How is Machine Learning Impacting Neuroscience?

A Case Study into the Impact of ML on Science

David Carlson
Assistant Professor
Dept. of Civil & Env. Engineering
Dept. of Biostatistics & Bioinformatics
Duke University

Why add ML to Neuroscience?

- Neuroscience is experiencing a data deluge
  - Like many scientific fields
- More and higher quality data than ever is being produced
- Data is at a scale where manual pre-processing is too time-consuming
- Need advanced methods to make sense of complex data

Machine Learning in Neuroscience

- Machine learning needs to be incorporated into scientific applications in a *useful* manner

A Shared Vision for Machine Learning in Neuroscience

Mai-Anh T. Vu, Tülay Adalı, Demba Ba, György Buzsáki, David Carlson, Katherine Heller, Conor Liston, Cynthia Rudin, Vikaas S. Sohal, Alik S. Widge, Helen S. Mayberg, Guillermo Sapiro, and Kafui Dzirasa

Journal of Neuroscience 14 February 2018, 38 (7) 1601-1607; DOI: https://doi.org/10.1523/JNEUROSCI.0508-17.2018

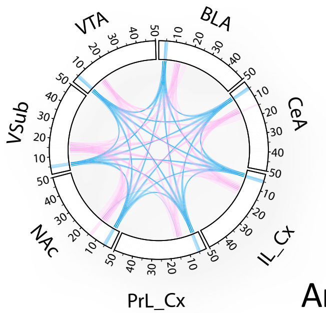
Goal of Today's Lecture

- Go through a couple of case studies:
  - Analyzing physiological data (adversarial methods for domain adaptation)
  - Automatically capturing behavior (extracting more information with CNNs)
  - Relationship between CNNs and the visual cortex

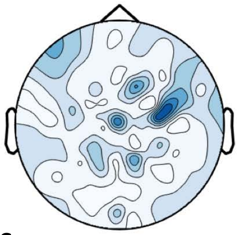
Analysis and Predictions from Electrophysiological Data

Learning *Interpretable* Neural Biomarkers for Clinical Conditions or Outcomes

- Neural activity is frequently used to try to understand the basis of neuropsychiatric disorders and effects of treatment
- Neural activity is complex
  - Can we use machine learning to break the complex signals into *interpretable* patterns?
  - Are these signals related to susceptibility or treatment outcomes?
  - Can they be used to develop novel treatments?



Animal Studies

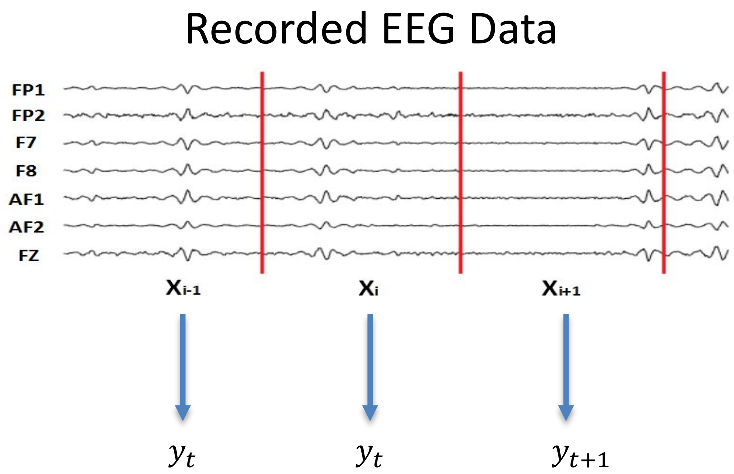


Clinical Trials

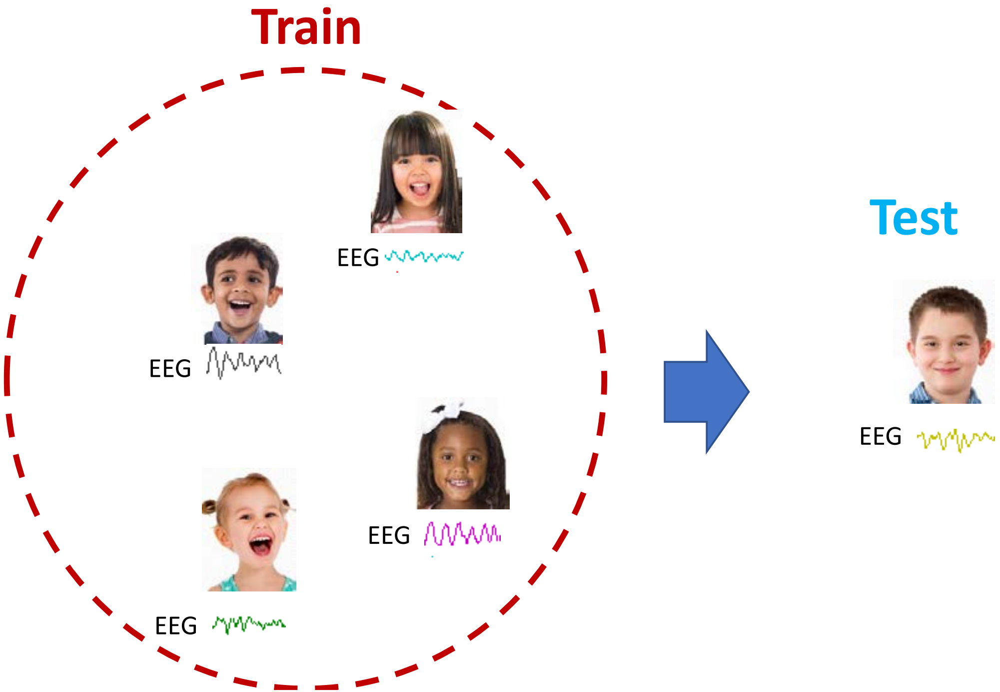
Can Machine Learning Help Understand the Brain?

What is our Scientific Question?

- Often want to relate recorded EEG data to:
  - Current behavior
  - Clinical variables
  - Seizure/not seizure
  - etc

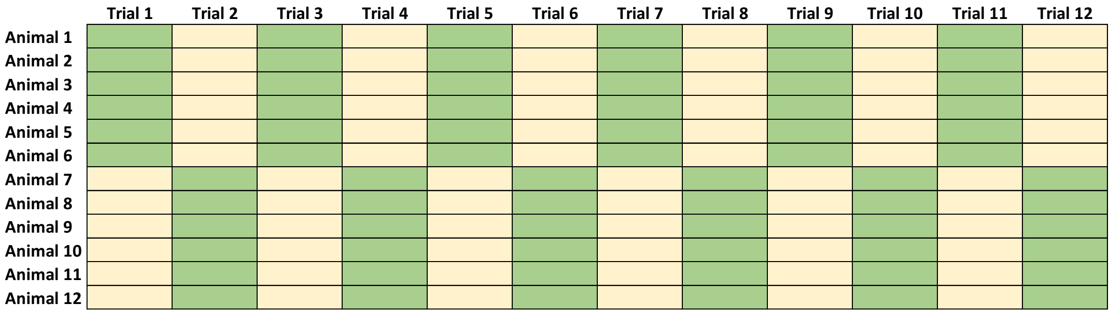


Goal (often) is to Predict on New Subjects

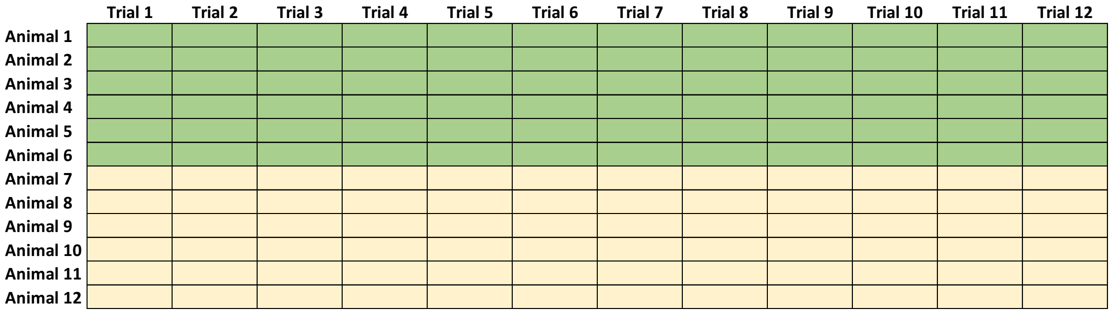


Validation over Animals or Subjects

We can split all available data into multiple groups, train (green) and validation (yellow).



