CSE527

- Introduction to Computer Vision
- **http://www.cs.sunysb.edu/~cse527**
- Instructor: Prof. Dimitris Samaras
- Ta: Vu Nguyen
- Fall 2020: Tue-Thu 3:00 – 4:20 in person and on Zoom
- Use Blackboard and Piazza

Slide credits: D Forsythe, I. Kokkinos. S. Lazebnic, S. Seitz, J. Hays, A. Berg

1

Textbooks

Main: Computer Vision: Algorithms and Applications 2nd ed. by Richard Szeliski, Microsoft Research: draft at http://szeliski.org/Book/

Computer Vision: A Modern Approach, Forsyth and Ponce, Prentice Hall 2012.

Introductory Techniques for 3D Computer Vision, Trucco and Verri, Prentice Hall 1998.

Computer Vision: Models, Learning, and Inference by S. Prince, Cambridge University Press, 2012: draft at http://www.computervisionmodels.com/

2

Logistics

- 5 Homeworks will be 60%,
  - Even if you discuss with classmates, turn in your own code and write-up. Do not share code!
  - There will be 4 free late dates for the semester. After that 10% penalty per day.
- midterms are 10% each, Sep 22, Oct 27 2022,
- the final is 20%, Dec 13th 2020
- the optional final project 25%, Dec 15th 2020.
  - Up-to 3 people teams, and will require a significant programming and documentation effort.
  - Two or three people projects will be scaled accordingly.
  - Counts for 2 homeworks

3

What is Computer Vision?

- Inverse Problem of Image Formation
- Compute properties of a world (either 2D or 3D from one or more digital images
- Geometry
- Motion
- Recognition

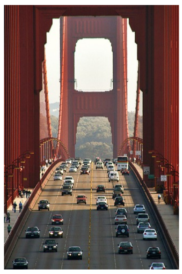
4

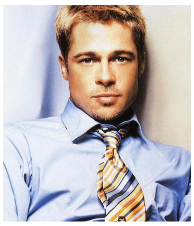
Why Vision?

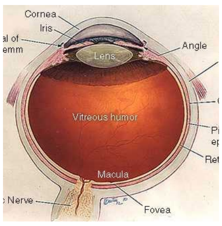
5

Why Vision? Light!

It is how we see other people,
navigate our environment,
communicate ideas, entertain,
and measure the world around us.



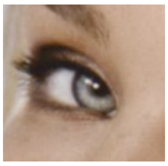




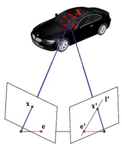
6

Why is light good for measurement?

Microscopy



Surveillance



3D Analysis / Navigation

Remote Sensing

- Plentiful, sometimes free
- Interacts with many things, but not too many
- Goes generally straight over distance
- Very small → high spatial resolution
- Fast, but not too fast → time of flight sensors
- Easy to detect → cameras work, are cheap
- Comes in flavors (wavelengths)

7

Why study Computer Vision?

- Images and movies are everywhere
- Fast-growing collection of useful applications
  - representations of the 3D world from pictures
  - Self-driving cars
  - automated surveillance (who's doing what)
  - movie post-processing
  - face finding
- Various deep and attractive scientific mysteries
  - how does object recognition work?
- Greater understanding of human vision

8

Computer Vision as a sensor

- Information about distant objects
- Passive Sensor
- High bandwidth

  1 picture = ? words
- Corresponds to the most complex human sensory function
  - Eat it? Run from it? Mate with it? +more…
- Computer Vision is not Animate Vision
  – Can be inspired though

9

Properties of Vision

- One can "see the future"
  – Cricketers avoid being hit in the head
    - There's a reflex --- when the right eye sees something going left, and the left eye sees something going right, move your head fast.

  – Gannets pull their wings back at the last moment
    - Gannets are diving birds; they must steer with their wings, but wings break unless pulled back at the moment of contact.
    - Area of target over rate of change of area gives time to contact.

10

Properties of Vision

- 3D representations are easily constructed
  - There are many different cues.
  - Useful
    - to humans (avoid bumping into things; planning a grasp; etc.)
    - in computer vision (build models for movies).
  - Cues include
    - multiple views (motion, stereopsis)
    - texture
    - shading

11

Properties of Vision

- People draw distinctions between what is seen
  - "Object recognition"
  - This could mean "is this a fish or a bicycle?"
  - It could mean "is this George Washington?"
  - It could mean "is this poisonous or not?"
  - It could mean "is this slippery or not?"
  - It could mean "will this support my weight?"
  - Great mystery
    - How to build programs that can draw useful distinctions based on image properties.

