The Algonauts Project

Explaining the Human Visual Brain

Workshop and Challenge

Dates: July 19-20, 2019

Place: MIT, Cambridge, MA

algonaunts.csail.mit.edu
Team and Sponsors

Team Leader: Radoslaw Cichy
Research Group Leader, Freie Universität Berlin

Team Leader: Aude Oliva
Principal Research Scientist, MIT

Team Leader: Gemma Roig
Assistant Professor, SUTD

Alex Andonian
Research Assistant, MIT

Kshitij Dwivedi
PhD Student, SUTD

Benjamin Lahner
Research Assistant, MIT

Kandan Ramakrishnan
Postdoctoral Researcher, MIT

Yalda Mohsenzadeh
Postdoctoral Researcher, MIT

Alex Lascelles
Research Assistant, MIT

Fern Keniston
Program Coordinator and Assistant to the Directors, MIT

Samantha Smiley
Administrative Assistant, MIT

Kim Martineau
Communications Officer, MIT

NSF
MIT-IBM Watson AI Lab
MIT Quest for Intelligence
The Algonauts Project

Explaining the Human Visual Brain

Workshop and Challenge

Dates: July 19-20, 2019
Place: MIT, Cambridge, MA

Workshop for Students Day

19 July 2019
<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>12:30 pm - 1:00 pm</td>
<td>Registration / Welcome</td>
</tr>
<tr>
<td>1:00 pm - 2:00 pm</td>
<td>Introduction to Neural Networks</td>
</tr>
<tr>
<td></td>
<td>Gemma Roig</td>
</tr>
<tr>
<td>2:00 pm - 2:15 pm</td>
<td>BREAK</td>
</tr>
<tr>
<td>2:15 pm - 3:15 pm</td>
<td>Introduction to Brain Imaging: fMRI and MEG/EEG</td>
</tr>
<tr>
<td></td>
<td>Yalda Mohsenzadeh</td>
</tr>
<tr>
<td>3:15 pm - 3:30 pm</td>
<td>BREAK</td>
</tr>
<tr>
<td>3:30 pm - 4:30 pm</td>
<td>Comparing Brains and DNNs: Methods and Findings</td>
</tr>
<tr>
<td></td>
<td>Martin Hebart</td>
</tr>
<tr>
<td>4:30 pm - 4:45 pm</td>
<td>BREAK</td>
</tr>
<tr>
<td>4:45 pm - 5:45 pm</td>
<td>Comparing Brains and DNNs: Theory of Science</td>
</tr>
<tr>
<td></td>
<td>Radoslaw Cichy</td>
</tr>
<tr>
<td>5:45 pm - 6:00 pm</td>
<td>Summary</td>
</tr>
</tbody>
</table>
The Algonauts Project
Explaining the Human Visual Brain

<table>
<thead>
<tr>
<th>20 July Schedule</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:30 am – 9:00 am</td>
<td>Breakfast</td>
</tr>
<tr>
<td>9:00 am – 9:15 am</td>
<td>Introduction by Radoslaw Cichy</td>
</tr>
<tr>
<td>9:15 am – 9:35 am</td>
<td>Matt Botvinick</td>
</tr>
<tr>
<td>9:35 am – 9:55 am</td>
<td>Aude Oliva</td>
</tr>
<tr>
<td>9:55 am – 10:15 am</td>
<td>Thomas Naselaris</td>
</tr>
<tr>
<td>10:15 am – 11:00 am</td>
<td>Posters and Coffee</td>
</tr>
<tr>
<td>11:00 am – 11:20 am</td>
<td>David Cox</td>
</tr>
<tr>
<td>11:20 am – 11:40 am</td>
<td>James DiCarlo</td>
</tr>
<tr>
<td>11:40 am – 12:00 pm</td>
<td>Kendrick Kay</td>
</tr>
<tr>
<td>12:00 pm – 1:30 pm</td>
<td>Lunch on Your Own</td>
</tr>
<tr>
<td>1:30 pm – 1:50 pm</td>
<td>Introduction to the Algonauts Challenge by Radoslaw</td>
</tr>
<tr>
<td>1:50 pm – 2:50 pm</td>
<td>Invited Talks: Challenge Winners</td>
</tr>
<tr>
<td>2:50 pm – 3:30 pm</td>
<td>Posters and Coffee</td>
</tr>
<tr>
<td>3:30 pm – 3:50 pm</td>
<td>Talia Konkle</td>
</tr>
<tr>
<td>3:50 pm – 4:10 pm</td>
<td>Nikolaus Kriegeskorte</td>
</tr>
<tr>
<td>4:10 pm – 4:30 pm</td>
<td>Jack Gallant</td>
</tr>
<tr>
<td>4:30 pm – 5:00 pm</td>
<td>Panel Discussion with Speakers</td>
</tr>
<tr>
<td>5:00 pm – 6:30 pm</td>
<td>Reception</td>
</tr>
</tbody>
</table>
Introduction to Deep Neural Networks