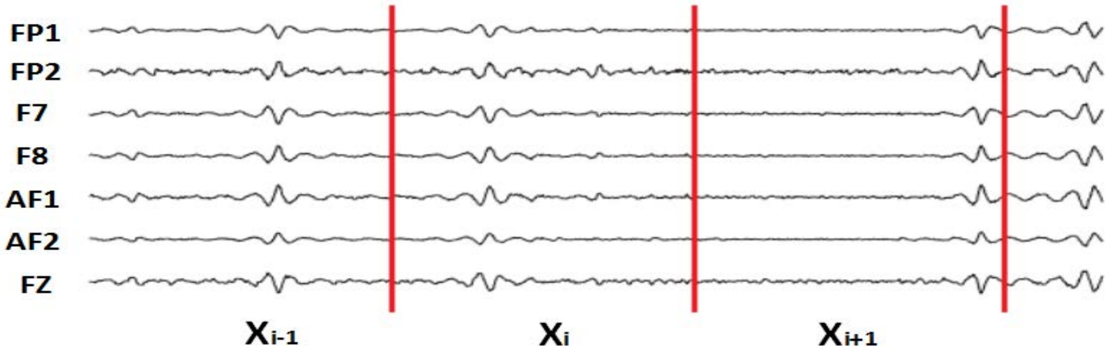
The top arrangement estimates how well we can predict within the current subjects



The bottom arrangement estimates how much we can predict on future subjects

Vu et al 2018

Windowing data makes it seem much bigger than it really is

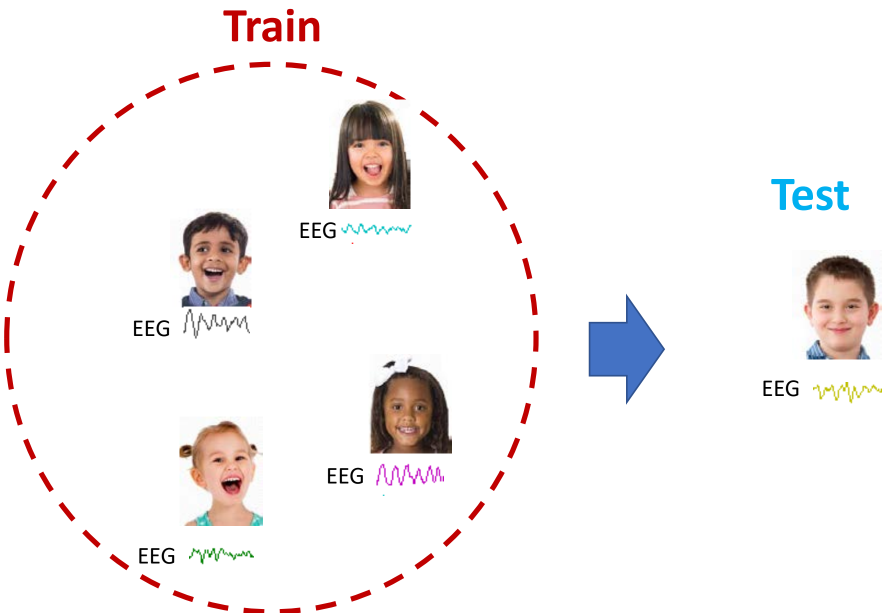


- Standard practice is to predict over sub-segments of EEG data, which makes *N* seem large
- Often only a few participants in any study or trial

“Little Big Data”

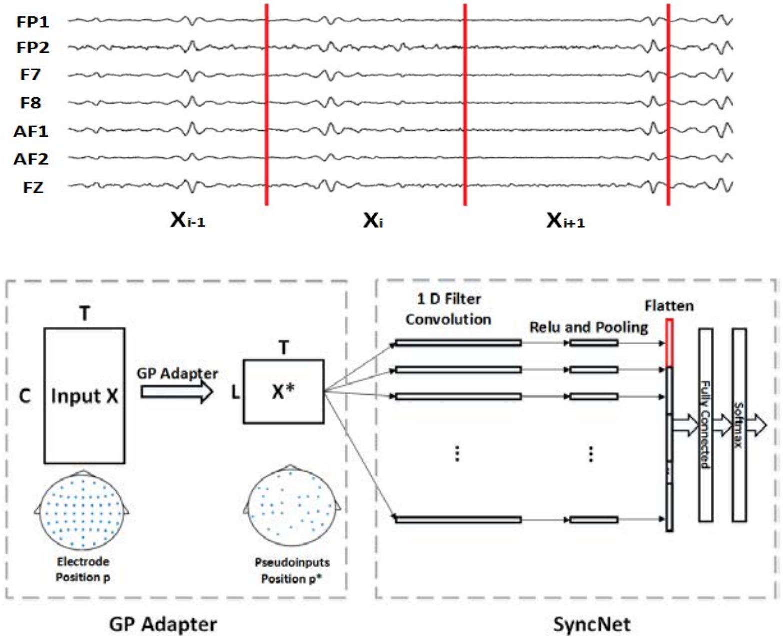
Many repeats and labels (e.g high *N*), especially true when looking at instantaneous EEG

Only a few participants to represent the entire population



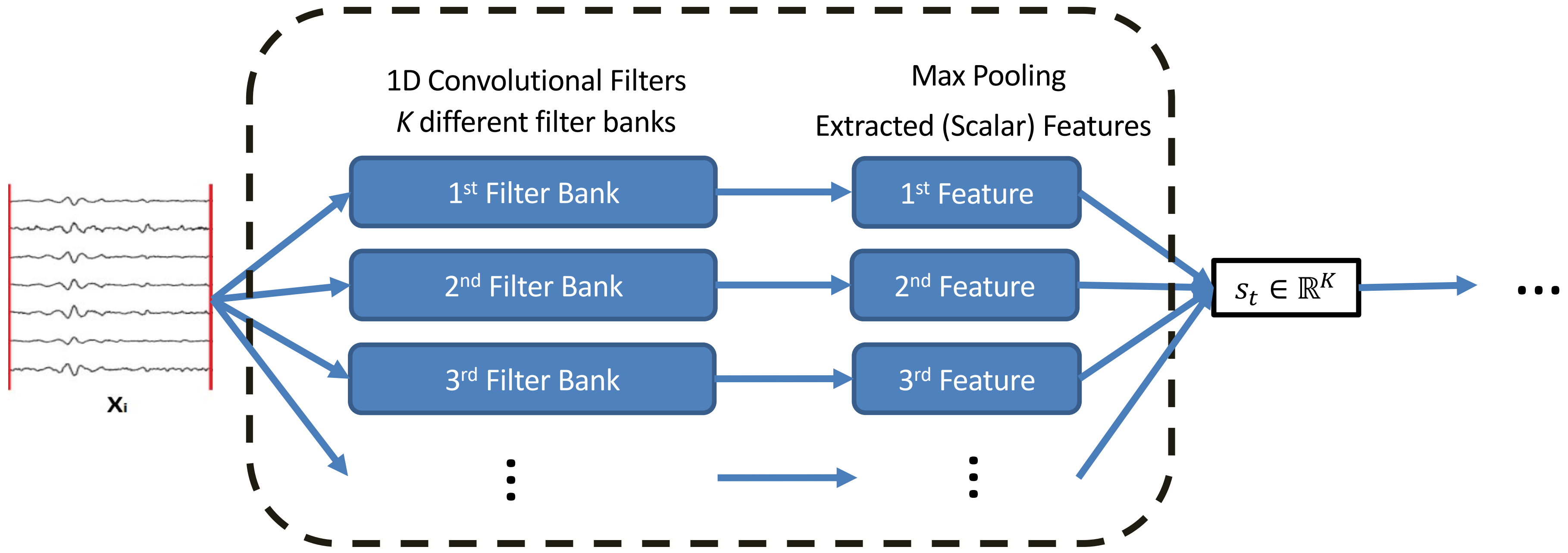
Transparent Machine Learning for EEG

- We have developed a custom Convolutional Neural Network (CNN) approach for electroencephalography (EEG) data, SyncNet
- Learns a mapping to a pseudo-input space; able to incorporate studies with differing electrode layouts into the same analysis (GP Adapter)
- Is this CNN interpretable?



Li et al, NeurIPS 2017

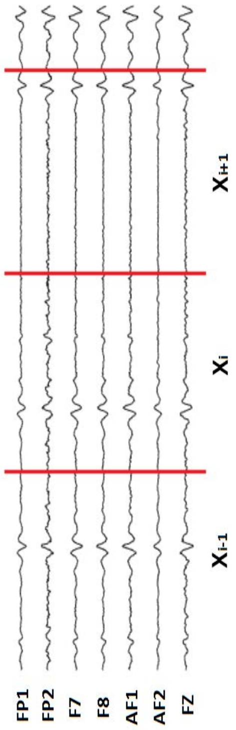
Shallow Convolutional Neural Network



Only a single convolutional layer

Can view this as a nonlinear feature extraction step.

Viewing our CNN as a Nonlinear Feature Extraction



Nonlinear Feature Extraction → Low-Dimensional Extracted Feature → Logistic Regression

Can interpret logistic regression weights.

Can we *understand* and *interpret* the extracted features?

