CS375 / Psych 249:
Large-Scale Neural Network Models for Neuroscience

Lecture 1: Motivations
2017.09.25

Daniel Yamins
Stanford Neurosciences Institute
Stanford Artificial Intelligence Laboratory
Departments of Psychology and Computer Science
Stanford University
The origin of the problem

What: Neuroscience.
The origin of the problem

What: Neuroscience.

Why: Behavior (cognition).
The origin of the problem

What: Neuroscience.

Why: Behavior (cognition). [Skinner vs Chomsky notwithstanding.]
The origin of the problem

What: Neuroscience.

Why: Behavior (cognition).
The origin of the problem

What: Neuroscience.

Why: Behavior (cognition).
The origin of the problem

What: Neuroscience.

Why: Behavior (cognition).
The origin of the problem

What: Neuroscience.

Why: Behavior (cognition).

How: At-scale computational models — mostly neural networks.

... because “word models” & toy problems aren’t going to be good enough for the neuroscience of the future
Understanding complex, noisy data streams is a critical part of cognition.
Problem: Entity Extraction

Understanding complex, noisy data streams is a critical part of cognition.

William James

The baby, assailed by eyes, ears, nose, skin, and entrails at once, feels it all as one great blooming, buzzing confusion.

Without sophisticated parsing and entity extraction, the world would be “as one great blooming, buzzing confusion” (for babies or otherwise).
Understanding complex, noisy data streams is a critical part of cognition.

William James

The baby, assailed by eyes, ears, nose, skin, and entrails at once, feels it all as one great blooming, buzzing confusion.

Without sophisticated parsing and entity extraction, the world would be “as one great blooming, buzzing confusion” (for babies or otherwise).

* actually not clearly true for babies ...
Understanding complex, noisy data streams is a critical part of cognition.

“Mercedes behind Lamborghini, on a field in front of mountains.”
Problem: Entity Extraction

View: position, size, pose, illumination

Distortion & Noise

Geometric variation

Background variation

Car identities:
- Beetle
- BMW Z3
- Clio
- Celica
- Alfa
- VW Bora
- BMW 325
- Astra
“Visual object recognition”

- Fast, effortless, & accurate
- Domain general
- Tolerant to high variation

Image adapted from MIT Street Scenes Database (Courtesy of Tommy Poggio)
Problem: Entity Extraction

from Hong et al (2016)
Problem: Entity Extraction from Hong et al (2016)
Problem: Entity Extraction

from Hong et al (2016)
Problem: Entity Extraction

Aspect Ratio and Angle

from Hong et al (2016)
Problem: Entity Extraction

We can quickly assess the scene as a whole.

- **Category**
- **Identity**
- **3-D Object Scale**
- **Perimeter**
- **2-D Retinal Area**
- **Bounding Box**
- **Aspect Ratio**
- **Major Axis Length**
- **Major Axis Angle**
- **X and Y Axis Position**
- **Pose in each axis**

from Hong et al (2016)
A working definition of an “explicit” representation = a basis in which a problem is linearly separable

The same concept applies to higher dimensional spaces
You need **SELECTIVITY** for different objects

You need **TOLERANCE** to changes in the retinal image

- Computationally easy (e.g. templates)
- Computationally easy (e.g. simply integrate)
- BOTH -- computationally difficult!
Any population representation

an object identity manifold
A “good” population representation

The Computational Crux of the problem

A “bad” population representation

The Tangling of Object Manifolds

Actual pixel representation
(~ retinal image representation)

object manifolds are "tangled"
(Due to identity-preserving image variation.)

Axes of natural variation of natural “physics” representation of world

e.g.

retinal photoreceptor voltage
Axes of natural variation for natural behavioral events

e.g.

deforming face moving in complex-lighted environment
Axes of natural variation for natural **behavioral** events (e.g. deforming face moving in complex-lighted environment)

*are misaligned with*

Axes of natural variation of natural “**physics**” representation of world e.g. retinal photoreceptor voltage
Axes of natural variation for natural **behavioral** events (e.g. deforming face moving in complex-lighted environment)

*are misaligned with*

Axes of natural variation of natural “**physics**” representation of world e.g. retinal photoreceptor voltage
Why is the problem hard computationally?

1. Nonlinear misalignment between physical and behavioral dimensions
Problem: Entity Extraction

Why is the problem hard computationally?

