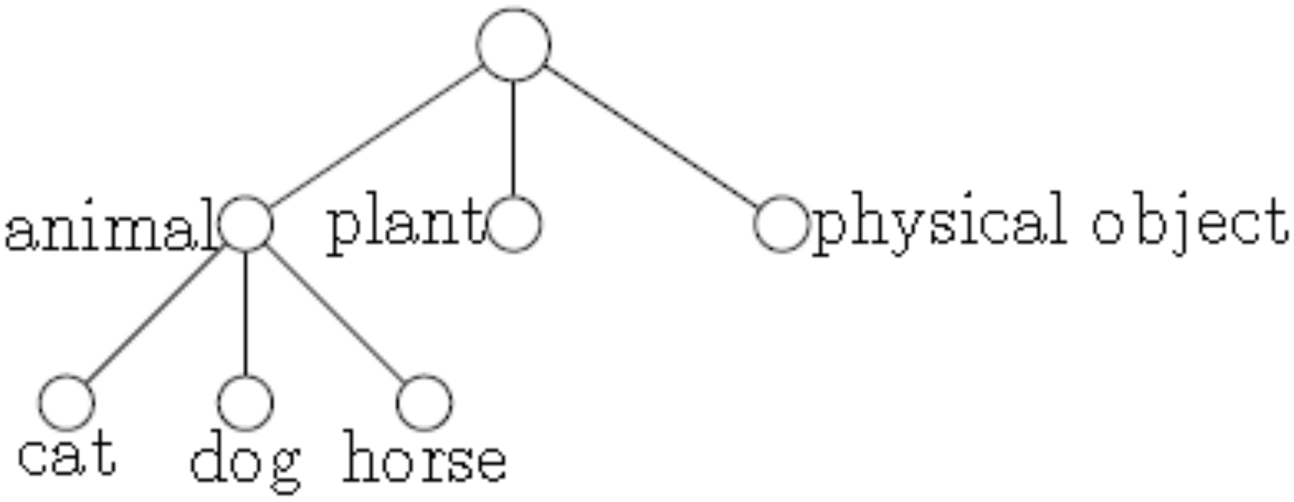
Learning from a single image




CLAY TWP.
CLAY TWP.
FIRE DEPT
ENGINE
241

Learning from a single image

Maybe a child learns how to learns the category of an image from a single image from millions of images.



Time to start developing the theory of deep learning.

Thank you