Duke UNIVERSITY

Parameterized Convolutional Filters

- Uses *parameterized* convolutional filters based on Morlet wavelets

$$f_c^{(k)}(\tau) = b_c^{(k)} cos\left(\omega^{(k)}\tau + \phi_c^{(k)}\right) e^{-\beta^{(k)}\tau^2}$$

- c, k are channel (which electrode) and filter index, respectively
  - $\omega^{(k)}$ and $\beta^{(k)}$ control frequency properties
  - $b_c^{(k)}$ and $\phi_c^{(k)}$ are channel-specific amplitudes and phase shifts
- Frequency properties are well-understood from wavelets, so we can borrow that knowledge

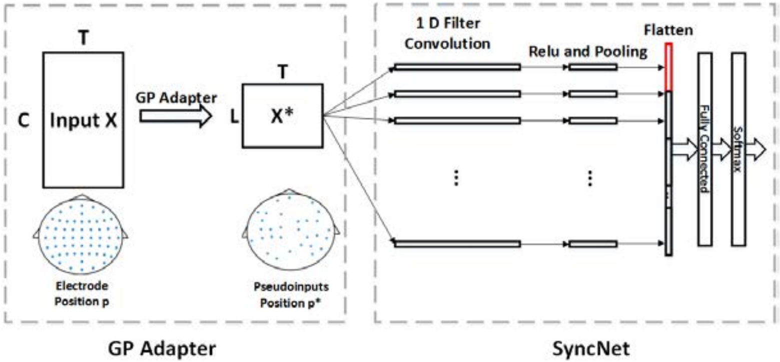
Extracting Features

- Can extract the output of the convolutional filters:

$$\boldsymbol{h}_{tk} = \sum_c x_{tc} * f_c^{(k)}(\tau)$$

$$\tilde{h}_{tk} = \max(\boldsymbol{h}_{tk})$$

- This is a max pooling over the complete time window
  - Each convolutional filter bank is reduced to a single output
  - *K* distinct filter banks will convert an EEG window into *K* features.
- Can view this as a nonlinear feature extractor



Li et al, NeurIPS 2017

Learning Treatment Biomarkers

- Want to learn neural dynamics that change post-treatment
- To evaluate this, we attempt to classify the treatment stage of the autologous umbilical cord blood clinical trial (0 months/baseline, 6 months post-treatment, and 12-months post-treatment) *using EEG alone*
- Proof-of-concept for applying to methodology to larger datasets and also learning diagnostic classification and treatment efficacy biomarkers from EEG signals
  - No controls yet
  - Closed-label clinical trial (N=180) (placebo vs. treatment) just finished and will be analyzed soon.

Convolutional Filter Visualization

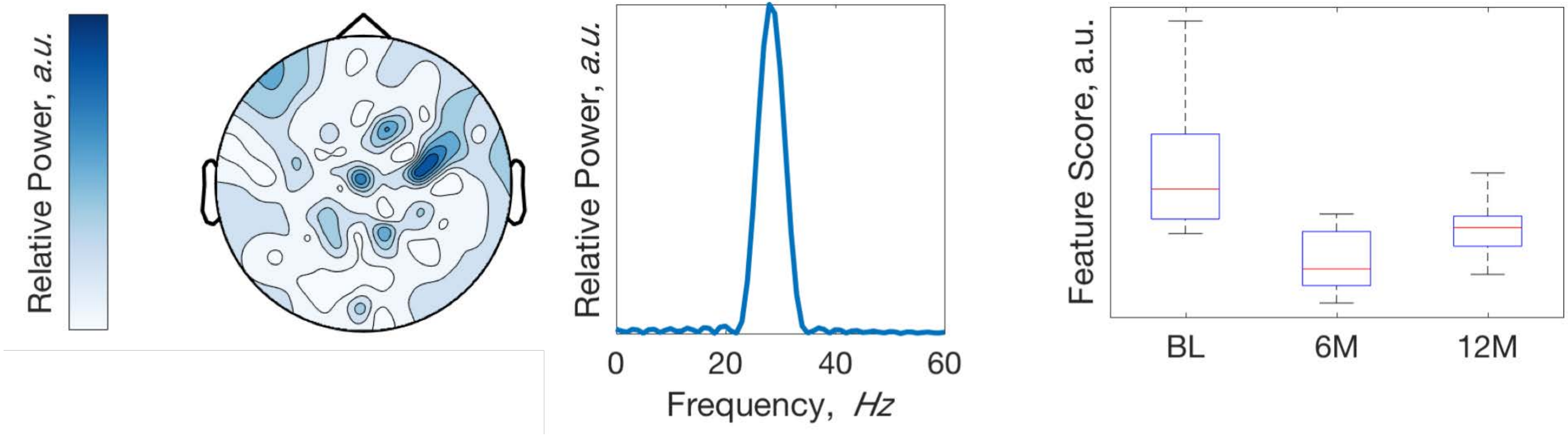


Figure: One of the ten features learned in the neural network. (Left) This figure shows relative power in arbitrary units (a.u.) defined by the learned variables in the network. (Middle) This figure shows the frequency range used by the learned filter defined by the learned variables. (Right) To demonstrate the effect of the learned feature, one can visualize its value from each data sample.

Do CNNs Learn Better Biomarkers?

- Evaluated by leave-one-participant out cross-validation, average over windows
- Predict one of (baseline, 6 months post-treatment, 12 months post-treatment)
- Most common analysis is based on Power Spectral Density features

	Accuracy
Random Guessing	33.3
Dominant Class	41.1
Diff. Ent. + SVM	50.4
Power Spec. Density+SVM	49.9
MC-DCNN*	58.4
SyncNet*	**60.1**

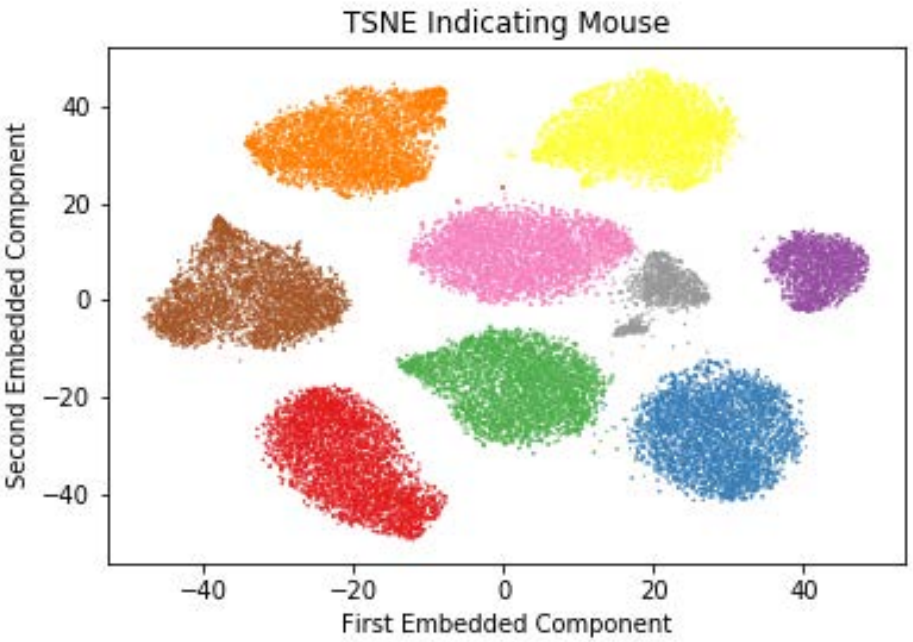
*Custom CNN-based approaches

Biological and Medical Measurements are Heterogenous

Often measurements have more differences between individuals than between outcomes/labels.

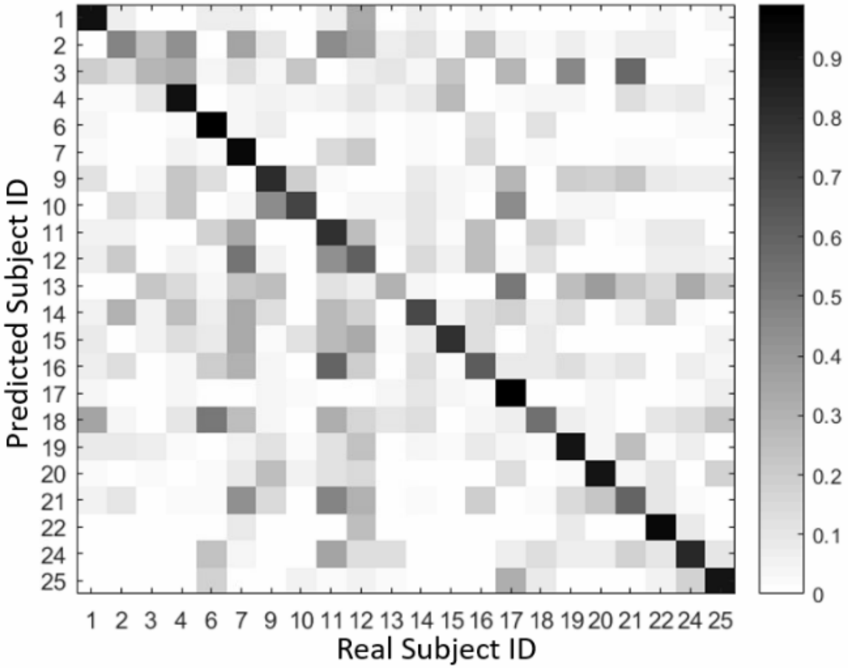
On the right is a TSNE representation of a behavioral test of a mouse model of ASD. Separate mice are coded by a different color.

