ABV-Indian Institute of Information Technology and Management Gwalior

==========

Course on Computer Vision (ITIT-9507)

==========

Instructor – Dr. Sunil Kumar

Office – 206, F-Block (V), Tel No – 0751-2449710 (O), Email - snk@iiitm.ac.in

Mob - 8472842090

Computer Vision

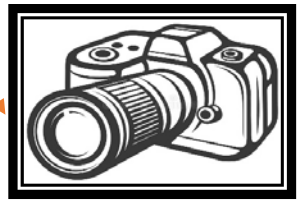
- The ability of Computer to see images and/or videos in a sense that
  - it can understand
  - Helps to take action based on what it infers

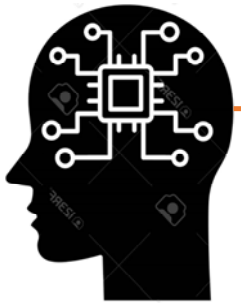
- Analogous to : Human vision system + Human brain

Vision System Vs. Computer Vision

Image

Sensing Device





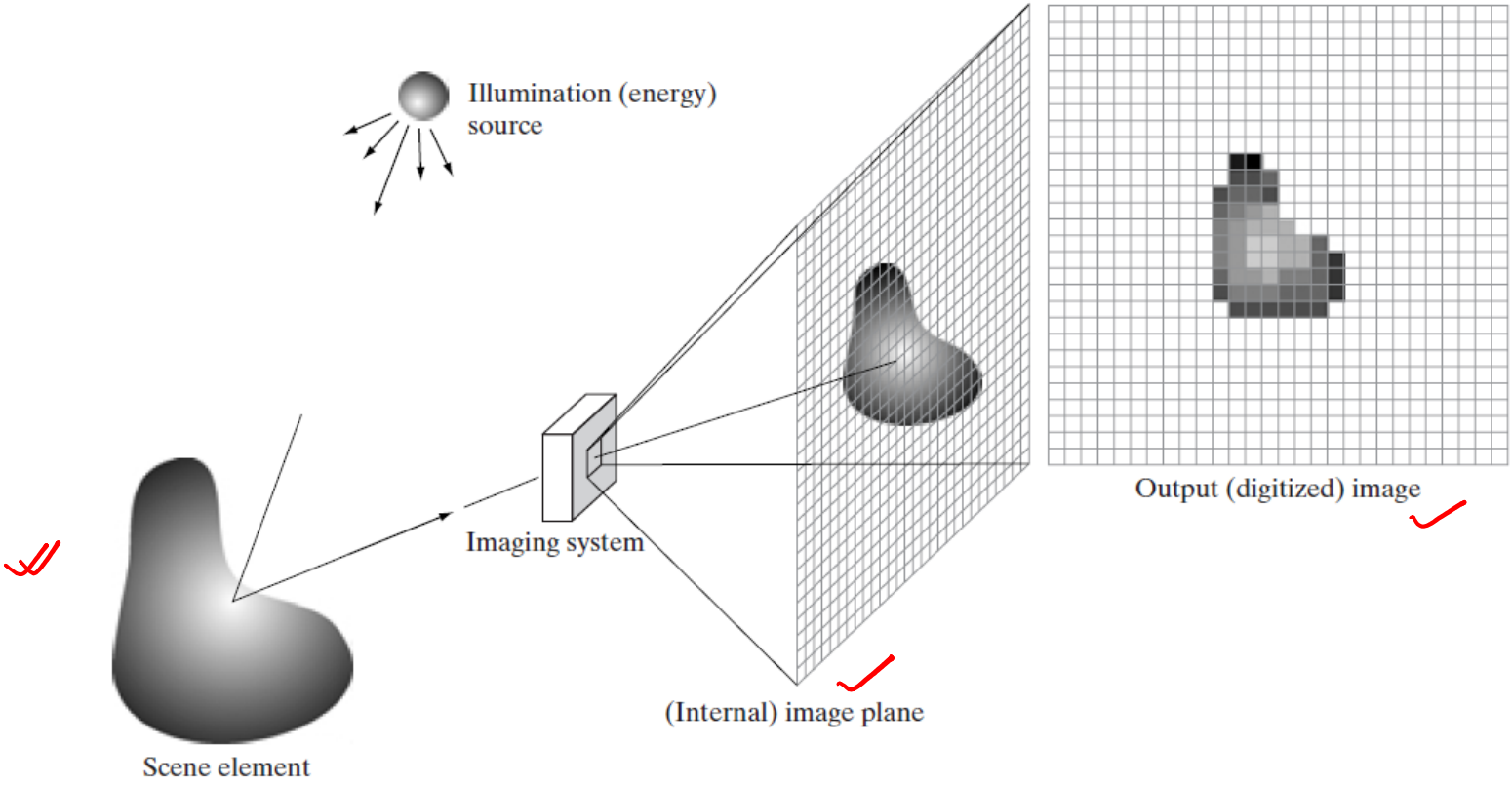
Interpreting Device

❑ **Inferences**

- ❖ It can answers many questions
  - ❖ How many persons are there in image?
  - ❖ Where is Mr. Barack obama and what is he doing?
  - ❖ Many more.
- ❖ In fact, one can write a story on this image.

- ❖ Computer vision enables computer to do all such tasks that a human can do.
- ❖ Still, it has to do a lot many ….

What is an Image (Digital Image)?



a
b c d e

FIGURE 2.15 An example of the digital image acquisition process. (a) Energy ("illumination") source. (b) An element of a scene. (c) Imaging system. (d) Projection of the scene onto the image plane. (e) Digitized image.

Digital Image Representation

Figure- 1 : ABV-IIITM Gwalior sports

	0	1	2	3	4	5	6	7	• • •	N-1
0	5	20	3	0	88	47	255	156	20	90
1	50	210	30	10	128	40	27	61	17	25
2	7	23	43	137	18	51	43	29	89	86
3	42	42	63	96	128	48	127	161	127	203
4	50	210	30	10	128	40	27	61	17	26
5	56	0	30	77	256	52	224	58	66	83
⋮										
M-1	15	21	14	53	128	40	27	61	17	222

y-axis

x-axis

Digital Image Representation

5	20	3	0	88	47	255	156	20	90
50	210	30	10	128	40	27	61	17	25
7	23	43	137	18	51	43	29	89	86
42	42	63	96	128	48	127	161	127	203
50	210	30	10	128	40	27	61	17	26
56	0	30	77	256	52	224	58	66	83
15	21	14	53	128	40	27	61	17	222
15	21	14	53	128	40	27	61	17	222

Figure- 1 : ABV-IIITM Gwalior sports

Resolution : 356 x 512 = No. of rows * No. of columns

Types of a Digital Image

Grayscale image : 0 to 255 for 8-bits

Color image : [R-plane, G-plane, B-plane]

- Grayscale image with intensity value 0 (black) and 1 (white) Called “black and white image”.

Types of a Digital Image

RGB Color image : [R-plane, G-plane, B-plane]

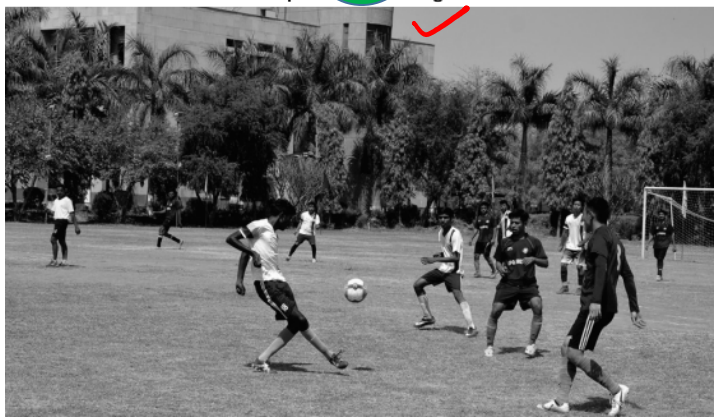
R

G

B

R-plane of the image

G-plane of the image

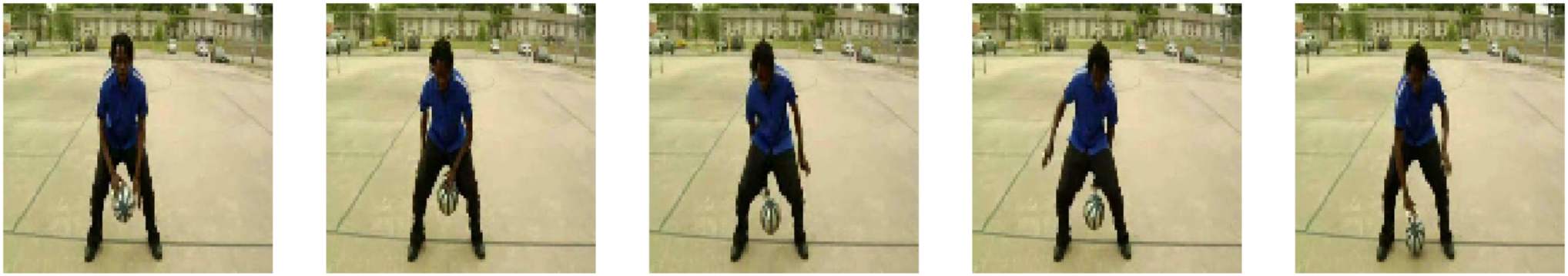


B-plane of the image

Video

- Defined by a sequence of images (also called frames).
- Frame rate : 30 frames/second