1. Nonlinear misalignment between physical and behavioral dimensions

2. Needs to be done *fast*, and thus, presumably, massively in parallel
Problem: Entity Extraction

“Core object perception” regime

Chance is 50%

All the data I will show you today

Typical primate fixation duration during natural viewing

Amount of variation

Med
Lesions in IT cortex produce deficits in shape discrimination tasks
(Gross et al, 1973, Mishkin 1982)

Lesions in parietal cortex produce deficits in landmark task
(Pohl et al. 1973)
Background: Ventral visual stream
Decision and action

Memory

• Tolerance to identity-preserving transforms
• Ability to support visual recognition
• Correlation with perceptual report
• Sensitivity to behavioral state (e.g. attention)
• Visually-evoked latency
• Selectivity to visual “feature” conjunctions
• Effects of experience (plasticity)

Background: Ventral visual stream
Background: Ventral visual stream

Neuroanatomical, cytoarchitectonic, and latency evidence:

Adapted from DiCarlo et al. 2012
Background: Ventral visual stream

Horel et al. (1987); Freiwald and Tsao (2010), Pitcher, et al. (2009)
Rust (2010), Freiwald (2010), Lehky (2007)
Majaj (2012)
Background: Ventral visual stream

Ventral visual stream

rhesus macaque (macaca mulatta)
Background: Ventral visual stream

rhesus macaque (macaca mulatta)
Background: Ventral visual stream

rhesus macaque (Macaca mulatta)

Ventral visual stream
Multi-array Electrophysiology Experiment

Multi-array electrophysiology in macaque V4 and IT.

About 300 total sites

Ha Hong
Jim DiCarlo
Multi-array Electrophysiology Experiment

5760 images

64 objects

8 categories

uncorrelated photo backgrounds

Animals  Boats  Cars  Chairs  Faces  Fruits  Planes  Tables

Low variation

Medium variation

High variation

Pose, position, scale, and background variation
Multi-array Electrophysiology Experiment

About 300 total sites

Output = Binned spike counts 70ms-170ms post stimulus presentation averaged over 25-50 reps of each image.
Multi-array Electrophysiology Experiment

Responses to 1600 test images of two example units

IT unit 53

Images sorted first by **category**, then **variation level**.
Neural-Behavior Decoding

linear combination of units
Neural-Behavior Decoding

Animal or not?

linear combination of units
Neural-Behavior Decoding

- Animal or not?
- Car or not?

Different linear combination

Linear combination of units
Neural-Behavior Decoding

Animal or not?
linear combination of units

different linear combination
Car or not?

Chair or not?

Face or not?
Neural-Behavior Decoding

Animal or not?

linear combination of units

linear coefficients discovered on training data, evaluated on separate test data

different linear combination

Car or not?

Chair or not?

Face or not?
V4 loses out at higher variation:
Range of Human Behavior

Variation Level

Low

Medium

High

Animals
Boats
Cars
Chairs
Faces
Fruits
Planes
Tables

Cars Lo var
Cars Med var
Cars Hi var

Car 1
Car 2
Car 3
Car 4
Car 5
Car 6
Car 7
Car 8

Faces Lo var
Faces Med var
Faces Hi var

Face 1
Face 2
Face 3
Face 4
Face 5
Face 6
Face 7
Face 8

at ceiling ...

... at chance
V4 loses out at higher variation:

... but humans are much less affected.

Yamins* and Hong* et. al. PNAS (2014)
V4 loses out at higher variation:

... but humans are much less affected.

... as is the IT neural population.
V4 loses out at higher variation:

... but humans are much less affected.

... as is the IT neural population.

At high variation levels, IT much better than V4 and existing models.
IT Neurons Track Human Performance

IT matches human error patterns as well as raw performance.

- **Low-Variation Face subordinate tasks.**
Comparison of Object Recognition Behavior in Human and Monkey

Upshot: human and non-human primate basic level core object perception (sp. identification) are indistinguishable.