Each point is one observation from a single mouse.

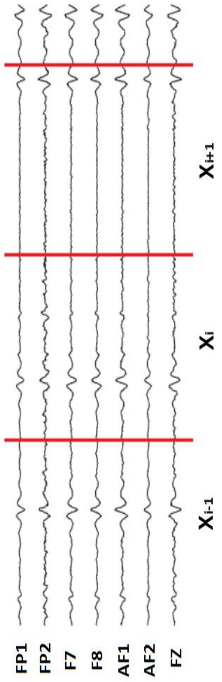


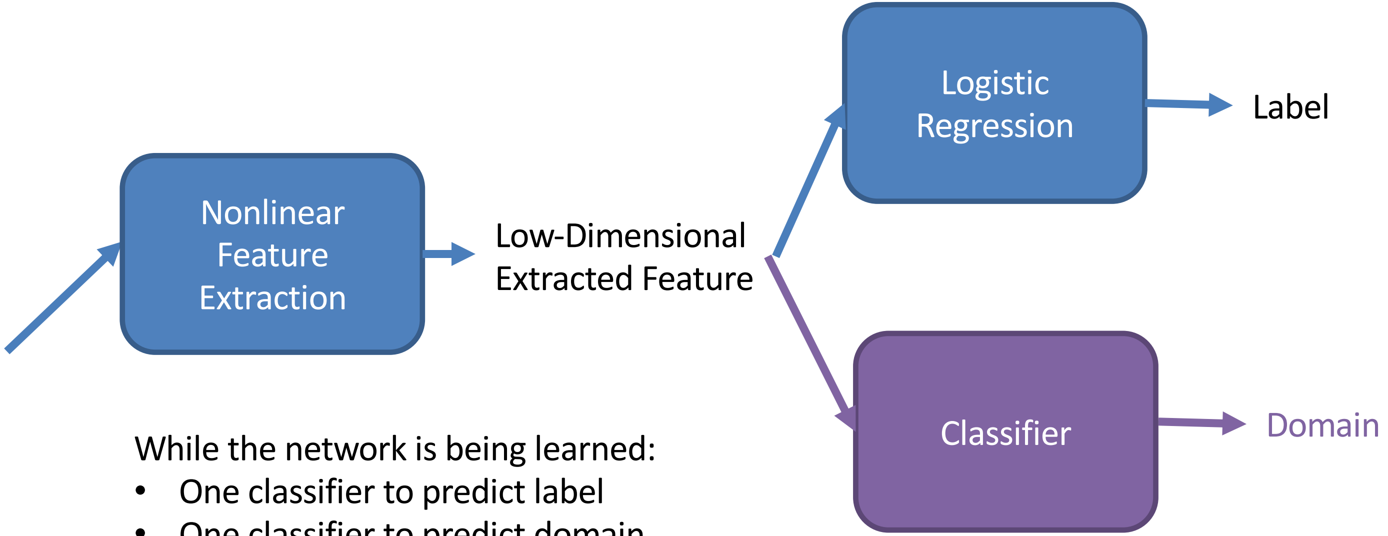
Do we Learn Participant-Specific Features?

- If we learn SyncNet to predict class, it learns participant-specific patterns
- The confusion matrix on the right takes the extracted features and predicts participant ID
  - High accuracy *despite not being trained for this task*
- Can we tell the network it shouldn't do this?



An Existing Approach: Domain Adversarial Neural Networks





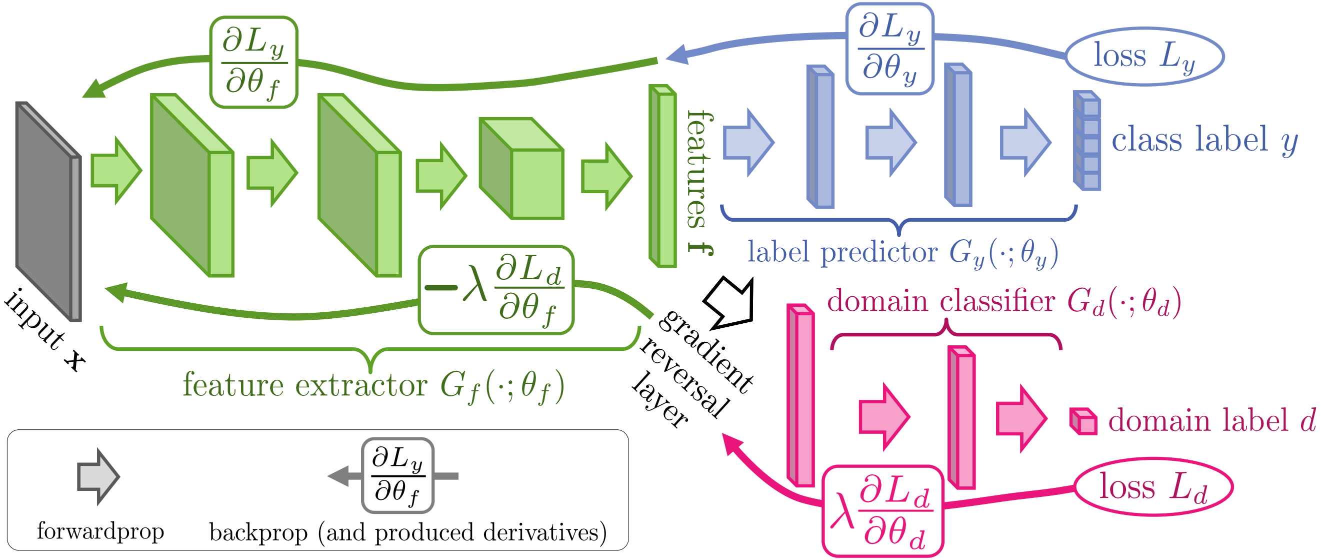
While the network is being learned:

- One classifier to predict label
- One classifier to predict domain
- The features should work well to predict the label *but trick the domain classifier*

Ganin et al, 2016

Removes participant-specific information in the feature space!

An Alternative Use of Adversarial Learning



Ganin et al, 2016

Adapting Between Domains

MNIST

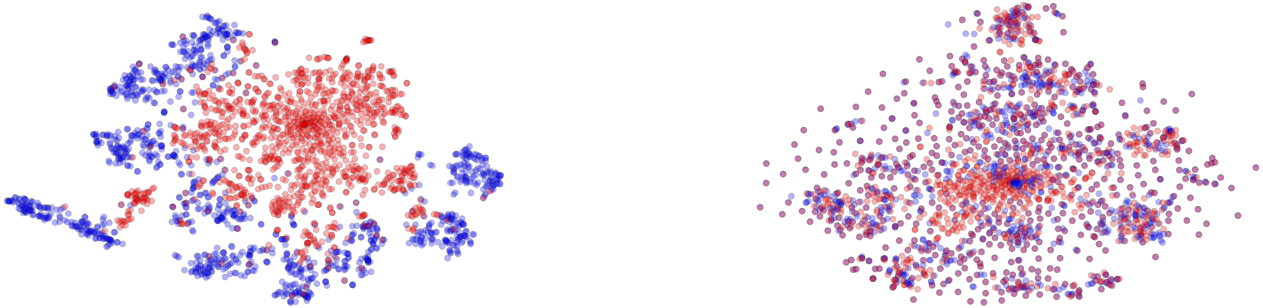
Source

Target

MNIST-M

Method	Source / Target	MNIST / MNIST-M
Source only		.5225
SA (Fernando et al., 2013)		.5690 (4.1%)
DANN		**.7666** (52.9%)
Train on target		.9596

MNIST → MNIST-M: top feature extractor layer



(a) Non-adapted (b) Adapted

Does this help in EEG?

- Removes participant-specific information
- Learned features much less predictive of identity
- The assumption requires that all participants are the same in the feature space
  - Works well in images and text data
  - Bad assumption in medical data—every child is unique!

	Accuracy
Random Guessing	33.3
Dominant Class	41.1
Diff. Ent. + SVM	50.4
Power Spec. Density+SVM	49.9
MC-DCNN	58.4
SyncNet	**60.1**
SyncNet+DANN	58.7

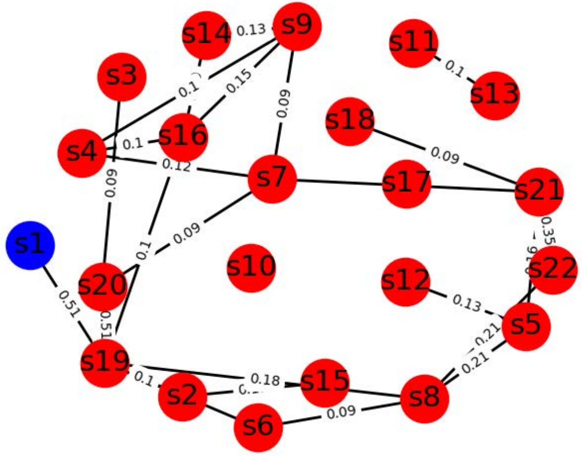
A Less Stringent Assumption

- Instead we require that everyone is similar to at least one other person, but not everyone!
- This can also be trained in an adversarial framework
- The network has the following properties:
  - The label prediction tries to predict well on the labels/outcomes
  - The domain classifier tries to predict which individual, but the loss is modified to only penalize if you can differentiate between similar individuals
  - The learned features try to do well on label prediction and trick the domain classifier

Li et al, NeurIPS 2018, AISTATS 2019

Learning Participant Relationships

- Instead, we want to learn relationships between participants
  - Should have similar features to a few similar participants
- As data size increases, we want to find cliques of similar individuals
- Large practically important gains on the Autism Center data



Li et. al., NeurIPS 2018, Li et. Al., AISTATS 2019

Does Multiple Domain Adaptation Help Us?