Class: dribble



Class: kick_ball

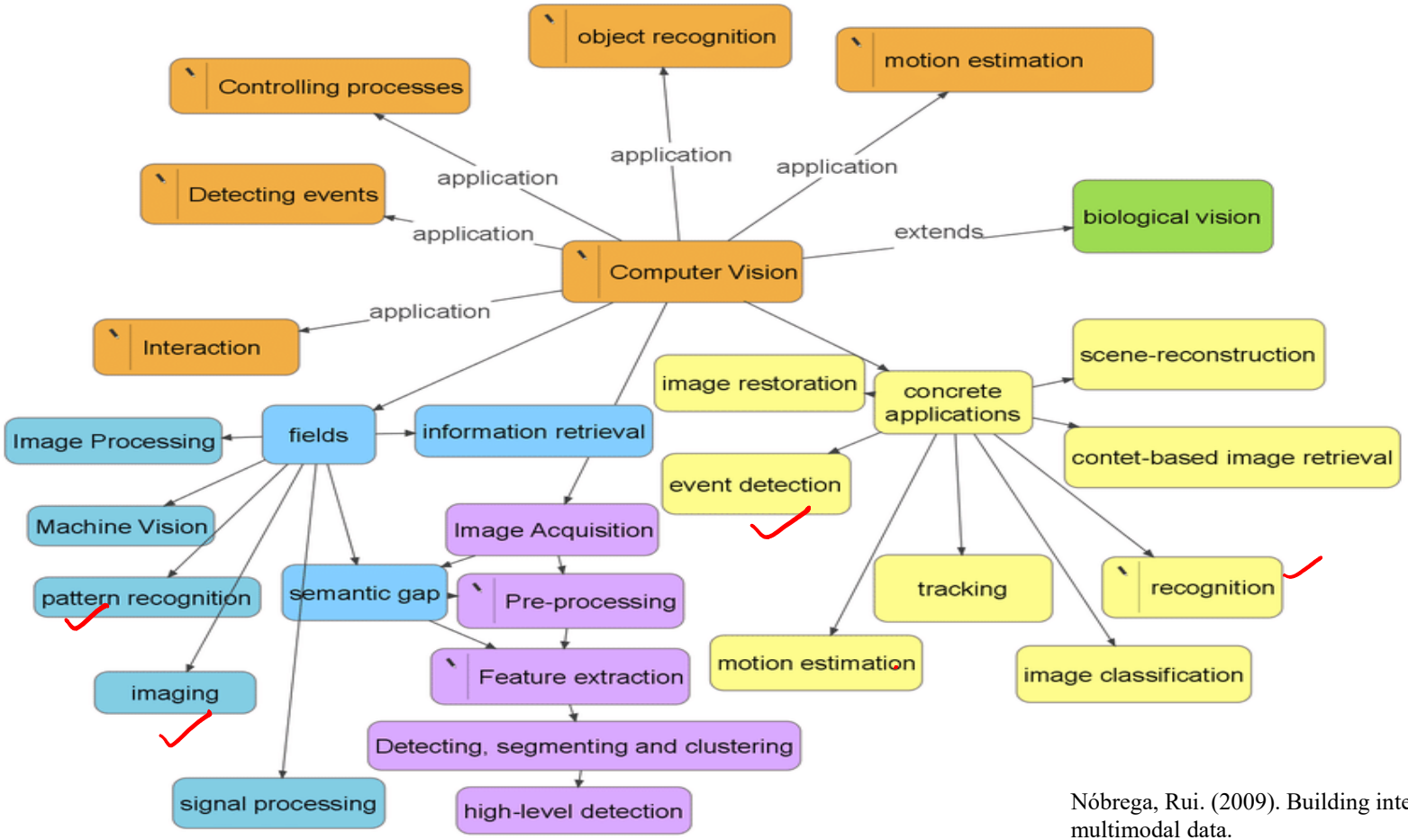
Motivation of this Course

- Why do you need to study this course?
  - Billions of images/videos captured per day

  - Huge number of useful applications

Motivation of this Course

- Why do you need to study this course?

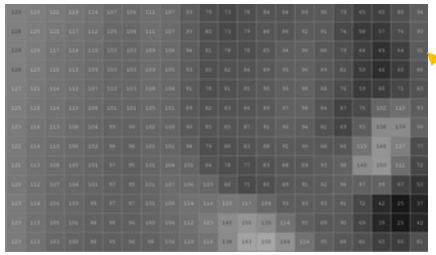


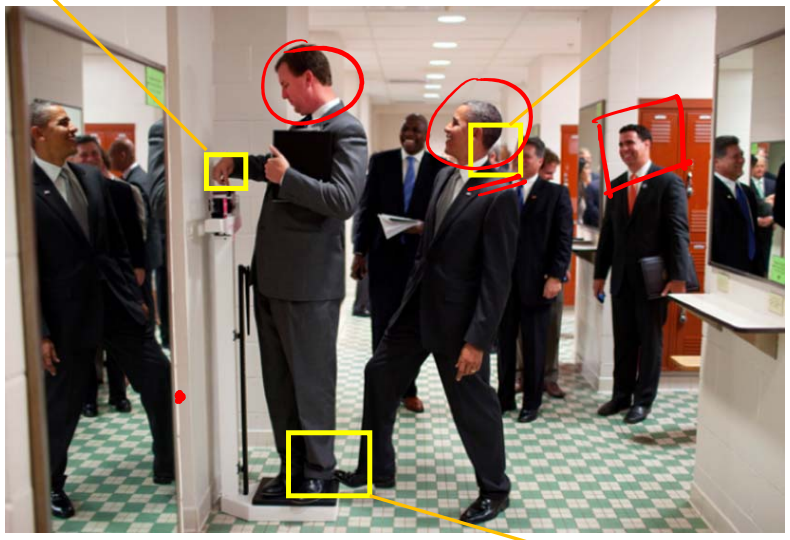
Wide range of Applications

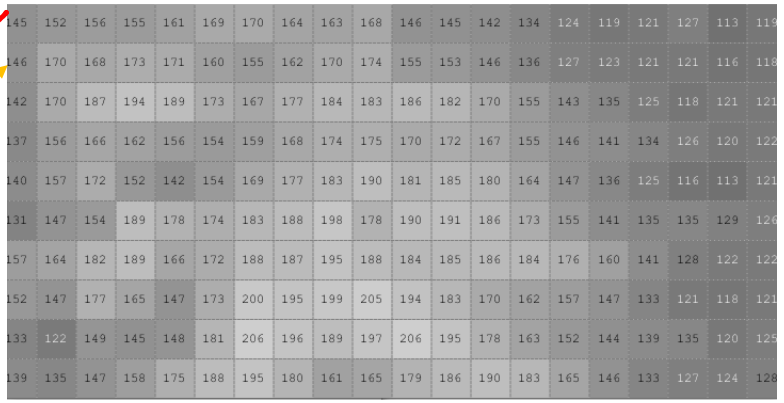
Nóbrega, Rui. (2009). Building interactive spatial and temporal models using multimodal data.

Goal of Computer Vision

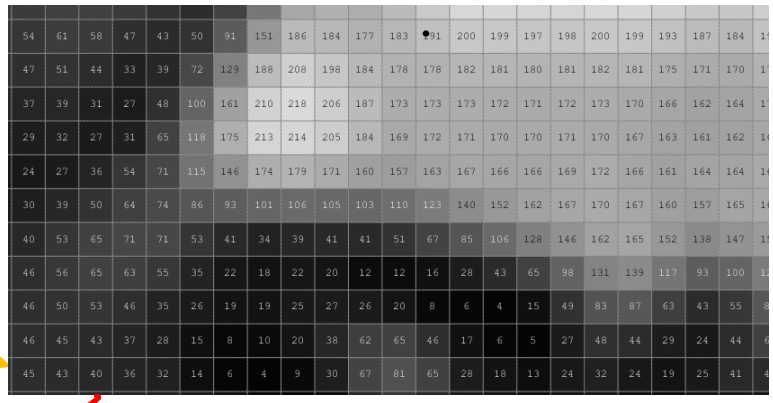
- To bridge the gap between "image pixels" and its "sematic meaning"