Does not depend on reporting effector (touch vs. eye movement)
Feature Space as Encoding

Pixel space: $\mathbb{R}^{1000000}$

Feature space: $\mathbb{R}^{4000(?)}$
Behavior = Feature space + Simple decision rule
= encoding + decoding

Feature Space as Encoding

Pixel space: $\mathbb{R}^{1000000}$
Output

Feature space: $\mathbb{R}^{4000(?)}$

Behavioral

Category Judgement

Linear Classifier

Localization

Linear Regressor

“Subjective” Similarity judgement

Distance Function

CHAIR
Behavior = Feature space + Simple decision rule
= encoding + decoding
Encoding & Decoding

Stimulus $\rightarrow$ Neurons $\rightarrow$ Behavior

representation

read-out
Encoding & Decoding

Stimulus → Neurons → Behavior

Representation

Neurons

Read-out

Category
Location
Size
Pose
Depth relationships

visual representation
Encoding & Decoding

Stimulus → Neurons → Behavior

Stimulus representation → Neurons read-out → Behavior

visual representation

Category, Location, Size, Pose, Depth relationships

Encoding & Decoding

Stimulus → Neurons → Behavior

Stimulus representation → Neurons read-out → Behavior

visual representation

Category, Location, Size, Pose, Depth relationships

Encoding & Decoding

Stimulus → Neurons → Behavior

Stimulus representation → Neurons read-out → Behavior

visual representation

Category, Location, Size, Pose, Depth relationships
Encoding & Decoding

Stimulus representation → Neurons read-out → Behavior

Stimulus: visual representation

Category, Location, Size, Pose, Depth relationships

very nonlinear*

*which is presumably why so much brainmeat needs to be devoted to it.
GOAL: Predictive model of single-neuron responses throughout the ventral stream to arbitrary image stimuli.

Key questions:
(a) how many layers?
GOAL: Predictive model of single-neuron responses throughout the ventral stream to arbitrary image stimuli.

Key questions:
(a) how many layers?
(b) what's in each layer, specifically?
**GOAL:** Predictive model of single-neuron responses throughout the ventral stream to arbitrary image stimuli.

Key questions:
(a) how many layers?

(b) what's in each layer, specifically?

(c) what behavioral goals and biophysical facts constrain it to be as it is?
Understanding complex, noisy data streams is a critical part of cognition.
Problem: Entity Extraction

Understanding complex, noisy data streams is critical part of cognition.

“Hannah is good at compromising.”
Understanding complex, noisy data streams is critical part of cognition.

Problem: Entity Extraction

“Hannah is good at compromising.”

variation sources: speaker identity, background noise, reverberation...
Human speech recognition is remarkably invariant:
Human speech recognition is remarkably invariant:

- Recordings from standard speech recognition databases (TIMIT, WSJ) with words spoken at least 20 times
- Combined with significant background noise

- **auditory scenes**
  - “She **had** your **suit** in **wash** water **all year** …”

- **speech babble**
  - ‘had’
  - ‘suit’
  - ‘wash’
  - ‘year’
Common sounds …

Man speaking  Road traffic  Guitar
Flushing toilet  Zipper  Coughing
Pouring liquid  Cellphone vibrating  Crumpling paper
Tooth-brushing  Water dripping  Siren
Woman speaking  Scratching  Splashing water
Car accelerating  Car windows  Computer speech
Biting and chewing  Telephone ringing  Alarm clock
Laughing  Chopping food  Walking with heels
Typing  Telephone dialing  Vacuum
Car engine starting  Girl speaking  Wind
Running water  Car horn  Boy speaking
Breathing  Writing  Chair rolling
Keys jangling  Computer startup sound  Rock song
Dishes clanking  Background speech  Door knocking
Ringtone  Songbird  
Microwave  Pouring water  
Dog barking  Pop song  
  Water boiling

*Sam Norman-Haignere, Nancy Kanwisher, and Josh McDermott
The Cocktail Party Problem

Real-world settings often involve concurrent sounds.
What happens to sound in a room:
• Presence of other speakers obscures much structure of target utterance, but speech remains intelligible.

• Present-day speech recognition algorithms (e.g. in your iPhone) fall apart in such circumstances.
Auditory Cortex and Audition

How are circuits making sense of complex sound patterns?
Auditory Cortex and Audition

Auditory Cortex and Audition

Auditory Cortex and Audition


*monkey*
Spatiotemporal filtering? *Shamma, 2005*


*monkey*
Auditory Cortex and Audition

Spatiotemporal filtering? *Shamma, 2005*

*monkey*

---

*Tramo et. al, Curr. Opin. Neuro. (1999)*
Problem: Entity Extraction

“Mercedes behind Lamborghini, on a field in front of mountains.”

“Hannah is good at compromising”
“Mercedes behind Lamborghini, on a field in front of mountains.”

“Hannah is good at compromising.”
“Mercedes behind Lamborghinis, on a field in front of mountains.”

