6.819 / 6.869: Advances in Computer Vision

Website: http://6.869.csail.mit.edu/fa15/

Aude Oliva

Lecture TR 9:30AM – 11:00AM
(Room 34-101)
Instructors & TAs

Aude Oliva
Principal Research Scientist
oliva@mit.edu
Office hours:
32-D432
Tuesday 11-12 pm
or
by appointment T/Th

Yusuf Aytar
Postdoctoral Researcher
yusuf@csail.mit.edu
Office hours:
Room TBD
Tuesday 5-6 pm

Aditya Khosla
EECS graduate student
khosla@csail.mit.edu
Office hours:
Room TBD
Monday 5-6 pm

Bolei Zhou
EECS graduate student
bolei@mit.edu
Office hours:
Room TBD
Wednesday 5-6 pm
Contacting us by email

- Put **course number in Subject line:**
  - 6.819 (or 6.869)
- Put **topic in subject line:**
  
  Ex: 6.819 Missing Tuesday lecture
  6.869 Meeting request: Tuesday 2 pm?
  6.819 Late for PS2: Interview
  6.869 Project Option 2: Suggestion
  6.819 Project Option 3: Literature survey
## Schedule at a glance

<table>
<thead>
<tr>
<th>#</th>
<th>Day</th>
<th>Date</th>
<th>Topic</th>
<th>Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Thursday</td>
<td>9/10/2015</td>
<td>Introduction</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Tuesday</td>
<td>9/15/2015</td>
<td>Early vision: Basic of Image processing I</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Thursday</td>
<td>9/17/2015</td>
<td>Early vision: Basic of Image processing II</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Tuesday</td>
<td>9/22/2015</td>
<td>Early vision: Basic of Image processing II</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Thursday</td>
<td>9/24/2015</td>
<td>Early vision: Image Statistics</td>
<td>PS1 given</td>
</tr>
<tr>
<td>6</td>
<td>Tuesday</td>
<td>9/29/2015</td>
<td>Early vision: Segmentation</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Thursday</td>
<td>10/1/2015</td>
<td>Early vision: Texture synthesis</td>
<td>PS2 given - PS1 due</td>
</tr>
<tr>
<td>8</td>
<td>Tuesday</td>
<td>10/6/2015</td>
<td>Mid level: Image formation, lenses</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Thursday</td>
<td>10/8/2015</td>
<td>Mid level: Basics 3D, Shape from X</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Thursday</td>
<td>10/15/2015</td>
<td>Mid level: Motion: continuous</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Tuesday</td>
<td>10/20/2015</td>
<td>Mid level: Motion: discrete</td>
<td>PS3 given - PS2 due</td>
</tr>
<tr>
<td>13</td>
<td>Tuesday</td>
<td>10/27/2015</td>
<td>Learning: Intro. to Deep Learning (CNN)</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Thursday</td>
<td>10/29/2015</td>
<td>Session on Project Topics</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Tuesday</td>
<td>11/3/2015</td>
<td>High level : Object and Scene Recognition I</td>
<td>Project summary due</td>
</tr>
<tr>
<td>16</td>
<td>Thursday</td>
<td>11/5/2015</td>
<td>High level : Object and Scene Recognition II</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Tuesday</td>
<td>11/10/2015</td>
<td>Applications: Deep Learning for Vision I</td>
<td>PS4 given - PS3 due</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11/11/2015</td>
<td>Tutorial for CNN</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Thursday</td>
<td>11/12/2015</td>
<td>Applications: Deep Learning for Vision II</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Tuesday</td>
<td>11/17/2015</td>
<td>Applications: Deep Learning for Vision III</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Thursday</td>
<td>11/19/2015</td>
<td>Applications: Image Retrieval</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Tuesday</td>
<td>11/24/2015</td>
<td>Tutorial: How to give a short talk</td>
<td>PS4 due</td>
</tr>
<tr>
<td>22</td>
<td>Tuesday</td>
<td>12/1/2015</td>
<td>Applications: Human and Artificial Visual Brains</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Thursday</td>
<td>12/3/2015</td>
<td>Project Presentation: Mini Places Challenge</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Tuesday</td>
<td>12/8/2015</td>
<td>Project presentation</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Thursday</td>
<td>12/10/2015</td>
<td>Project presentation</td>
<td>Final report due</td>
</tr>
</tbody>
</table>

32-D463 - Monday 5-6 pm
Assignments

• Problem sets (60%)
• Final project (40%)
  – Summary project (5%)
  – Final presentation (5%)
  – Research component of final project (30%)
• No exams or quizzes
Materials

• **Piazza**: for student collaboration and finding project groups. Instructors and TAs will not be active on Piazza.

• **Stellar**: for turning in late assignments and receiving grades

• Readings: see class website

Problem sets (60%)

• Four problem sets: 15% each
• Collaboration policy
  – Psets are due individually
  – Done individually but you can talk to people
  – Writing always individually
• Turn a printed version in class. Late due on Stellar.
• Up to 4 days late total, for the 4 Psets altogether
  (i.e. if you use all the 4 days on PS1 for instance, you have none left for the other PSets).
If you are late after that, the grade of the late PS will be zero.
Projects (40%)

Three Project Options
1) Summary of final project proposal (5%): 1 page (template)
   – Individually
   – Due the first week of November (earlier, better!)
2) Research component of final project (30%, template) and final presentation (5%).
   – Presentation (2-5 minutes each): Dec 3, 8, 10
   – Everybody presents.