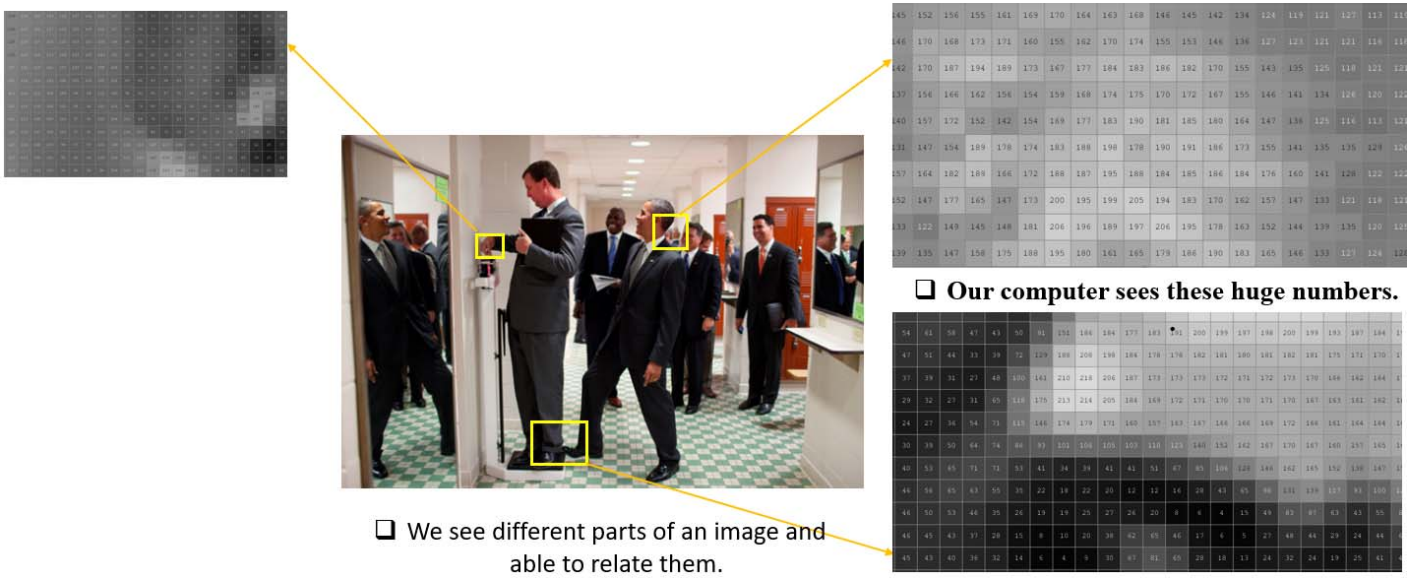
- **Our computer sees these huge numbers.**



- We see different parts of an image and able to relate them.

Goal of Computer Vision

- To bridge the gap between "image pixels" and its "sematic meaning"



- Our computer sees these huge numbers.
- We see different parts of an image and able to relate them.

- In other words: The goal of Computer Vision is "to enable computers to extract meaning of every pixel in a image (frames of a given video) , relate them with every other pixels in order to interprets the image or a video.
- So, Computer Vision mimics the human perception.

Goal of Computer Vision

- Every image tells a story

A picture is worth a thousand words

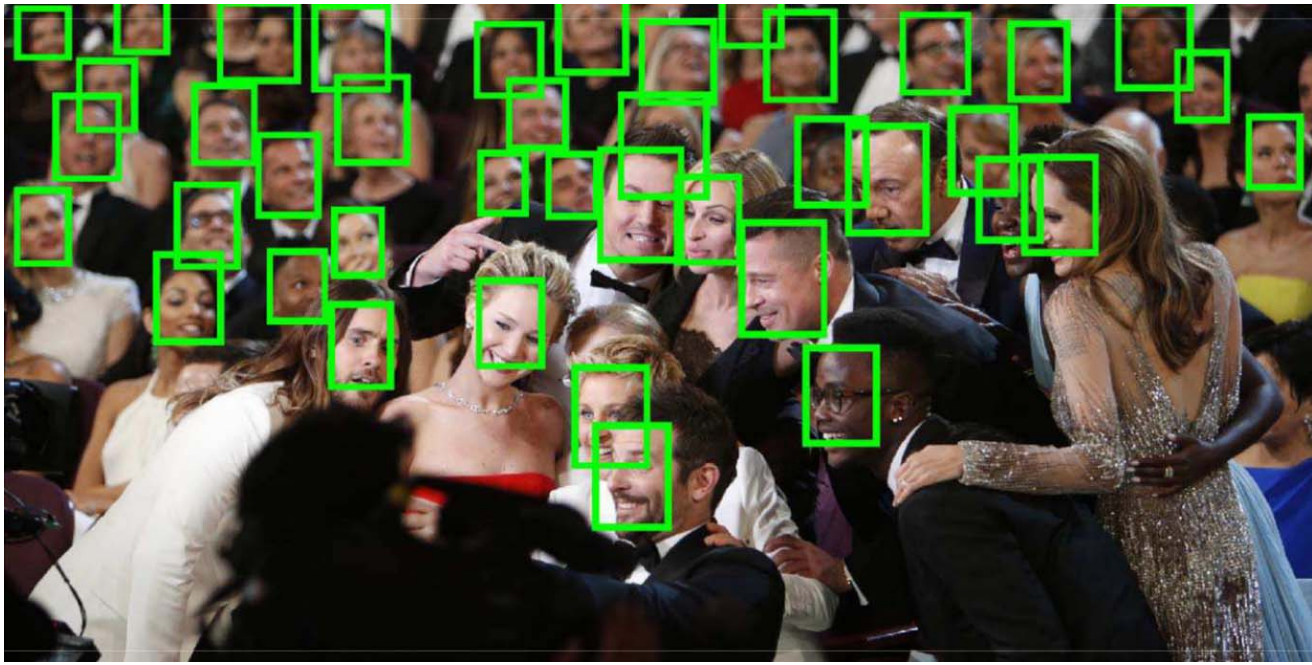
- Can computer do so? … It is still harder… and things are in progress

Applications of Computer Vision

- There are many applications …
  - Face detection
  - Human identification
  - Facial expression recognition
  - Object recognition and localization
  - Segmentation : Medical imaging
  - Robotics
  - Self-driving car
  - Virtual and augmented reality
  - Lip reading
  - Disease recognition (Covid-19)
  - Person identification in crowd
  - Multi-object tracking
  - Action recognition
  - Video surveillance
  - Image and video synthesis
  - Vision-based biometrics
  - Many more …

CV Tasks : Face Detection

Problem -1 : Does an image or a video contain one or more faces? If yes, localize them.

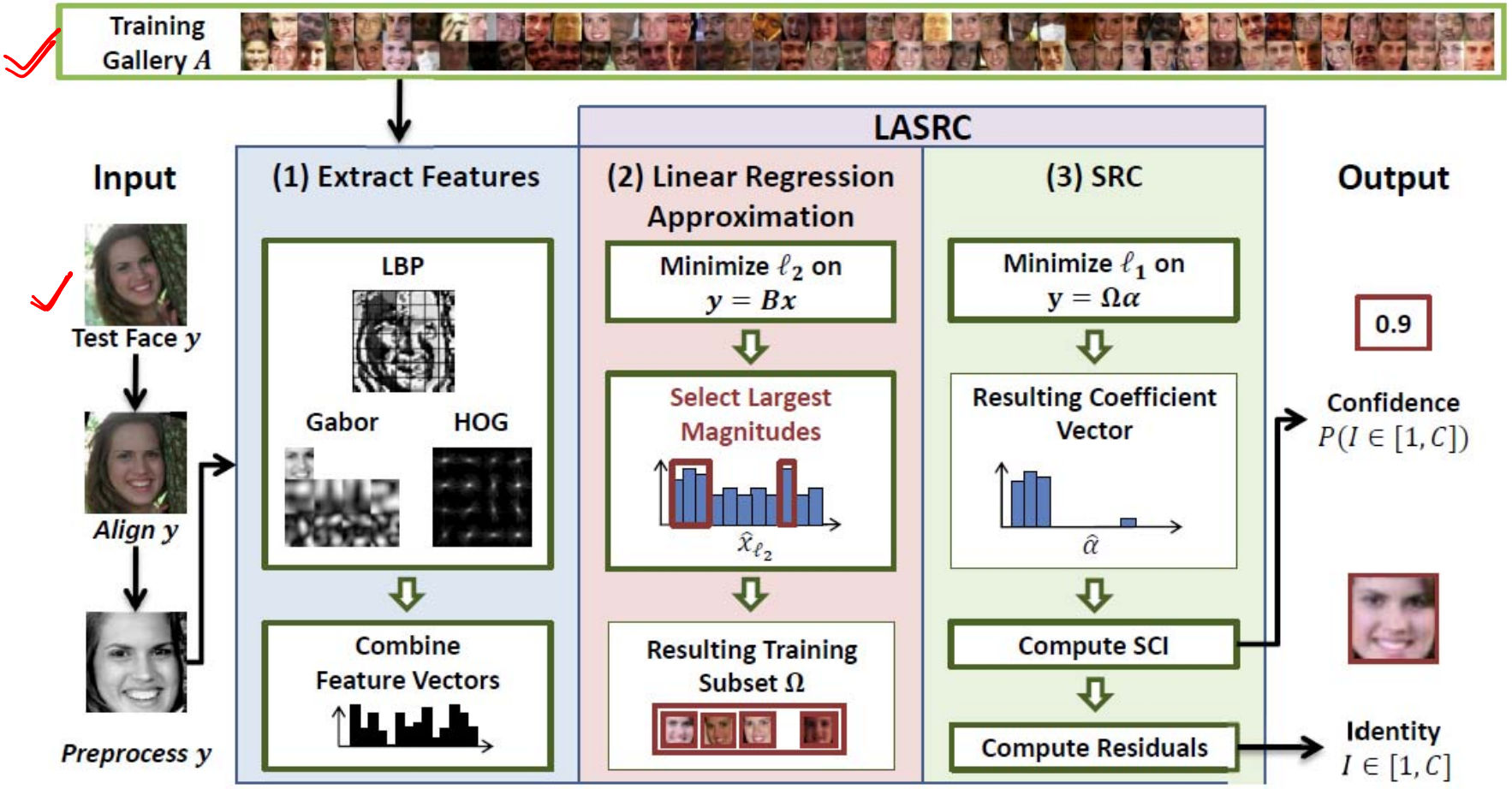


https://modelcards.withgoogle.com/face-detection

Yang, Shuo, et al. "From facial parts responses to face detection: A deep learning approach." *Proceedings of the IEEE international conference on computer vision (ICCV)*. 2015.