- Combining our previous interpretable neural network approach (SyncNet) with our Multiple Domain Matching Network (MDMN) yields significant gains
- Statistically significant improvement (p=.002, Wilcoxon Paired-Rank Test)
- This types of biomarkers can be used to explore how the brain is changing post-treatment

	Accuracy
Random Guessing	33.3
Dominant Class	41.1
Diff. Ent. + SVM	50.4
Power Spec. Density+SVM	49.9
MC-DCNN	58.4
SyncNet	**60.1**
SyncNet+DANN	58.7
SyncNet+MDMN	**67.8**

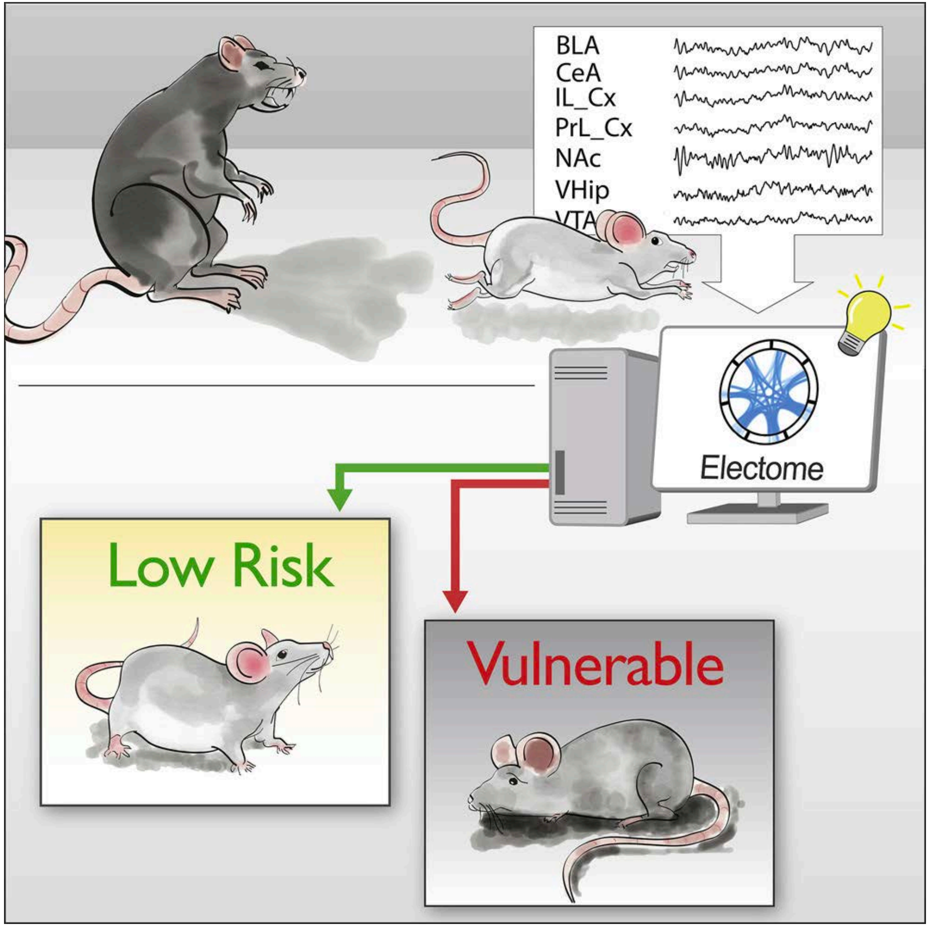
Comments

- Machine learning and deep learning techniques can help us get stronger scientific links to data
- “Little big data” structures are common in physiological data
- Important to consider the structure of the data

Extracting Neural Biomarkers for Stress Susceptibility

Similar methods can be used to break down neural signals into component parts (or *electrical connectomes*).

Used previously to discover biomarkers for susceptibility to stress (depression susceptibility biomarker).



Hultman et al, Cell 2018

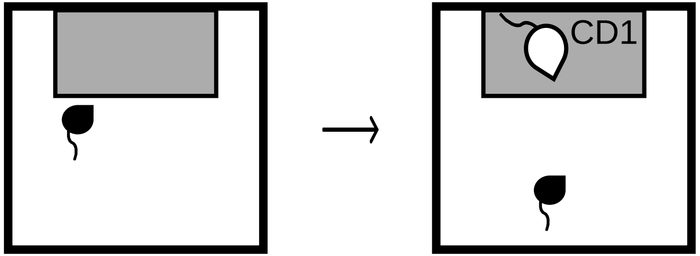
AUTOMATICALLY CAPTURING BEHAVIOR?

Why Behavior?

- Many studies in neuroscience examine behavioral responses
  - e.g., behavioral responses to treatment
- Essential component of many clinical diagnoses and outcomes
  - e.g., ASD, ADHD

Can Deep Learning Capture More Detailed Behavior?

- Behavioral protocols are tightly controlled
- Utility in direct understanding of the measurement
- Animals do much more than what is typically measured…



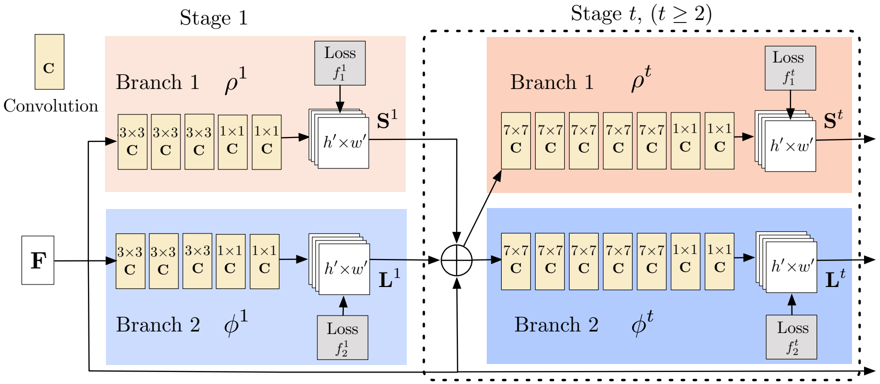
Choice Interaction Test
Measures % of time that a test mouse chooses to be near another mouse.

Can we capture more detailed behavior?

Example: Pose Estimation

Pose Estimation is Built on CNNs

- Convolutional neural networks provide the backend for pose estimation
- Complicated extensions, but the convolutional layers are a key block

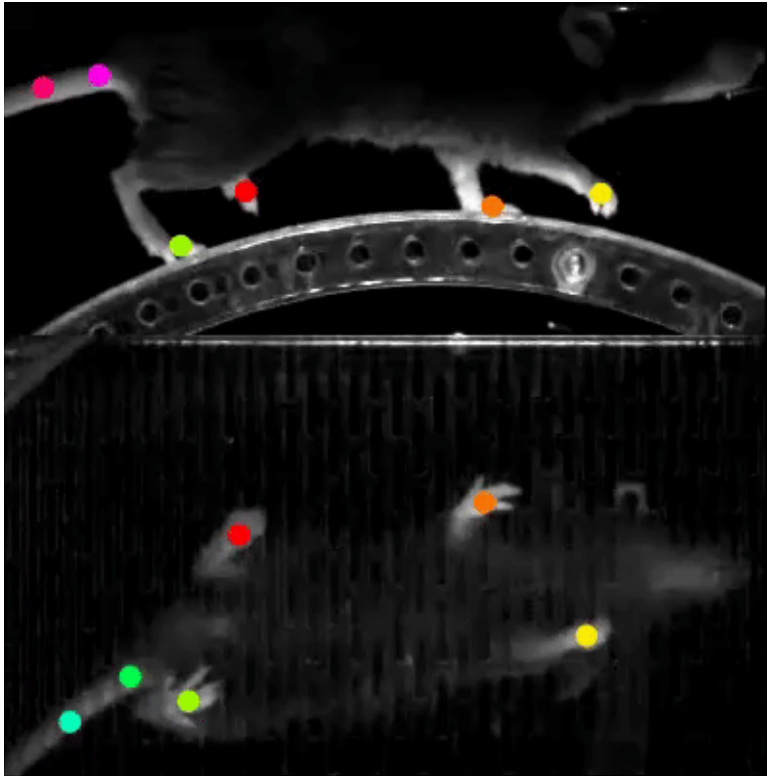


Network behind OpenPose based on repeating convolutional elements. From Cao et. al., 2017

Capturing Limbs on Animals?