Tutorial

Gemma Roig

The Algonauts Project

Explaining the Human Visual Brain

Workshop and Challenge

Dates: July 19-20, 2019

Place: MIT, Cambridge, MA
Overview

- Introduction
- Artificial Neural Networks
- Computational Models of Object Recognition
- Artificial Neural Networks for Object Recognition
- Applications
"Can machines think?"
Recognition

Object recognition  What is in the image?

Bike

Train

Bird
Recognition

We want the algorithms to **learn** to do object recognition given examples of object categories.

**Training phase:**
The model learns with examples.

**Testing phase:**
Automatic labelling of instances never seen before by the algorithm.

There are different modalities of supervision: fully supervised, unsupervised, semi-supervised, etc.
Overview

- Introduction
- Artificial Neural Networks
  - Computational Models of Object Recognition
  - Artificial Neural Networks for Object Recognition
- Applications
Computational Principles

Simplified neuroscience: a neuron computes a dot product between its inputs and the synaptic weights

\[ \langle x, t \rangle = \sum_{i=1}^{n} x_i t_i \]
Simple Perceptron

F. Rosenblatt 1957

\[ \text{Out} = \text{sgn} \left( \sum_{i=0}^{n} x_i w_i \right) \]

One layer NN

Image credit: missinglink.ai
Perceptron

Types of Nonlinearities

Step function
\[ f(x) = \begin{cases} 
0 & : x < 0 \\
1 & : x \geq 0
\end{cases} \]

Linear Rectifier (ReLu)
\[ f(x) = \begin{cases} 
0 & : x < 0 \\
x & : x \geq 0
\end{cases} \]

Sigmoid
\[ \sigma(x) = \frac{1}{1 + e^{-x}} \]

etc.
The Perceptron Learning Rule

Given training samples $\{x_i, y_i\}_{\forall i}$

$x_i$ -> input of example $i$, $y_i$ -> groundtruth target of example $i$
The Perceptron Learning Rule

Given training samples \( \{x_i, y_i\} \forall i \)

- \( x_i \rightarrow \) input of example \( i \),
- \( y_i \rightarrow \) groundtruth target of example \( i \)

**Initialization:**

Initialize the weights \( w \) to 0 or small random numbers.
The Perceptron Learning Rule

Given training samples \( \{x_i, y_i\} \forall i \)

- \( x_i \rightarrow \) input of example \( i \),
- \( y_i \rightarrow \) groundtruth target of example \( i \)

**Initialization:**

Initialize the weights \( w \) to 0 or small random numbers.

**Iterate:**

For each training sample \( x_i \):

1. Calculate the output value: \( out = sgn(\sum_{i=0}^{n} x_i w_i) \)
The Perceptron Learning Rule

Given training samples \( \{x_i, y_i\} \forall i \)

- \( x_i \rightarrow \) input of example \( i \),
- \( y_i \rightarrow \) groundtruth target of example \( i \)

Initialization:
Initialize the weights \( w \) to 0 or small random numbers.