12

Why Computational Visual Recognition?

13

Why Computational Visual Recognition?

14

Why Computational Visual Recognition?

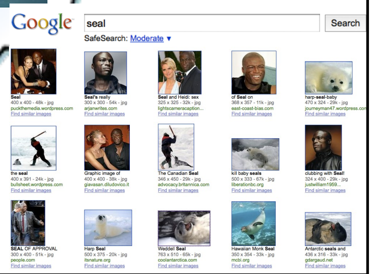
15

Why Computational Visual Recognition?

Billions of images, many with no text / meta data

16

Why Computational Visual Recognition?



Billions of images, many with no text / meta data

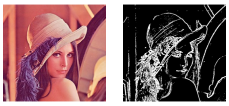
17

Range of recognition tasks

- Duplicate detection
- Edge detection
- Same (rigid) object
- Face detection
- Face Identification
- General category recognition

18

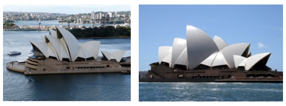
Range of recognition tasks



- Duplicate detection
- Edge detection
- Same (rigid) object
- Face detection
- Face Identification
- General category recognition

19

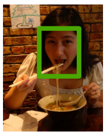
Range of recognition tasks



- Duplicate detection
- Edge detection
- Same (rigid) object
- Face detection
- Face Identification
- General category recognition

20

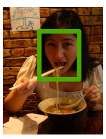
Range of recognition tasks



- Duplicate detection
- Edge detection
- Same (rigid) object
- Face detection
- Face Identification
- General category recognition

21

Range of recognition tasks



Tamara

- Duplicate detection
- Edge detection
- Same (rigid) object
- Face detection
- Face Identification
- General category recognition

22

Range of recognition tasks

- Duplicate detection
- Edge detection
- Same (rigid) object
- Face detection
- Face Identification
- General category recognition

23

Vision for perception, interpretation

Slide credit: Kristen Grauman

24

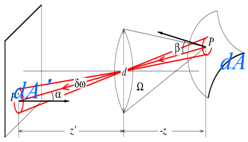
Image Sciences

Image processing

Image to Image

Imaging

Physics to Image

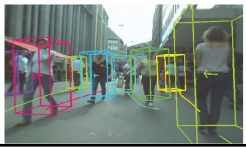


Graphics

Symbols to Image

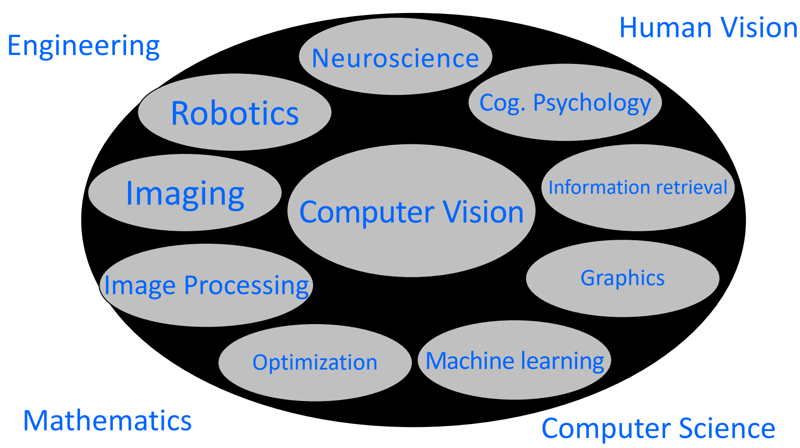
Computer Vision

Image to Symbols (INVERSE PROBLEM)



25

Relationship with other fields



26

Computer Vision vs. Computer Graphics

- Graphics
  - Produce "plausible" images
  - You choose the models, conditions, imaging parameters, etc.
- Computer Vision
  - Given real images with noise, sampling artifacts …
  - Estimate physically quantities
  - Ill-posed ---- what is the minimum world knowledge we need?