“Hannah is good at compromising”
Problem: Entity Extraction

Petersen, 2007
Problem: Entity Extraction

- **a)** 
imaged with a diagram showing neural pathways including Trigeminal Ganglion and Thalamus.

- **b)** 
  - **Input Shapes:** Cube, Chair, Duck
  - **Artificial Vibrissal Array:** Sweeps

- **c)** 
  - **Task-Optimized Neural Network Architecture(s):** Matched to real morphology

- **d)** 
  - **Shape Category Recognition Output:** Cube, Chair, Duck
Many Different Computational Goals

- temporal sequence processing
  - motion processing, action recognition (parietal cortex?)
  - navigation (hippocampus?)
Many Different Computational Goals

- temporal sequence processing
  - motion processing, action recognition (parietal cortex?)
  - navigation (hippocampus?)

- motor command production & execution (motor cortex)
Many Different Computational Goals

- temporal sequence processing
  - motion processing, action recognition (parietal cortex?)
  - navigation (hippocampus?)

- motor command production & execution (motor cortex)

- memory and strategic planning (hippocampus, prefrontal cortex)
Many Different Computational Goals

- temporal sequence processing
  - motion processing, action recognition (parietal cortex?)
  - navigation (hippocampus?)

- motor command production & execution (motor cortex)

- memory and strategic planning (hippocampus, prefrontal cortex)

- language
Many Different Computational Goals

- temporal sequence processing
  - motion processing, action recognition (parietal cortex?)
  - navigation (hippocampus?)

- motor command production & execution (motor cortex)

- memory and strategic planning (hippocampus, prefrontal cortex)

- language

- emotions, theory-of-mind
The Neuron Doctrine

The “Strong” Neuron Doctrine:

Fig. 2. A drawing done by Cajal to show some of the neurons of the retina in vertical section.
The Neuron Doctrine

The “Strong” Neuron Doctrine:

a neuron receives signals at its dendrites and cell body and transmits them, as action potentials, along the axon in one direction: away from the cell body
The Neuron Doctrine

The “Strong” Neuron Doctrine:

1. The nervous system is made up of discrete cells (“neurons”), connected by extracellular junctions (synapses) into a directed graph.
The “Strong” Neuron Doctrine:

i. The nervous system is made up of discrete cells (“neurons”), connected by extracellular junctions (synapses) into a directed graph

ii. Neurons are “excitable cells” that “fire” by a mechanism of electrochemical (de)polarization [Hodgkin-Huxel model]
The Neuron Doctrine

The “Strong” Neuron Doctrine:

i. The nervous system is made up of discrete cells (“neurons”), connected by extracellular junctions (synapses) into a directed graph

ii. neurons are “excitable cells” that “fire” by a mechanism of electrochemical (de)polarization [Hodgkin-Huxel model]

iii. the firing pattern of a neuron is a parameterized function that “integrates” the firing patterns of the neurons that synapse onto it
The Neuron Doctrine

The “Strong” Neuron Doctrine:

i. The nervous system is made up of discrete cells (“neurons”), connected by extracellular junctions (synapses) into a directed graph.

ii. Neurons are “excitable cells” that “fire” by a mechanism of electrochemical (de)polarization [Hodgkin-Huxel model].

iii. The firing pattern of a neuron is a parameterized function that “integrates” the firing patterns of the neurons that synapse onto it.

iv. The parameters of the function are plastic and therefore learnable.
Artificial Neural Networks (ANNs)

McCulloch and Pitts (1943)
McCulloch and Pitts (1943)

\[ y_k = \phi \left( \sum_{j=0}^{m} w_{kj} x_j \right) \]

\[ \phi : \mathbb{R} \longrightarrow \mathbb{R} \]

some (nonlinear) activation function

\[ w_{kj} \in \mathbb{R}^{m+1} \]

“synaptic strengths”
Artificial Neural Networks (ANNs)

McCulloch and Pitts (1943)

\[
y_k = \phi \left( \sum_{j=0}^{m} w_{kj} x_j + b_j \right)
\]

\( \phi : \mathbb{R} \rightarrow \mathbb{R} \)

some (nonlinear) activation function

\( w_{kj} \in \mathbb{R}^{m+1} \)

“synaptic strengths”

\( b_j \in \mathbb{R} \)

“biases”
Artificial Neural Network (ANN) Models of the Brain

Core (obvious) idea: Model brain systems with ANNs
Core (obvious) idea: Model brain systems with ANNs

But how to find the correct parameters?*

*both continuous parameters like weights and discrete parameters of the architecture
Artificial Neural Network (ANN) Models of the Brain

Core (obvious) idea: Model brain systems with ANNs

But how to find the correct parameters?*
*both continuous parameters like weights and discrete parameters of the architecture

How to measure model correctness? (and model “understanding”?)

