

# EECS 4422/5323 Computer Vision

## Lecture 1: Introduction

Calden Wloka

4 September, 2019

# Outline

- Motivation
- Related Disciplines
- Computational Tractability
- Introduction to Course Topics
- Course Structure

# What is Computer Vision?



A digital image is nothing more than a structured matrix of numbers, and computer vision seeks to extract useful information from that input.

$$\begin{array}{c}
 [21,11,9], [11, 1, 0] \cdots [72,50,36] \\
 [4, 0, 0] \cdots \vdots \\
 \vdots \cdots \vdots \\
 \vdots \cdots \vdots \\
 [38,42,54] \cdots \cdots \cdots [12,11,19]
 \end{array}$$


- A tortoiseshell cat
- The cat is wearing a Santa costume
- There is a window in the background
- The cat looks somewhat concerned

# Computer Vision Goals

Common problem domains include:

- Extracting semantic information
  - Object recognition
  - Object localization, segmentation
  - Content understanding, visual question answering

# Computer Vision Goals

Common problem domains include:

- Extracting semantic information
  - Object recognition
  - Object localization, segmentation
  - Content understanding, visual question answering
- Supporting environment interaction
  - Localization and mapping
  - Obstacle detection and navigation
  - Augmented reality (AR) support

# Why Vision?

- Powerful sensory modality
  - Operates in many environments over long range
  - Captured by a minimally invasive, passive sensor
  - Good spatial and temporal resolution

# Why Vision?

- Powerful sensory modality
  - Operates in many environments over long range
  - Captured by a minimally invasive, passive sensor
  - Good spatial and temporal resolution
- Intuitive for most people
  - Humans are highly visual, so generally know what information is available in the visual domain
  - This is a double-edged sword - we lack introspection on many aspects of vision

# Why Vision?

- Powerful sensory modality
  - Operates in many environments over long range
  - Captured by a minimally invasive, passive sensor
  - Good spatial and temporal resolution
- Intuitive for most people
  - Humans are highly visual, so generally know what information is available in the visual domain
  - This is a double-edged sword - we lack introspection on many aspects of vision

It's important to note that vision is not always the best suited solution to a given problem!



# How to Approach Computer Vision

Computer vision is a highly interdisciplinary field.

- Biological Vision
  - Generally works well and provides a strong proof of concept
  - Psychology - behaviour
  - Physiology - functional anatomy

# How to Approach Computer Vision

Computer vision is a highly interdisciplinary field.

- Biological Vision
  - Generally works well and provides a strong proof of concept
  - Psychology - behaviour
  - Physiology - functional anatomy
- Physics
  - Important in understanding image formation
  - Optics - how light interacts with the world and forms an image
  - Geometry - images are 2D renderings of a 3D world

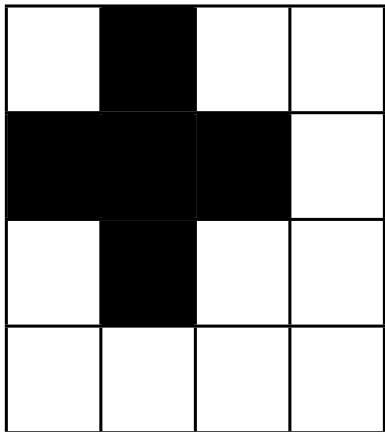
# How to Approach Computer Vision

Computer vision is a highly interdisciplinary field.

- Biological Vision
  - Generally works well and provides a strong proof of concept
  - Psychology - behaviour
  - Physiology - functional anatomy
- Physics
  - Important in understanding image formation
  - Optics - how light interacts with the world and forms an image
  - Geometry - images are 2D renderings of a 3D world
- Information Processing and Computational Theory
  - Provides a rigorous theoretical framework
  - Information processing - how to extract information from a signal
  - Computational theory - important for claims of generality, completeness, or real-time processing

## Claim: General Vision is Intractable

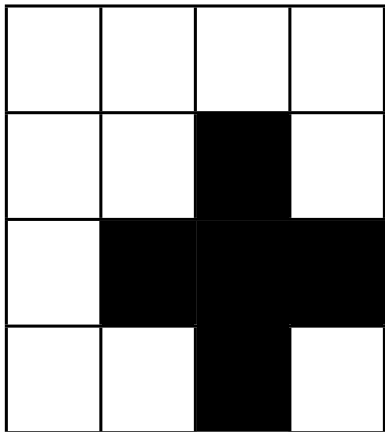
Take an  $4 \times 4$  binary pixel image. We can view each pixel as a possible component in building up a meaningful visual representation.



