EN3160 Image Processing and Machine Vision: Introduction

Ranga Rodrigo
ranga@uom.lk

The University of Moratuwa, Sri Lanka

August 1, 2023

Section 1

Introduction to the Course

What Is Image Processing?

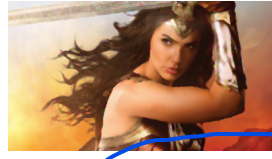
- In digital image processing, we manipulate a digital image to produce another digital image, which is an enhanced version of the original image.
- E.g., blurring, noise filtering, color enhancement, segmentation (?).

(a) Noisy image

use Gaussian filter to remove the noise. but it kind of blur the image

(b) Gaussian filtered image

(c) Image with salt and pepper noise



(d) Median filtered image

Figure: Filtering

What is computer vision?

- In computer vision, we analyze a digital images or videos to make a decision.
- E.g., face detection, object detection, semantic segmentation.

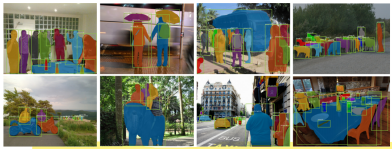


Figure 2. **Mask R-CNN** results on the COCO test set. These results are based on ResNet-101 [19], achieving a *mask* AP of 35.7 and running at 5 fps. Masks are shown in color, and bounding box, category, and confidences are also shown.

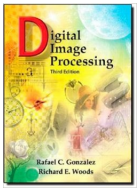
(a) Mask RCNN Object detection and semantic segmentation

(b) Face detection (from Wikipedia)

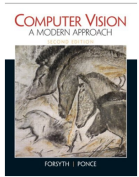
Figure: Examples of computer vision.

Text Books

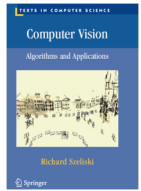
1. Gonzalez and Woods, Digital Image Processing
2. Forsyth and Ponce, Computer Vision: A Modern Approach
3. Richard Szeliski, Computer Vision: Algorithms and Applications (available online)
4. Milan Sonka, Image Processing, Analysis, and Machine Vision



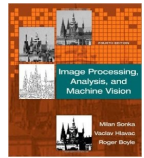
(a)



(b)



(c)



(d)

Figure

Assessment Criterion

Continuous assessments: 40%, final closed-book written examination: 60%.

Item	Date	Weight	
In-class quizzes	Surprise	10%	Easy, be
Assignment A01: Intensity Transformations and Neighborhood Filtering	Aug. 30	15%	
Assignment A02: on Fitting and Alignment (RANSAC, homographic transformation, image stitching)	Sept 13	15%	
Assignment A03: Measurements, Tracking and Counting Objects along a Conveyor Belt	Sept. 20	15%	Mode
Project: on deep learning for vision, groups of two, topics will be given	Oct. 18	45%	Mode

Tips for Successful Completion

To extract meaning from pixels.

1. Mindset: “I want to solve this vision problem. What are the tools available? How have others solved it? How should I go about solving it?”
2. Attend every single lecture.
3. Go through the material the night before and engage in a good discussion in class.
4. Implement at least one algorithm discussed in class before the next class. This is in addition to the assignments.

Academic Integrity Policy

- You may discuss the assignments with each other. However, you must code on your own, and make the submissions on your own.
- You may benefit from the code and information on the internet, as long as you do not borrow code for the main theme of the assignment.
- Acknowledge your sources.

Source: S. Lazebnik

The Goal of Computer Vision

(a) What we see.

0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

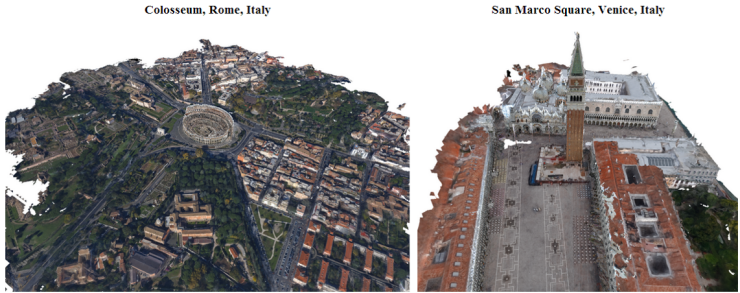
(b) What computer sees.