*both continuous parameters like weights and discrete parameters of the architecture
Artificial Neural Network (ANN) Models of the Brain

Core (obvious) idea: Model brain systems with ANNs

But how to find the correct parameters?*

*both continuous parameters like weights and discrete parameters of the architecture

How to measure model correctness? (and model “understanding”?)

This course is teach you how to do these things.
But how to find the correct parameters?
Two strategies to find the correct parameters.

1. Fit neural data
Two strategies to find the correct parameters.

1. Fit neural data
   - ex: fit model to categorization performance, check against ventral stream data

2. Solve a high-level ecological task
   - ... compare to neural data
Two strategies to find the correct parameters.

1. Fit neural data
   - ex: fit model to categorization performance, check against ventral stream data

2. Solve a high-level ecological task
   - ... compare to neural data
Two strategies to find the correct parameters.

less normative theory

1. Fit neural data

1.5 Fit behavioral error pattern (Turing test)

... check against neural data

more normative theory

2. Solve a high-level ecological task

... compare to neural data and Turing Test

ex: fit model to categorization performance, check against ventral stream data
Heuristic of “Goal-Driven Modeling”

“Mercedes behind Lamborghini, on a field in front of mountains.”

“Hannah is good at compromising”
Heuristic of "Goal-Driven Modeling"

"Mercedes behind Lamborghini, on a field in front of mountains."

"Hannah is good at compromising"
1. Formulate comprehensive model class (CNNs)
1. Formulate comprehensive model class (CNNs)

2. Choose challenging, ethologically-valid tasks (categorization)

1. Formulate comprehensive model class (CNNs)

2. Choose challenging, ethologically-valid tasks (categorization)

3. Implement generic learning rules (gradient descent)

1. Formulate comprehensive model class (CNNs)

2. Choose challenging, ethologically-valid tasks (categorization)

3. Implement generic learning rules (gradient descent)

> Map to brain data. (ventral stream)

argmin\left[ L(p^*) \right]_{a \in A}

where \( p^* \) is result of

\[
\frac{dp_a}{dt} = -\lambda(t) \cdot \langle \nabla_{p_a} L(x) \rangle_{x \in D}
\]

**A** = architecture class

**L** = loss function

**D** = dataset
1. \( A = \text{architecture class} \)

2. \( L = \text{loss function} \quad D = \text{dataset} \)

3. \[
\text{argmin} \left[ L(p_a^*) \right] \quad \forall a \in A
\]

where \( p^* \) is result of

\[
\frac{dp_a}{dt} = -\lambda(t) \cdot \langle \nabla_{p_a} L(x) \rangle \quad x \in D
\]

"learning rule"

"task"
Two strategies

1. Fit neural data

1.5 Fit behavioral error pattern (Turing test)

... check against neural data

Models of visual cortex, motor system

more normative theory

2. Solve a high-level ecological task

... compare to neural data and Turing Test

less normative theory
Two strategies

Models of the Retina, LFADS

1. Fit neural data

1.5 Fit behavioral error pattern (Turing test)

...check against neural data

Models of visual cortex, motor system

more normative theory

2. Solve a high-level ecological task

...compare to neural data and Turing Test
**Theorem 1.** Let $\phi(\cdot)$ be any nonconstant, bounded, monotonically increasing continuous real-valued function. Let $C_m$ denote the space of continuous functions on the $m$-dimensional unit cube. Then for any $\epsilon > 0$ and $f \in C_m$, there exists an $N$, $v_i, b_i \in \mathbb{R}$, and $w_i \in \mathbb{R}^m$ such that the function

$$\hat{f}(x) \triangleq \sum_{i=1}^{N} v_i \phi(w_i^T x + b_i)$$

satisfies

$$|\hat{f}(x) - f(x)| < \epsilon \text{ for all } x \in [0, 1]^m.$$
The Universal Approximation Theorem