Iterate:
For each training sample \( x_i \):

1. Calculate the output value: \( out = sgn \left( \sum_{i=0}^{n} x_i w_i \right) \)

2. Update the weights. \( w = w + \eta x_i (y_i - out) \)
Multi-layer Perceptron

Rumelhart et al. 1986

possibly many more layers

learning with back-propagation
Overview

- Introduction
- Artificial Neural Networks
- Computational Models of Object Recognition
  - Artificial Neural Networks for Object Recognition
- Applications
Hubel and Wiesel

Nobel prize (1981)
Hubel and Wiesel

LGN-type cells

Simple cells

Complex cells

(Hubel & Wiesel 1959)
The visual ventral stream

The ventral stream hierarchy: V1, V2, V4, IT
A gradual increase in the receptive field size, in the complexity of the preferred stimulus, in tolerance to position and scale changes

Kobatake & Tanaka, 1994
Two operations (~OR, ~AND): disjunctions of conjunctions

- Tuning operation (Gaussian-like, AND-like)
  \[ y = e^{-|x-w|^2} \]

- Simple units

- Max-like operation (OR-like)
  \[ y = \max\{x_1, x_2, \ldots\} \]

- Complex units

Each operation ~microcircuits of ~100 neurons
Overview

- Introduction
- Artificial Neural Networks
- Computational Models of Object Recognition
- Artificial Neural Networks for Object Recognition
- Applications
Convolutional Neural Networks (CNNs)

Convolutional assumption

LeCun et al. 98
Deep CNN (2012)

Learned with back propagation on GPUs  (7 days)
ImageNet dataset (1 million labeled images available)
Techniques to avoid overfitting
Learned with back propagation on GPUs (7 days)
Results on ImageNet

ImageNet competition results

Error rate

Year


AlexNet w/o DNN

Image credit: wikipedia
Results on ImageNet

ImageNet Classification Error (Top 5)

- 2011 (XRCE): 26.0
- 2012 (AlexNet): 16.4
- 2013 (ZF): 11.7
- 2014 (VGG): 7.3
- 2014 (GoogLeNet): 6.7
- Human: 5.0
- 2015 (ResNet): 3.6
- 2016 (GoogLeNet-v4): 3.1

Image credit: von Zitzewitz, 2017
Object classification

AlexNet 12 | VGG 14 | GoogLeNet 14 | ResNet 15
For each image patch $p$, $x^p$ and kernel $k$, $w^k$

\[ i=4, j=4, c=3 \]
\[ \sum_{i=1, j=1, c=1} x^p_{i,j,c} w^k_{i,j,c} \]
Convolution in Deeper Layers

$W_1[3, 3, 4, 6]$

$W_2[2, 2, 6, 10]$

$W_2[1, 1, 10, \ldots]$

Width x height x channels x # k

Image credit: codelabs google
Max or Average Pooling
Max or Average Pooling

2 × 2 Max-Pool

Stride 2

Image credit: computersciencewiki.org
11 layers 8K weights

- conv 3x3 x 32
- conv 1x1 x 32
- maxpool
- conv 3x3 x 32
- conv 1x1 x 32
- maxpool
- conv 3x3 x 32
- conv 1x1 x 32
- maxpool
- conv 3x3 x 16
- conv 1x1 x 8
- flatten
- softmax 5

192x192 x 3
192x192 x 32
192x192 x 32
96x96 x 32
96x96 x 32
48x48 x 32
48x48 x 32
24x24 x 32
24x24 x 32
12x12 x 16
12x12 x 8
1152
5 classes
Avoid Overfitting

- Architecture of the network as prior:
  - Convolutions
  - Non-linear activation, e.g., ReLU

- Use data augmentation in the training
  - Affine transformations

- Dropout

- Batch Normalization
Rectified Linear Unit

**ReLU (blue line)**

![Image]