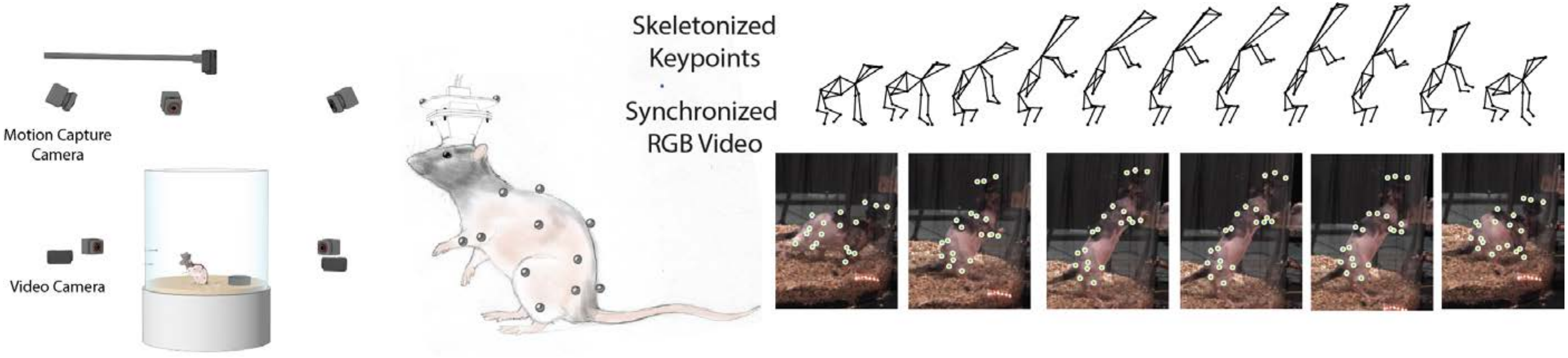
http://www.mousemotorlab.org/deeplabcut

Captures Fine-Grained Behavior



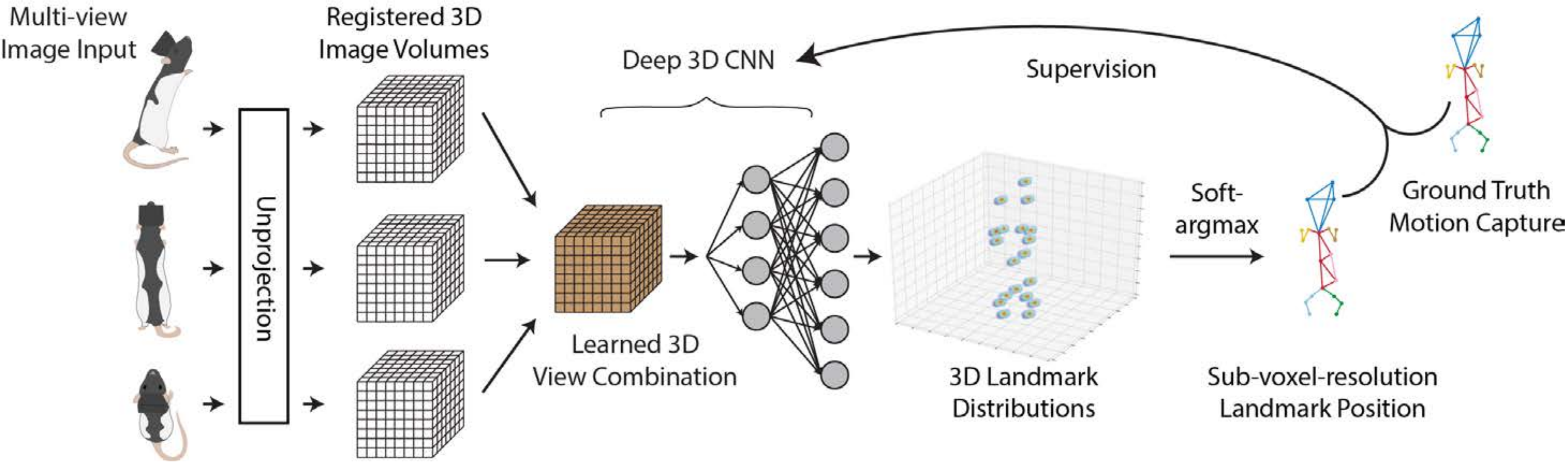
http://www.mousemotorlab.org/deeplabcut

Using Multiple Cameras



Dunn et al. Under review.

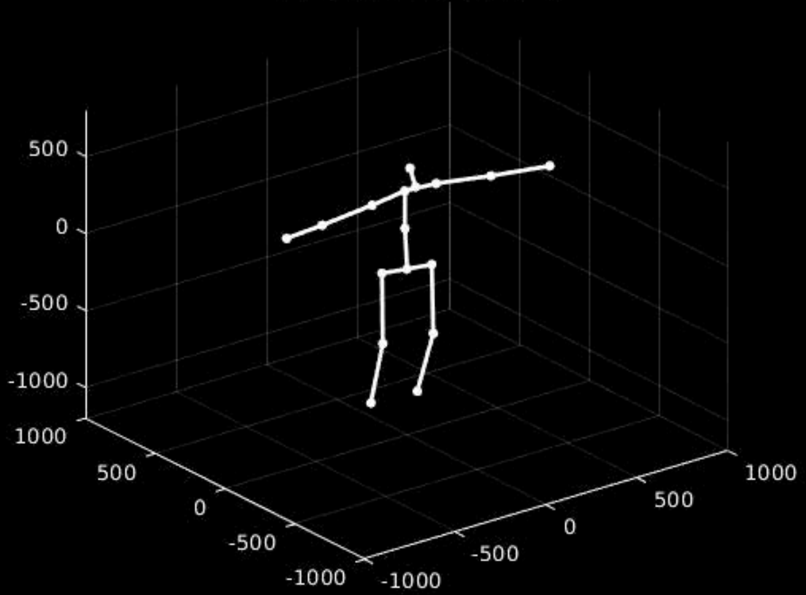
Using 3D CNNs



Dunn et al. Under review.

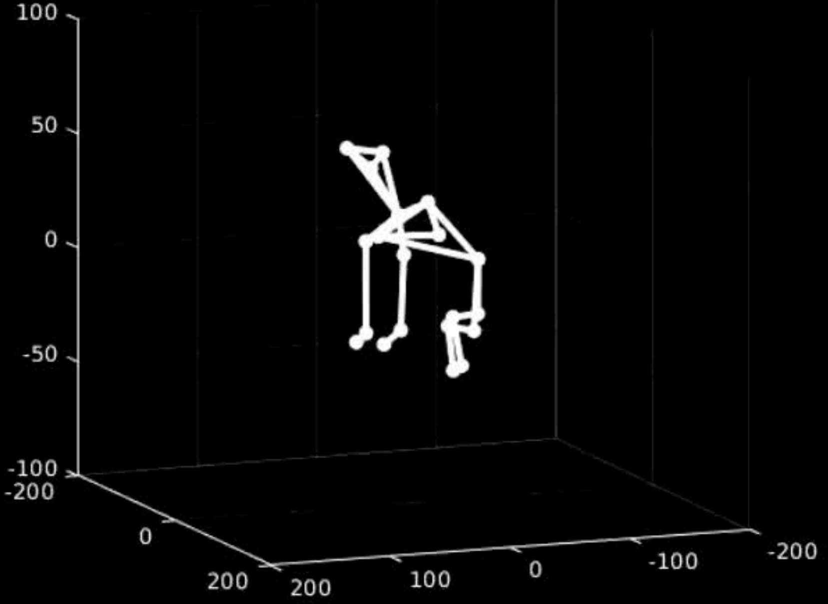
Video (1 of 4 views)

DANNCE 3D Prediction



Video (1 of 6 views)

DANNCE 3D Prediction



How to use this Information?