27

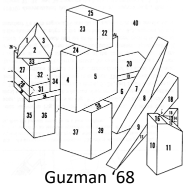
Computer Vision vs. Image Processing

- Image Processing
  - Mostly concerned with image-to-image transformations
  - Filtering
  - Enhancement
  - Compression
- Computer Vision
  - Concerned with how images reflect the 3D world
  - Filtering for feature extraction
  - Enhancement for recognition/detection
  - Compression that preserves geometric information in images

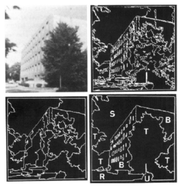
28

Ridiculously brief history of computer vision

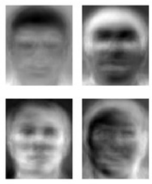
- 1966: Minsky assigns computer vision as an undergrad summer project
- 1960's: interpretation of synthetic worlds
- 1970's: some progress on interpreting selected images
- 1980's: ANNs come and go; shift toward geometry and increased mathematical rigor
- 1990's: face recognition; statistical analysis in vogue
- 2000's: broader recognition; large annotated datasets available; video processing starts; augmented reality
- 2010's: Deep learning with ConvNets
- 2030's: robot uprising?



Guzman '68



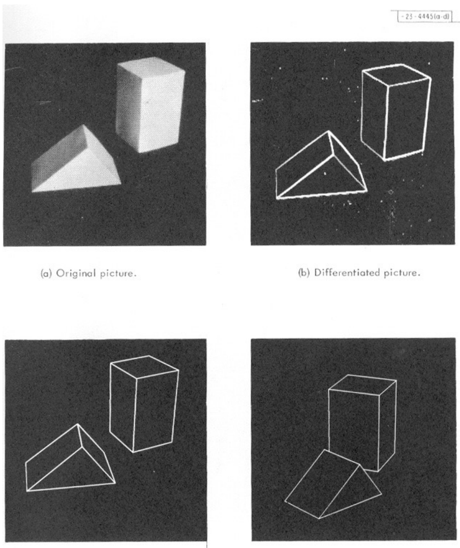
Ohta Kanade '78



Turk and Pentland '91

29

Visual data in 1963



(a) Original picture. (b) Differentiated picture.

(c) Line drawing. (d) Rotated view.

L. G. Roberts, Machine Perception of Three Dimensional Solids, Ph.D. thesis, MIT Department of Electrical Engineering, 1963.

30

Slide credit: Kristen Grauman

30

Visual data today

Svetlana Lazebnik

31

Where is computer vision useful?

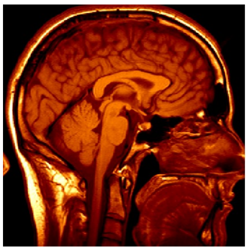
- Medical Image Analysis
- Robotics
- Automobile industry
- Optical Character Recognition
- Visual aids for the blind
- Industrial inspection
- Surveillance
- Military
- Film industry
- Entertainment
- Human-Computer Interaction
- Image Search

...

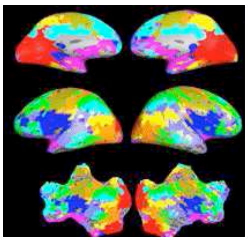
- http://www.cs.ubc.ca/spider/lowe/vision.html

32

Medical Image Analysis



Analysis of images acquired with Computerized Tomography, Magnetic Resonance Imaging, Ultrasound…



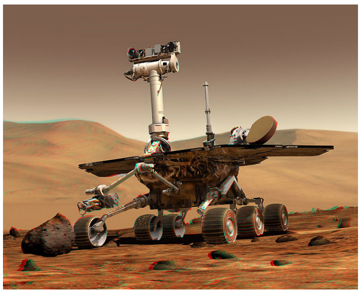
33

Image Guided Surgery

FMRI data anaysis

33

Robotics



DARPA Grand Challenge

NASA's Mars Rover

Toy industry

34

34

Industrial robots

Vision-guided robots position nut runners on wheels

35

Biometrics

Iris Recognition

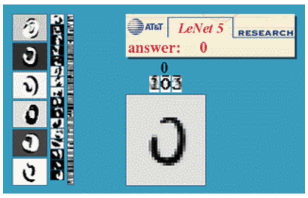
Fingerprint Recognition

Face Recognition

36

36

Text Detection and Recognition



Optical Character Recognition (OCR)

[redacted]

Licence Plates

37

`Blindsight'