Figure 1: A four-layer convolutional neural network with ReLUs (solid line) reaches a 25% training error rate on CIFAR-10 six times faster than an equivalent network with tanh neurons (dashed line). The learning rates for each net-
Avoid Overfitting

Dropout

- training phase:
  - remove stochastically hidden units

- Hidden units set to 0 with a probability (0.5, changes stochastically)

- Hidden units can not co-adapt to other hidden units

(a) Standard Neural Net
Avoid Overfitting

 Dropout

(a) Standard Neural Net
(b) After applying dropout.
Avoid Overfitting

- Dropout

  testing phase:
  all hidden units used

  - Multiply hidden layers by the dropout probability (0.5, not stochastic)

  - Better generalization
Learning

\[ L(y, f(x; w)) \]

back-propagation

\[ f(x; w) \]

stochastic gradient descent
Back-propagation

Learning based on iterating between:

1. Propagation
   1.1. Forward pass through NN
   1.2. Backward pass using partial derivatives

2. Weights updates

(stochastic gradient descend — with mini-batches)
Visualization of learned filters
Visualization of learned filters

http://people.csail.mit.edu/torralba/research/drawCNN/drawNet.html
Invariance Properties

Figure 5. Analysis of vertical translation, scale, and rotation invariance within the model (rows a-c respectively). Col 1: 5 example images undergoing the transformations. Col 2 & 3: Euclidean distance between feature vectors from the original and transformed images in layers 1 and 7 respectively. Col 4: the probability of the true label for each image, as the image is transformed.
Overview

- Introduction
- Artificial Neural Networks
- Computational Models of Object Recognition
- Artificial Neural Networks for Object Recognition

Applications
Applications

- Use a pre-trained CNN as a feature extractor
- Fine-tune on limited data
- Train from scratch on big data
Applications

- Use a pre-trained CNN as a feature extractor
- Fine-tune on limited data
- Train from scratch on big data
Object classification

AlexNet 12  VGG 14  GoogLeNet 14  ResNet 15
Object Detection

Faster Region CNN

Ren et al. 16
Object Detection
Applications

- Use a pre-trained CNN as a feature extractor
- Fine-tune on limited data
- Train from scratch on big data
Reducing the Semantic Gap in Saliency Prediction by Adapting Neural Networks
Applications

- Use a pre-trained CNN as a feature extractor
- Fine-tune on limited data
- Train from scratch on big data
## Places Recognition

10 million images with 400+ unique scene categories

[places2.csail.mit.edu](http://places2.csail.mit.edu)
Semantic Segmentation

Learning Deconvolution Network for Semantic Segmentation

Noh et al. 15
Semantic Segmentation

Noh et al. 15
Depth Map Prediction

Depth Map Prediction from a Single Image using a Multi-Scale Deep Network

Input Image  →  Depth  →  Normals  →  Labels

Eigen and Fergus 14
Multiple tasks predictions

Zamir et al., CVPR 2018
Applications

not only for vision...

Statistical parametric speech synthesis using deep neural networks

Zen et. al 13
Applications

End-to-End Deep Neural Network for Automatic Speech Recognition

phonemes recognition
Exploring vision tasks representation in the brain

Can we assess functions of a brain area by comparing the correlation of its responses with a large set of diverse models trained on different computer vision tasks?

