EN3160 Fundamentals of Image Processing and Computer Vision

Ranga Rodrigo
Department of Electronic and Telecommunication Engineering,
University of Moratuwa, Sri Lanka.
ranga@uom.lk. https://ranga.staff.uom.lk/

Western

UNIVERSITY OF MORATUWA

ENTC

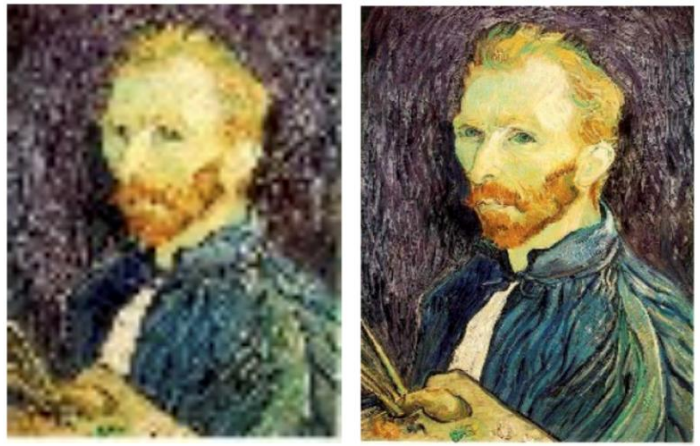
Electronic and Telecommunication Engineering

https://www.facebook.com/WesternUniversity/posts/its-hard-to-be-upset-about-the-weather-when-it-makes-campus-look-like-a-magical-once/10155573317922795/

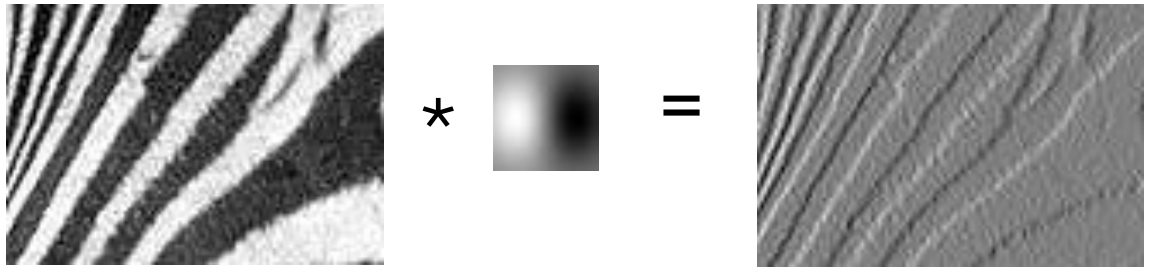
Course Outline

1. Image processing and early vision
   - Sampling, interpolation, point operations, filtering (e.g., convolutions), edge detection, feature extraction, optical flow
2. Fitting and alignment
   - Least squares, voting methods, transforms, stitching
3. Image formation
   - Perspective projection, cameras, light and shading, color
4. 3D vision
   - Camera calibration, two-view geometry, structure from motion, stereo, dense 3D reconstruction (neural radiance fields, Gaussian splatting)
5. Segmentation
   - Thresholding, region growing, snakes, grab cuts, introduction to deep leaning based methods
6. Introduction to deep learning for vision
   - Linear classifiers, image classification, object detection, generative methods (auto-regressive, diffusion)
7. Recent topics (time permitting)