Source: S. Narasimhan

Section 2

Sate-of-the-Art in Vision

Reconstruction: 3D from Photo Collections



Q. Shan, R. Adams, B. Curless, Y. Furukawa, and S. Seitz, The Visual Turing Test for Scene Reconstruction, 3DV 2013. `https://www.youtube.com/watch?v=NdeD4cjLI0c`

Source: S. Lazebnik

Reconstruction: 4D from Photo Collections

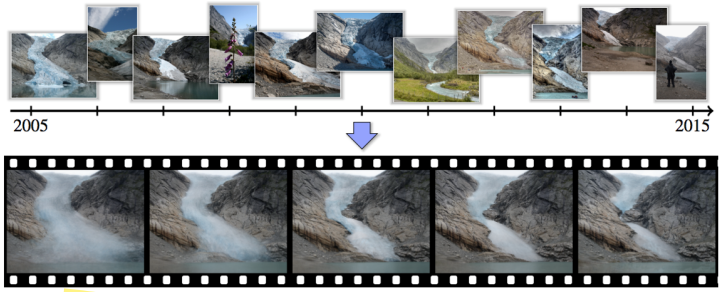


Figure 1: *We mine Internet photo collections to generate time-lapse videos of locations all over the world. Our time-lapses visualize a multitude of changes, like the retreat of the Briksdalsbreen Glacier in Norway shown above. The continuous time-lapse (bottom) is computed from hundreds of Internet photos (samples on top).* Photo credits: Aliento Más Allá, jirihnidek, mcxurxo, elka.cz, Juan Jesús Orío, Klaus Wißkirchen, Daikrieg, Free the image, dration and Nadav Tobias.

R. Martin-Brualla, D. Gallup, and S. Seitz, Time-Lapse Mining from Internet Photos, SIGGRAPH 2015. https://www.youtube.com/watch?v=wptzVm0tngc&feature=youtu.be

Source: S. Lazebnik

Reconstruction: 4D from Depth Cameras

Figure 1: Real-time reconstructions of a moving scene with DynamicFusion; both the person and the camera are moving. The initially noisy and incomplete model is progressively denoised and completed over time (left to right).

R. Newcombe, D. Fox, and S. Seitz, DynamicFusion: Reconstruction and Tracking of Non-rigid Scenes in Real-Time, CVPR 2015.

`https://www.youtube.com/watch?v=i1eZekcc_lM&feature=youtu.bee`

Source: S. Lazebnik

Reconstruction in Construction Industry

https://www.reconstructinc.com/#product-demo

Source: D. Hoiem

Recognition: “Simple Patterns”

(a)

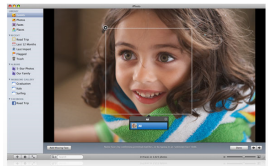
(b)

(c)

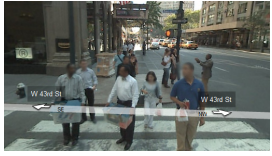
(d)

Recognition: Faces

(a)



(b)



(c)

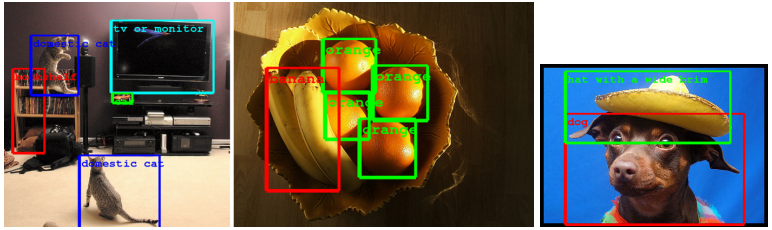
Figure

Concerns about Face Recognition

Beijing bets on facial recognition in a big drive for total surveillance — Washington Post, 1/8/2018