A + formed by a  
combination of five pixels.

## Claim: General Vision is Intractable

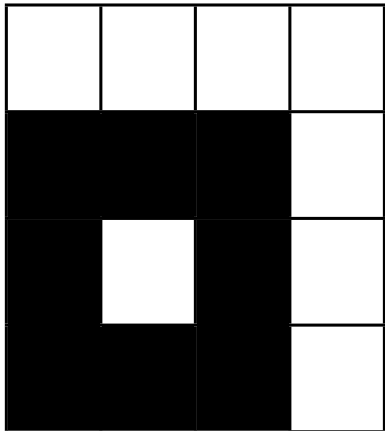
Take an  $4 \times 4$  binary pixel image. We can view each pixel as a possible component in building up a meaningful visual representation.



The same shape can occur in a different location.

## Claim: General Vision is Intractable

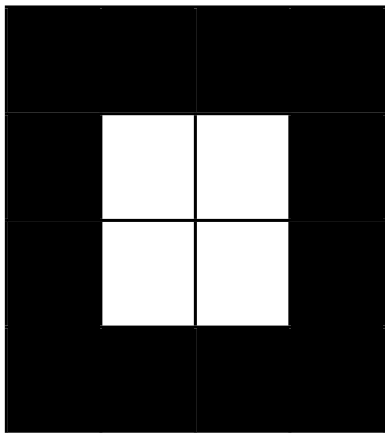
Take an  $4 \times 4$  binary pixel image. We can view each pixel as a possible component in building up a meaningful visual representation.



We could also represent an O using eight pixels.

## Claim: General Vision is Intractable

Take an  $4 \times 4$  binary pixel image. We can view each pixel as a possible component in building up a meaningful visual representation.



Or we could represent an  $O$  at a different spatial scale using twelve pixels.

# Claim: General Vision is Intractable

Pixels need not be contiguous to represent meaningful content (this is easier to see in a larger image).

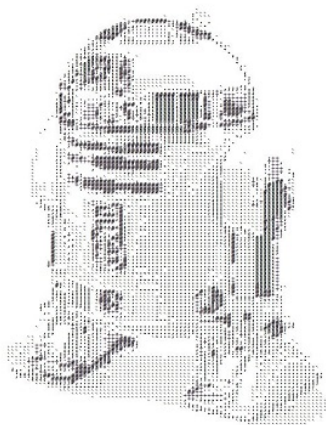


Image source: Original source unknown.



# Claim: General Vision is Intractable

Thus, for an arbitrary  $m \times n$  binary image, any subset of pixels may potentially represent a meaningful pattern.

- For a binary image, there are still  $2^{mn}$  possible combinations of pixels

# Claim: General Vision is Intractable

Thus, for an arbitrary  $m \times n$  binary image, any subset of pixels may potentially represent a meaningful pattern.

- For a binary image, there are still  $2^{mn}$  possible combinations of pixels
- For a modestly sized image, e.g.  $640 \times 480$ , there are more than  $10^{92476}$  possible combinations

# Claim: General Vision is Intractable

Thus, for an arbitrary  $m \times n$  binary image, any subset of pixels may potentially represent a meaningful pattern.

- For a binary image, there are still  $2^{mn}$  possible combinations of pixels
- For a modestly sized image, e.g.  $640 \times 480$ , there are more than  $10^{92476}$  possible combinations
- The fastest supercomputer in the world has achieved peak speeds of  $10^{18}$  operations per second

# Claim: General Vision is Intractable

Thus, for an arbitrary  $m \times n$  binary image, any subset of pixels may potentially represent a meaningful pattern.