CV Tasks : Open-Universe Face Identification

Problem -2(a)：STILL-IMAGE, OPEN-UNIVERSE FACE IDENTIFICATION

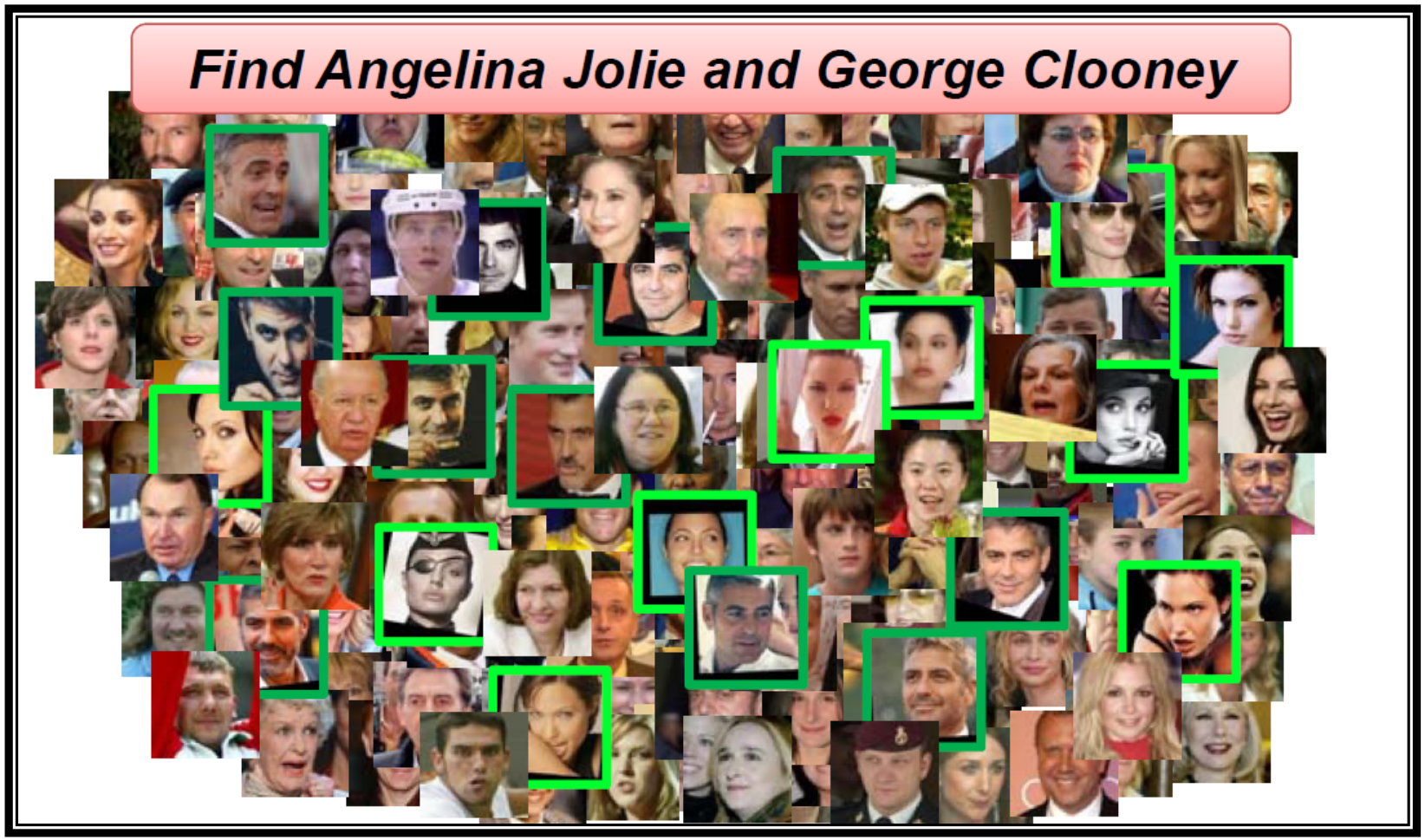


CV Tasks : Open-Universe Face Identification

Problem -2(a) : STILL-IMAGE, OPEN-UNIVERSE FACE IDENTIFICATION

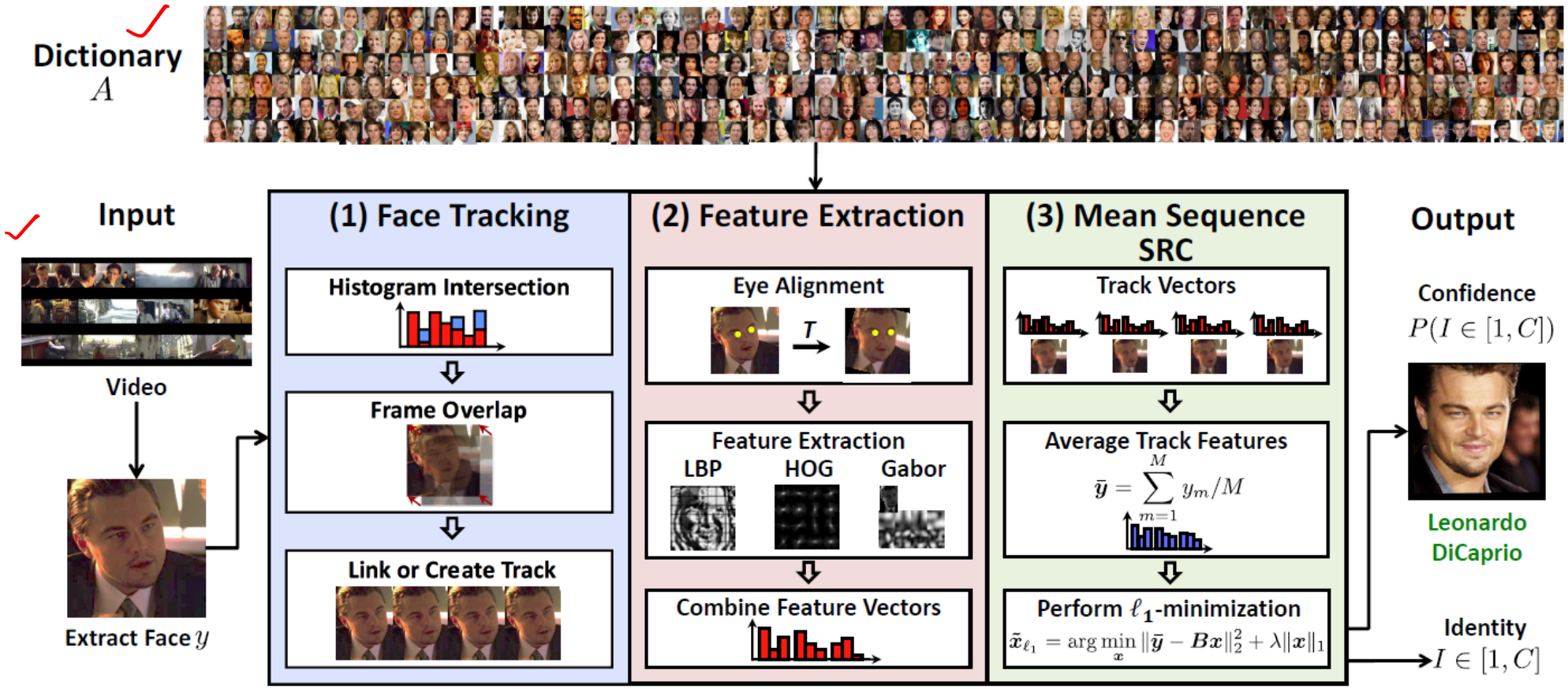
CV Tasks : Open-Universe Face Identification

Problem -2(a) : STILL-IMAGE, OPEN-UNIVERSE FACE IDENTIFICATION



CV Tasks : Open-Universe Face Identification

Problem -2 (b) : Video, OPEN-UNIVERSE FACE IDENTIFICATION



Thesis : - Taming Wild Faces: Web-Scale, Open-Universe Face Identification in Still and Video Imagery

CV Tasks : Object Recognition

Problem -3 : Given an image I, does it contain image of a person?