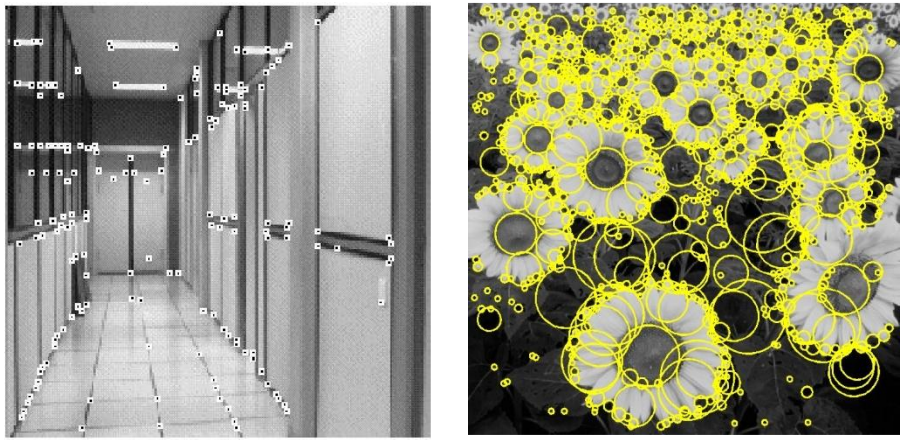
Image Processing and Early Vision



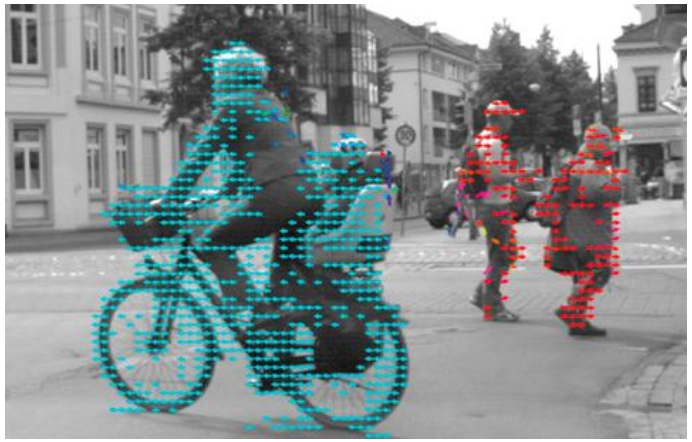
Sampling, interpolation, Fourier analysis