- For a binary image, there are still  $2^{mn}$  possible combinations of pixels
- For a modestly sized image, e.g.  $640 \times 480$ , there are more than  $10^{92476}$  possible combinations
- The fastest supercomputer in the world has achieved peak speeds of  $10^{18}$  operations per second
- Thus, brute force analysis of this single image would take more than  $10^{92449}$  years (for reference, the age of the universe is  $\approx 1.4 \times 10^{10}$  years)

# But Animals Can See!

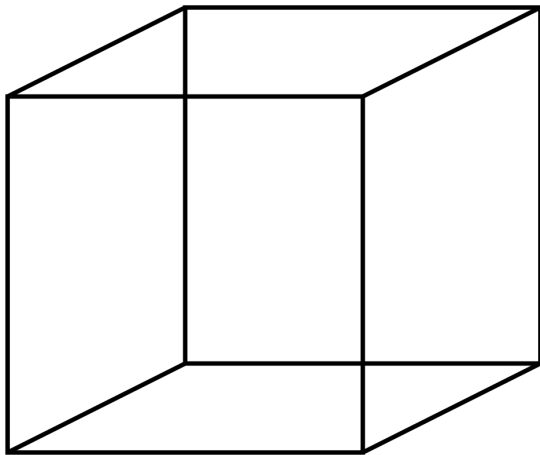
Animals can see, but their visual perception is tailored to solve a specific set of evolutionary imperatives. Heuristic approximations, physical constraints, and learned optimizations can all narrow the field of potential solutions.

For example, consider your interpretation of these images:



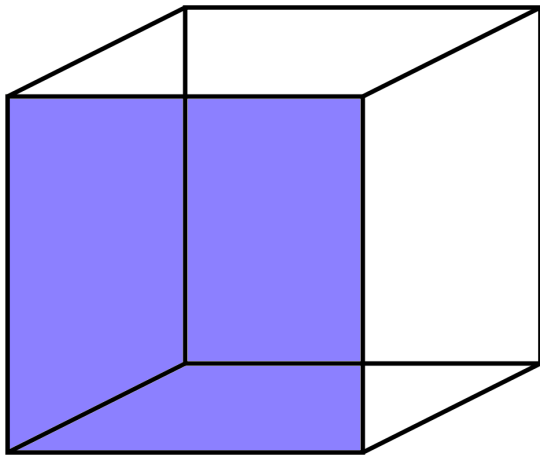
# Attention Tunes Vision Dynamically

A classic example of the affects of attention is the Necker Cube:



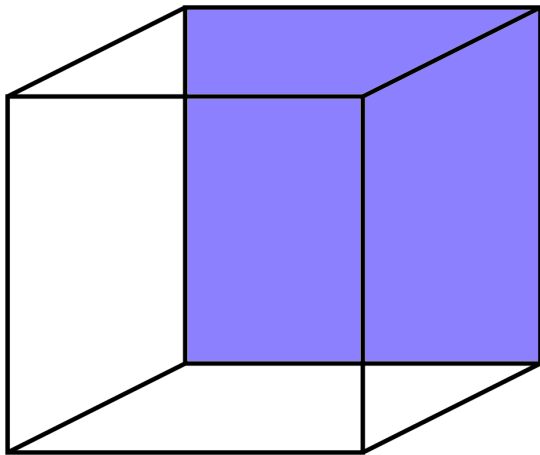
# Attention Tunes Vision Dynamically

The Necker Cube can be interpreted as having the lower face in front:



# Attention Tunes Vision Dynamically

Or the upper face:





# Task Matters

The visual task which we are currently trying to solve can dramatically alter what we see.

[Magic Trick Video](#)

# The Take-Away

Domain knowledge matters!

Despite the words of Krizhevsky *et al.* (2012),

“...our results can be improved simply by waiting for faster GPUs and bigger datasets to become available.”