YES

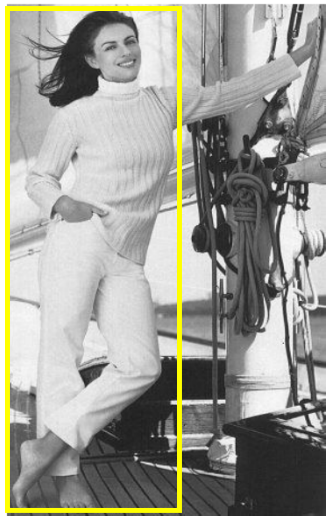
CV Tasks : Object Recognition

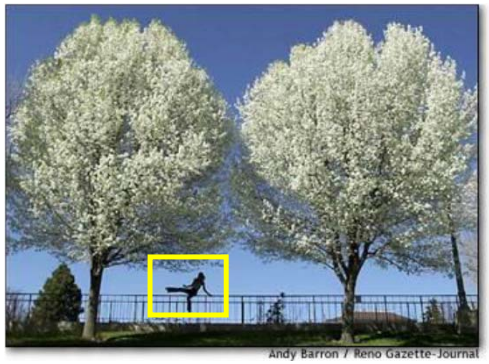
Problem -3 : Given an image I, does it contain image of a person?

NO

CV Tasks : Object Localization

Problem -4 : Where is the object (say "person") in a given image I?





CV Tasks : Image Segmentation

- Image segmentation is the process of assigning a label to every pixel in an image in such way that pixels with the same label share certain characteristics.
- In image segmentation :
  - Input – image
  - Output – regions, structures

Example-1

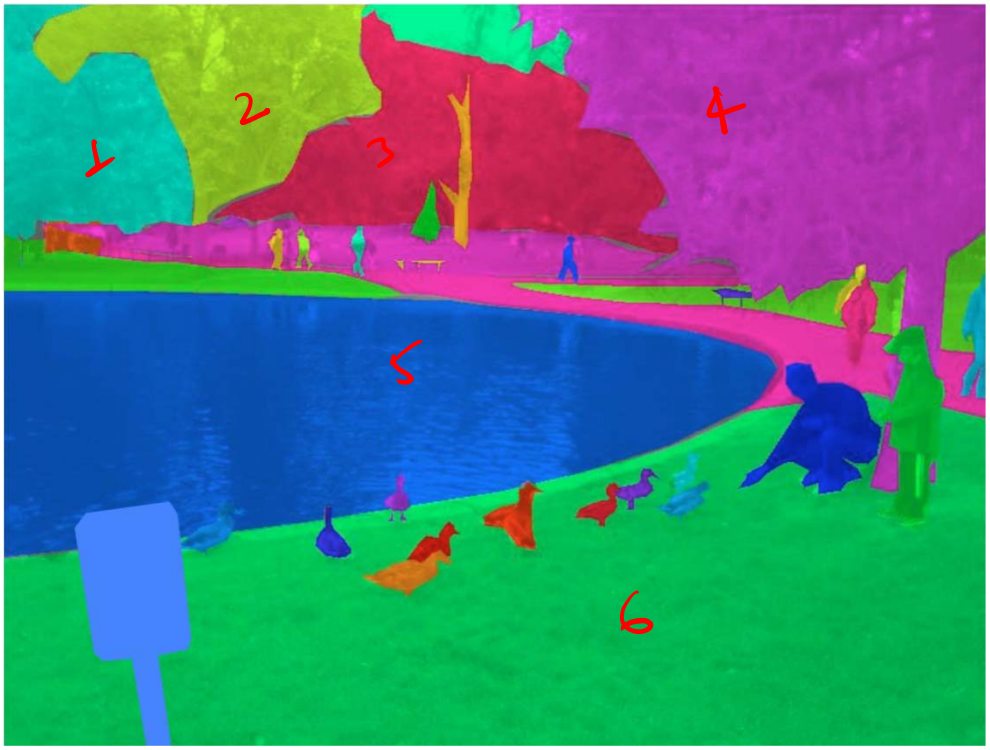
https://medium.com/cogitotech/what-is-the-difference-between-image-segmentation-and-classification-in-image-processing-303d1f660626

CV Tasks : Image Segmentation

Problem -5 :

- In image segmentation :
  - Input – image
  - Output – regions, structures

(a) : Input image



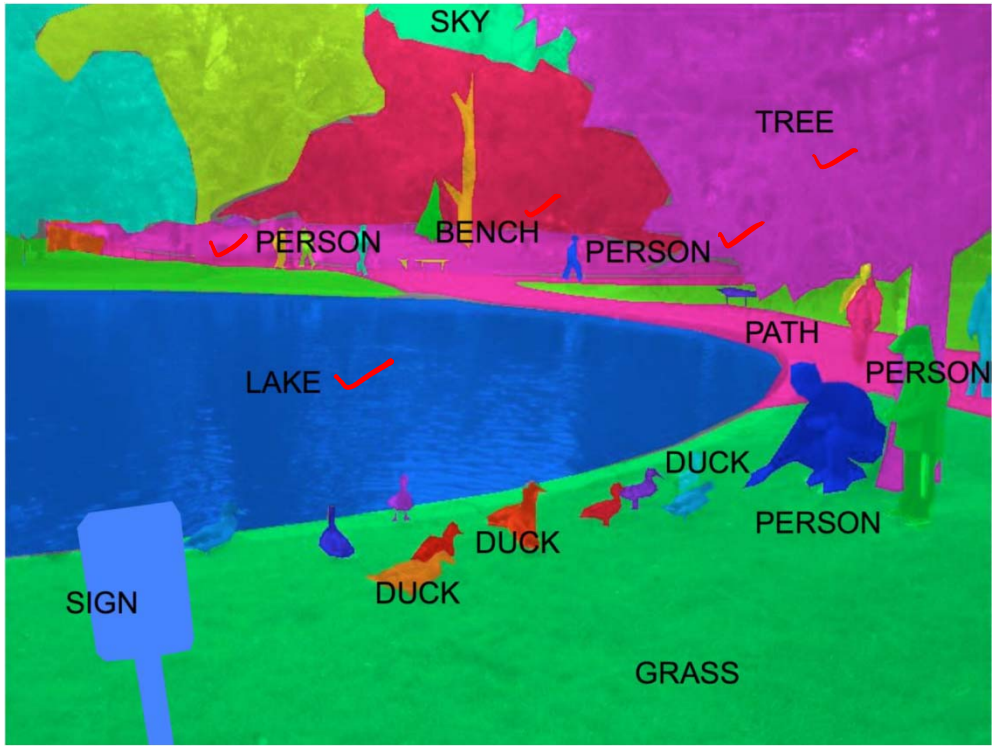
(b) Output labelled image

Example-2

CV Tasks : Semantic Segmentation

- Problem -6 : In semantic image segmentation,
- Idea : recognizing, understanding what is in the image at pixel level.
- It is a challenging task due to many reasons.

(a) : Input image



(b) Output : recognized different objects with their labels

CV Tasks : Semantic Segmentation

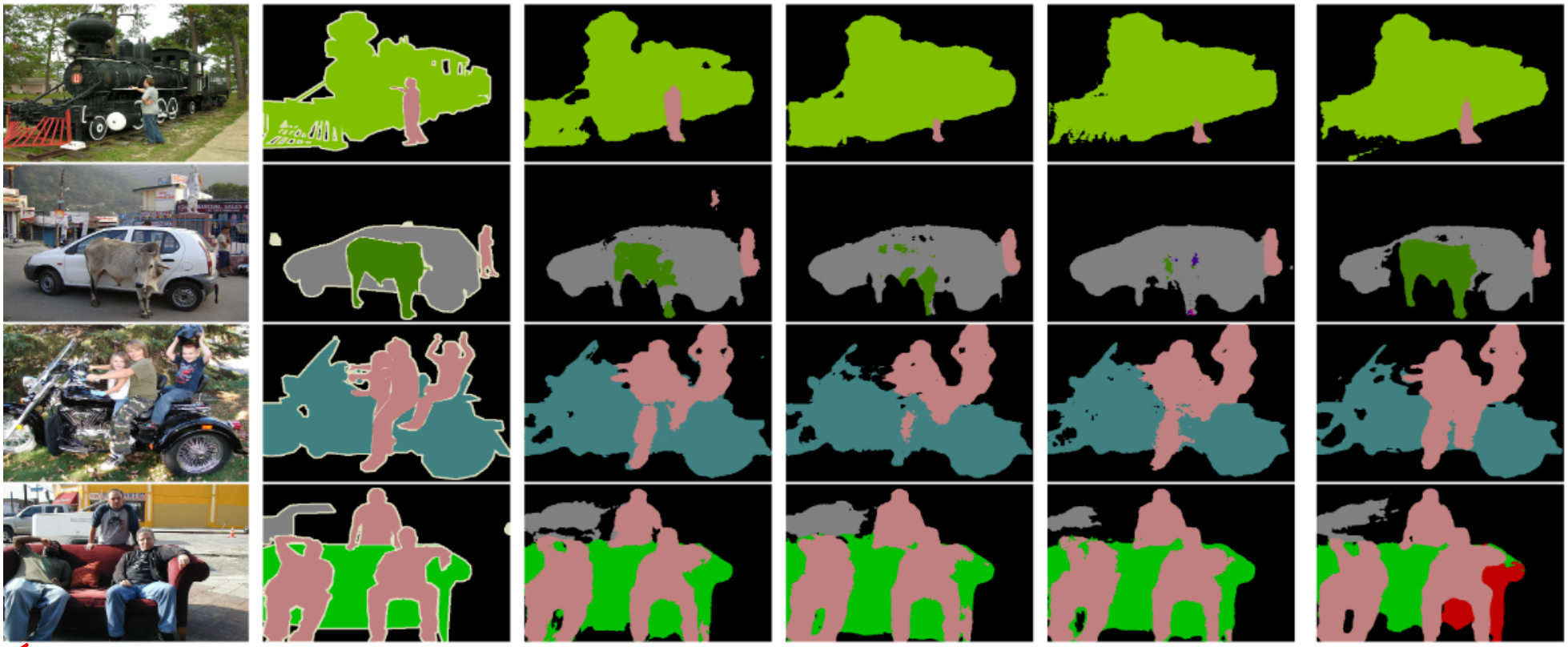
- Problem -7 : In semantic image segmentation,
- Idea : recognizing, understanding what is in the image at pixel level.
- Going one more deeper level and try to understand what is happening in the image

(a) : Input image

(b) Output Understating from the image

CV Tasks : Semantic Segmentation

- Problem -7 : In semantic image segmentation,
- Idea : recognizing, understanding what is in the image at pixel level.