37

Automobile Industry

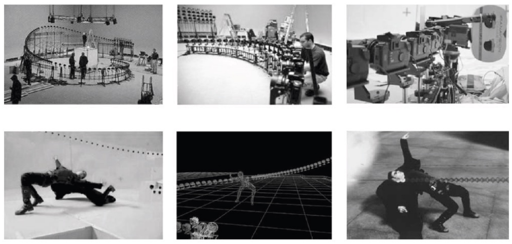
Pedestrian and Car Detection

Monitoring Driver Alertness

38 Lane Detection

38

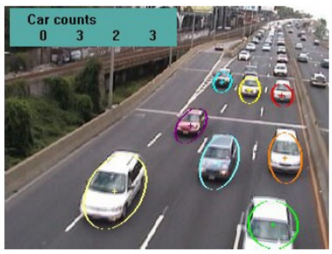
Visual Effects

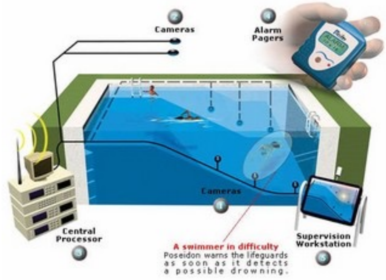


39

39

Security & Surveillance





40

40

Object recognition

Google Glass

Bing Vision

41

Added value to commercial products

- Digital Cameras

Point & find

42

kooaba

42

Motion capture

Microsoft's XBox Kinect

43

43

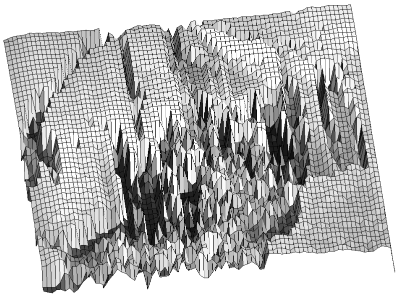
Augmented and Virtual Reality

Magic Leap, Oculus, Hololens, etc.

44

Why can't computers see (yet)?

- Imagine describing `red' or `ugly' to a blind man
- Input to a computer: 2D/3D function



45

45

Current state of the art

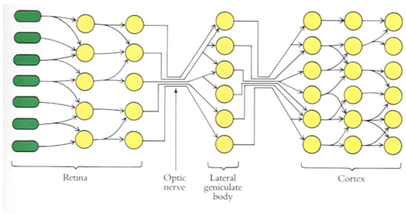
- You just saw examples of current systems.
  - Many of these are less than 5 years old
- This is a very active research area, and rapidly changing
  - Many new apps in the next 5 years

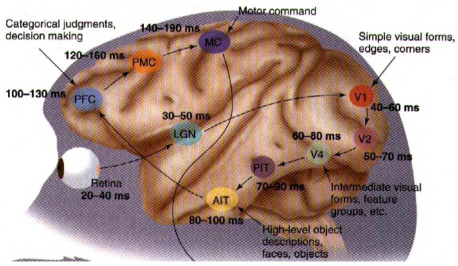
46

How do we solve vision?

We perform the vision task with amazing speed and accuracy

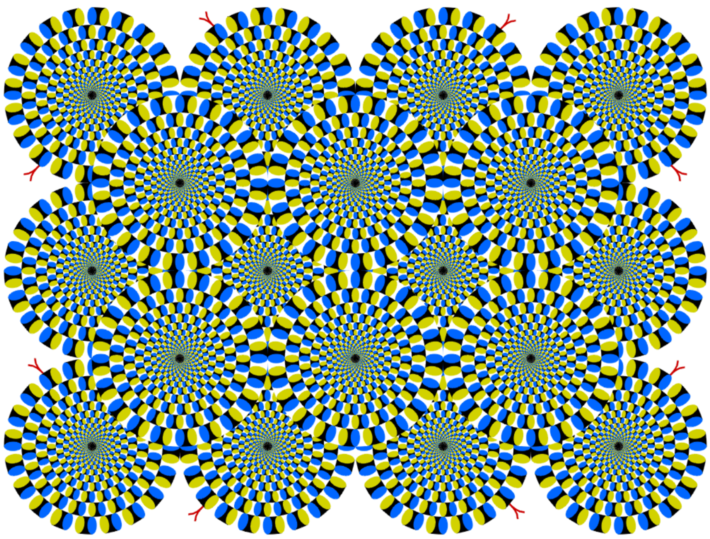
- But not effortlessly:
  - Almost 50% of your brain is doing vision
    - Substantially more than what is involved in doing math!





