Instructor

Name: Jean Oh
(Preferred to be called “Jean”)
Office: NSH 4521
Email: jeanhoh@cmu.edu
Prefix subject line with [16-785]
Course objectives

• Learn how to develop intelligence for robots
  – Learn good old technologies that are still useful
  – Catch up with new technologies that have been published recently

• Learn how to start and complete a project from proposal to execution
Academic integrity

• CMU academic integrity policy

• Acknowledge if you receive help from others
About intellectual property

• Software is easy to steal
  – Copy & paste
  – Rewriting

• Acknowledge sources of information

• Learn to read license terms before using open-sourced software

• Never take credit for someone else’s work
Reasonable Person Principle (RPP)

• Base culture of CMU School of Computer Science
• Everyone gives/gets the benefit of doubt for trying to be reasonable
  o Everyone will be reasonable.
  o Everyone expects everyone else to be reasonable.
  o No one is special.
  o Do not be offended if someone suggests you are not being reasonable.
Extensions and late assignments

• Everything is due BEFORE class begins on due date
• 5 no-penalty late days: use them anyway you want except for final project presentation & report
• 50% deduction within 24 hours from due
• 0% after 24 hours past due
Grading

• Homework 30%
• Class participation 30%
• Class project 40%
Homework (30%)

• 10 (+1) reading assignments (3 pt each)
• 1 free pass; if used student must attend the class that discusses the papers
• For each paper write 1-2 sentences for:
  – Problem definition
  – Technical challenges
  – Summary of approach/main ideas
  – Evaluation methods and results
  – Discussion points
Class participation (30%)

• Present paper & lead review classes (20%)
  – At least two papers per student
  – Sign up on Canvas (Google doc)

• Active participation during class (10%)
  – Attending
  – Raising or answering insightful questions
  – Sharing additional literature review on related topic
Questions about class participation?
Class project (40%)

• Proposal (10%)
  – Page limits: 5  (1 extra page for references)
• Midterm report (10%)
  – Page limits: 5  (1 extra page for references)
• Final presentation (10%)
  – 20 minutes
• Final report (10%)
  – Page limits: 10
Project topics

• Requirement
  – Must be interdisciplinary
  – 1-3 members per team

• Examples
  – Vision + language
  – Planning + language
  – Vision + planning
  – Vision + language + planning
Class project: proposal (10%)

- Maximum 5 pages
- Project title (Use self-explanatory titles)
- Problem definition
- Technical challenges
- Proposed approach
- Milestone schedule (include time commitment of each member)
- Expected outcomes
- Team members
  - Contact information
  - Bio sketch for each member describing one's technical background and intended contributions to the project
- 10-minute team presentation in class
Class project: midterm report (10%)

• In-depth literature review on existing work
• Progress report (maximum 5 pages)
  – Detailed technical approach
  – Schedule update (what has/hasn’t been done)
  – Detailed plan for experiments
  – Preliminary results
• 10-minute team presentation in class
Project final report (10%)

• Maximum 10 pages
• Yes, it can be build from your midterm report but additionally include:
  – Abstract
  – Complete description of technical approach
  – Evaluation methods and/or experimental setup
  – Final results
  – Technical contributions
  – Limitations
  – Future directions
• Project website (Extra points)
Class project: final presentation (10%)

• 20 minutes per team
• Motivation
• Problem definition
• Existing work
• Proposed approach
• Results
• Future directions
Questions about projects?
Course topics

• Planning, scheduling, and learning
• Vision: image understanding
• Language: sequence representation
• Vision + language
• Vision + planning
• Vision + language + planning
## Schedule

<table>
<thead>
<tr>
<th>Date</th>
<th>Introduction</th>
<th>Intelligence design for problem solving</th>
<th>Reinforcement learning</th>
<th>Human guidance</th>
<th>Project proposal</th>
<th>IRL</th>
<th>Deep learning</th>
<th>Language-planning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/17</td>
<td>Introduction</td>
<td>1/24 Overview papers</td>
<td>1/31 Applications</td>
<td>2/7 Human guidance</td>
<td>2/14 IRL</td>
<td>2/21 Deep learning basics</td>
<td>2/28 Scheduling (Steve Smith)</td>
<td>3/7 Deep IRL</td>
</tr>
<tr>
<td></td>
<td>Image understanding</td>
<td>Sequence learning</td>
<td>Vision-language</td>
<td>Image synthesis, plan to language</td>
<td>Architecture</td>
<td>Evaluation</td>
<td>Project final presentations</td>
<td></td>
</tr>
<tr>
<td>3/19</td>
<td>CNN</td>
<td>3/26 RNN LSTM</td>
<td>4/2 Image captioning</td>
<td>4/9 GANs</td>
<td>4/16 Intelligence architecture</td>
<td>4/23 User study design (Minkyung Lee)</td>
<td>4/30 Project final presentations</td>
<td></td>
</tr>
</tbody>
</table>
Planning, Scheduling, & Learning

“Convert high-level specification of tasks from humans into low-level descriptions of how to [move]” (LaValle’06)