- More detailed information alone doesn’t lead to breakthroughs
- Need to make *sense* of the large-scale information

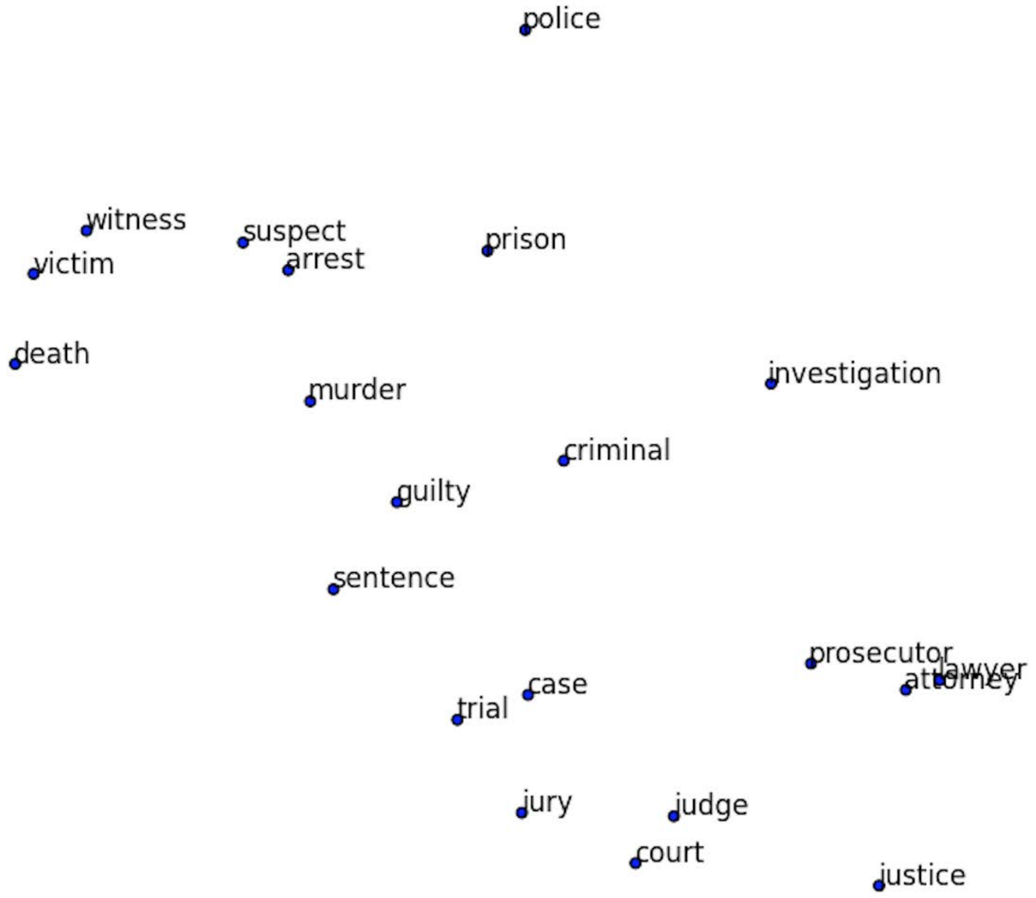
Recall: Word Geography

Here we show the learned geography of many different vocabulary words.

Recall: Word Geography

Here we show the learned geography of many different vocabulary words.

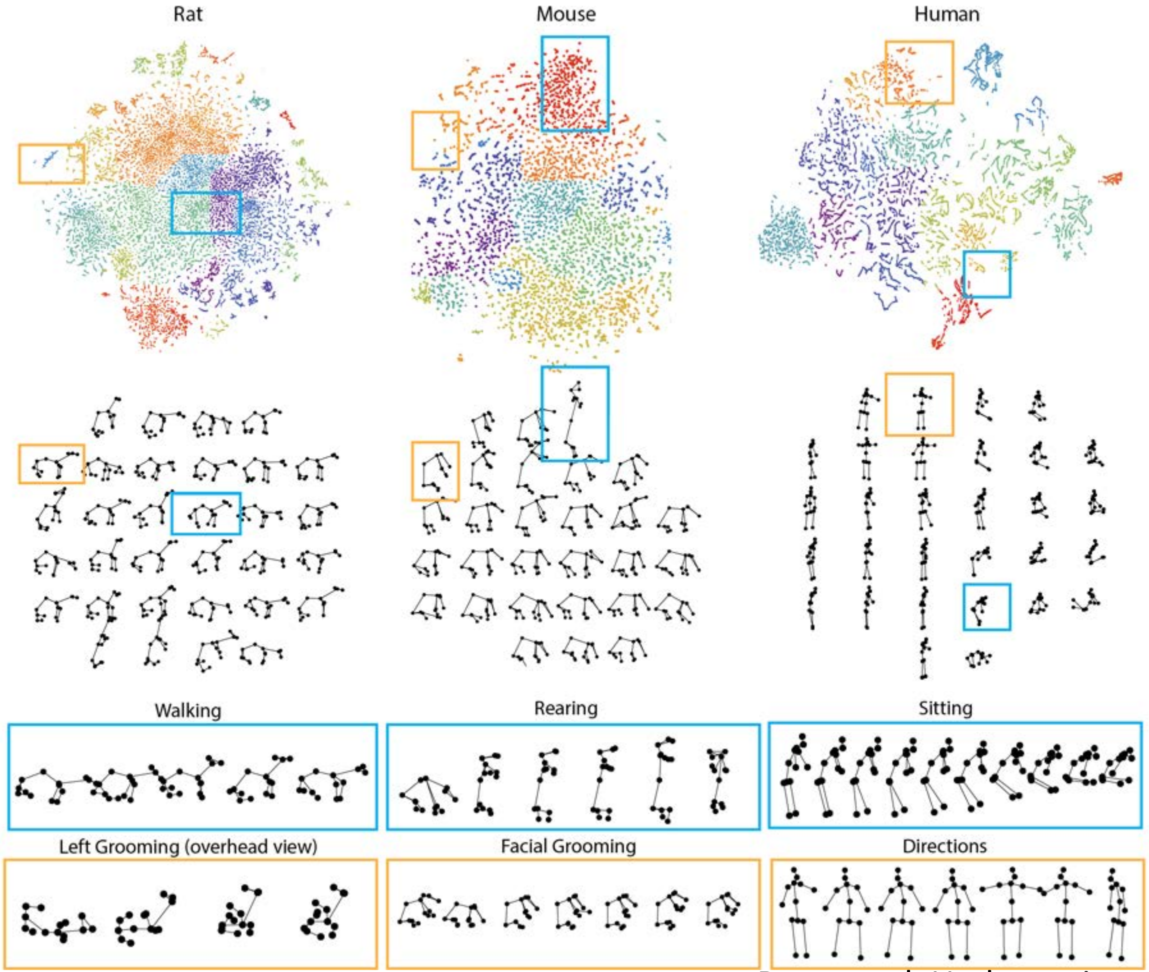
A key idea was that we could learn methods of embedding similar items into nearby space.



Learning a Behavior Map

Much like our learned map for word embeddings, we can learn a map of behaviors.

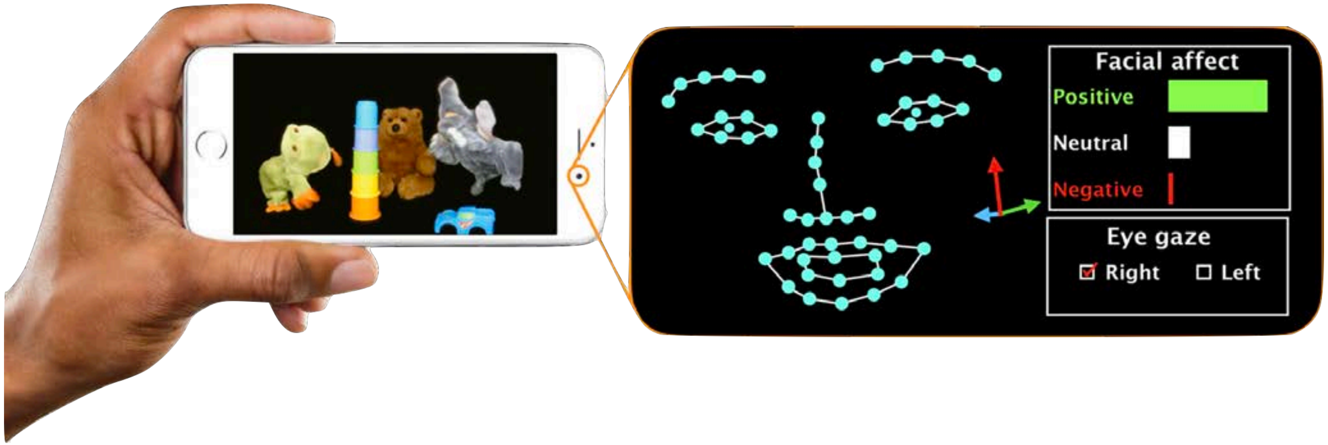
Similar behaviors appear at the same location in space, so we can visualize traces



Dunn et al. Under review.

What about clinical data?