47

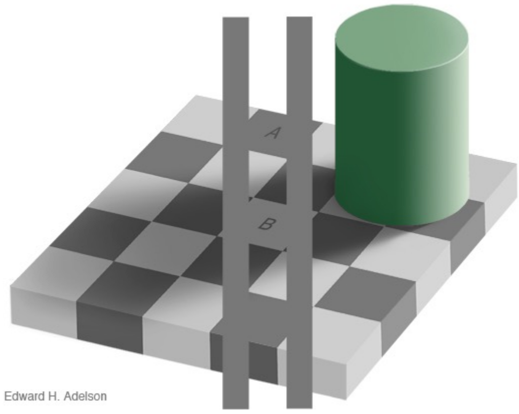
Is seeing trivial?



48

48

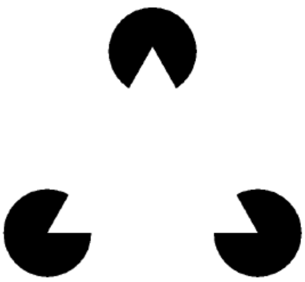
Is seeing trivial?



http://web.mit.edu/persci/people/adelson/checkershadow_illusion.html

49

Is seeing trivial?



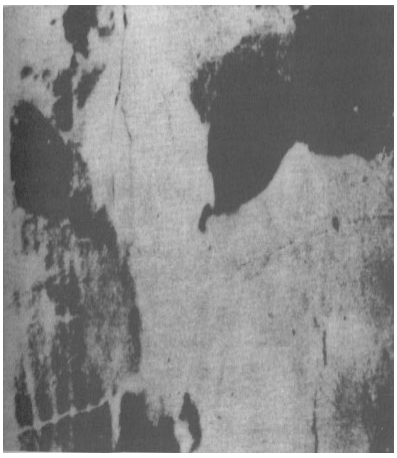
50

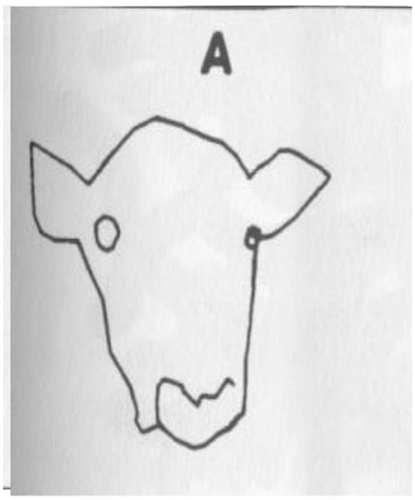
50

Is seeing trivial?

51

Is seeing trivial?

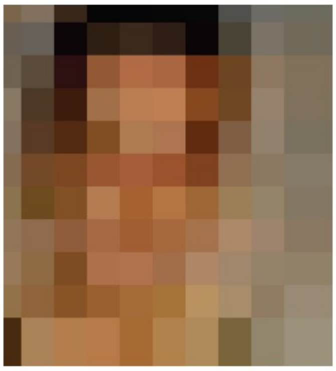




52

52

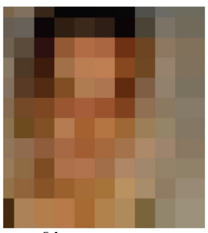
Face or non-face?



53

53

Face or non-face?



54

54

Context

55

55

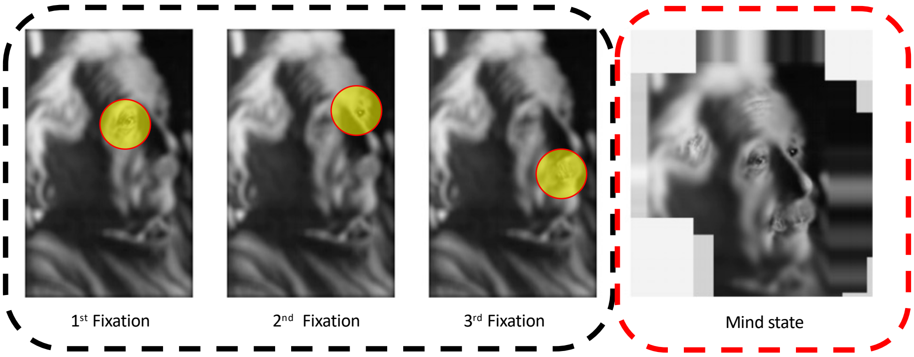
Understanding how humans look at an image in visual search and free viewing tasks

56

Foveal Vision

57

How do we view the world?