Source: Lazebnik

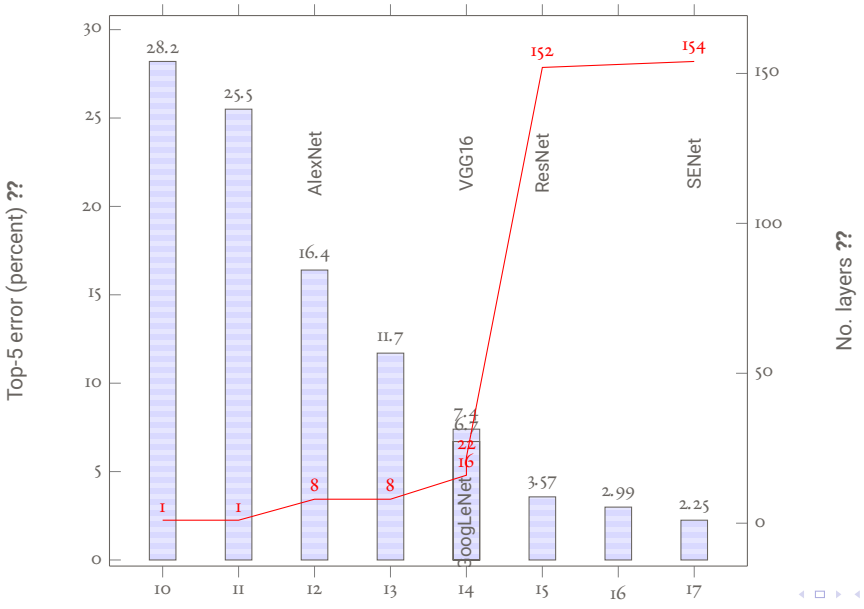
Recognition: General Categories



Computer Eyesight Gets a Lot More Accurate, NY Times Bits blog, August 18, 2014
Building A Deeper Understanding of Images, Google Research Blog, September 5, 2014

Source: Lazebnik

ImageNet Challenge



Bengio, Hinton and LeCun Win the Turing Award

Fathers of the Deep Learning Revolution Receive ACM A. M. Turing Award

Recognition: Instance Segmentation

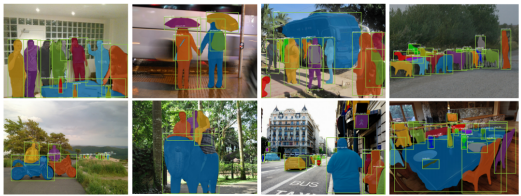


Figure 2. **Mask R-CNN** results on the COCO test set. These results are based on ResNet-101 [19], achieving a *mask* AP of 35.7 and running at 5 fps. Masks are shown in color, and bounding box, category, and confidences are also shown.

(a)

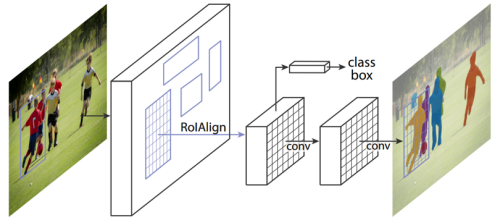


Figure 1. The **Mask R-CNN** framework for instance segmentation.

(b)

Figure

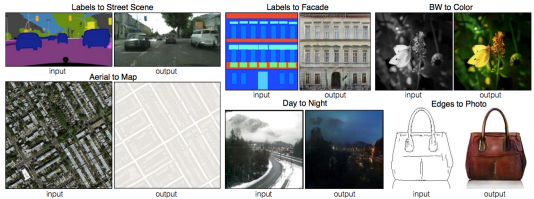
K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” in IEEE International Conference on Computer Vision, Venice, Italy, 2017, pp. 2980–2988.

Image Generation

T. Karras, T. Aila, S. Laine, and J. Lehtinen, Progressive Growing of GANs for Improved Quality, Stability, and Variation, ICLR 2018.