Can use Smartphones or Tablets to Measure Behavioral Responses



Geri Dawson

Guillermo Sapiro

Currently being researched for potential in screening for Autism Spectrum Disorder

Comments

- Computational approaches to behavior are under development
- Lots of potential to improve data available in the clinic
- Can build stronger representations of neurobehavioral patterns

QUICK NOTE ON THE CNN AND THE BRAIN

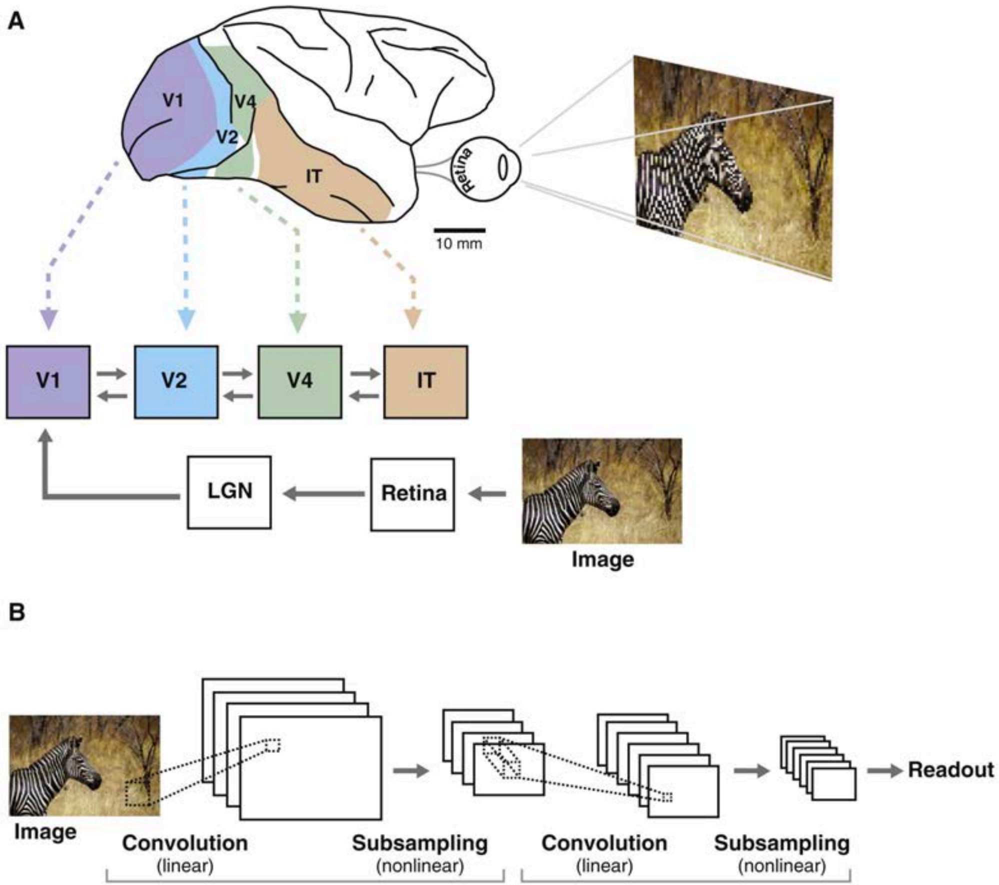
What does a cat have to do with convolutional neural networks?

- The cat visual cortex is often given credit for inspiring deep learning…

The Brain Inspired the Convolutional Neural Network

The visual cortex is roughly structured in layers, with each layer corresponding to more and more complex representations.

The famous Hubel and Wiesel results on the cat visual cortex served as a *loose* inspiration for the convolutional neural network.



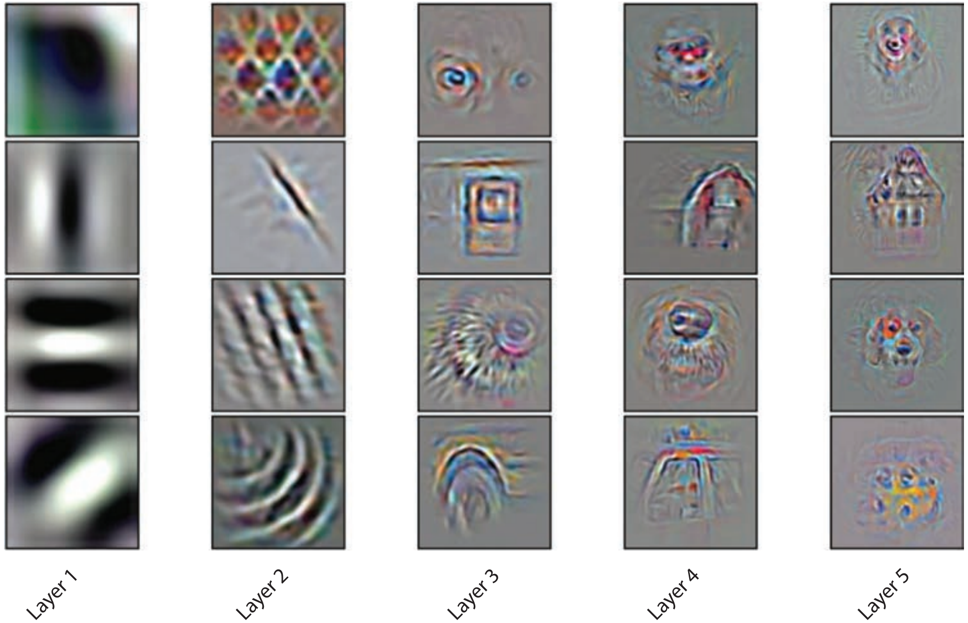
Cox and Dean, Current Biology 2014

The Learned Filters on Convolutional Neural Networks Qualitatively Match The Brain

The visual cortex is roughly structured in layers, with each layer corresponding to more and more complex representations.

The first layer mimics the “simple cell” patterns found by Hubel and Wiesel.

Later layers mimic how how deeper layers capture more complex structures.



Kriegeskorte, Ann. Rev. Vis. Sci. 2015

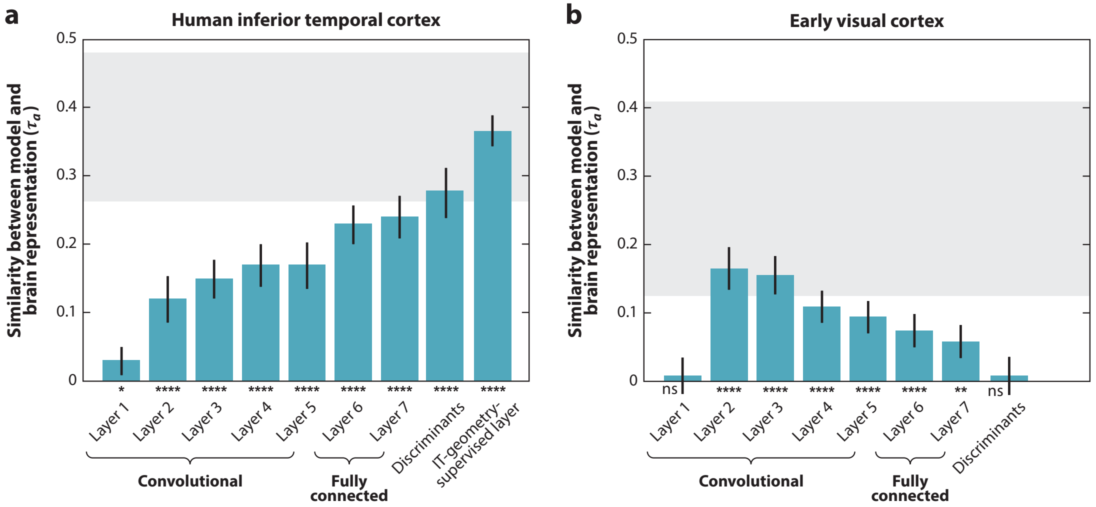
The Learned Filters on Convolutional Neural Networks Qualitatively Match The Brain

The visual cortex is roughly structured in layers, with each layer corresponding to more and more complex representations.

The first layer mimics the “simple cell” patterns found by Hubel and Wiesel.

Later layers mimic how how deeper layers capture more complex structures.

Can predict neural firing in different layers; early convolutional layers predict early visual layers



Kriegeskorte, Ann. Rev. Vis. Sci. 2015

Conclusions

- Machine learning is increasingly used in scientific and medical analysis
- *Many* ways to use:
  - Automate tasks
  - Bring new information to the table
  - Improved statistical predictions
- Must consider the scientific question and evaluation

Some Additional Comments

arXiv.org > cs > arXiv:1906.01998

Search...

Help | Advar

Computer Science > Machine Learning

The Secrets of Machine Learning: Ten Things You Wish You Had Known Earlier to be More Effective at Data Analysis

Cynthia Rudin, David Carlson

(Submitted on 4 Jun 2019)

Despite the widespread usage of machine learning throughout organizations, there are some key principles that are commonly missed. In particular: 1) There are at least four main families for supervised learning: logical modeling methods, linear combination methods, case-based reasoning methods, and iterative summarization methods. 2) For many application domains, almost all machine learning methods perform similarly (with some caveats). Deep learning methods, which are the leading technique for computer vision problems, do not maintain an edge over other methods for most problems (and there are reasons why). 3) Neural networks are hard to train and weird stuff often happens when you try to train them. 4) If you don't use an interpretable model, you can make bad mistakes. 5) Explanations can be misleading and you can't trust them. 6) You can pretty much always find an accurate-yet-interpretable model, even for deep neural networks. 7) Special properties such as decision making or robustness must be built in, they don't happen on their own. 8) Causal inference is different than prediction (correlation is not causation). 9) There is a method to the madness of deep neural architectures, but not always. 10) It is a myth that artificial intelligence can do anything.