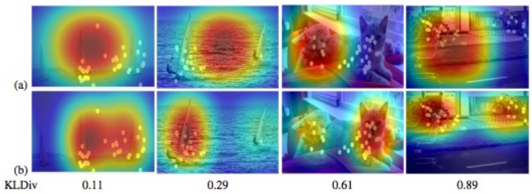


[Barlow, Horace B. "Unsupervised learning." *Neural computation* 1.3 (1989): 295-311.]
[Olshausen, Bruno A. "Emergence of simple-cell receptive field properties by learning a sparse code for natural images." *Nature* 381.6583 (1996): 607-609.]
[Osindero, Simon, Max Welling, and Geoffrey E. Hinton. "Topographic product models applied to natural scene statistics." *Neural Computation* 18.2 (2006): 381-414.]
[Lee, Honglak, Chaitanya Ekanadham, and Andrew Y. Ng. "Sparse deep belief net model for visual area V2." *Advances in neural information processing systems*. 2008.]

58

58

Sparsity predicts Visual Attention



Human vision behavior in search tasks emerges when imposing IoR

Implemented by means of Non-Maximum Suppression
Better predicts human attention in visual search tasks

- Wei, Zijun, et al. "Learned Region Sparsity and Diversity Also Predicts Visual Attention." *Advances in Neural Information Processing Systems*. 2016.

59

Low- & high-level vision: Chicken & egg

60

60

Challenges: What is an object?