There is no single solution. Understanding the fundamentals of vision is vital to knowing how best to approach a problem (even if you are applying learning).

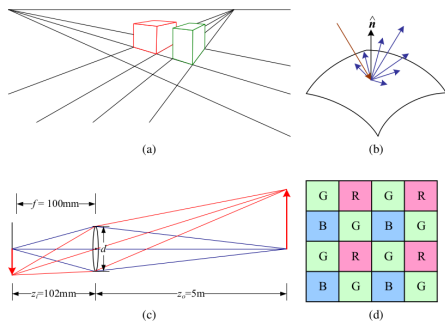
For a more complete discussion of the complexity of vision, see: Tsotsos, J. (1990). Analyzing vision at the complexity level. *Behavioral and Brain Sciences*, 13(3), 423-445. doi:10.1017/S0140525X00079577

# A Quick Overview of Course Topics

1. Image Formation
2. Image Representation
3. Feature Detection
4. Image Understanding
5. Stereopsis
6. Motion Analysis

# Image Formation

Image formation covers how the interaction of light with the environment is captured to form images.



Pertinent topics include:

- projection (3D to 2D)
- light interactions with surfaces
- lens optics
- sensor properties

Image source: Richard Szeliski, *Computer Vision: Algorithms and Applications*, Springer 2011

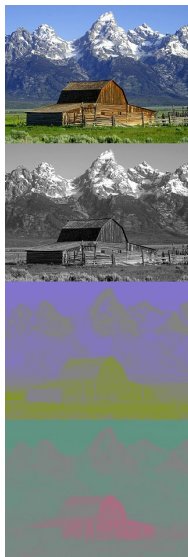
# Image Representation

Digital images are sets of numbers organized according to an agreed upon standard.

There are many standards, however, and different representations are better suited to certain tasks or provide different visualization options to observers. Common transformations between and within different representation spaces will be discussed.

On right: Example of channel separation for YCbCr image representation.

Image source: Public domain [Wikipedia example](#)



# Image Representation Topics

Topics in image representation include:

- Colour spaces and transformations

# Image Representation Topics

Topics in image representation include:

- Colour spaces and transformations
- Basic image processing

# Image Representation Topics

Topics in image representation include:

- Colour spaces and transformations
- Basic image processing
  - Point operators



# Image Representation Topics

Topics in image representation include:

- Colour spaces and transformations
- Basic image processing
  - Point operators
  - Neighbourhood operators

# Image Representation Topics

Topics in image representation include:

- Colour spaces and transformations
- Basic image processing
  - Point operators
  - Neighbourhood operators
- Frequency domain representation

# Image Representation Topics

Topics in image representation include:

- Colour spaces and transformations
- Basic image processing
  - Point operators
  - Neighbourhood operators
- Frequency domain representation
- Image pyramids

# Feature Detection

A *feature* is a general concept, and refers to some piece of information relevant to a particular application or reasoning process.

Some pertinent properties of features include:

- Binary vs. graded

# Feature Detection

A *feature* is a general concept, and refers to some piece of information relevant to a particular application or reasoning process.

Some pertinent properties of features include:

- Binary vs. graded
- Can be combined into *feature vectors* or hierarchically arranged

# Feature Detection

A *feature* is a general concept, and refers to some piece of information relevant to a particular application or reasoning process.

Some pertinent properties of features include:

- Binary vs. graded
- Can be combined into *feature vectors* or hierarchically arranged
- May vary in degree of abstraction, e.g.
  - An edge
  - A face
  - A bedroom

# Classic Features

Classic features typically are proscribed and have semantic meaning, such as edges, corners, and lines.



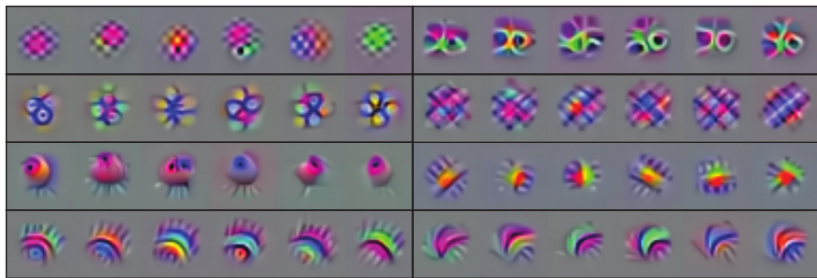
Input Image



Sobel Edge Detection

# Learned Features

Features may also be learned from data, though these features are often difficult to semantically interpret.