• AI planning
• Motion planning
• Reinforcement learning
• Inverse reinforcement learning
• Scheduling
Vision: image understanding

• Convolutional neural nets
• State of the art networks in 2012 – 2017
  – AlexNet’12, Inception’15, VGG’15, U-Net’15, DeepLab’17, DenseNet’17
Vision + Language

• Image region $\rightarrow$ keywords / labels
  – Image classification
  – Object detection
  – Scene classification
Image captioning

“boy is doing backflip on wakeboard.”
[Kaparthy & Fei-Fei ’15]

“robot is doing backflip in a garage.”
[Image: Boston Dynamics Atlas]
Vision + Language

• Vision $\rightarrow$ Language
  – Image captioning with domain knowledge
  – Describing images with various sentiments

• Language $\rightarrow$ Vision
  – Semantic understanding based on description, e.g., object detection, semantic map building
  – Image synthesis given language description
CAT
There is a cat sitting by the window. A kitten is looking out the window from the red brick house. The cat in the green window frame looks still as if it is in a painting. The cat is positioned perfectly in the center of the window pane, looking outside attentively. A little cat is sitting by the window of an old brick house. The wooden window pane looks painted over many times. A curtain is closed behind the cat, so nothing else can be seen from the outside.
Sketch demo

How Do Humans Sketch Objects?

Mathias Eitz*
TU Berlin

James Hays†
Brown University

Marc Alexa‡
TU Berlin

Figure 1: In this paper we explore how humans sketch and recognize objects from 250 categories – such as the ones shown above.

• Interactive recognition
• Sketch-based image retrieval
• Art recognition
Image-to-Image Translation with Conditional Adversarial Networks

Phillip Isola  Jun-Yan Zhu  Tinghui Zhou  Alexei A. Efros

Berkeley AI Research (BAIR) Laboratory, UC Berkeley
{isola,junyanz,tinghuiz,efros}@eecs.berkeley.edu

Isola et al., CVPR’17

• Scene classification
• Map generation
• Sketch-based image retrieval
• B&W to color
• Day to night

Figure 1: Many problems in image processing, graphics, and vision involve translating an input image into a corresponding output image. These problems are often treated with application-specific algorithms, even though the setting is always the same: map pixels to pixels. Conditional adversarial nets are a general-purpose solution that appears to work well on a wide variety of these problems. Here we show results of the method on several. In each case we use the same architecture and objective, and simply train on different data.
Language to image

Create composite sketch from description

A tall male in his mid-40s wearing a dark-colored hat…

[Google image search results for “composite sketch”]
Language to image

Creating art work from description [Songeun Lee, 2011]

https://vimeo.com/17456055

“It’s got quite a small head and quite a big body.”

“It has two pointy ears.”
More vision-language applications

Understanding and describing environments, people, and activities

Military

Fashion

Social robots

Law enforcement
Vision + planning

Crusher [Silver et al., IJRR 2010]
Vision + planning

Figure 3: **Left:** Raw perception data from Crusher, in the form of camera images and colorized LiDAR. **Right:** 2D costs derived from perception data, and a resulting planned path. Brighter pixels indicate higher cost.

Silver et al., “Learning from Demonstration for Autonomous Navigation in Complex Unstructured Terrain,” IJRR 2010
“Navigate left of the building to the traffic barrel that is behind the building.”
Plans made by humans

“Navigate left of the building to the traffic barrel that is behind the building.”
"Navigate left of the building to a traffic cone behind the building."

1. When the command is given, the robot sees only 4 front walls of the church building.
2. From these wall segments, a complete geometry of a building is predicted.
3. Symbols in TBS (building and traffic cone) are grounded. Because traffic cones haven’t been detected, the robot uses the contextual clue from the command, e.g., “behind the building” to select an unknown object behind the building.
4. Initial plan is generated, considering “left of building” spatial constraint.
5. The robot hasn’t found any traffic cones behind the building; it continues with the initial plan.
6. The robot detects a traffic cone and re-plans towards the correct goal.

[Oh et al., AAAI 2015]
Semantic Navigation (robot view)

Navigate left of the building to a traffic barrel behind the building

“Toward mobile robots reasoning like humans” in AAAI 2015,
Language + planning

• Incorporating human inputs into algorithmic planner
• Generating planning heuristics from manual
• Hypothesizing environment from description
“Stay off the grass; go to the back of the shed.”
Explaining decisions

• Describing plans in natural language
• Explaining the rationale behind decisions
  – I think that it is a cat because ...
Instruction following in manipulation

Joint work with Katharina Muelling, Tushar Chugh, Shiyu Dong, Bikram Hanzra, Tae-Hyung Kim, & William Seto

Using speech, even children can easily interact with robots.
Conversational mobile robots

Joint work with Matthew Wilson & Ralph Hollis
AI + Art

• GANGogh

https://github.com/rkjones4/GANGogh
Robotics + Art

[Niklas Roy, 2011]
Questions about course topics?
Homework & logistics

• Sign up for paper presentation on Canvas (Google doc)
• Read overview papers and write summary
• No office hour this week (email me for any questions)