61

61

Challenges: Many Nuisance Parameters

Illumination

Object pose

Clutter

Occlusions

Intra-class appearance

Viewpoint

62

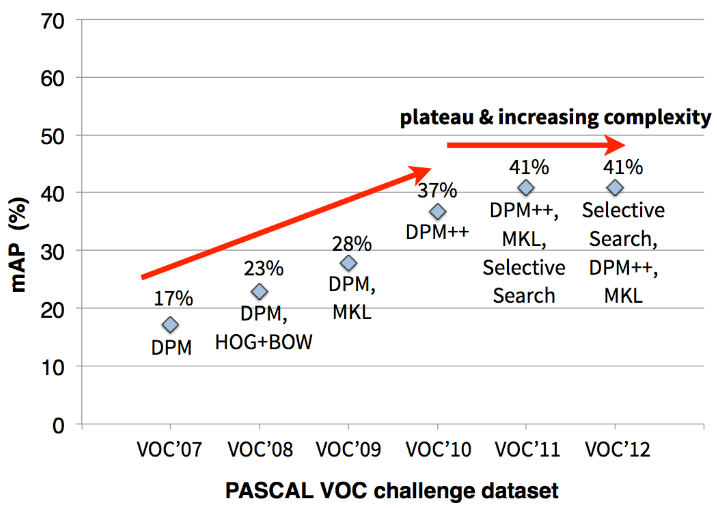
62

Challenges: Intra-Category Variation

63

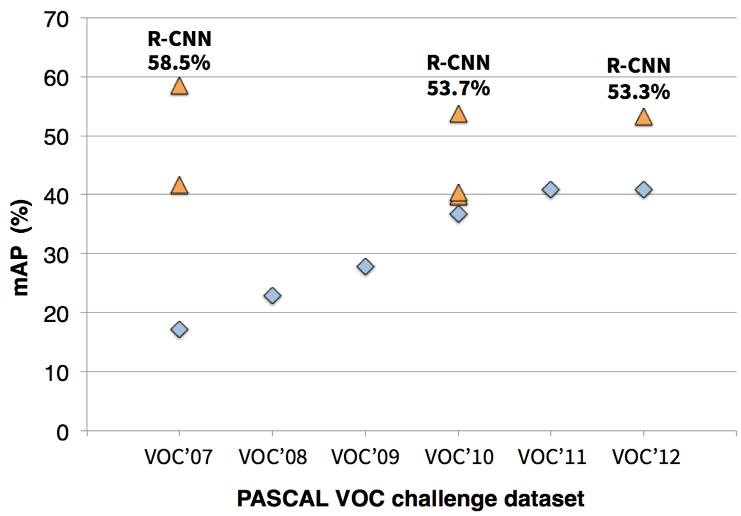
64

Object recognition 2010-2012: saturation



65

Object recognition, 2013



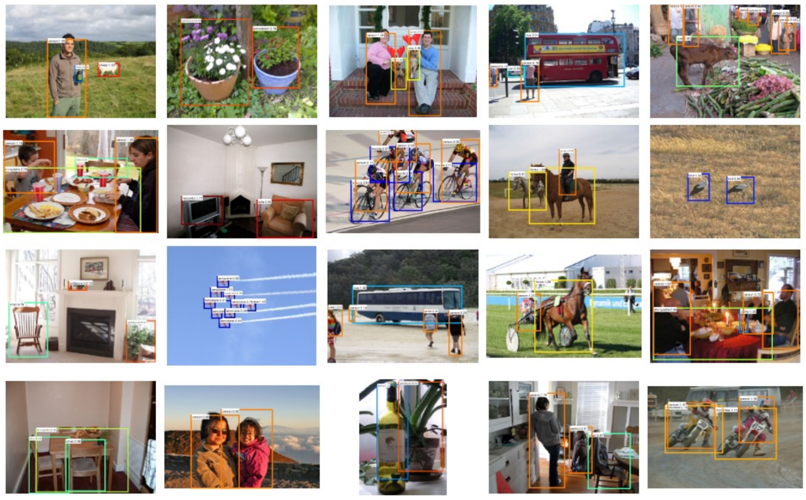
A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. *NIPS*13

P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. ICLR 14.

R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *CVPR* 14

66

Object recognition, 2014



A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. *NIPS*13
P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. ICLR 14.
R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *CVPR* 14

67

Object recognition, 2015

Evolution of Performance on PASCAL VOC-2007 over the recent years

68

State of the art today?

With enough training data, computer vision nearly matches human vision at most recognition tasks

Deep learning has been an enormous disruption to the field. More and more techniques are being “deepified”.

Question:
Who won 2019 Turing Award?

69

Computer Vision at Stony Brook

Face Recognition/Modeling
Graphics
Geometric Modeling
Medical Image Analysis
Transportation/Infrastructure
Tracking/Motion analysis
Video Analysis
Machine Learning

Strong collaborators in CS and other departments on campus Psychology, Music, Art, BNL, BMI, Ecology, Civil Engineering

70

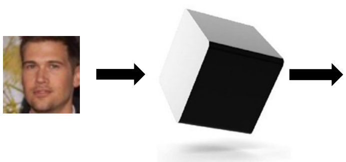
Image Formation
Basic facts about light
Anatomy of a camera
Matting
Image Noise
Modeling image noise
Convolution
Smoothing images
Image Pyramids
Image Features
Point Features, Corners
Edge Features
Scale, Orientation
Model Fitting
Lines, Curves
The Hough Transform
Deformation
RANSAC
Perspective Projection
Homogeneous Coordinates
Image Warping, Mosaics
Multiple View Geometry
Stereo Viewing and Reconstruction
3D Range Scanning
Machine Learning/Object Recognition
Object representation
PCA for Image Patches
Classifiers
Object Categories

Deep Learning
Convolutional Neural Networks
Architectures
Applications
Deep Learning Practice
Pre-Training
Data Augmentation
Illumination
Shading, Shadows,
Reflectance properties
Deep Generative Models
Autoencoders, VAEs,
Generative Adversarial Networks
Motion
Motion Capture
Tracking in 2D and 3D
Recurrent Neural Networks
Action Recognition
Segmentation
Grouping, SuperPixels
Nearest Neighbors
UNET,
Semantic Segmentation
Big Data
Annotated Data Sets
Crowdsourcing
Weakly Supervised and
Unsupervised Learning

71

Neural Face Editing

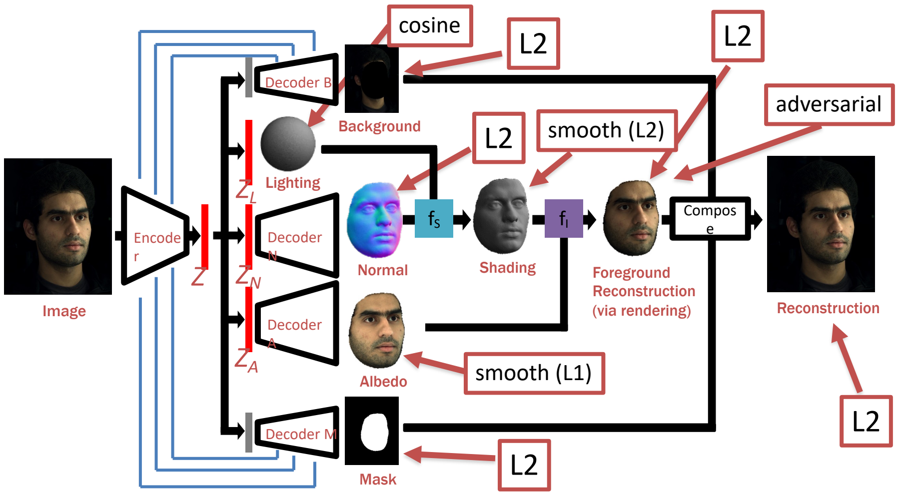
- Motivation
  - A general framework for multi-tasks, semantic face editing.
  - Exploring the power of generative deep neural networks.