Example visualization of unit activation in the `conv3_2` layer of VGG-19.

Image source: Cadena *et al.*, (2018) Diverse feature visualizations reveal invariances in early layers of deep neural networks



# Image Understanding

Image understanding refers to the task of interpreting features to derive semantic understanding of the image contents. This is a notoriously difficult problem.



IN CS, IT CAN BE HARD TO EXPLAIN  
THE DIFFERENCE BETWEEN THE EASY  
AND THE VIRTUALLY IMPOSSIBLE.

Image source: [xkcd](#)

# Deep Learning Looks Impressive

In recent years deep learning has made impressive strides in many classically difficult problems like recognition, leading some to claim that we have solved (or are close to solving) this problem domain.

PARK or BIRD

Want to know if your photo is from a U.S. national park? Just drag it into the box to the left, and we'll tell you. We'll use the GPS embedded in your photo (if it's there) to see whether it's from a park, and we'll use our super-cool computer vision skills to try to see whether it's a bird (which is a hard problem, but we do a pretty good job at it).

To try it out, just drag any photo from your desktop into the upload box, or try dragging any of our example images. We'll give you your answers below!

Want to know more about PARK or BIRD, including why the heck we did this? Just click here for more info →

PARK???

No idea. There's no GPS info in that photo.

BIRD? YES

Dude, that is such a bird.

EXAMPLE PHOTOS

Photo credits

Image source: [Flickr](#) via [The Laughing Squid](#)

# Deep Learning Challenges

Deep learning performance can be extremely brittle, and often in ways which are difficult to predict.



A herd of sheep grazing on a lush green hillside

Tags: grazing, sheep, mountain, cattle, horse

Image source: [Janelle Shane, Nautilus, 2018](#)

# Deep Learning Challenges

If a situation not present in the training data occurs, output is highly unpredictable.



Left: A man is holding a dog in his hand  
Right: A woman is holding a dog in her hand  
Image: @5ouperSarah


Image source: [Janelle Shane, Nautilus, 2018](#)

For more examples and an interesting discussion, I recommend you read Shane's entire article, [This Neural Net Hallucinates Sheep](#).

# Challenging Images

Even state-of-the-art methods still struggle with challenges like camouflage.

clarifai [Demo](#) [Solutions](#) [Pricing](#) [Developer API](#) [Resources](#) [Contact Sales](#)



General [VIEW DOCS](#)

LANGUAGE  
English (en)

PREDICTED CONCEPT	PROBABILITY
no person	0.972
nature	0.959
snow	0.946
outdoors	0.935
winter	0.933
landscape	0.932
water	0.926
travel	0.917
desktop	0.883
cold	0.879
wild	0.865
sky	0.849
rock	0.847

Image source: [Clarifai Demo](#)

# More Challenges in Image Understanding

It's not just brittleness which remains an open problem, however. Many images can have multiple interpretations.

Take, for example, a picture of airplanes.



Image source: Original source unknown.

## More Challenges in Image Understanding

It's not just brittleness which remains an open problem, however. Many images can have multiple interpretations.

Take, for example, a picture of airplanes.



“Just two planes having a laugh.”

Image source: Original source unknown.

# Let's Get Philosophical for a Moment...

At some level, all images are abstractions.



Translation: "This is not a pipe"

Image source: [The Treachery of Images, Wikipedia](#).



# Let's Get Philosophical for a Moment...

At some level, all images are abstractions.



Translation: "This is not a pipe"

Image source: [The Treachery of Images, Wikipedia](#).