Source: Lazebnik

Image Generation



P. Isola, J.-Y. Zhu, T. Zhou, A. Efros, Image-to-Image Translation with Conditional Adversarial Networks, CVPR 2017.

Source: Lazebnik

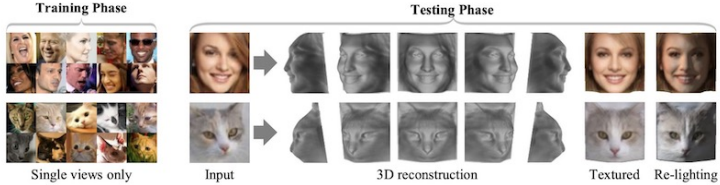
Deep Video Portraits (DeepFakes)

in zoom we can see this

Kim, H. and Garrido, P. and Tewari, A. and Xu, W. and Thies, J. and Nießner, N. and Pérez, P. and Richardt, C. and Zollhöfer, M. and Theobalt, C. Deep Video Portraits, Siggraph 2018.

Source: Lazebnik

SinGAN: Learning a Generative Model From a Single Natural Image



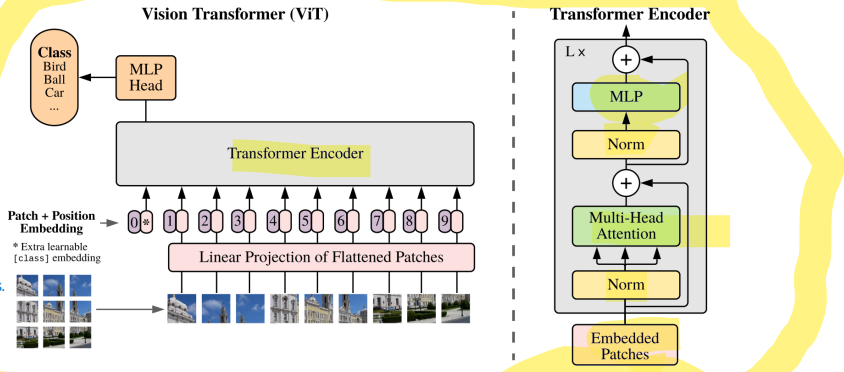
Unsupervised Learning of Probably Symmetric Deformable 3D Objects From Images in the Wild

Tamar Rott Shaham, Tali Dekel, and Tomer Michaeli. "SinGAN: Learning a Generative Model From a Single Natural Image". In: *IEEE/CVF International Conference on Computer Vision*. Seoul, Korea, 2019, pp. 4569–4579 Shangzhe Wu, Christian Rupprecht, and Andrea Vedaldi. "Unsupervised Learning of Probably Symmetric Deformable 3D Objects From Images in the Wild". In: *IEEE/CVF Conference on Computer Vision and Pattern*

Vision Transformer

The Vision Transformer, often abbreviated as ViT, is a deep learning architecture designed for solving computer vision tasks using the transformer architecture, which was originally introduced for natural language processing tasks. The transformer architecture gained significant attention and success in NLP due to its ability to capture long-range dependencies in sequences of data. The Vision Transformer extends this concept to image data.

Traditional convolutional neural networks (CNNs) have been the dominant architecture for image recognition tasks. CNNs excel at capturing local spatial patterns but can struggle with capturing long-range relationships between image elements. The Vision Transformer aims to address this limitation by applying the transformer's attention mechanism to images. (more details chatgpt)



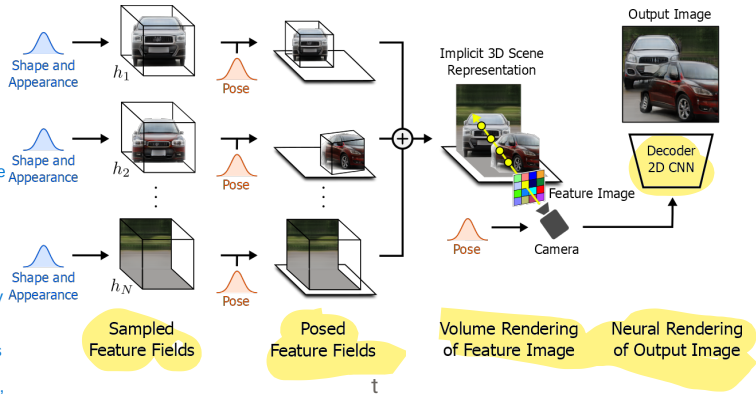
Vision Transformer

Alexey Dosovitskiy et al. “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale”. In: *International Conference on Learning Representations*. 2021