new lighting
new expression
wear glasses
younger/older
⋮

72

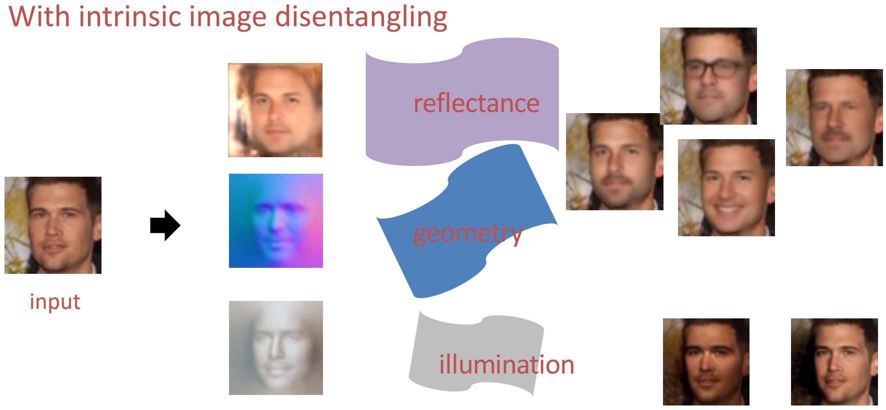
Neural Face Editing



73

Editing by manifold traversal

With intrinsic image disentangling

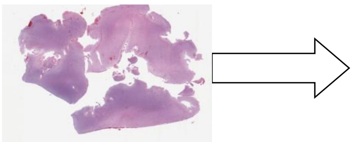


74

Automatic Glioma Classification using Deep Learning

Joint work with Le Hou Tahsin Kurc, Joel Saltz

- Gliomas are the most common brain cancers.
- Better classification is critical to the development of targeted therapies.
- We have achieved accuracy comparable to Inter-observer agreement

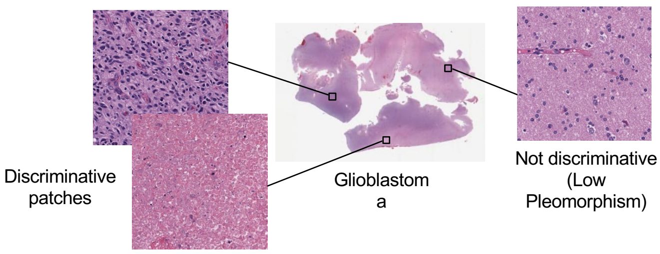


Glioblastoma?
Oligodendroglioma?
Astrocytoma?
Oligoastrocytoma?
Anaplastic Astro?
Anaplastic Oligo?

75

Challenges

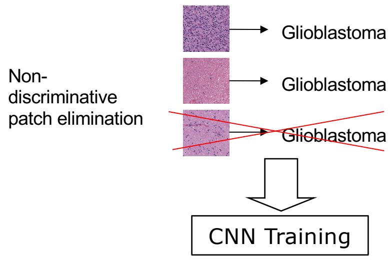
- Gigapixel resolution.
  - Classifiers need to be patch-based.
  - However, not all regions are discriminative.
- High intra-class heterogeneity.



76

Methods

- Multiple Instance Learning + Convolutional Neural Networks.
  - Eliminate non-discriminative patches in an EM fashion.
  - Train a CNN model on discriminative patches only.
- Given a test image, apply CNN on every patch then aggregate patch-level predictions using logistic regression.



77

Shadow detection and removal

From Shadow Segmentation to Shadow Removal

Hieu Le, Dimitris Samaras
Stony Brook University

Stony Brook University

ECCV'20 ONLINE
23-28 AUGUST 2020

78

Questions?

- Welcome to CSE527