Ibrahim, Mostafa S., et al. "Semi-supervised semantic image segmentation with self-correcting networks." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.

CV Tasks : Counting in Extremely Dense Crowd Images

- Problem -8 : Count the no. of people in a given extremely dense crowed image

Ground truth/proposed [1] = 673/**653**

Ground truth/proposed [1] = 2322/**2203**

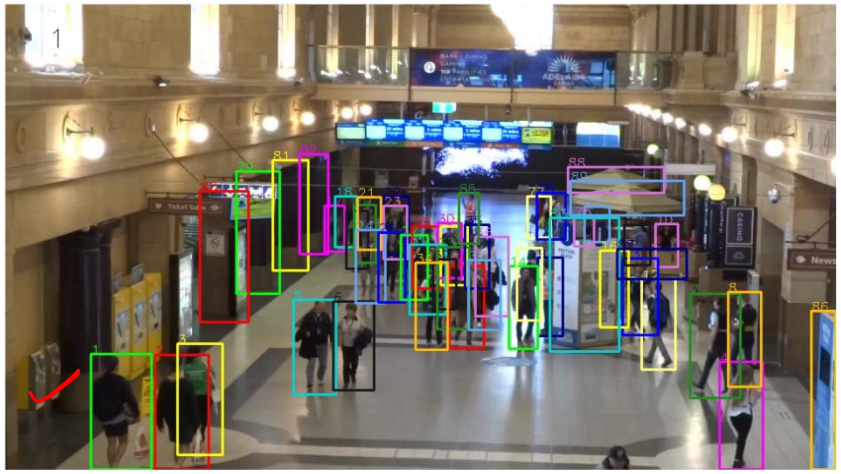
[1] Idrees, Haroon, et al. "Multi-source multi-scale counting in extremely dense crowd images." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2013.

CV Tasks : MOT20 Dataset

Problem -9 : A benchmark for multi object tracking in crowded scenes

(a)

(b)



(c)

(a) Image frame of a video
(b) Ground truth
(c) Faster-RCNN

CV Tasks : Multiple-Object Tracking

Problem -9 : Multiple Object Tracking with Correlation Learning

Wang, Qiang, et al. "Multiple Object Tracking with Correlation Learning." *Proceedings of the IEEE/CVF Conference on **Computer Vision and Pattern Recognition***. **2021**.

CV Tasks : Multiple-Object Tracking

Problem -9 : Discriminative Appearance Modeling with Multi-track Pooling for Real-time Multi-object Tracking

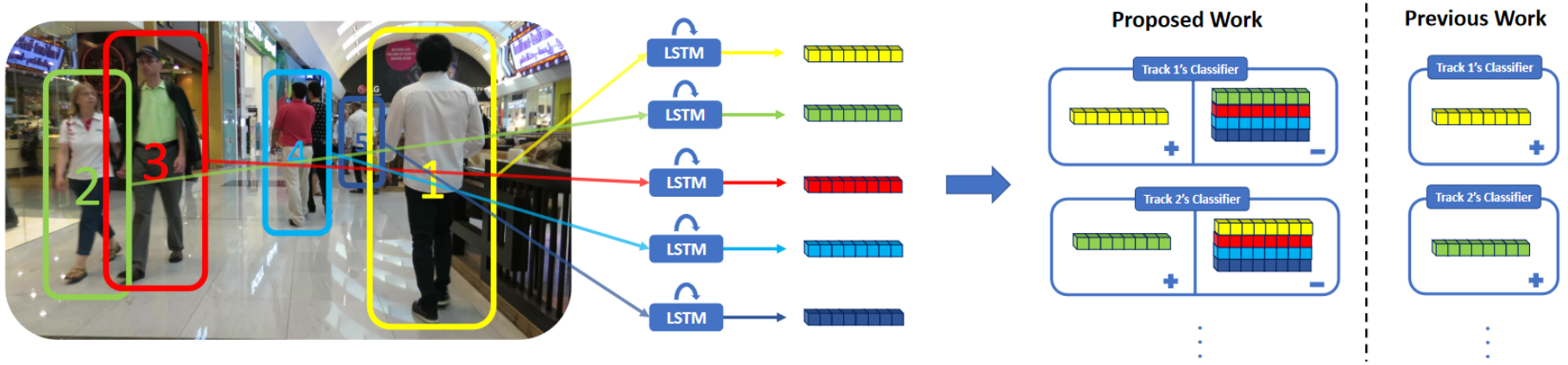


Figure 1: Existing recurrent neural network-based track classifiers used only matched detections for updating its appearance memory during tracking. This does not consider other objects in the scene (i.e. negative examples), which may have similar appearances. We propose to improve the predicted likelihood of such a classifier by augmenting its memory with appearance information about other tracks in the scene with multi-track pooling, leveraging the appearance information from the full set of tracks in the scene. The resulting classifier learns to adapt its prediction based on the information from other tracks in the scene.

Kim, Chanho, et al. "Discriminative Appearance Modeling with Multi-track Pooling for Real-time Multi-object Tracking." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

The code and trained models are available at `https://github.com/chkim403/blstm-mtp`.

CV Tasks : Multiple-Object Tracking

Problem -10 : Detection, Tracking, and Counting Meets Drones in Crowds: A Benchmark

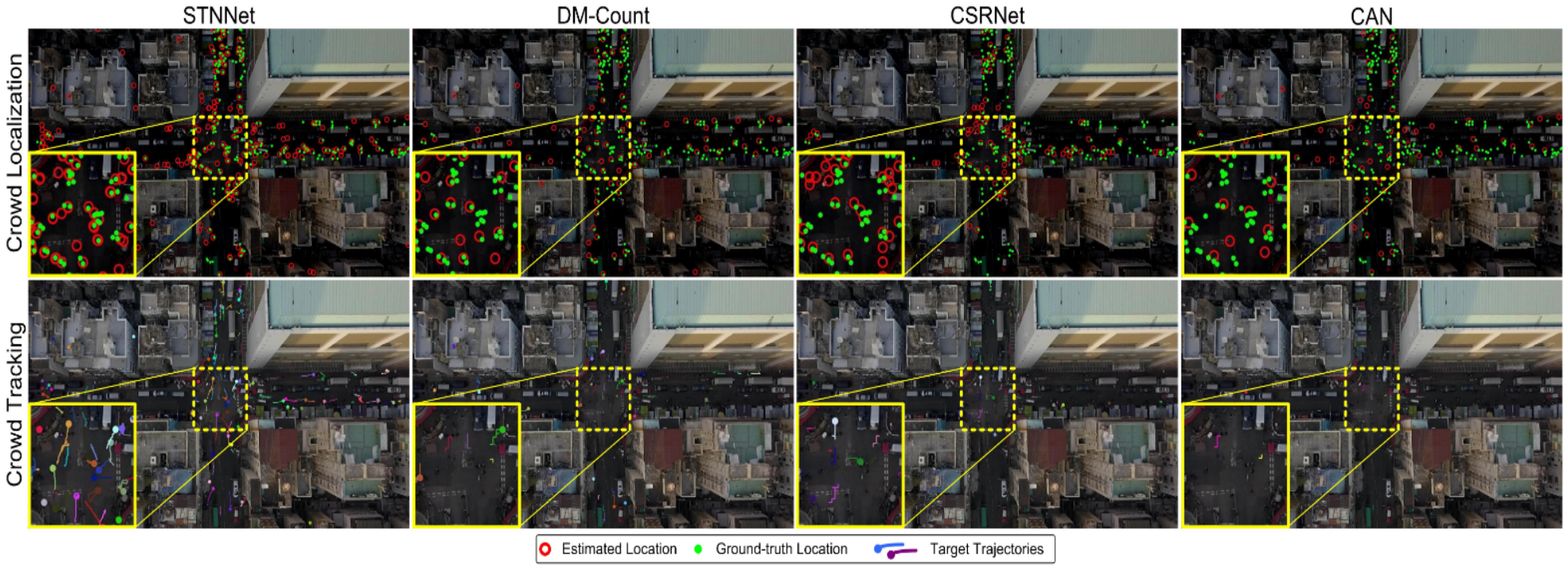


Figure 6. Qualitative results of DM-Count [34], CSRNet [15], CAN [17], and our STNNet on DroneCrowd. Best view in color version.