You are welcome to come to our office hours to brainstorm and suggest your project ideas.
Summary of Project Proposal

• The project proposal should be one page maximum following this template:
  
  • **What is the problem/question** that you will be investigating?
  • **What are the most relevant readings?** (2-4 papers)
  • **What data will you use?**
  • **What method or algorithm will you use?**

• **How will you evaluate your results?**
  
  Qualitatively, what kind of results do you expect (e.g. plots or figures)
  
  Quantitatively, what kind of analysis will you use to evaluate and/or compare your results (e.g. what performance metrics or statistical tests)?
Project : A survey (individual)

• Select a topic (to be discussed with one of us)
• Select 10-12 papers: send the list
• Read the papers
• Write a 2500 words survey article (a survey template will be given).
• You can opt for that option, change from a coding project to the survey, at any moment before Thanksgiving
Project: Your own project

2-4 people

• **Applications/Models.** If you have access to a specific large image dataset (e.g. biology, engineering, physics, neuroscience) and a categorization task, you can apply models to this problem.

• From what you learn in class, you can choose a topic/question and propose an approach/model (including questions related to neuroscience).
Project: Mini Places Challenge

2-4 people – Challenge *to be announced*

---

**Introduction**

The goal of this challenge is to identify the scene category depicted in a photograph. The data for this task comes from the Places2 dataset, which contains 10+ million images belonging to 400+ unique scene categories. Specifically, the challenge data will be divided into 8.1M images for training, 20k images for validation and 381k images for testing coming from 401 scene categories. Note that there is a non-uniform distribution of images per category for training, ranging from 4,000 to 30,000, mimicking a more natural frequency of occurrence of the scene.

For each image, algorithms will produce a list of at most 5 scene categories in descending order of confidence. The quality of a labeling will be evaluated based on the label that best matches the ground truth label for the image. The idea is to allow an algorithm to identify multiple scene categories in an image given that many environments have multi-labels (e.g. a bar can also be a restaurant) and that humans often describe a place using different words (e.g. forest, path, forest, woods).

**Dates**

- **August 15, 2015**: Development kit, data, and evaluation software made available
- **November 13, 2015, 5pm PST**: Submission deadline
- **December 10, 2015**: Challenge results released
- **December 17, 2015**: Winner presents at ICCV 2015 Workshop

**Organizers**

- Aditya Khosla
- Bolei Zhou
- Agata Lapedriza
- Antonio Torralba
- Aude Oliva
Vision: High-Powered Engine
Brain dynamics of seeing
Human Visual System as a Model
Predictions:

- Type of environment: outdoor
- Semantic categories:
  - picnic_area: 0.74
  - yard: 0.13
Computation: learning from millions of instances
Two extremes of visual learning

Extrapolation problem
- Generalization
- Diagnostic features

Interpolation problem
- Correspondence
- Finding the differences

Number of training samples
80.000.000 images

75.000 non-abstract nouns from WordNet

Online image search engines

And after 1 year downloading images

80 million images

A. Torralba, R. Fergus, W.T. Freeman. PAMI 2008
The importance of having
Lots Of Images
What can we do with a good similarity metric and a lot of data?

The space of world images: As large databases become available, this opens the door to effective data driven methods.

Hays, Efros, Siggraph 2006
Russell, Liu, Torralba, Fergus, Freeman. NIPS 2007
The evolution of vision databases

- COIL-20
- PASCAL (2005)
- MNIST (1998)
- Object-centered datasets
- ImageNet (2009)
- 2 year old kid

# images
IMAGENET

22K categories and 14M images

- Animals
  - Bird
  - Fish
  - Mammal
  - Invertebrate
- Plants
  - Tree
  - Flower
  - Food
  - Materials
- Structures
  - Artifact
  - Tools
  - Appliances
  - Structures
- Person
  - Scenes
    - Indoor
    - Geological Formations
  - Sport Activities

Deng, Dong, Socher, Li, Li, & Fei-Fei, 2009

Slide from Fei-Fei Li & Andrej Karpathy
The Image Classification Challenge:
1,000 object classes
1,431,167 images

Top 5 categories
- Output: Scale
- T-shirt
- Steel drum
- Drumstick
- Mud turtle

Output:
- Scale
- T-shirt
- Giant panda
- Drumstick
- Mud turtle

Russakovsky et al. arXiv, 2014

Slide from Fei-Fei Li & Andrej Karpathy
The evolution of vision databases

**Object-centered datasets**
- COIL-20
- Caltech 101
- MNIST (1998)
- IMAGENET (2009)
- 2 year old kid

**Scene-centered datasets**
- PASCAL (2005)
- 15 scenes database (2006)
- SUN database (2010)
- places
- 8 scenes database (2001)

# images
- $10^3$
- $10^4$
- $10^5$
- $10^6$
- $10^7$
- $10^8$
- $10^9$
High coverage across categories

High diversity within category
Deep Learning in Computer Vision

Deep Convolutional Neural Network (CNN)

- stack of several layers of computations.
- computations include convolution, max-pooling, fully connecting, etc.

- First proposed in 1989, rediscovered to revolutionize computer vision in 2012.
Deep Learning for Object Recognition

- **ILSVRC**: large-scale recognition on ImageNet
  1000 object classes
  1.2 million training data (~1000 images per class)
  100 test images per class

places

Convolutional Neural Network

Convolution  Max pooling  Normalization  Fully connected

Layer 1  Layer 2  Layer 3  Layer 4  Layer 5  FC6  FC7  FC8

pool1  pool2  conv4  pool5

savannah  field  lake  ...  ...  kitchen
Places

Smart phone: places.csail.mit.edu
Predictions:

- **Type of environment**: outdoor
- **Semantic categories**: swimming_pool/outdoor:0.80,
- **SUN scene attributes**: man-made, nohorizon, naturallight, bathing, warm, directsun, sunny, swimming, diving, stillwater, openarea
- **Informative region for the category "swimming_pool_outdoor" is: