COMPUTER VISION
CS-6350

Prof. Sukhendu Das
Deptt. of Computer Science and Engg.,
IIT Madras, Chennai – 600036.

Email: sdas@cse.iitm.ac.in
URL: //www.cse.iitm.ac.in/~sdas
//www.cse.iitm.ac.in/~vplab/computer_vision.html

JULY. – 2023.

INTRODUCTION

Contents to be covered

Use slides as brief :
Points, concepts, links

These are not substitute for materials in books

References

1. "Digital Image Processing"; R. C. Gonzalez and R. E. Woods; Addison Wesley; 1992+.
2. "Computer Vision: Algorithms and Applications"; by Richard Szeliski; Springer-Verlag London Limited 2011.
3. "Multiple View geometry"; R. Hartley and A. Zisserman, 2002 Cambridge university Press.
4. "Pattern Recognition and Machine Learning"; Christopher M. Bishop; Springer, 2006.
5. "Digital Image Processing and Computer Vision"; Robert J. Schallkoff; John Wiley and Sons; 1989+.
6. "Pattern Recognition: Statistical. Structural and Neural Approaches"; Robert J. Schallkoff; John Wiley and Sons; 1992+.
7. "3-D Computer Vision"; Y. Shirai; Springer-Verlag, 1984
8. "Computer Vision: A Modern Approach"; D. A. Forsyth and J. Ponce; Pearson Education; 2003+.

References (Contd..)

Journals:

- IEEE-T-PAMI (Transactions on Pattern Analysis and Machine Intelligence)
- IEEE-T-IP (Transactions on Image processing)
- PR (Pattern Recognition)
- PRL (Pattern Recognition Letters)
- CVIU (Computer Vision, Image Understanding)
- IJCV (International Journal of Computer Vision)

Online links

1. CV online: **http://homepages.inf.ed.ac.uk/rbf/CVonline**
2. Computer Vision Homepage:

http://www-2.cs.cmu.edu/afs/cs/project/cil/ftp/html/vision.html

Typical Distribution of marks for Evaluation/grading

Quiz (50 mins.) -	**15 - 20**
End Sem exam (120-150 mins.) -	**35 – 40**
TPA -	**30 - 35**
TUTs -	**10 - 15**
Total	**100**

+/- 05 marks variation at any part;
To be finalized well before End Sem Exam.

Pre-Req: - Linear Algebra; Geometry; Stat&Prob basics; Calculus basics; DSP, Programming, Data Structure basics

July-Nov '23

<table>
<tr><th>Days</th><th>8.00 - 8.50</th><th>9.00 - 9.50</th><th>10.00 - 10.50</th><th>11.00 - 11.50</th><th>12.00 – 12.50</th><th rowspan="11">LUNCH</th><th>14.00 - 15.15</th><th>15.30 - 16.45</th><th>17.00 - 17.50</th></tr>
<tr><td rowspan="2">Mon</td><td rowspan="2">A</td><td rowspan="2">B</td><td rowspan="2">C</td><td rowspan="2">D</td><td rowspan="2">G</td><td colspan="2">P</td><td rowspan="2">J/J3</td></tr>
<tr><td>H/H1</td><td>M/M2</td></tr>
<tr><td rowspan="2">Tue</td><td rowspan="2">B</td><td rowspan="2">C</td><td rowspan="2">D</td><td rowspan="2">E</td><td rowspan="2">A</td><td colspan="2">Q</td><td rowspan="2">F</td></tr>
<tr><td>M/M1</td><td>H/H2</td></tr>
<tr><td rowspan="2">Wed</td><td rowspan="2">C</td><td rowspan="2">D</td><td rowspan="2">E</td><td rowspan="2">F</td><td rowspan="2">B</td><td colspan="2">R</td><td rowspan="2">G</td></tr>
<tr><td>J/J1</td><td>K/K2</td></tr>
<tr><td rowspan="2">Thu</td><td rowspan="2">E</td><td rowspan="2">F</td><td rowspan="2">G</td><td rowspan="2">A</td><td rowspan="2">D</td><td colspan="2">S</td><td rowspan="2">H/H3</td></tr>
<tr><td>L/L1</td><td>J/J2</td></tr>
<tr><td rowspan="2">Fri</td><td rowspan="2">F</td><td rowspan="2">G</td><td rowspan="2">A</td><td rowspan="2">B</td><td rowspan="2">C</td><td colspan="2">T</td><td rowspan="2">E</td></tr>
<tr><td>K/K1</td><td>L/L2</td></tr>
</table>

- TUTs – Altn. weeks; Mid-sem etc.

May be held Online Occasionally

What is CVPR ?

http://cvpr2022.thecvf.com/

https://openaccess.thecvf.com/menu

https://openaccess.thecvf.com/CVPR2022

Also, check **ICCV *(26*), ECCV (*21*), NIPS**

https://scholar.google.com/citations?view_op=top_venues&hl=en

	Publication	h5-index	h5-median
1.	Nature	467	707
2.	The New England Journal of Medicine	439	876
3.	Science	424	665
4.	IEEE/CVF Conference on Computer Vision and Pattern Recognition	422	681
5.	The Lancet	368	688
6.	Nature Communications	349	456
7.	Advanced Materials	326	415
8.	Cell	316	503
9.	Neural Information Processing Systems	309	503
10.	International Conference on Learning Representations	303	563
11.	JAMA	286	476
12.	Science of The Total Environment	273	375
13.	Nature Medicine	268	459
14.	Proceedings of the National Academy of Sciences	268	394
15.	Angewandte Chemie International Edition	266	362
16.	Chemical Reviews	264	459
17.	International Conference on Machine Learning	254	463

CVPR-20 - CFP

- 3D computer vision
- Action and behavior recognition
- Adversarial learning, adversarial attack and defense methods
- Biometrics, face, gesture, body pose
- Computational photography, image and video synthesis
- Datasets and evaluation
- Efficient training and inference methods for networks
- Explainable AI, fairness, accountability, privacy, transparency ethics in vision
- Image retrieval
- Low-level and physics-based vision
- Machine learning architectures and formulations
- Medical, biological and cell microscopy
- Motion and tracking
- Neural generative models, auto encoders, GANs
- Optimization and learning methods
- Recognition (object detection, categorization)
- Representation learning, deep learning
- Scene analysis and understanding
- Segmentation, grouping and shape
- Transfer, low-shot, semi- and un- supervised learning
- Video analysis and understanding
- Vision + language, vision + other modalities
- Vision applications & systems, vision for robotics & autonomous vehicles
- Visual reasoning and logical representation

CVPR-22 - CFP

3D from multi-view & sensors
3D from single images
Action/event recognition
Adversarial attacks & defense
Behavior analysis
Biometrics
Computational photography
Computer vision theory
Computer vision for social good
Datasets and evaluation
Deep learning architectures & techniques
Document analysis and understanding
Efficient learning and inference
Explainable computer vision
Face and gesture
Image and video synthesis and generation
Low-level vision
Machine learning
Medical, biological and cell microscopy
Motion and tracking
Navigation and autonomous driving
Optimization methods

Photogrammetry and remote sensing
Physics-based vision and shape-from-X
Pose estimation and tracking
Privacy & federated learning
Recognition: detection, categorization, retrieval
Representation learning
RGBD sensors and analytics
Robot vision
Scene analysis and understanding
Segmentation, grouping & shape analysis
Statistical methods
Transfer/low-shot/long-tail learning
Transparency, fairness, accountability, privacy & ethics in vision
Self-/semi-/meta-/unsupervised learning
Video analysis and understanding
Vision + graphics
Vision + language
Vision + X
Vision applications and systems
Visual reasoning
Others

3D from single images
Adversarial attack and defense
Autonomous driving
Biometrics
Computational imaging
Computer vision for social good
Computer vision theory
Datasets and evaluation
Deep learning architectures and techniques
Document analysis and understanding
Efficient and scalable vision
Embodied vision: Active agents, simulation
Explainable computer vision
Humans: Face, body, pose, gesture, movement
Image and video synthesis and generation
Low-level vision
Machine learning (other than deep learning)
Medical and biological vision, cell microscopy

Multi-modal learning
Optimization methods (other than deep

Photogrammetry and remote sensing
Physics-based vision and shape-from-X
Recognition: Categorization, detection, retrieval

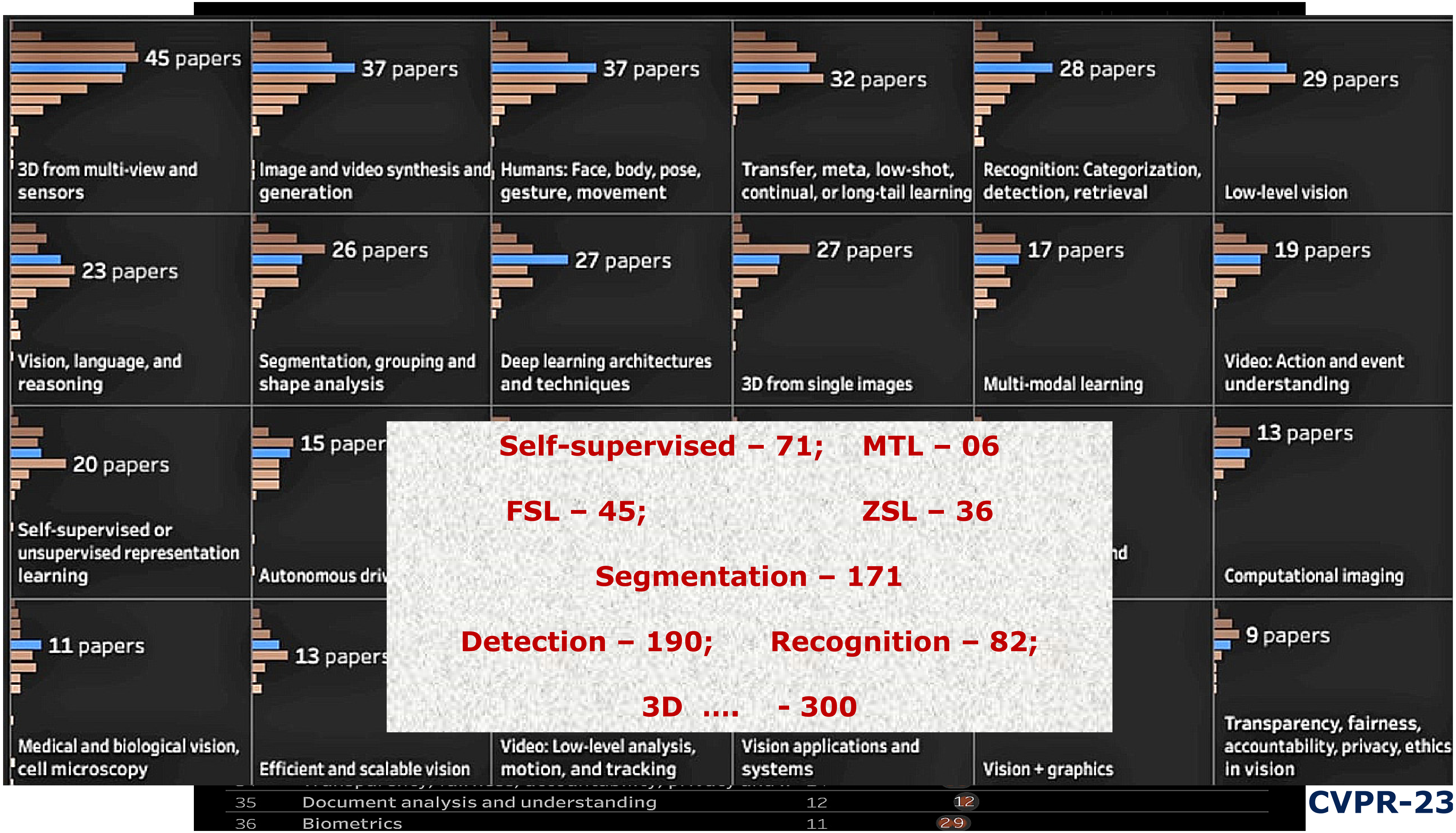
Robotics
Scene analysis and understanding
Segmentation, grouping and shape analysis
Self-supervised or unsupervised representation learning

Transfer, meta, low-shot, continual, or long-tail learning

Transparency, fairness, accountability, privacy, ethics in vision

Video: Action and event understanding
Video: Low-level analysis, motion, and tracking
Vision + graphics
Vision, language, and reasoning
Vision applications and systems

CVPR – 2022-3



Self-supervised – 71; MTL – 06

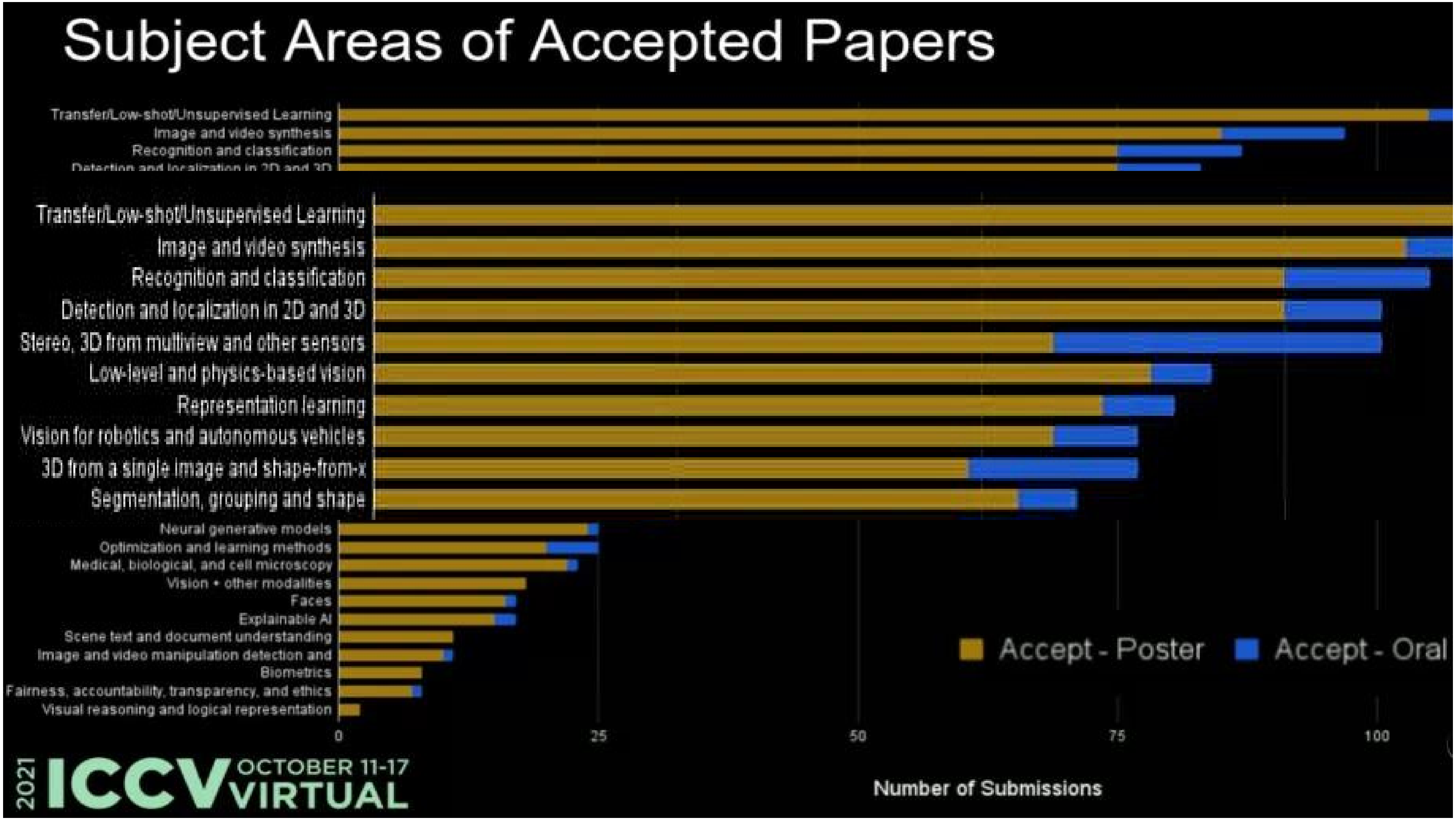
FSL – 45; ZSL – 36

Segmentation – 171

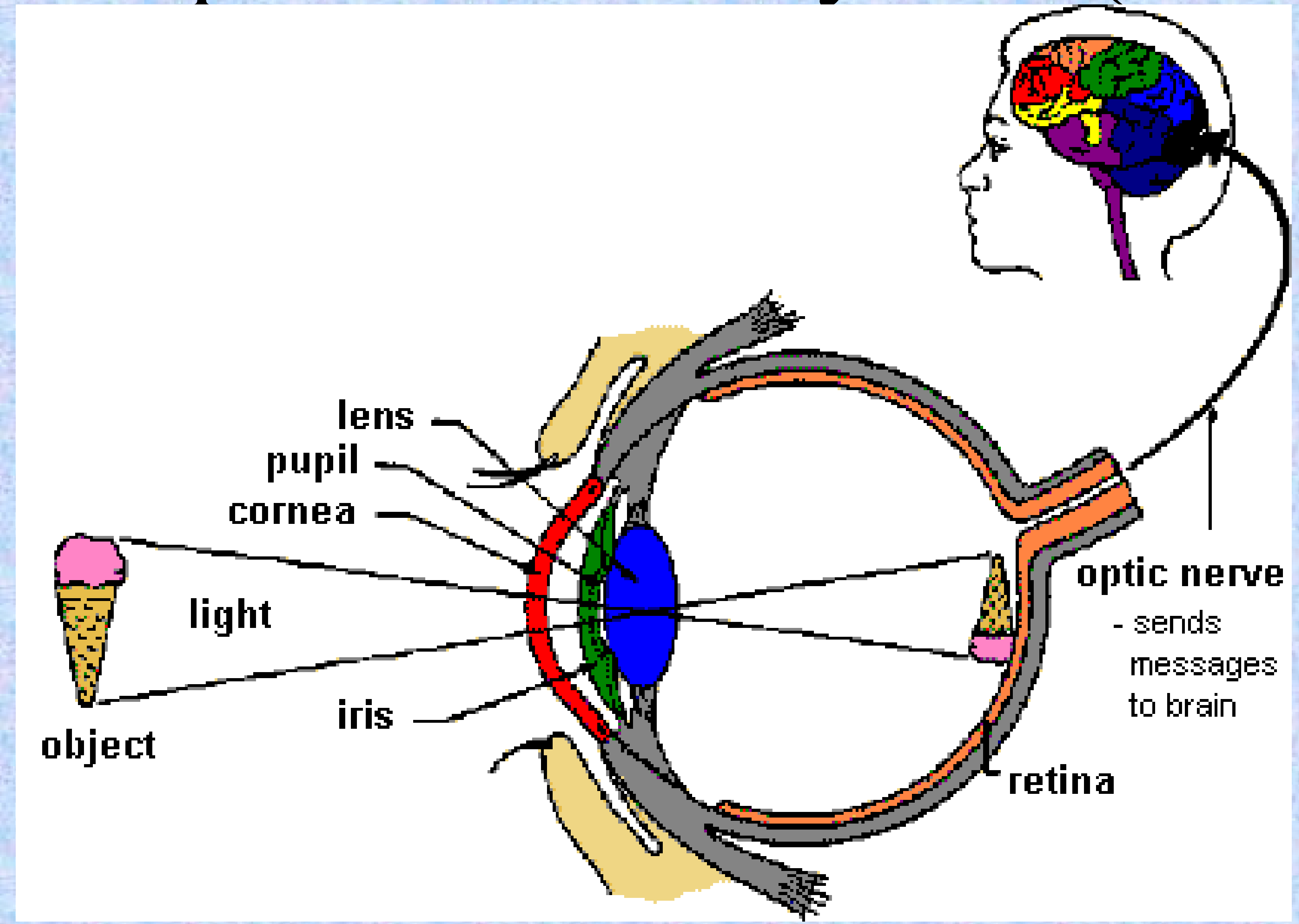
Detection – 190; Recognition – 82;

3D .... - 300

CVPR-23

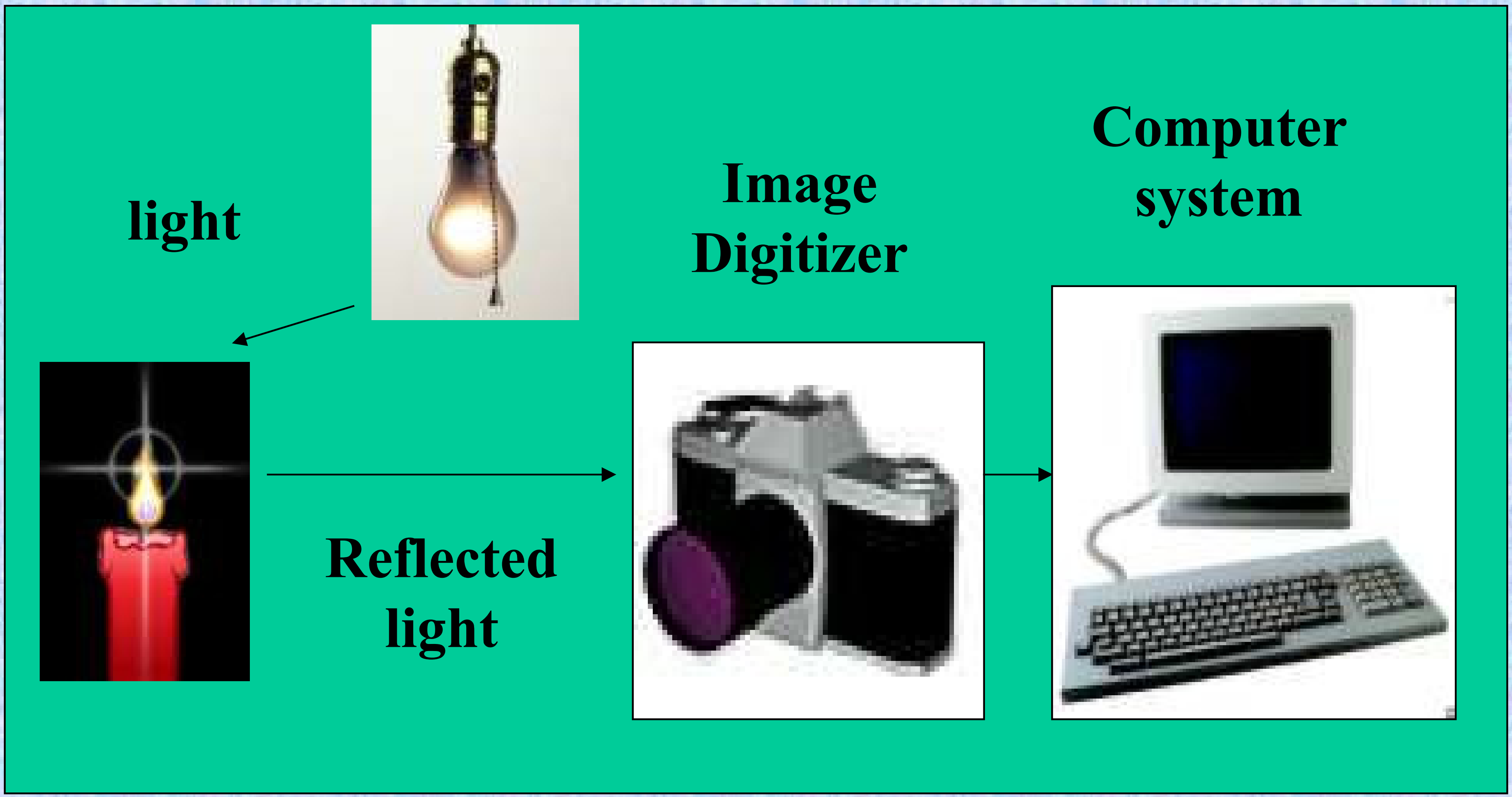

Subject Areas of Accepted Papers
Transfer/Low-shot/Unsupervised Learning
Image and video synthesis
Recognition and classification
Detection and localization in 2D and 3D
Stereo, 3D from multiview and other sensors
Low-level and physics-based vision
Representation learning
Vision for robotics and autonomous vehicles
3D from a single image and shape-from-x
Segmentation, grouping and shape
Neural generative models
Optimization and learning methods
Medical, biological, and cell microscopy
Vision + other modalities
Faces
Explainable AI
Scene text and document understanding
Image and video manipulation detection and
Biometrics
Fairness, accountability, transparency, and ethics
Visual reasoning and logical representation
0
25
50
75
100
Number of Submissions
Accept - Poster
Accept - Oral
2021 ICCV OCTOBER 11-17 VIRTUAL

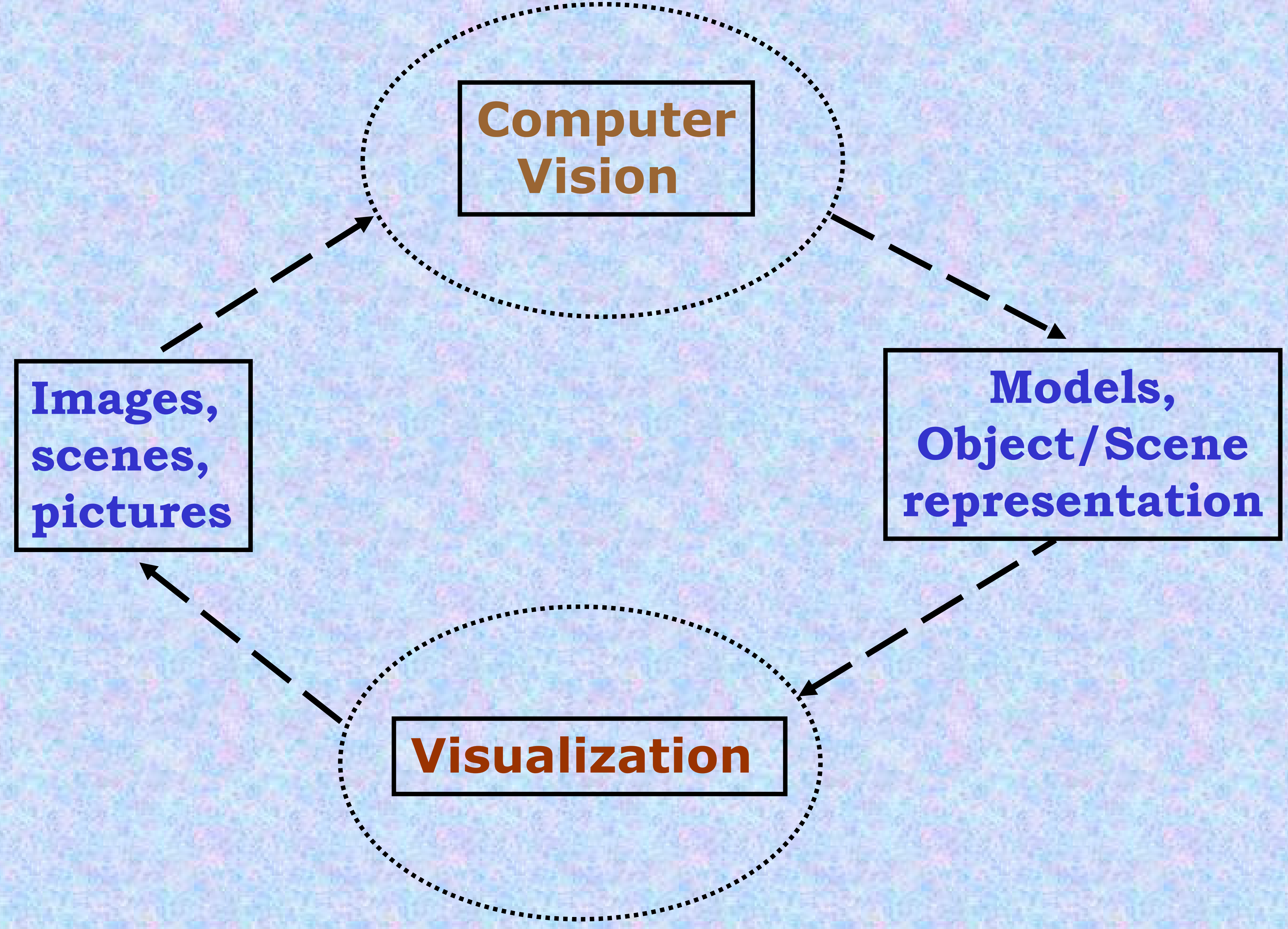
Human Vision System (HVS) Vs. Computer Vision System (CVS)



The Optics of the eye

A computer Vision System (CVS)



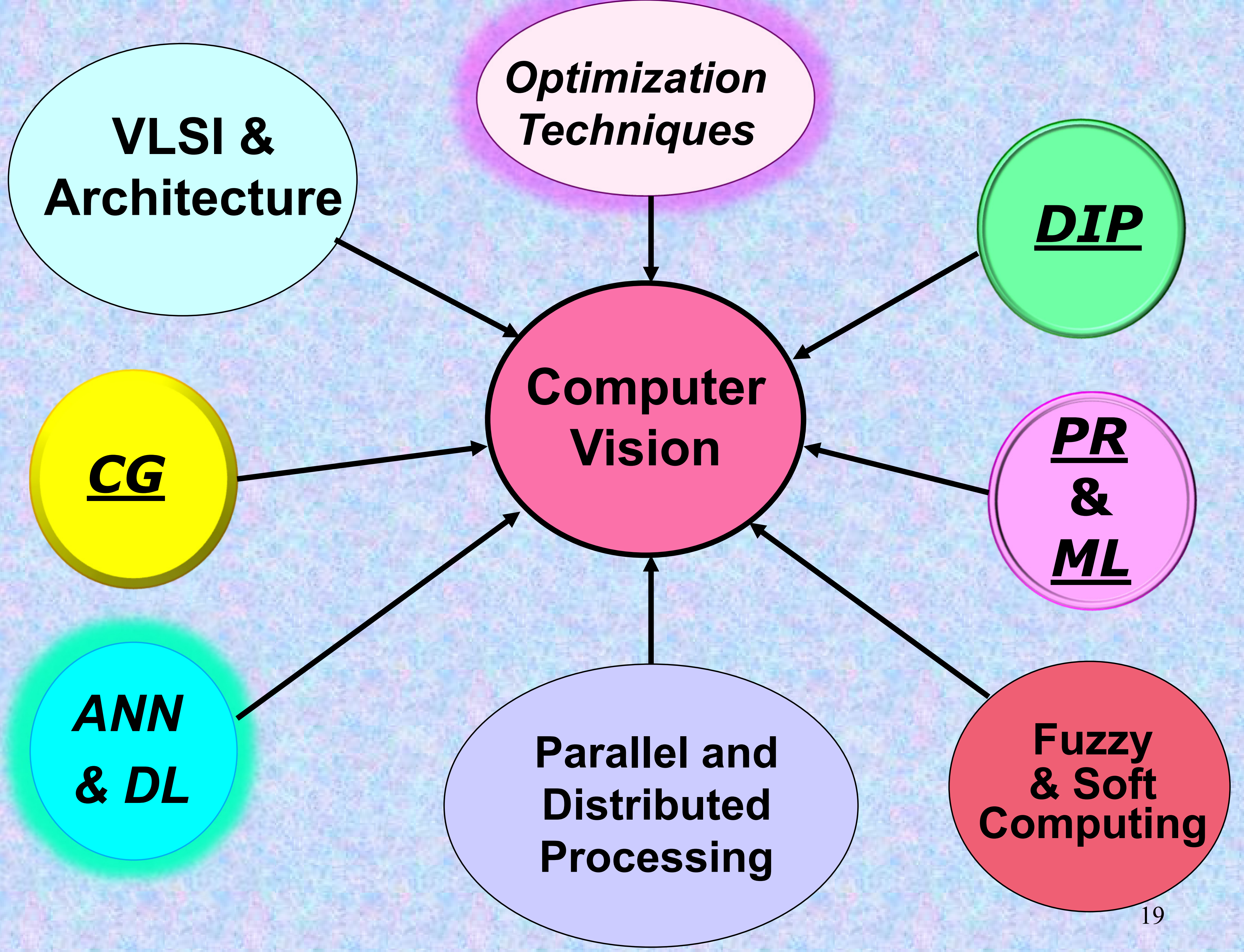

Computer Vision
Models, Object/Scene representation
Visualization
Images, scenes, pictures

Computer Vision is an area of work, which is a combination of concepts, techniques and ideas from Digital Image Processing, Pattern Recognition, Artificial Intelligence and Computer Graphics.

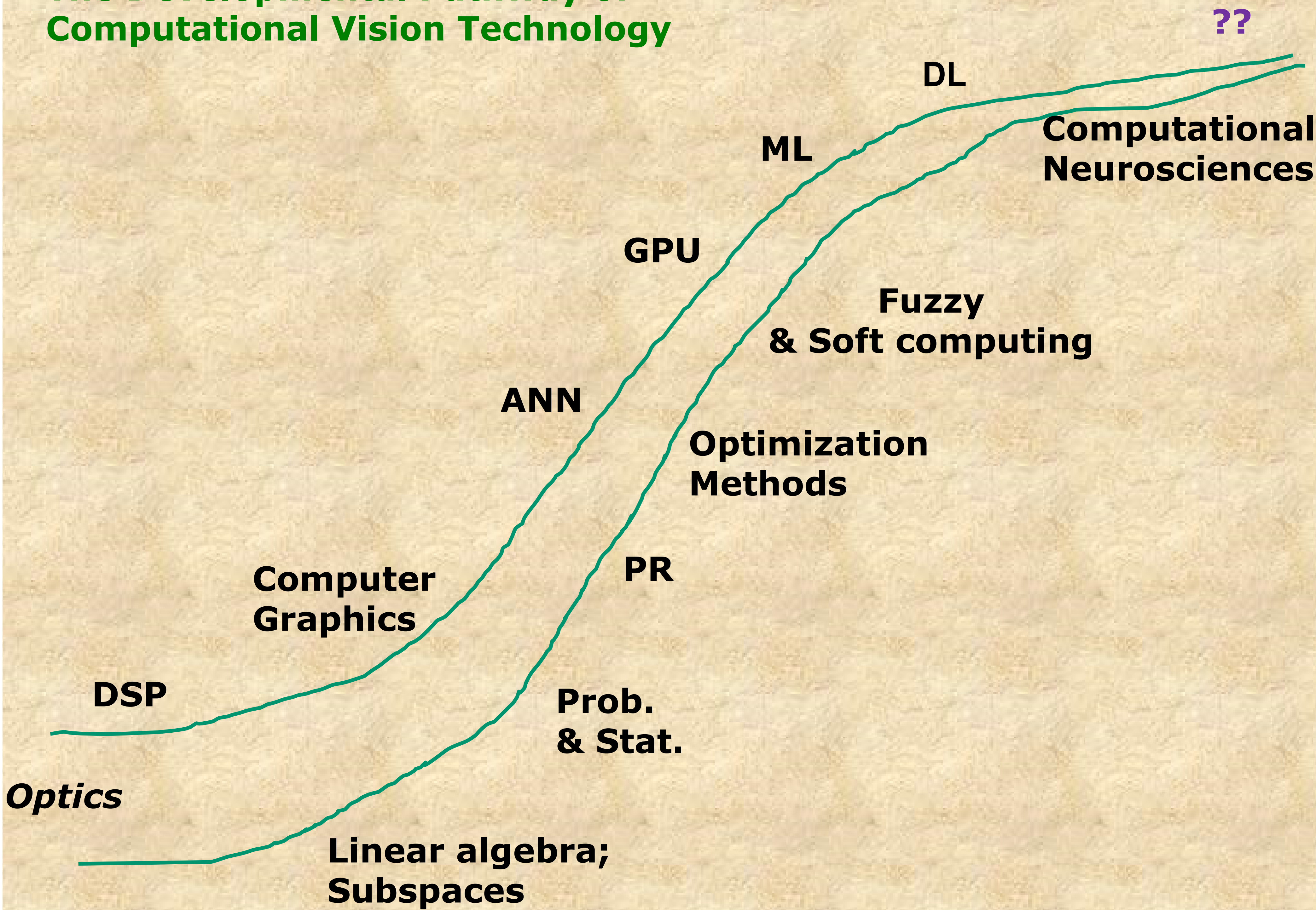
Majority of the tasks in the fields of Digital Image Processing or Computer Vision deals with the process of understanding or deriving the scene information or description, from the input scene (digital image/s). The methods used to solve a problem in digital image processing depends on the application domain and nature of data being analyzed.

Analysis of two-dimensional pictures are generally not applicable of processing three-dimensional scenes, and vice-versa. The choice of processing, techniques and methods and 'features' to be used for a particular application is made after some amount of trial and error, and hence experience in handling images is crucial in most of these cases.

For example, analysis of remote sensed or satellite imagery involves techniques based on classification or analysis of texture imagery. These techniques are not useful for analyzing optical images of indoor or outdoor scenes.


VLSI & Architecture
Optimization Techniques
DIP
Computer Vision
CG
PR & ML
ANN & DL
Parallel and Distributed Processing
Fuzzy & Soft Computing

The Developmental Pathway of Computational Vision Technology



Digital Image processing **is in many cases concerned with taking one array of pixels as input and producing another array of pixels as output which in some way represents an improvement to the original array.**

Purpose:

1. **Improvement of Pictorial Information**
   - **improve the contrast of the image,**
   - **remove noise,**
   - **remove blurring caused by movement of the camera during image acquisition,**
   - **it may correct for geometrical distortions caused by the lens.**

2. **Automatic Machine perception** **(termed Computer Vision, Pattern Recognition or Visual Perception) for intelligent interpretation of scenes or pictures.**

Elements of a Digital Image Processing System

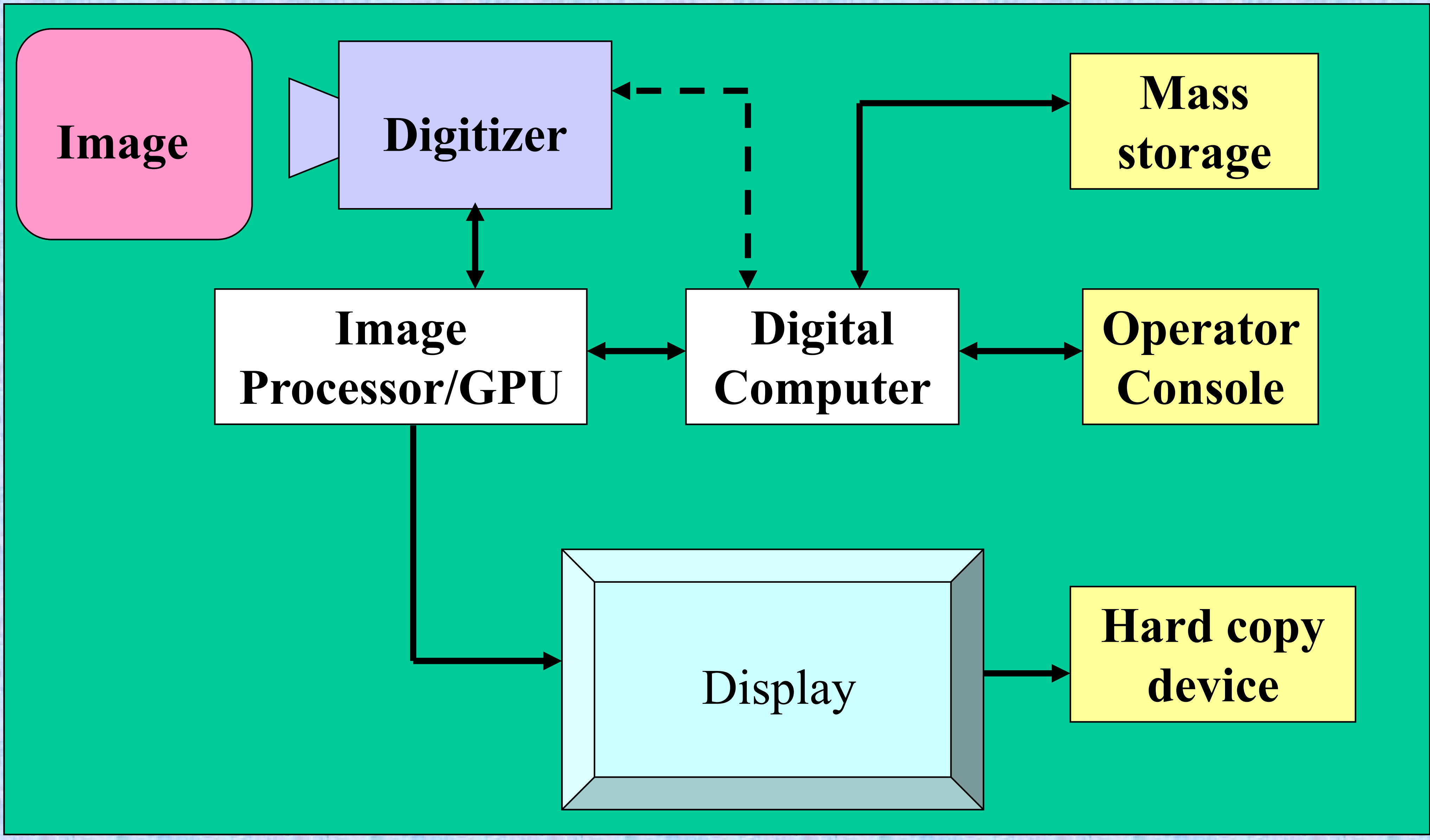
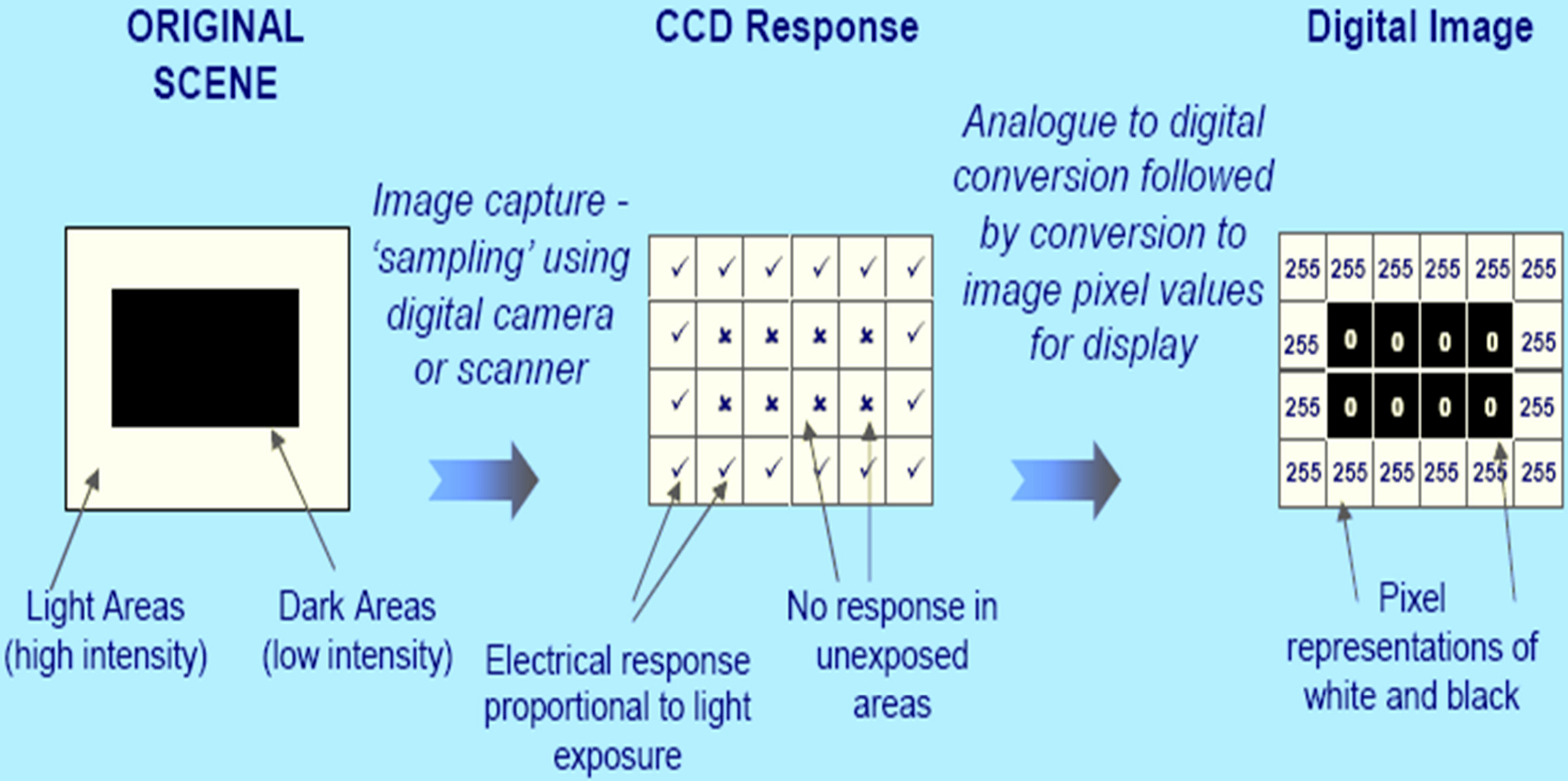


Image processors: Consists of set of hardware modules that perform 4 basic functions:

- Image acquisition: frame grabber
- Storage: frame buffer
- Low-level processing: specialized hardware device designed to perform Arithmetic Logic operations on pixels in parallel
- Display: read from image memory (frame buffer) and convert to analog video signal

- **Digitizers:** Converts image into numerical representation suitable for input to a digital computer
- **Digital Computers:** Interfaced with the image processor to provide versatility and ease of programming.
- **Storage Devices:** For bulk storage. e.g:- Magnetic disks, magnetic tapes, optical disks
- **Display and Recording devices :** Monochrome and Color Television monitors, CRT, Laser printers, heat-sensitive paper devices, and ink spray systems.

Image acquisition using a CCD camera



Resolution standards: HDMI - 1024*768; UHD -

A digital Image

Image is an array of integers: $f(x,y) \in \{0,1,\ldots,I_{max}-1\}$,
where, $x,y \in \{0,1,\ldots,N-1\}$

- N is the resolution of the image and I_{max} is the level of discretized brightness value
- Larger the value of N, more is the clarity of the picture (larger resolution), but more data to be analyzed in the image
- If the image is a gray-level (8-bit per pixel - termed raw, gray) image, then it requires N^2 Bytes for storage
- If the image is color - RGB, each pixel requires 3 Bytes of storage space.

Image Size (resolution)	Storage space required	
	Raw - Gray	Color (RGB)
64*64	4K	12K
256*256	64K	192K
512*512	256K	768K

2048×1536 = ☐ megapixels → ☐ MB for RGB

A ***digital image*** is a two-dimensional (3-D image is called range data) array of intensity values, f(x, y), which represents 2-D intensity function discretized both in spatial coordinates (**spatial sampling**) and brightness (**quantization**) values.

The elements of such an array are called **pixels** (picture elements).

The storage requirement for an image depends on the **spatial resolution** and number of bits necessary for **pixel quantization**.

The processing of an image depends on the application domain and the methodology used to solve a problem. There exists four broad categories of tasks in digital image processing:

(i) Compression, (ii) Segmentation,
(iii) Recognition and (iv) motion.

Segmentation deals with the process of fragmenting the image into homogeneous meaningful parts, regions or sub-images. Segmentation is generally based on the analysis of the histogram of images using gray level values as features. Other features used are edges or lines, colors and textures.

Recognition deals with identification or classification of objects in an image for the purpose of interpretation or identification. Recognition is based on models, which represent an object. A system is trained (using HMM, GMM, ANN etc.) to learn or store the models, based on training samples. The test data is then matched with all such models to identify the object with a certain measure of confidence.

Compression **involves methodologies for efficient storage and retrieval of image data, which occupies large disk space. Typical methods are, JPEG-based, Wavelet based, Huffman Coding, Run length coding etc. for still images and MPEG-I, II, IV & VII for digital video or sequence of frames.**

Motion analysis (or dynamic scene analysis) **involves techniques for the purpose of tracking and estimation of the path of movement of object/s from a sequence of frames (digital video). Methods for dynamic scene analysis are based on (i) tracking, (ii) obtaining correspondence between frames and then (iii) estimating the motion parameters and (iv) structure of moving objects. Typical methods for analysis are based on optical flow, iterative Kalman filter and Newton/Euler's equations of dynamics.**

There are generally three main categories of tasks involved in a complete computer vision system. They are:

• ***Low level processing***: **Involves image processing tasks in which the quality of the image is improved for the benefit of human observers and higher level routines to perform better.**

• ***Intermediate level processing***: **Involves the processes of feature extraction and pattern detection tasks. The algorithms used here are chosen and tuned in a manner as may be required to assist the final tasks of high level vision.**

• ***High level vision***: **Involves autonomous interpretation of scenes for pattern classification, recognition and identification of objects in the scenes as well as any other information required for human understanding.**

A top down approach, rather than a bottom-up approach is used in the design of these systems in many applications. The methods used to solve a problem in digital image processing depends on the application domain and nature of data being analyzed.

Different fields of applications include:

- Character Recognition,
- Document processing,
- Commercial (signature & seal verification) application,
- ***Biometry and Forensic*** ***(authentication: recognition and verification of persons using face, palm & fingerprint),***
- **Pose** and gesture identification,
- **Automatic inspection of industrial products,**
- Industrial process monitoring,
- Biomedical Engg. (Diagnosis and surgery),
- ***Military surveillance and target identification,***
- **Navigation and mobility (for robots and unmanned vehicles - land, air and underwater),**
- ***Remote sensing (using satellite imagery),***
- ***GIS***
- Safety and security (night vision),
- Traffic monitoring,
- ***Sports (training and incident analysis)***
- VLDB (organization and retrieval)
- Entertainment and virtual reality.

TARGETED INDUSTRIAL APPLICATIONS

Intelligent Traffic Control

Anti-forging Stamps

Card Counting Systems

Drive Quality Test

Camera Flame Detection

CCTV Fog Penetration

Key Image Search/Index

Security Monitoring

Robust Shadow Detection

Vehicle Segmentation

Visual Tracking Systems

Illegal content (adult) Filter

Scratch Detection

Smart Traffic Monitoring

Vehicle Categorization

Vehicle Wheel alignment

Number Plate Identification

Referrals for Line calls

Different categories of work being done in CV, to solve problems:

2-D image analysis – segmentation, target detection, matching, CBIR;

3-D multi-camera calibration; Correspondence and stereo; Reconstruction of 3-D Objects and surfaces;

Pattern Recognition for Objects, scenes;

Video and motion analysis; Video analytics; CBVR; Compression;

Feature extraction: Canny, GHT, Snakes, DWT, Corners, SIFT, GLOH, LESH;

Multi-sensor data, Decision and feature fusion;

Image and Video-based Rendering;

Steganography and Watermarking;

The various sub-categories of technology in these related fields are:

- *image enhancement,*
- *image restoration and filtering,*
- *representation and description,*
- *feature extraction,*
- *image segmentation,*
- *image matching,*
- color image processing,
- image synthesis,
- image representation,
- *image reconstruction*
- *range data processing,*
- *stereo image processing*
- *computational geometry,*
- *image morphology,*
- *artificial neural networks,*
- *Neuro-fuzzy techniques,*
- *computational geometry,*
- *parallel architectures & algorithms.*

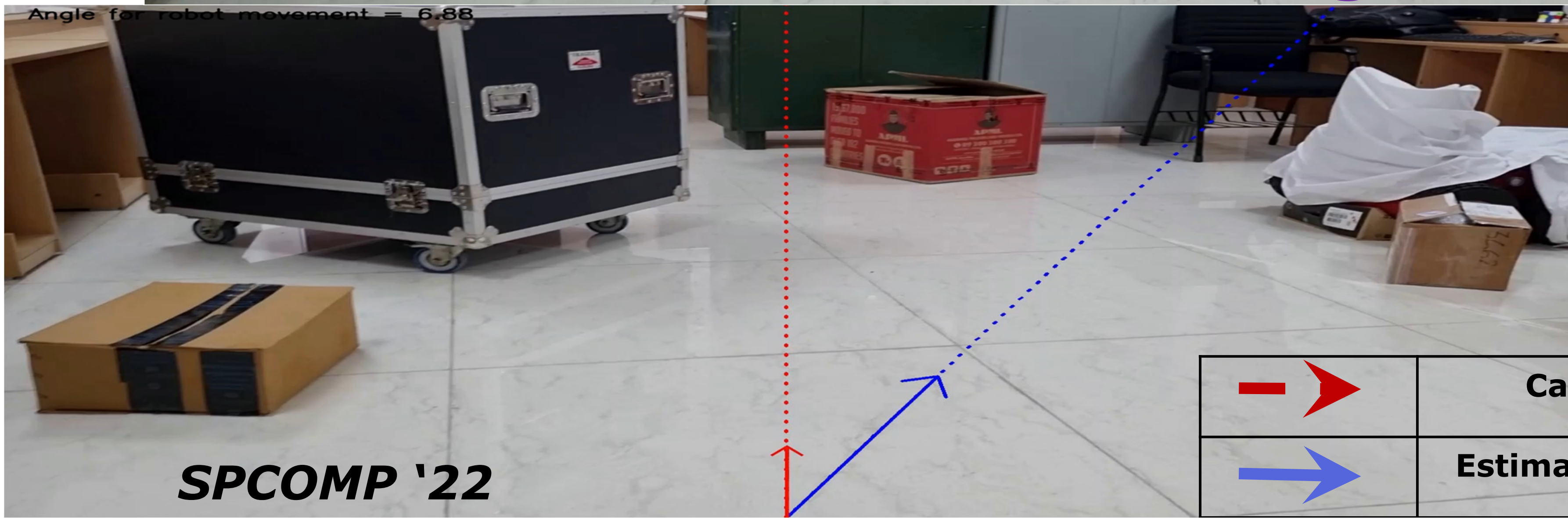
Few DEMOS and ILLUSTRATIONS

Courtesy: TA/students of VPLAB - CSE-IITM

Video Object Segmentation

Best Student Paper Award - "Motion-based Occlusion-aware Pixel Graph Network for Video Object Segmentation", Saptakatha Adak and Sukhendu Das; In 26th International Conference on Neural Information Processing (**ICONIP, Rank A**), Sydney, Australia, December 12-15, 2019.

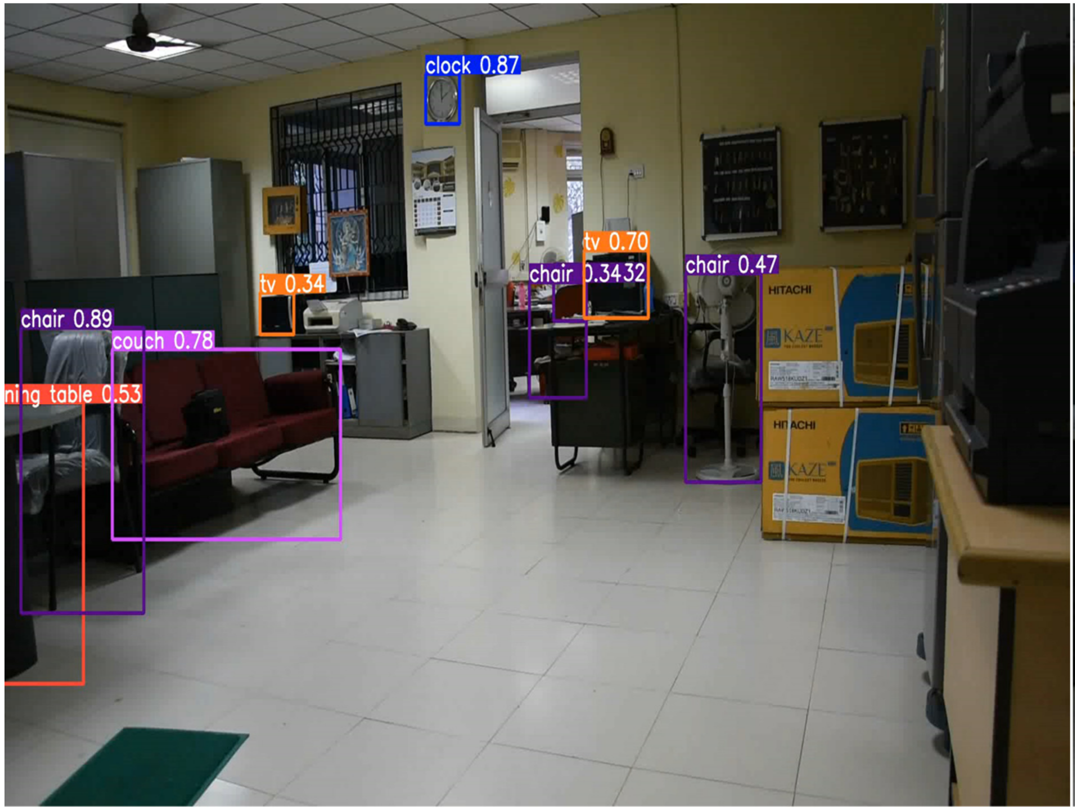
Floor Auto-Navigational aid



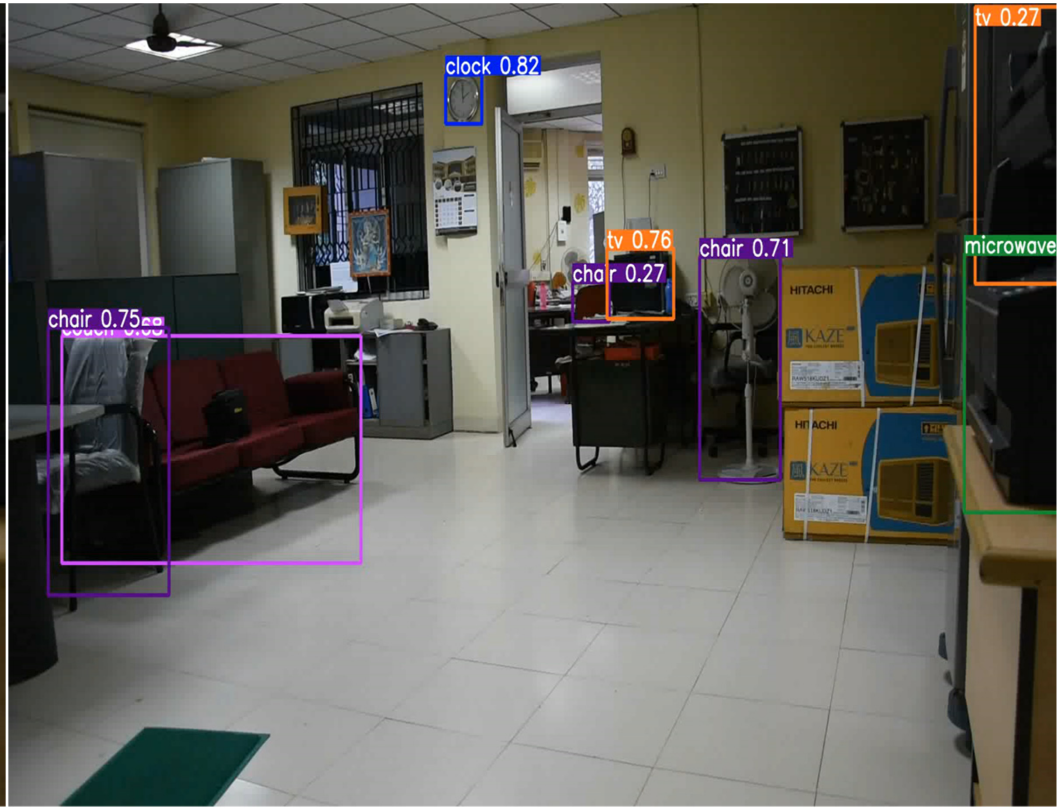
Red arrow	Camera View Axis
Blue arrow	Estimated Orientation

SPCOMP '22

Heavy Version of YOLOV5



Light Version of YOLOV5



Sl	Process	GPU - NVIDIA GeForce RTX 2080	CPU CORE i7 8th Generation
1	Yolov5 - Heavy	40 fps	-
2	Yolov5 – Light	149 fps	18 fps
3	Yolov7 (2022)	23 fps	-

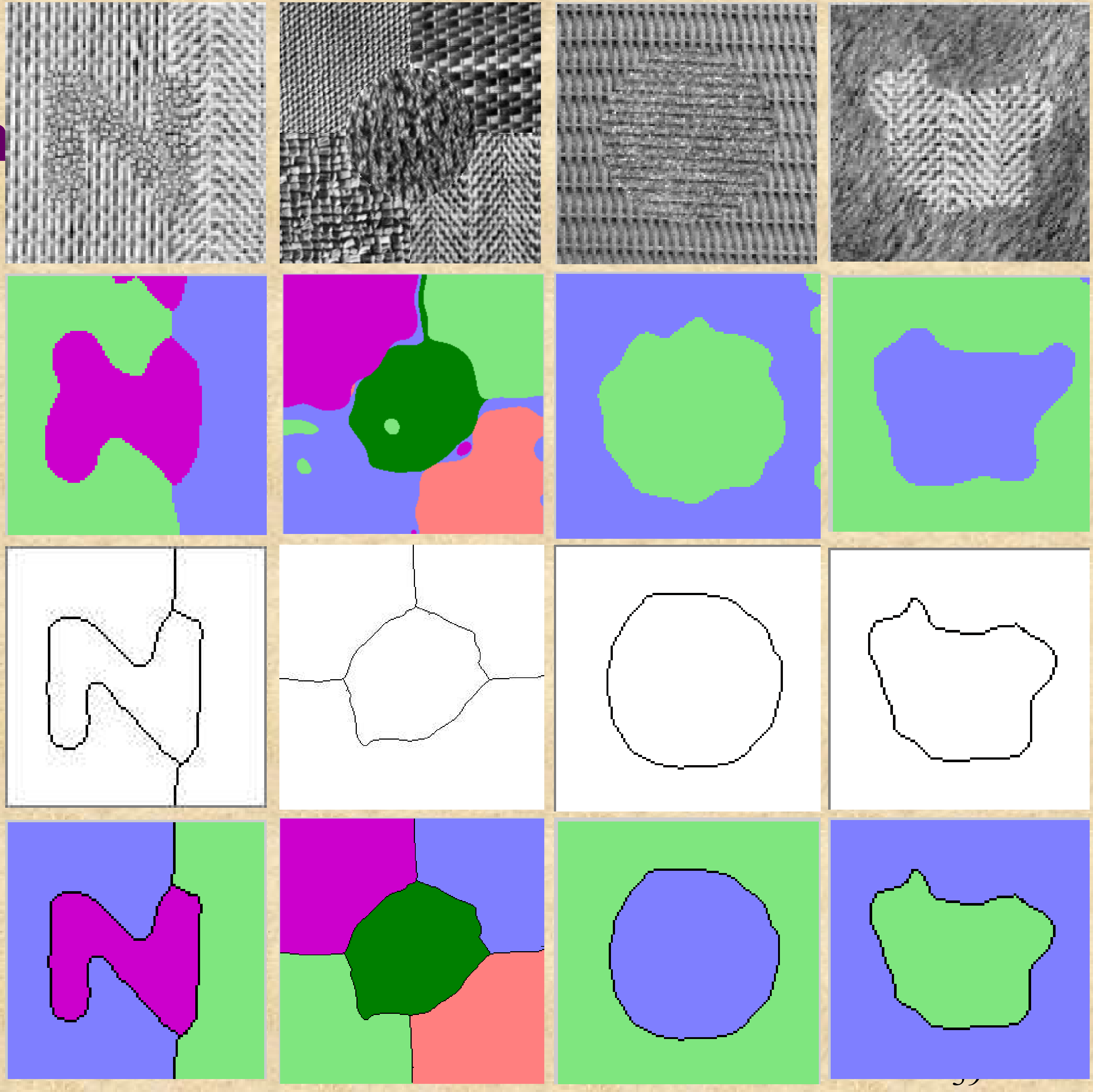
Results of Segmentation

Input Image

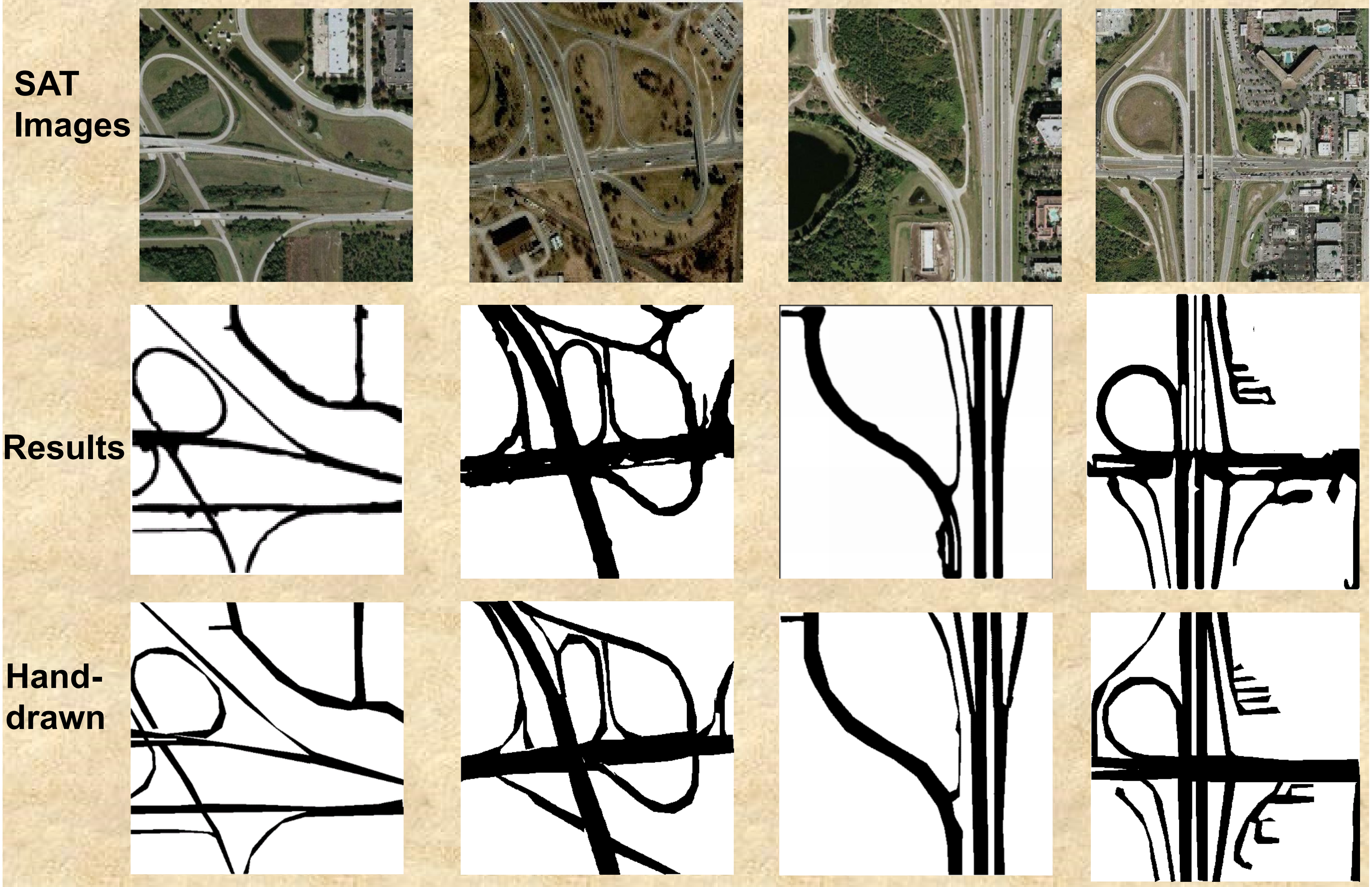
Segmented map before integration

Edge map before integration

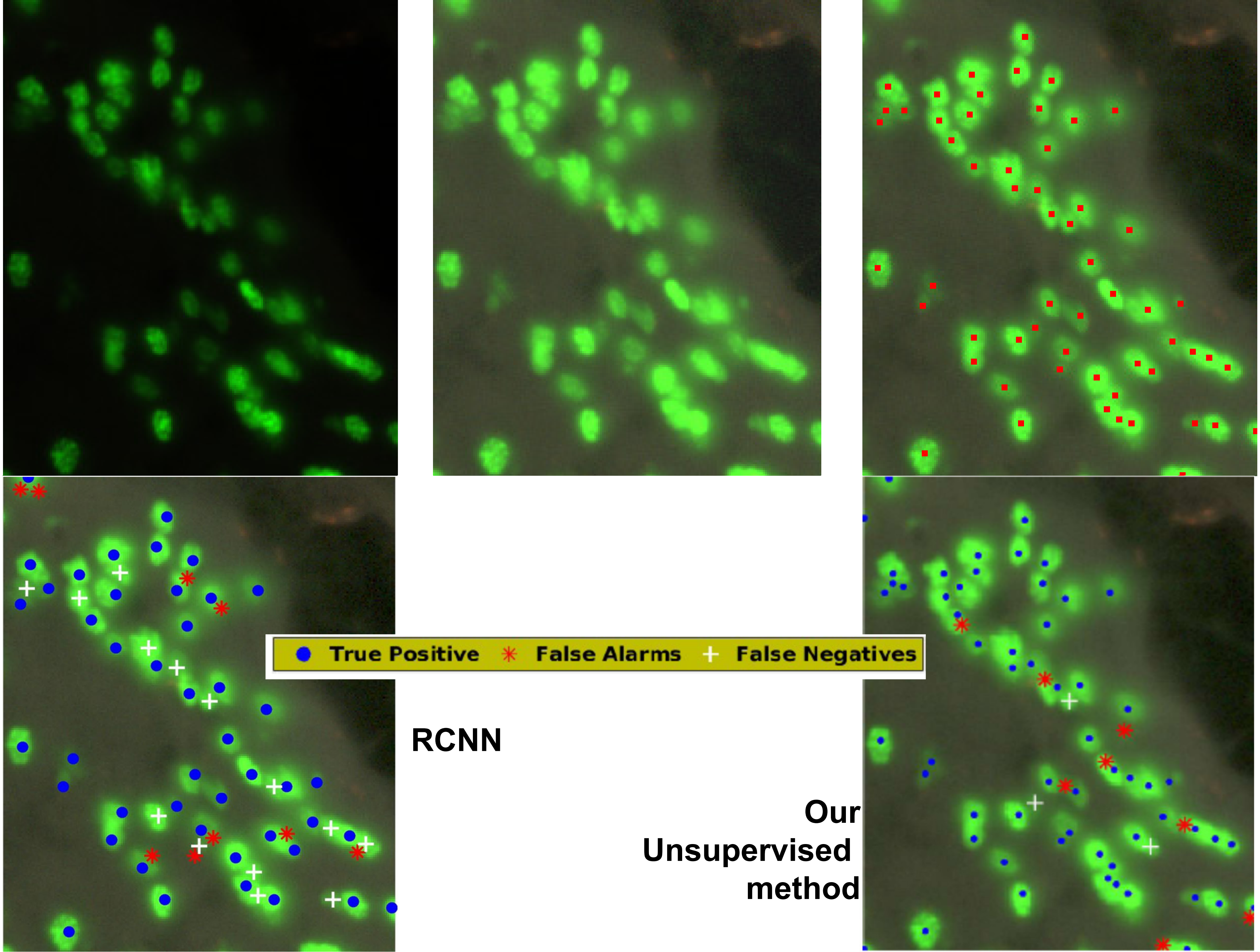
Segmented map and Edge map after integration



Road extraction from Satellite Images



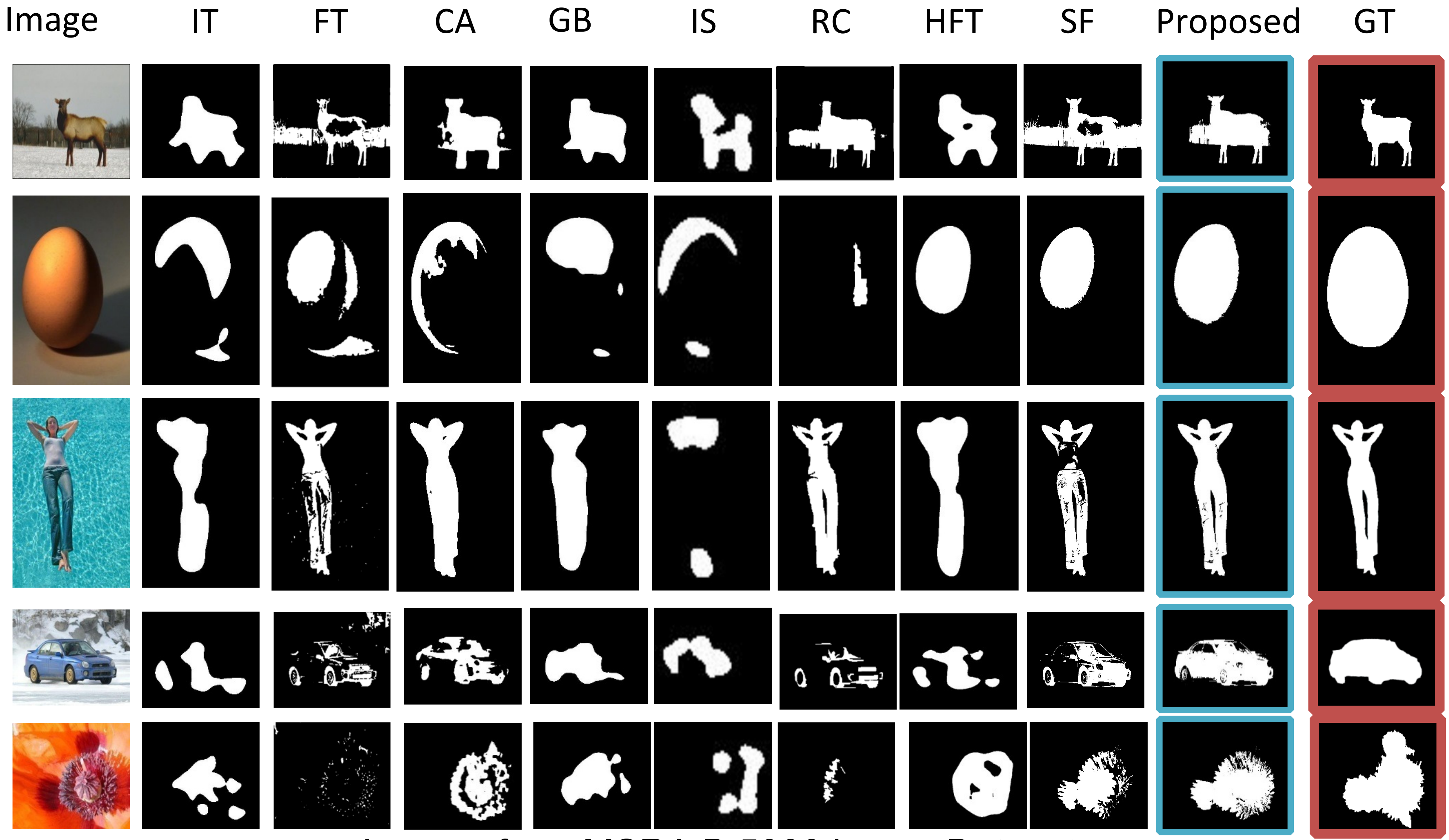
Object Extraction From an Image


True Positive
False Alarms
False Negatives
RCNN
Our
Unsupervised
method

Method 1

Salient Object Segmentation In Images

Unsupervised Saliency



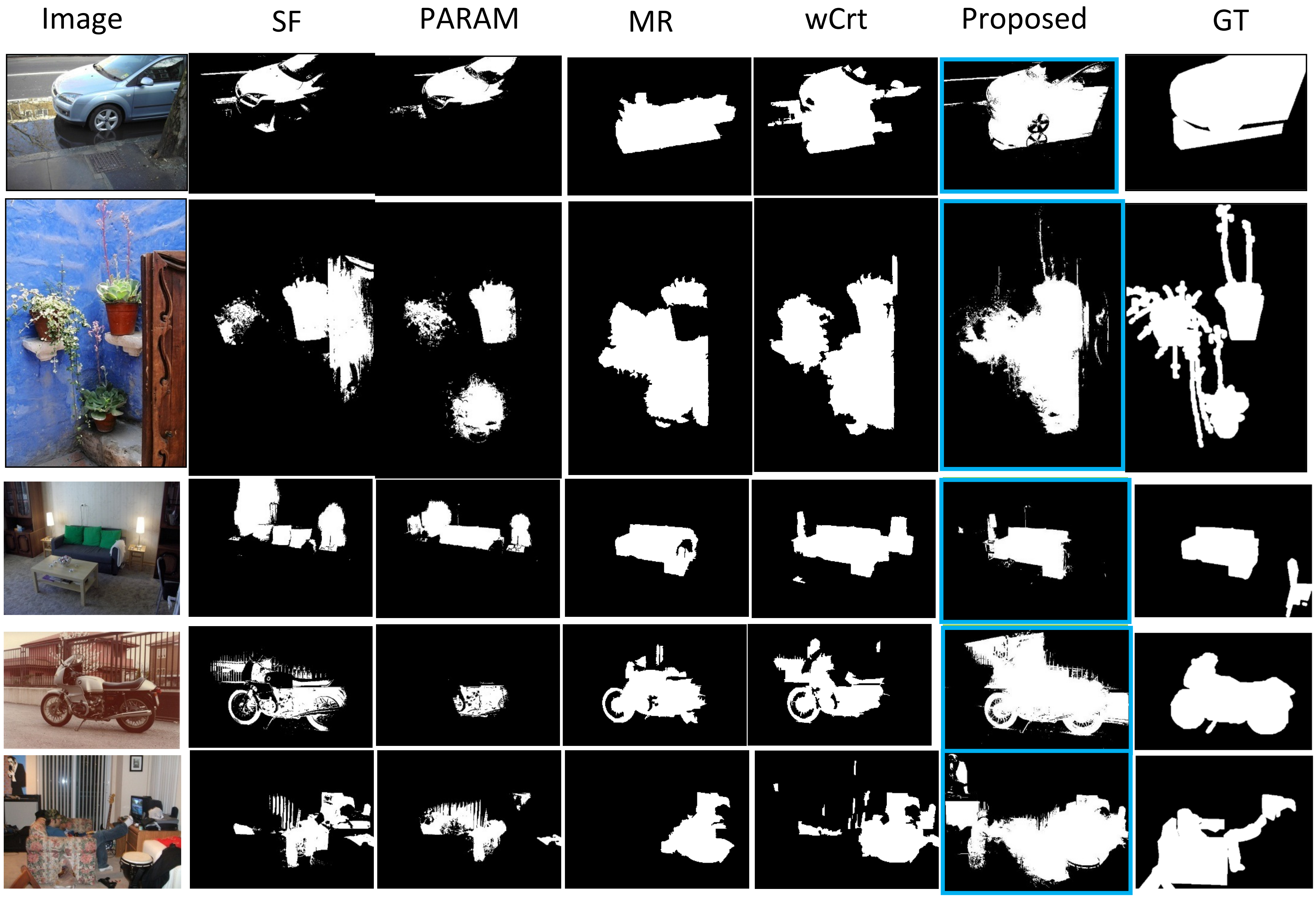
Images from MSRA B 5000 image Dataset

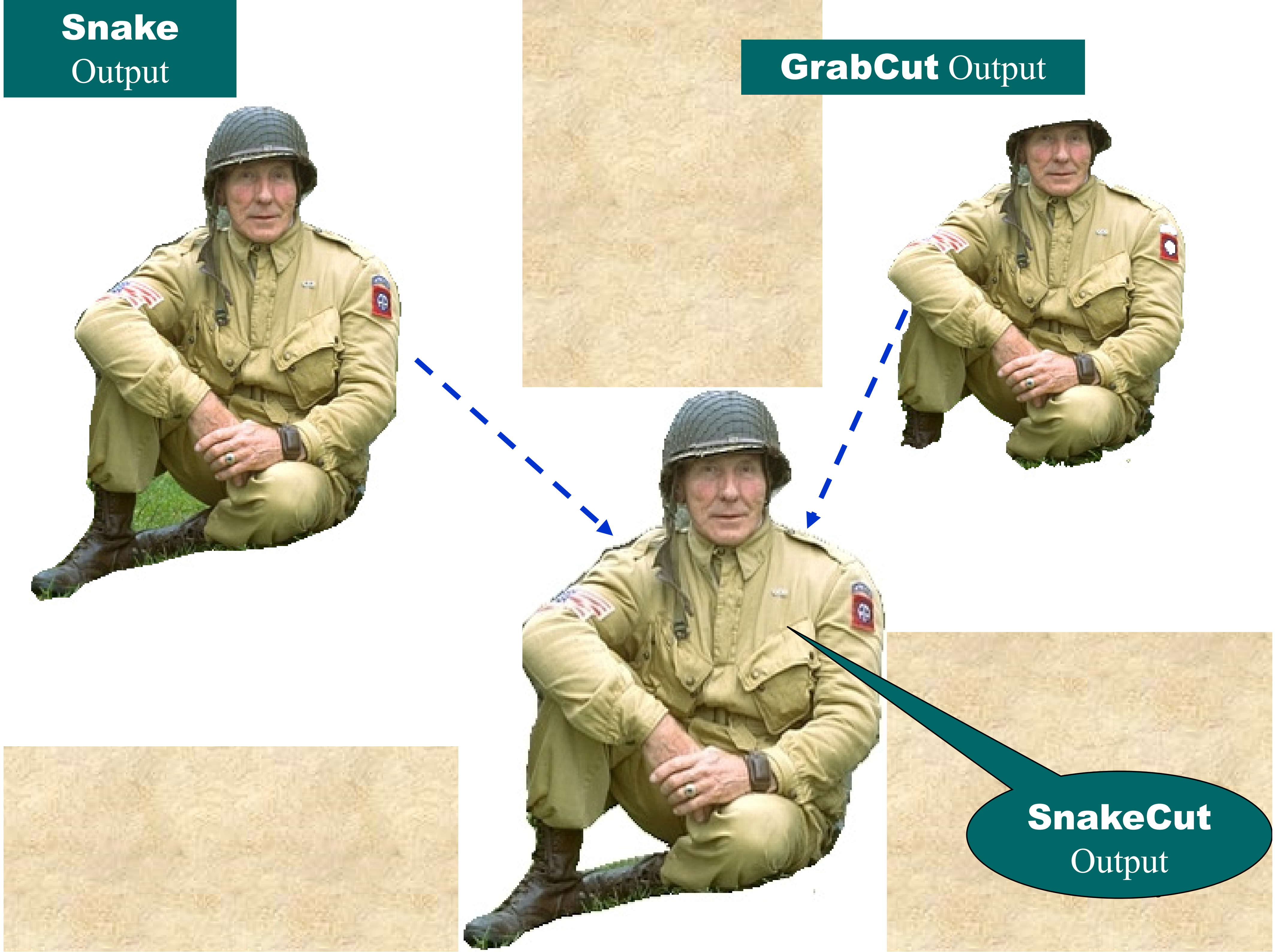
http://research.microsoft.com/en-us/um/people/jiansun/SalientObject/salient object.htm

Oct 24, 2014

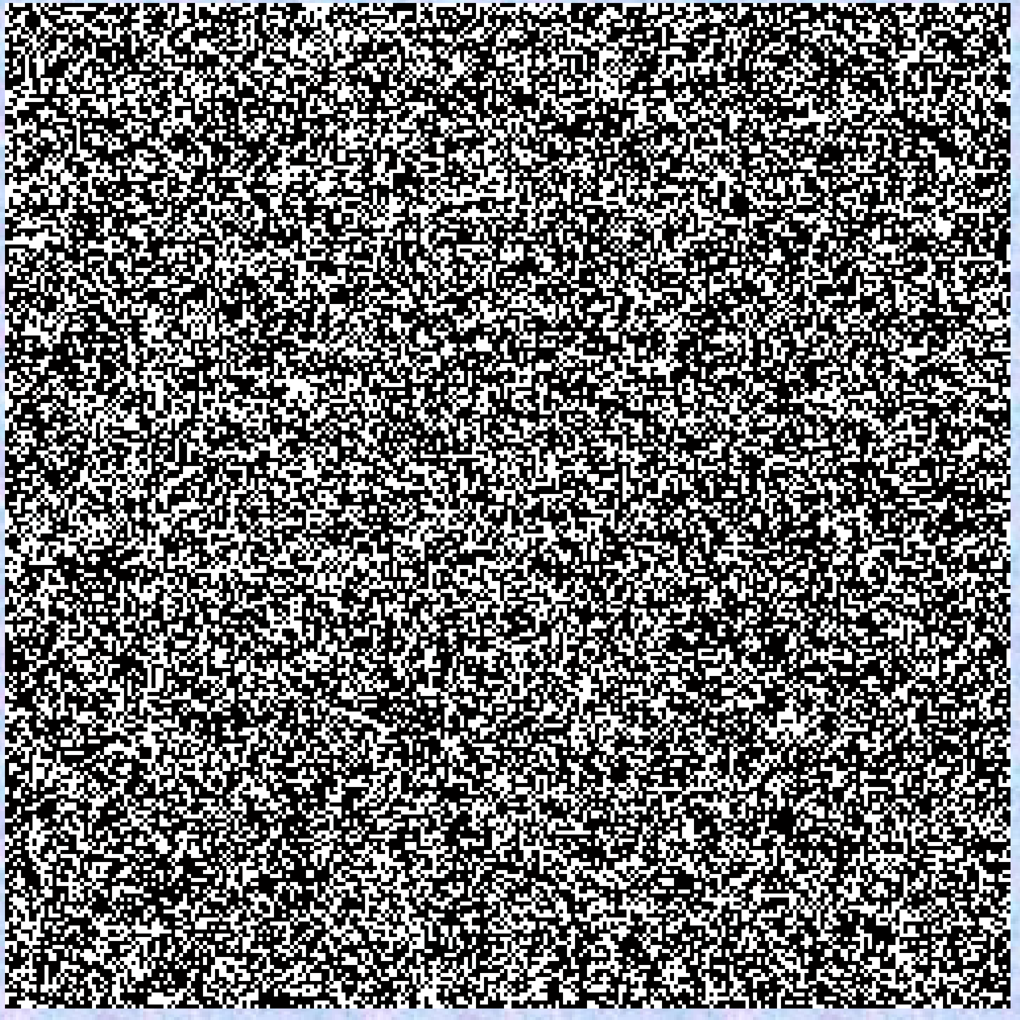
Method 2

Visual Results on PASCAL

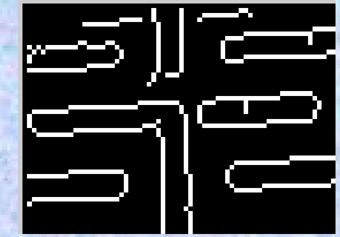



Snake Output
GrabCut Output
SnakeCut Output

The Problem Definition



IMRN



IMT

Given a bitmap template (IMT) and a noisy bitmap image IMRN which contains IMT (believe me):
FIND OUT the location of IMT in IMRN !

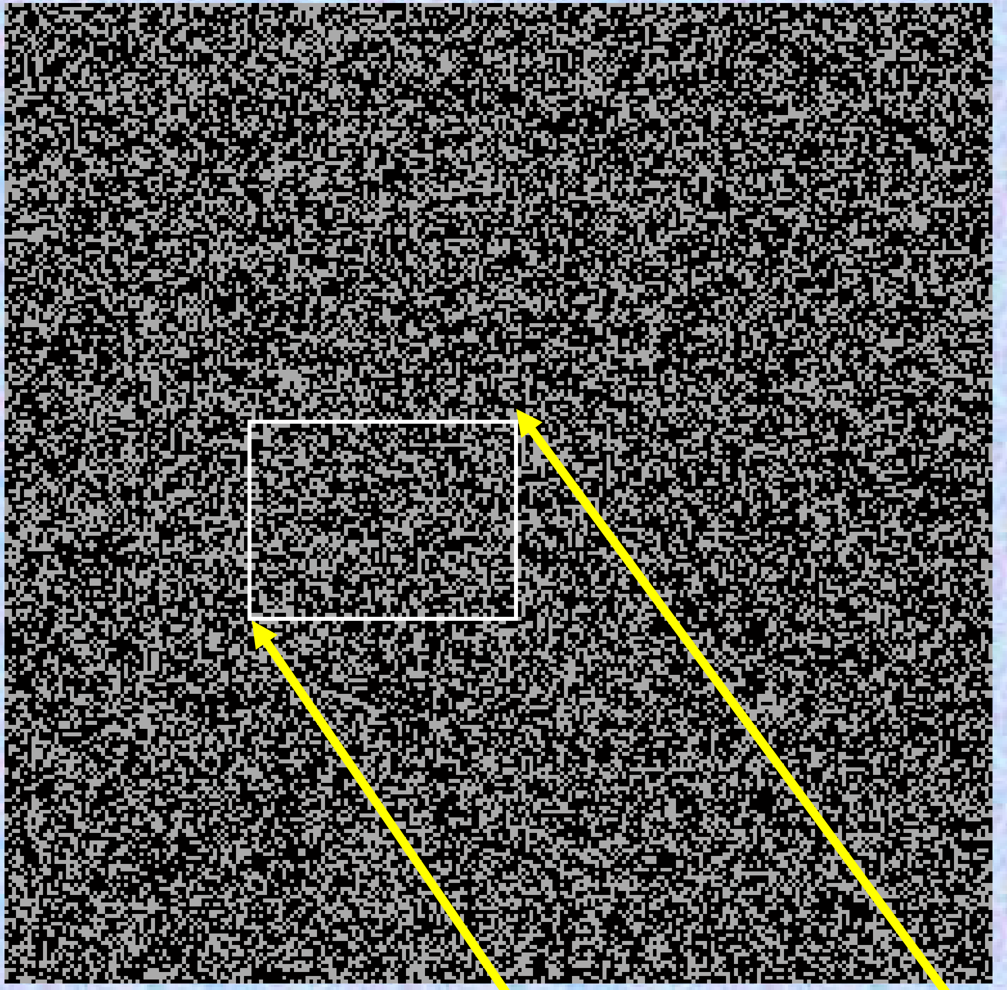
Go to the next page for more:

Problem explanation for pessimists.

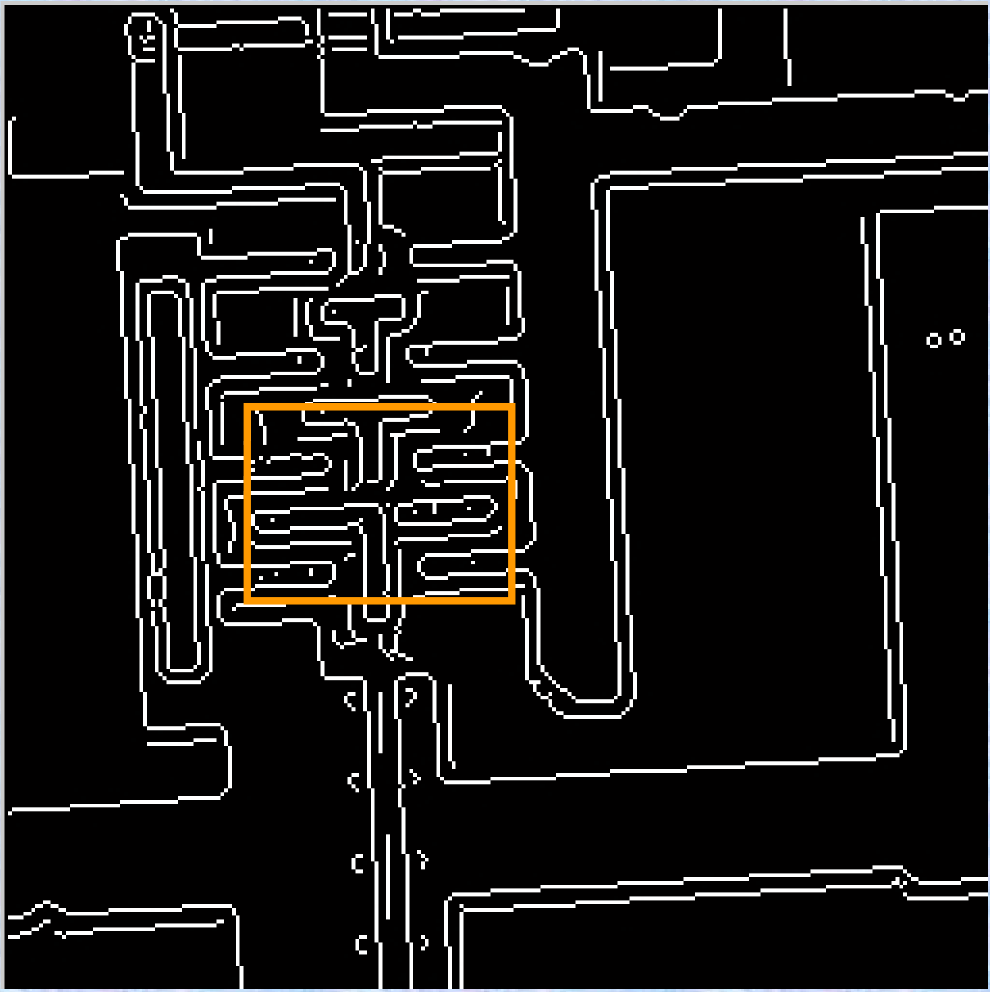
- IMRN (in previous page) is obtained by adding a large level of "Salt and Pepper" noise onto IMR bitmap image.

- IMT is also obtained from IMR as shown above.

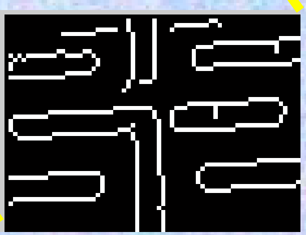
The RESULT beats the human EYE



IMRN



IMR



IMT

Published almost 3 decades ago;
Without GPU and DL

Thank you