STNNet : Wen, Longyin, et al. "Detection, Tracking, and Counting Meets Drones in Crowds: A Benchmark." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

CV Tasks : Image Dehazing

Problem -11 : Given a haze image I, How can we remove haze in order to get clear sharpened image.

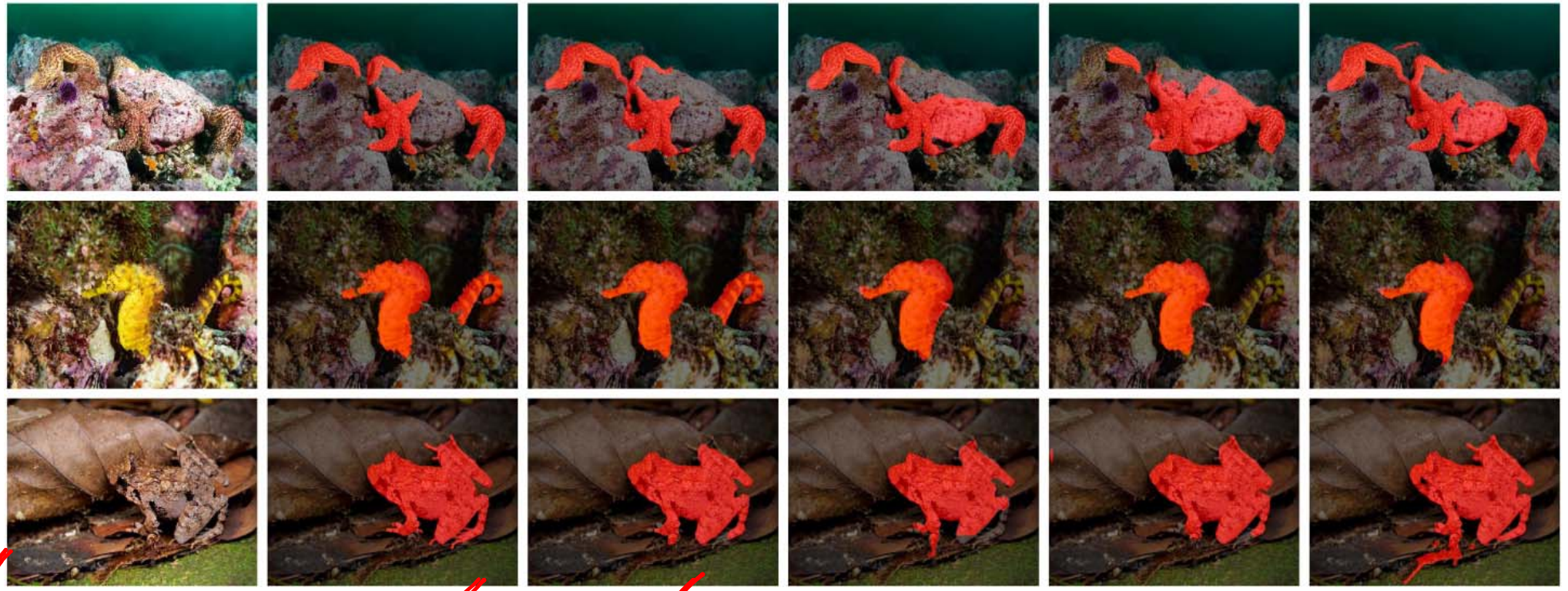
(a) Hazy input (b) DCP [17] (c) DehazeNet [5] (d) AOD-Net [25] (e) GridDehazeNet [30]

(f) FFA-Net [34] (g) MSBDN [10] (h) KDDN [23] (i) Ours (j) Ground-truth

Wu, Haiyan, et al. "Contrastive Learning for Compact Single Image Dehazing." *Proceedings of the IEEE/CVF Conference on **Computer Vision and Pattern Recognition**.* **2021**. (Code available)

CV Tasks : Camouflaged Object Segmentation

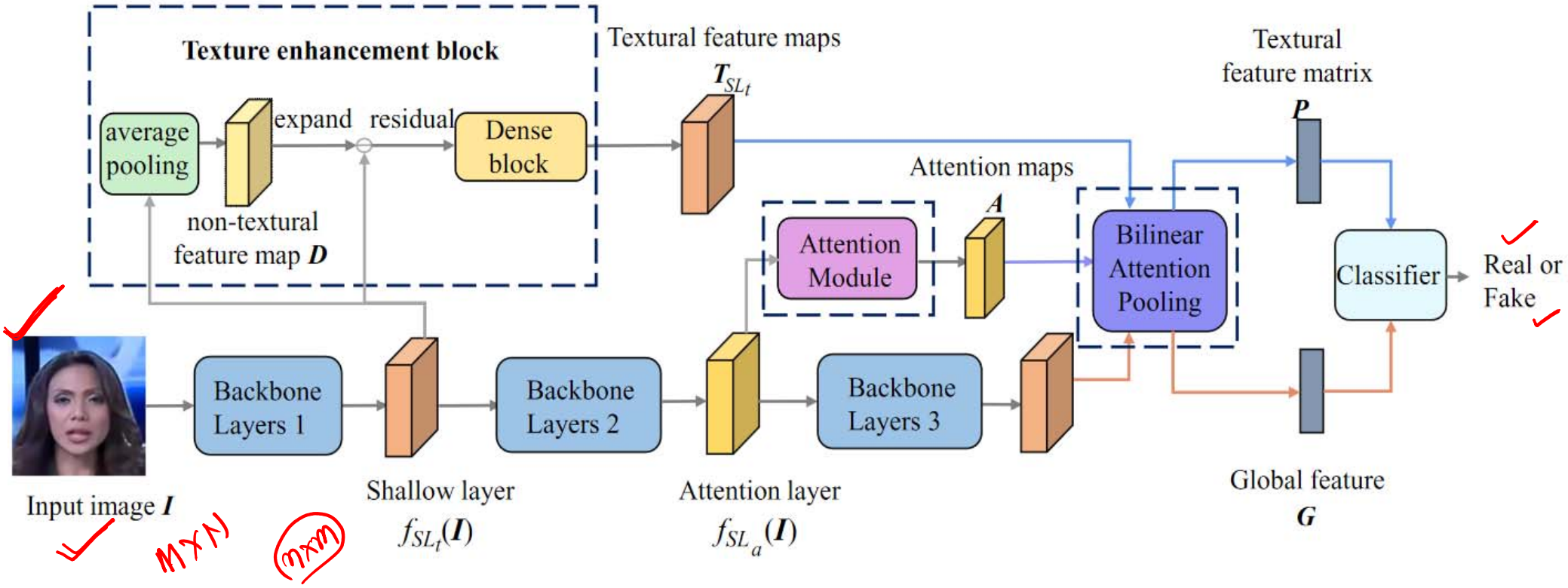
Problem -12 : What if the object color is similar to the background? How do we come up with methods which can solve segmentation problem in underline environment.



Mei, Haiyang, et al. "Camouflaged object segmentation with distraction mining." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021. (Code available)

CV Tasks : Image Forgery

Problem -13 : How to detect forgery (global and/or local) encountered in a given image.



Zhao, Hanqing, et al. "Multi-attentional deepfake detection." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021. . (Code available)