**Theorem 1.** Let \( \phi(\cdot) \) be any nonconstant, bounded, monotonically increasing continuous real-valued function. Let \( C_m \) denote the space of continuous functions on the \( m \)-dimensional unit cube. Then for any \( \varepsilon > 0 \) and \( f \in C_m \), there exists an \( N, v_i, b_i \in \mathbb{R}, \) and \( w_i \in \mathbb{R}^m \) such that the function

\[
\hat{f}(x) \triangleq \sum_{i=1}^{N} v_i \phi(w_i^T x + b_i)
\]

satisfies

\[
|\hat{f}(x) - f(x)| < \varepsilon \text{ for all } x \in [0, 1]^m.
\]
The Universal Approximation Theorem

**Theorem 1.** Let \( \phi(\cdot) \) be any nonconstant, bounded, monotonically increasing continuous real-valued function. Let \( C_m \) denote the space of continuous functions on the \( m \)-dimensional unit cube. Then for any \( \varepsilon > 0 \) and \( f \in C_m \), there exists an \( N, v_i, b_i \in \mathbb{R} \), and \( w_i \in \mathbb{R}^m \) such that the function

\[
\hat{f}(x) = \sum_{i=1}^{N} v_i \phi(w_i^T x + b_i)
\]

satisfies

\[
|f(x) - \hat{f}(x)| < \varepsilon \text{ for all } x \in [0, 1]^m.
\]

This is a neural network with one hidden layer.
The Universal Approximation Theorem is **BAD** . . . if you’re a neuroscientist.

**Theorem 1.** Let $\phi(\cdot)$ be any nonconstant, bounded, monotonically increasing continuous real-valued function. Let $C_m$ denote the space of continuous functions on the $m$-dimensional unit cube. Then for any $\epsilon > 0$ and $f \in C_m$, there exists an $N$, $v_i, b_i \in \mathbb{R}$, and $w_i \in \mathbb{R}^m$ such that the function

$$\hat{f}(x) \triangleq \sum_{i=1}^{N} v_i \phi(w_i^T x + b_i)$$

satisfies

$$\vert f(x) - \hat{f}(x) \vert < \epsilon \text{ for all } x \in [0, 1]^m.$$
Defeating the UAT.

Line of Attack #1: Size. There is no guarantee that N is reasonable.
Defeating the UAT.

Line of Attack #1: Size. There is no guarantee that $N$ is reasonable.

**Theorem 2.** For fixed size $N$, the complexity of a function $f$ approximated by a $d$-layer network is exponential in $d$. 
Line of Attack #1: Size. There is no guarantee that $N$ is reasonable.

**Theorem 2.** *For fixed size $N$, the complexity of a function $f$ approximated by a $d$-layer network is exponential in $d$.*

Line of Attack #2: Learnability. There is no guarantee that (e.g.) stochastic gradient descent with a reasonable amount of training data can discover the weight $(w_i, v_i, b_i)$ parameters.
Defeating the UAT.

Line of Attack #1: Size. There is no guarantee that $N$ is reasonable.

**Theorem 2.** For fixed size $N$, the complexity of a function $f$ approximated by a $d$-layer network is exponential in $d$.

Line of Attack #2: Learnability. There is no guarantee that (e.g.) stochastic gradient descent with a reasonable amount of training data can discover the weight $(w_i, v_i, b_i)$ parameters.

Theorem-ish: deeper networks are easier to train (less unstable) than large shallow networks.
Defeating the UAT.

Line of Attack #1:  Size.  There is no guarantee that N is reasonable.

**Theorem 2.** For fixed size $N$, the complexity of a function $f$ approximated by a $d$-layer network is exponential in $d$.

Line of Attack #2:  Learnability.  There is no guarantee that (e.g.) stochastic gradient descent with a reasonable amount of training data can discover the weight $(w_i, v_i, b_i)$ parameters.

Theorem-ish:  deeper networks are easier to train (less unstable) than large shallow networks.

…. so structure matters, and we will exploit it.
1. Formulate comprehensive model class (**CNNs**)

2. Choose challenging, ethologically-valid tasks (**categorization**)

3. Implement generic learning rules (**gradient descent**)

> Map to brain data. (**ventral stream**)
**Logistics**

**TA:** Damian Mrowca  

**Website:** cs375.stanford.edu

**Class Structure:** Mixture of Lectures and Tutorials

**Assignments:** 3 Homeworks, reproducing key results from the literature. Worked in small groups (3-4). Turned in as Lab Reports.

**Final Projects** (start 11/15, due 12/15).

**Tools:** http://cs375.stanford.edu/software-tools.html
THEY’RE MADE OUT OF MEAT