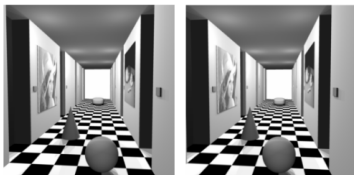
Magritte's explanation:  
The famous pipe. How people reproached me for it! And yet, could you stuff my pipe? No, it's just a representation, is it not? So if I had written on my picture "This is a pipe", I'd have been lying!

# Stereopsis

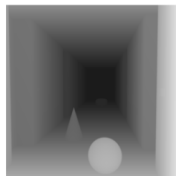
Stereopsis refers to the technique of using two (or more) images of a scene to recover 3D information.

Topics include:

- Camera calibration



Stereo pair



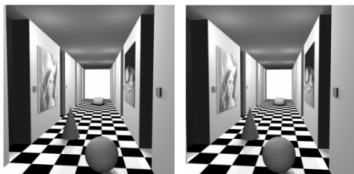
Range map

# Stereopsis

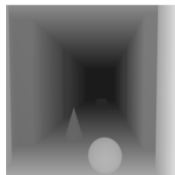
Stereopsis refers to the technique of using two (or more) images of a scene to recover 3D information.

Topics include:

- Camera calibration
- Calculating feature correspondence across views



Stereo pair



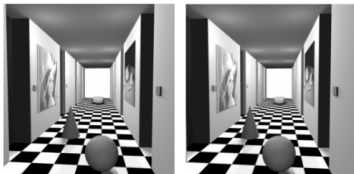
Range map

# Stereopsis

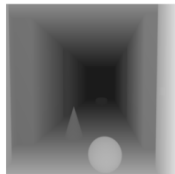
Stereopsis refers to the technique of using two (or more) images of a scene to recover 3D information.

Topics include:

- Camera calibration
- Calculating feature correspondence across views
- Dealing with noise or featureless surfaces



Stereo pair



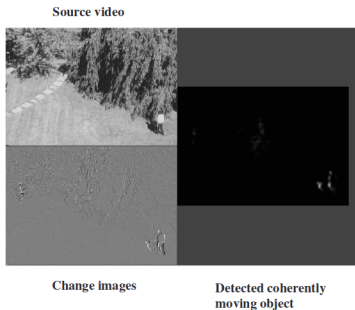
Range map

# Motion Analysis

Motion analysis refers to integrating information across time to detect coherent object motion, ego motion, or other dynamic scene properties.

Topics include:

- Spatiotemporal filters

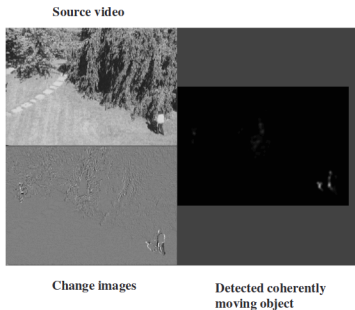


# Motion Analysis

Motion analysis refers to integrating information across time to detect coherent object motion, ego motion, or other dynamic scene properties.

Topics include:

- Spatiotemporal filters
- Optical flow

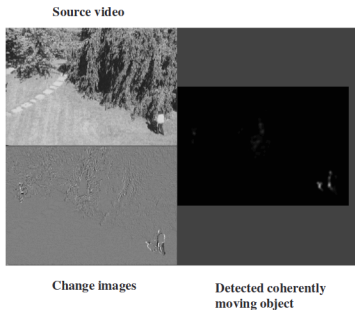


# Motion Analysis

Motion analysis refers to integrating information across time to detect coherent object motion, ego motion, or other dynamic scene properties.

Topics include:

- Spatiotemporal filters
- Optical flow
- Dealing with occlusion, appearance changes, or other challenging situations



# Basic Course Information

- The [course syllabus](#) is available on the website
- The website also includes additional useful information, and will be updated with announcements, lecture notes, and suggested readings
- This course is cross-listed; you should know if you are in 4422 or 5323
  - Lectures and lab content will be identical
  - Students in 5323 may be assigned additional readings, and will be asked additional assignment and midterm questions



# Email and Office Hours

You can reach me at [calden\(at\)eecs.yorku.ca](mailto:calden(at)eecs.yorku.ca).