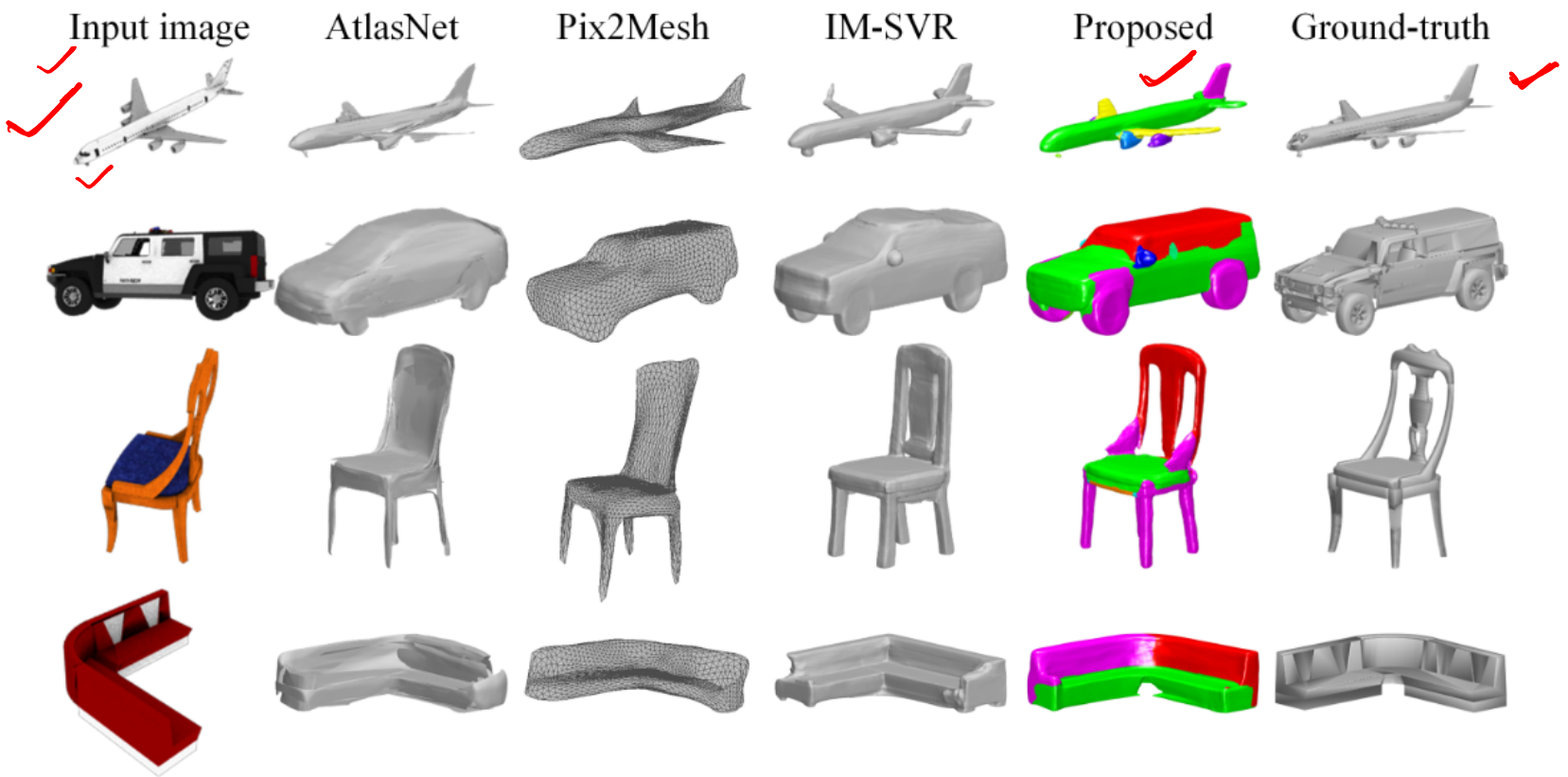
CV Tasks : Image Forgery

Problem -13 : How to detect forgery (global and/or local) encountered in a given image.

Wang, Chengrui, and Weihong Deng. "Representative Forgery Mining for Fake Face Detection." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021. (https://github.com/crywang/RFM.)

CV Tasks : 3D-Reconstruction from 2D-Images

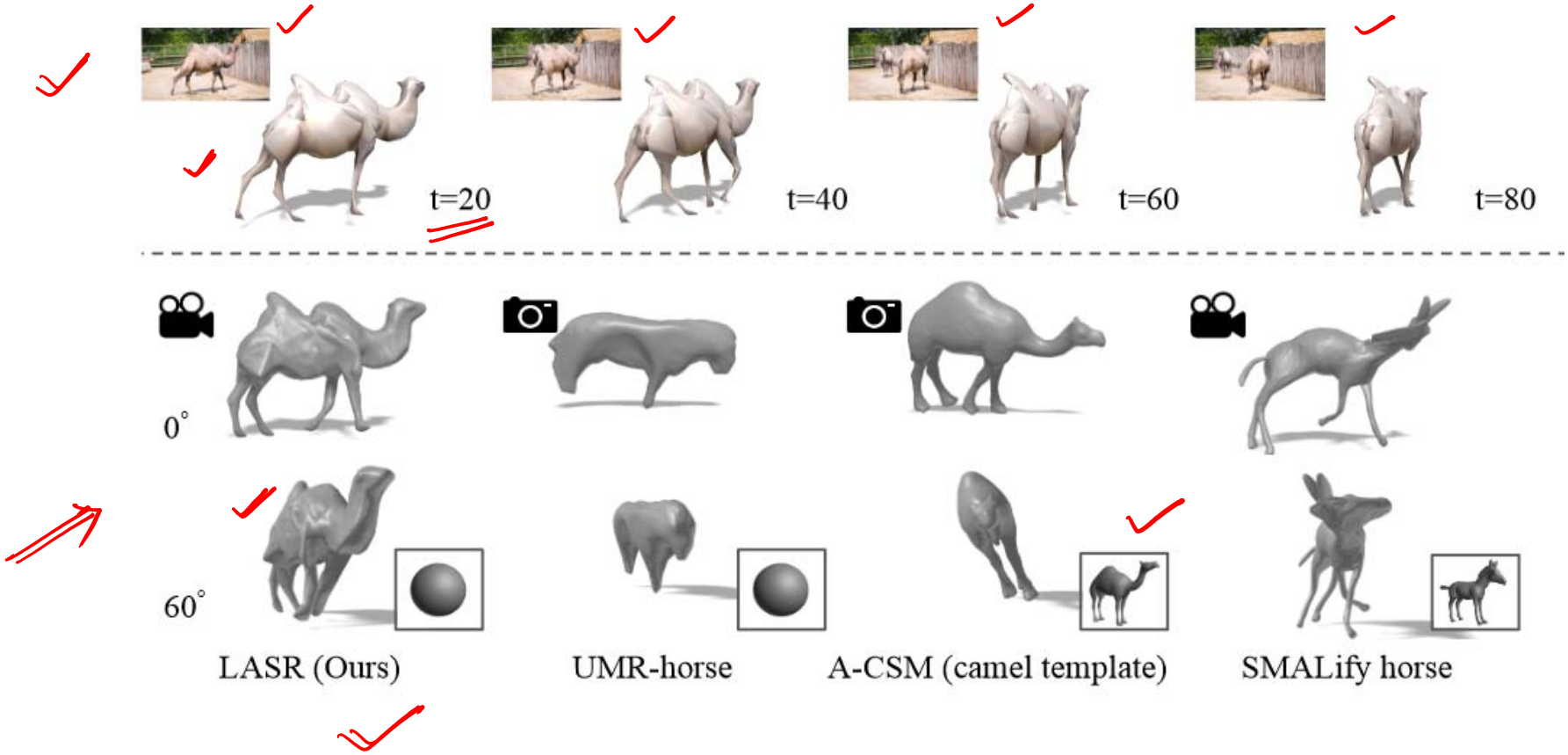
Problem -14 : 3D structure of objects observed from a single view is a fundamental computer vision problem



Liu, Feng, Luan Tran, and Xiaoming Liu. "Fully Understanding Generic Objects: Modeling, Segmentation, and Reconstruction." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021. Code is available at http://cvlab.cse.msu.edu/project-fully3dobject.html.

CV Tasks : Articulated 3D-shape Reconstruction

Problem -15 : Articulated 3D-shape reconstruction from video as input.



Yang, Gengshan, et al. "LASR: Learning Articulated Shape Reconstruction from a Monocular Video." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021. Code is available at lasr-google.github.io

CV Tasks : Self-Driving Car

Problem -16 : How does the Computer Vision algorithms help in self-driving cars?

Waymo

Tesla's Model S Autopilot is Amazing! - YouTube

Some of the Major Challenges in CV Tasks

1. Viewpoint variation : same object appears differently

View-1

View-2

View-3

2. Illumination – poor illumination is one of the major challenges in CV

Some of the Major Challenges in CV Tasks

- 3. Intra-class variability:

Car

Car

Car

- 4. Background clutter and Occlusion

Conferences and Journals in Computer Vision

- Conferences:
  - Computer Vision and Pattern Recognition (CVPR)
  - International Conference on Computer Vision (ICCV)
  - European Conference on Computer Vision (ECCV)
  - International Conference on Pattern Recognition (ICPR)
  - Asian Conference on Computer Vision (ACCV)
  - International Conference on Image Processing (ICIP)
- Journals :
  - IEEE Tran. on Pattern Analysis and Machine Intelligence (PAMI)
  - IEEE Transactions on Image Processing
  - International Journal of Computer Vision
  - Computer Vision and Image Understanding
  - …so on

Evaluation Pattern

- Major Exam ---------- 25
- Assignments / Quiz ---------- 25
- Project work -------- 50
  - Work done – 20/25
  - Report ------ 10
  - Paper (Journal/Conference) – 20/15
  - Group formation / problem selection : 5th Oct. 2021
  - Final submission date : 20th Nov. 2021

==========

Total = 100

==========

Pre-requisite of this course

- Fundamental of image processing
- Machine learning
- Basic probability and linear algebra
- Python

Reference

- Richard Szeliski, Computer Vision: Algorithms and Applications, Springer, 2010 (online draft),
- Mubarak Shah, "Fundamentals of Computer Vision" (Online available)
- Ian Goodfellow, Yoshua Bengio and Aaron Courville, "Deep Learning" (Online available)

Acknowledgement!

Sources for this lecture include materials from works by Szeliski, Abhijit Mahalanobis, Sedat Ozer, Ulas Bagci, Mubarak Shah, Antonio Torralba, and others. References are given for the source image contents.

Prof. Fei Fei Li

Queries!