Kshitij Dwivedi

Mick Bonner
Applications - Frameworks

- pyTorch
  - Python
  - [http://pytorch.org](http://pytorch.org)

- TensorFlow
  - Python, JavaScript
  - [https://www.tensorflow.org](https://www.tensorflow.org)

- Keras
  - Python, high level API on top of TensorFlow
  - [https://keras.io](https://keras.io)

- Caffe
  - C++ with Matlab and Python interfaces
  - [http://caffe.berkeleyvision.org](http://caffe.berkeleyvision.org)
<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>12:30 pm - 1:00 pm</td>
<td>Registration / Welcome</td>
</tr>
<tr>
<td>1:00 pm - 2:00 pm</td>
<td>Introduction to Neural Networks&lt;br&gt;<strong>Gemma Roig</strong></td>
</tr>
<tr>
<td>2:00 pm - 2:15 pm</td>
<td>BREAK</td>
</tr>
<tr>
<td>2:15 pm - 3:15 pm</td>
<td>Introduction to Brain Imaging: fMRI and MEG/EEG&lt;br&gt;<strong>Yalda Mohsenzadeh</strong></td>
</tr>
<tr>
<td>3:15 pm - 3:30 pm</td>
<td>BREAK</td>
</tr>
<tr>
<td>3:30 pm - 4:30 pm</td>
<td>Comparing Brains and DNNs: Methods and Findings&lt;br&gt;<strong>Martin Hebart</strong></td>
</tr>
<tr>
<td>4:30 pm - 4:45 pm</td>
<td>BREAK</td>
</tr>
<tr>
<td>4:45 pm - 5:45 pm</td>
<td>Comparing Brains and DNNs: Theory of Science&lt;br&gt;<strong>Radoslaw Cichy</strong></td>
</tr>
<tr>
<td>5:45 pm - 6:00 pm</td>
<td>Summary</td>
</tr>
</tbody>
</table>
The Algonauts Project

Explaining the Human Visual Brain

Workshop and Challenge

Dates: July 19-20, 2019

Place: MIT, Cambridge, MA

algonaughts.csail.mit.edu
<table>
<thead>
<tr>
<th>20 July Schedule</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:30 am – 9:00 am</td>
<td>Breakfast</td>
</tr>
<tr>
<td>9:00 am – 9:15 am</td>
<td>Introduction by Radoslaw Cichy</td>
</tr>
<tr>
<td>9:15 am – 9:35 am</td>
<td>Matt Botvinick</td>
</tr>
<tr>
<td>9:35 am – 9:55 am</td>
<td>Aude Oliva</td>
</tr>
<tr>
<td>9:55 am – 10:15 am</td>
<td>Thomas Naselaris</td>
</tr>
<tr>
<td>10:15 am – 11:00 am</td>
<td>Posters and Coffee</td>
</tr>
<tr>
<td>11:00 am – 11:20 am</td>
<td>David Cox</td>
</tr>
<tr>
<td>11:20 am – 11:40 am</td>
<td>James DiCarlo</td>
</tr>
<tr>
<td>11:40 am – 12:00 pm</td>
<td>Kendrick Kay</td>
</tr>
<tr>
<td>12:00 pm – 1:30 pm</td>
<td>Lunch on Your Own</td>
</tr>
<tr>
<td>1:30 pm – 1:50 pm</td>
<td>Introduction to the Algonauts Challenge by Radoslaw</td>
</tr>
<tr>
<td>1:50 pm – 2:50 pm</td>
<td>Invited Talks: Challenge Winners</td>
</tr>
<tr>
<td>2:50 pm – 3:30 pm</td>
<td>Posters and Coffee</td>
</tr>
<tr>
<td>3:30 pm – 3:50 pm</td>
<td>Talia Konkle</td>
</tr>
<tr>
<td>3:50 pm – 4:10 pm</td>
<td>Nikolaus Kriegeskorte</td>
</tr>
<tr>
<td>4:10 pm – 4:30 pm</td>
<td>Jack Gallant</td>
</tr>
<tr>
<td>4:30 pm – 5:00 pm</td>
<td>Panel Discussion with Speakers</td>
</tr>
<tr>
<td>5:00 pm – 6:30 pm</td>
<td>Reception</td>
</tr>
</tbody>
</table>
The Algonauts Project

Explaining the Human Visual Brain

Workshop and Challenge

Dates: July 19-20, 2019

Place: MIT, Cambridge, MA

algonautes.csail.mit.edu