Note that emails should include EECS 4422/5323 in the subject line to allow for priority response. I am more likely to respond quickly during business hours, but even then you may not get a response for up to 24 hours, so please plan accordingly.

Office hours will be held in LAS1004 from 15:00-16:00 on Mondays and Wednesdays (the hours before your lab), starting next week. You are encouraged to contact me ahead of time with questions so I can prepare more complete answers. If you need to meet at another time, email me to work out a meeting time and location.

Above all, remember to let me know if you are having issues. I can not fix things I don't know about!

# Grades

- 2 Assignments, 30% of final grade (15% each)

# Grades

- 2 Assignments, 30% of final grade (15% each)
- 1 Midterm, 30% of final grade
  - To be held in class October 28th
  - Let me know ASAP if you have a conflict (e.g. ICCV)

# Grades

- 2 Assignments, 30% of final grade (15% each)
- 1 Midterm, 30% of final grade
  - To be held in class October 28th
  - Let me know ASAP if you have a conflict (e.g. ICCV)
- 1 Project, 40% of final grade
  - White Paper (2%)
  - Proposal (8%)
  - Site Visit (5%)
  - Demo (7%)
  - Final Report (18%)

# Final Projects

The goal of the project component is twofold: to provide in-depth experience with a specific problem in computer vision, and to provide the opportunity for you to contribute to the field of computer vision.

Detailed information can be found on the [course website](#).

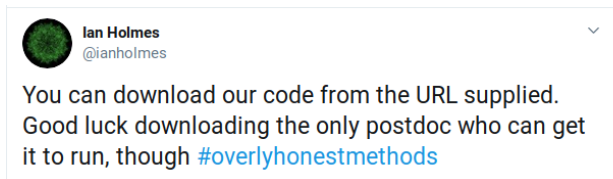
There are two primary project streams: *engineering* and *scientific*.

Please note that any student may pick either stream, regardless of program of study. Neither project is intended to be easier or harder, or to be more or less prestigious than the other. The only difference is on what manner of contribution each stream emphasizes.

# A Note About Project Groups

It is possible to work in groups of up to two people, but this must be discussed with me before the first project deadline and will necessitate an expansion of the project scope to justify the increased number of people. Groups must also clearly specify for each component of the project who is responsible for which part, and how the division of labour is equitable.

# Engineering Project Introduction



Research software is often messy and hacked together, and sometimes not even shared at all. This is an impediment to future research, and so a valuable contribution is the production of a (good) open source implementation of a model.

# Engineering Project Goal

The goal you should be aiming for with an engineering project is to be able to post a functional, stable, and as easy to use as possible implementation of a specific computer vision model to an open source repository (e.g. GitHub).

A good stretch goal is to also provide a detailed technical report of your implementation which can be posted to arXiv or even submitted to an open source software journal.



# Important Points about the Engineering Project

- The goal is to provide software which is useful to the larger research community. As such, code must be implemented in Python or C++.

# Important Points about the Engineering Project

- The goal is to provide software which is useful to the larger research community. As such, code must be implemented in Python or C++.
- The project may consist of either an implementation of a model which has not been released, or the conversion of a model which only exists in a closed format (e.g. MATLAB, or an executable binary) to one of the two accepted languages.

# Important Points about the Engineering Project

- The goal is to provide software which is useful to the larger research community. As such, code must be implemented in Python or C++.
- The project may consist of either an implementation of a model which has not been released, or the conversion of a model which only exists in a closed format (e.g. MATLAB, or an executable binary) to one of the two accepted languages.
- Any vision model which meets the above criteria may be selected so long as it contains a component which must be implemented which is semantically interpretable (*i.e.* the model is not just a single deep network).