Representing Scenes as Compositional Generative Neural Feature Fields

convolution in images,
Imagine you have a picture, and you want to find specific patterns or features in it. Convolution in digital signal processing is like sliding a small window called a " kernel" across the picture. At each position, you multiply the numbers in the kernel with the numbers in the picture that are currently under the window. Then you add up all these multiplied numbers to get a single value. This value tells you how much that specific pattern in the kernel matches the area of the picture under the window.

So, convolution helps you find patterns or details in an image by looking at small parts of the image and figuring out how well a specific pattern in the kernel fits those parts. This technique is widely used in image processing, computer vision, and other fields to analyze and enhance signals or images.

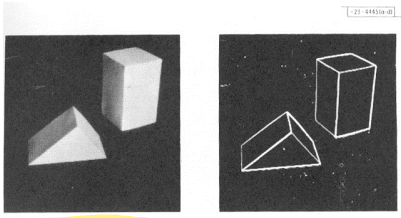


Michael Niemeyer and Andreas Geiger. "Giraffe: Representing scenes as compositional generative neural feature fields". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 11453–11464

More Best Paper Awards

1. CVPR 2022: Learning to Solve Hard Minimal Problems
2. CVPR 2023: Visual Programming: Compositional visual reasoning without training
3. CVPR 2023: Planning-oriented Autonomous Driving

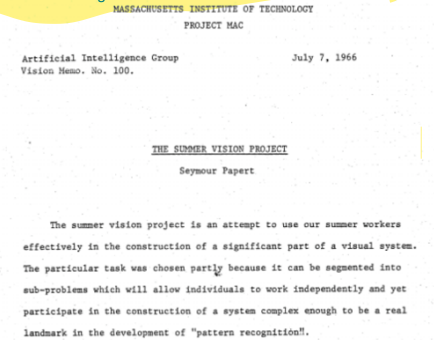
Origins of Computer Vision



(a) Machine perception of three-dimensional solids by Lawrence G. Roberts, MIT 1937

For example, when you see an object, your brain doesn't just process the visual information, but it also combines it with your past experiences, knowledge, and expectations to form a coherent perception of what the object is and what it means. Similarly, when you hear a sound, your brain processes the auditory information and makes sense of it by considering context and previous knowledge.

Perception is a complex cognitive process that involves the interaction of sensory input, cognitive processing, and our individual experiences. It plays a crucial role in our daily lives, influencing how we interact with our surroundings and make decisions based on our understanding of the world.



MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group — July 7, 1966
Vision Memo. No. 100.

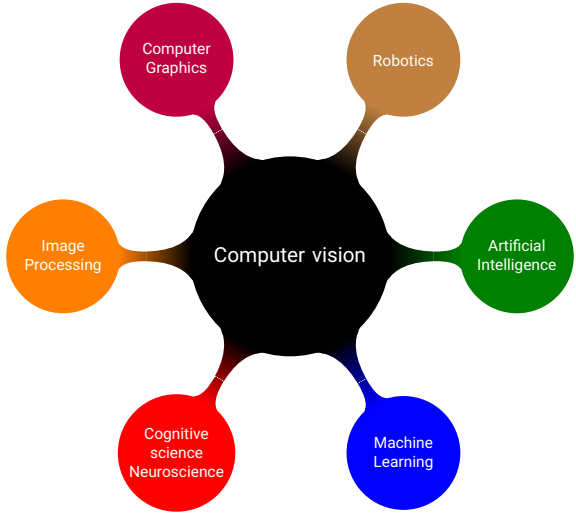
THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

(b) The Summer Vision Project at MIT, 1966.

Connections to Other Disciplines



Source: Lazebnik