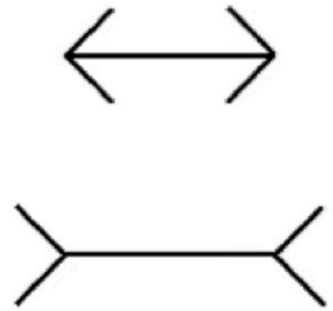
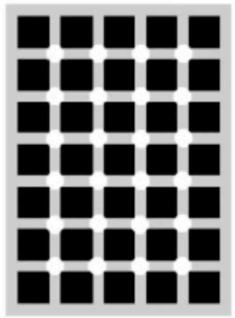
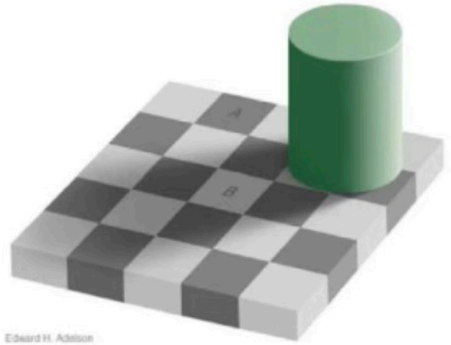
Why is it hard?[1]



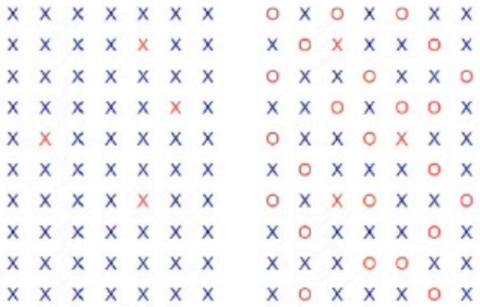
Müller-Lyer illusion: Which line is longer?



Variation of Hermann grid illusion: What do you see at the intersections?



Adelson's brightness constancy illusion: Which is brighter, A or B?



Count the red Xs in both figures, which is harder?

[1]Credit: Szeliski, Computer Vision: Algorithms and Applications, 2010

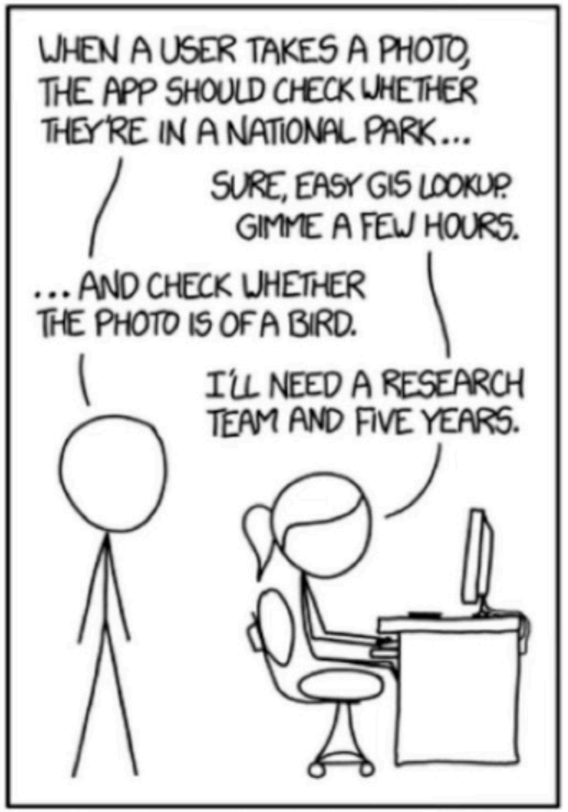
Why is it hard?

- Many practical use cases are **inverse model** applications
  - No knowledge of how an image was taken or camera parameters - but need to model the real world in which picture/video was taken (shape, lighting, color, objects, interactions). $\Longrightarrow$ Need to almost always model from incomplete/partial noisy information
  - *Forward models* are used in physics (radiometry, optics, and sensor design) and in computer graphics
- High-dimensional data $\Longrightarrow$ heavy computational requirements
- Computer vision is AI-complete

Credit: Anish Chopra, Medium.com

Why is it hard?

- No complete models of the human visual system exist
  - Existing models largely related to subsystems, not holistic
  - What is perceived, and what is cognized? When is an object important for a task, and when is the context important?
- Verifiability of mathematical/physical models non-trivial
  - How should similarity/dissimilarity between representations be defined? Is this a distance metric? Do all images follow such a distance metric?
  - How would a manipulation (counterfactual) in a given (potentially noisy) environment behave, w.r.t. the captured image/video? Can a physical model capture this?



Credit: Anish Chopra, Medium.com

Computer Vision: Topics

Learning-based Vision

Visual Recognition, Detection, Segmentation, Tracking, Retrieval, etc

Geometry-based Vision

Feature-based Alignment, Image Stitching, Epipolar Geometry, Structure from Motion, 3D Reconstruction, etc

Physics-based Vision

Computational Photography, Photometry, Lightfields, Color Spaces, Shape-from-X, Reflection, Refraction, Polarization, Diffraction, Interference, etc

Course Topics

Segment 1: The Journey So Far

- Image Formation, Linear Filtering
- Edges, Blobs, Features
- Visual Descriptors, Matching

Segment 2: The Building Blocks

- Review of Neural Networks
- Convolutional Neural Networks (CNNs)
- CNN Architectures, Visualizing and Understanding CNNs

Segment 3: The Many Forms and Uses

- Recognition, Verification, Retrieval, Detection, Segmentation

Segment 4: A Dimension Beyond

- Recurrent Neural Networks
- Spatio-Temporal Models
- Attention, Vision-Language Tasks

Segment 5: Staying Contemporary

- Deep Generative Models
- Learning with Limited Supervision
- Recent Trends