# Important Points about the Engineering Project

- The goal is to provide software which is useful to the larger research community. As such, code must be implemented in Python or C++.
- The project may consist of either an implementation of a model which has not been released, or the conversion of a model which only exists in a closed format (e.g. MATLAB, or an executable binary) to one of the two accepted languages.
- Any vision model which meets the above criteria may be selected so long as it contains a component which must be implemented which is semantically interpretable (*i.e.* the model is not just a single deep network).
- Because this project stream emphasizes software implementation, an evaluation of the software quality will be included in the grade assigned to the final report.

# Scientific Project Introduction

There are many open computer vision questions which can be pursued. This project stream emphasizes hypothesis-driven scientific exploration of some topic in computer vision.

Possible topics include:

- Algorithm comparisons

# Scientific Project Introduction

There are many open computer vision questions which can be pursued. This project stream emphasizes hypothesis-driven scientific exploration of some topic in computer vision.

Possible topics include:

- Algorithm comparisons
  - Performance benchmarking

# Scientific Project Introduction

There are many open computer vision questions which can be pursued. This project stream emphasizes hypothesis-driven scientific exploration of some topic in computer vision.

Possible topics include:

- Algorithm comparisons
  - Performance benchmarking
  - Interesting patterns of success and failure

# Scientific Project Introduction

There are many open computer vision questions which can be pursued. This project stream emphasizes hypothesis-driven scientific exploration of some topic in computer vision.

Possible topics include:

- Algorithm comparisons
  - Performance benchmarking
  - Interesting patterns of success and failure
- Adaptive or pre-processing pipelines



# Scientific Project Introduction

There are many open computer vision questions which can be pursued. This project stream emphasizes hypothesis-driven scientific exploration of some topic in computer vision.

Possible topics include:

- Algorithm comparisons
  - Performance benchmarking
  - Interesting patterns of success and failure
- Adaptive or pre-processing pipelines
  - Coarse-to-fine stereopsis

# Scientific Project Introduction

There are many open computer vision questions which can be pursued. This project stream emphasizes hypothesis-driven scientific exploration of some topic in computer vision.

Possible topics include:

- Algorithm comparisons
  - Performance benchmarking
  - Interesting patterns of success and failure
- Adaptive or pre-processing pipelines
  - Coarse-to-fine stereopsis
  - Screening of features by saliency or some other criterion

# Scientific Project Introduction

There are many open computer vision questions which can be pursued. This project stream emphasizes hypothesis-driven scientific exploration of some topic in computer vision.

Possible topics include:

- Algorithm comparisons
  - Performance benchmarking
  - Interesting patterns of success and failure
- Adaptive or pre-processing pipelines
  - Coarse-to-fine stereopsis
  - Screening of features by saliency or some other criterion
- Module combinations

# Scientific Project Goal

The goal you should be aiming for with a scientific project is to be able to produce a clearly written and interesting report of your results which reveal something new in the field of computer vision.

A good stretch goal is to aim for publication of your report on arXiv or in a conference or journal. If you are interested in this, it is good to establish a specific goal early, and feel free to speak to me for guidance on this topic.

# Important Points about the Scientific Project

- You are free to use any language and tools you can find to perform your study. Note that the TA and I will be better able to support your efforts if you use Python or C++, and I am also very familiar with MATLAB. These three languages also tend to have the best set of available computer vision tools.

# Important Points about the Scientific Project

- You are free to use any language and tools you can find to perform your study. Note that the TA and I will be better able to support your efforts if you use Python or C++, and I am also very familiar with MATLAB. These three languages also tend to have the best set of available computer vision tools.
- If you are aiming to publish your results, we can discuss a timeline beyond the scope of the course, but you must have sufficient work completed during the course to satisfy the course requirements.

# Important Points about the Scientific Project

- You are free to use any language and tools you can find to perform your study. Note that the TA and I will be better able to support your efforts if you use Python or C++, and I am also very familiar with MATLAB. These three languages also tend to have the best set of available computer vision tools.
- If you are aiming to publish your results, we can discuss a timeline beyond the scope of the course, but you must have sufficient work completed during the course to satisfy the course requirements.
- Deep learning is not always simple to use, has significant hardware requirements, and requires more data than you might think. Please keep this in mind.