Linear filtering, edge detection



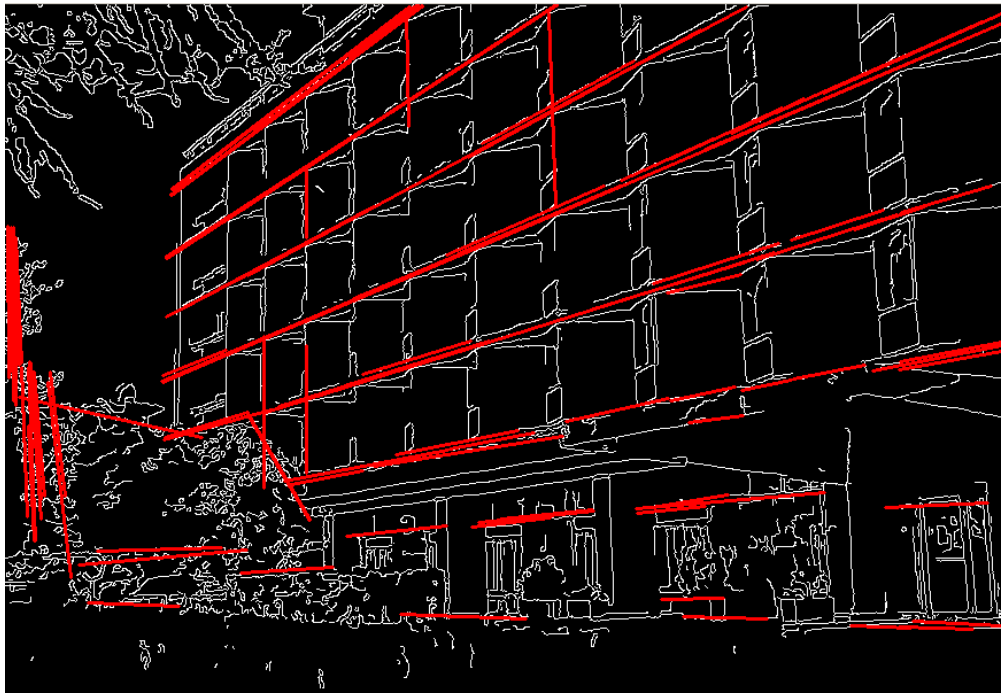
Feature extraction



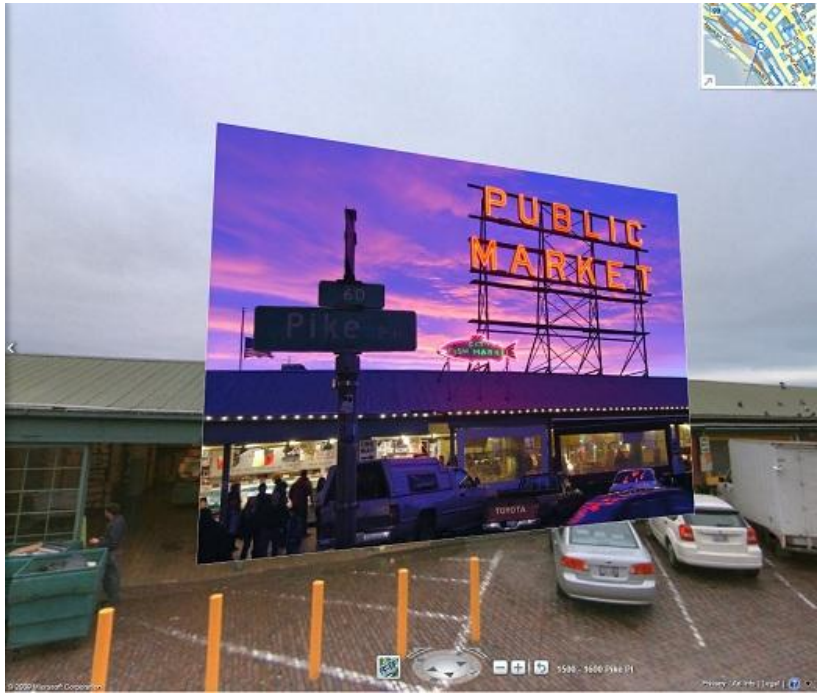
Optical flow

Source: Lazebnik

Fitting and Alignment



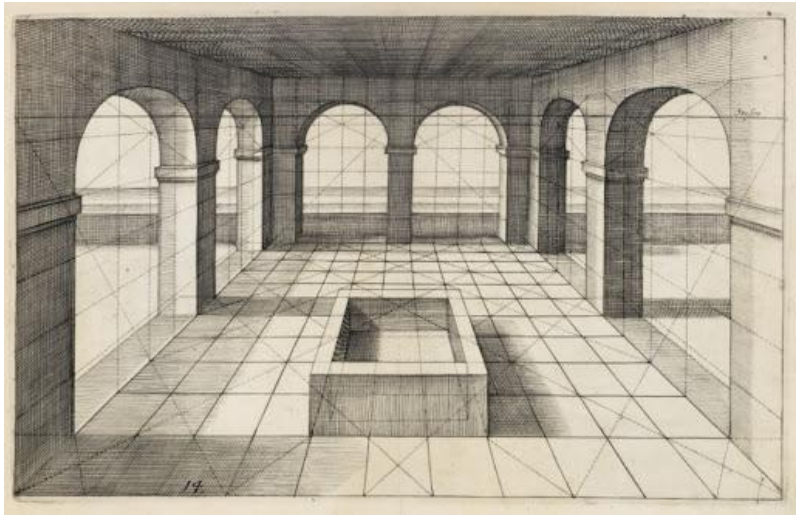
Fitting: Least squares, voting methods



Alignment, image stitching

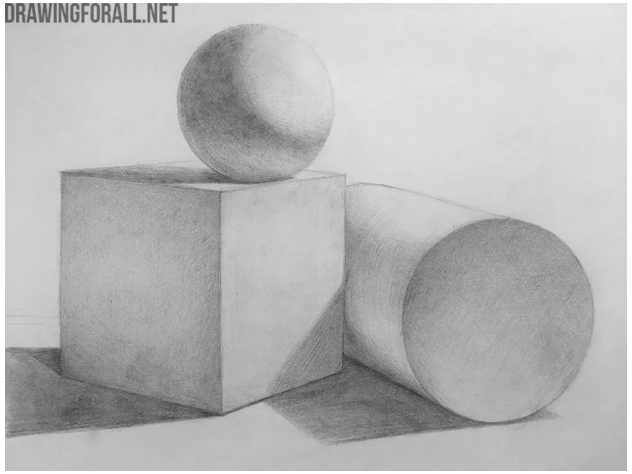
Source: Lazebnik

Image Formation

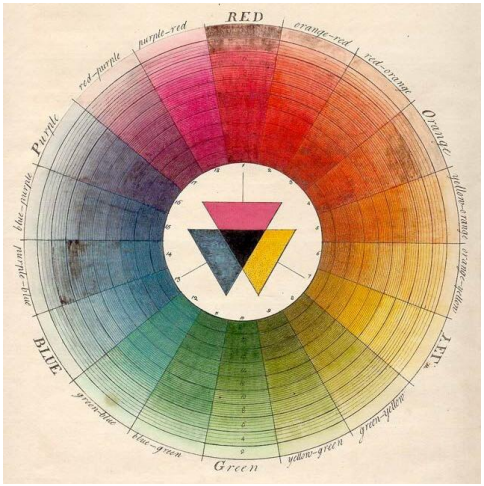


Perspective projection

Camera optics



Light and shading

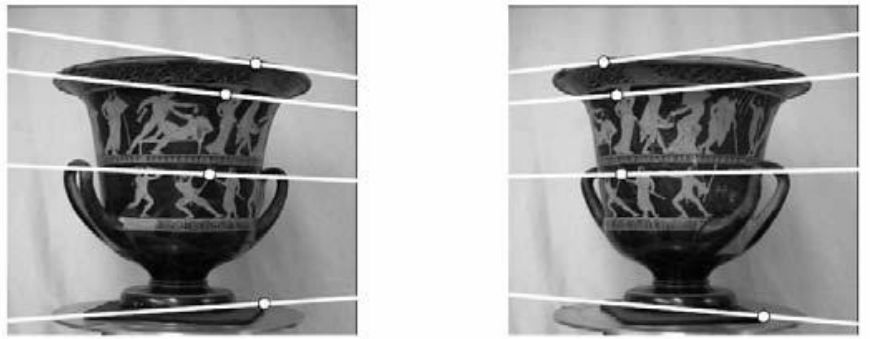


Color

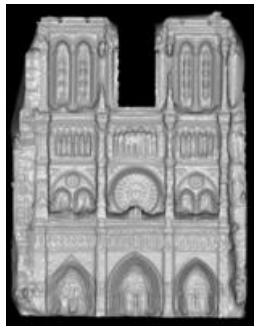
Source: Lazebnik

3D vision

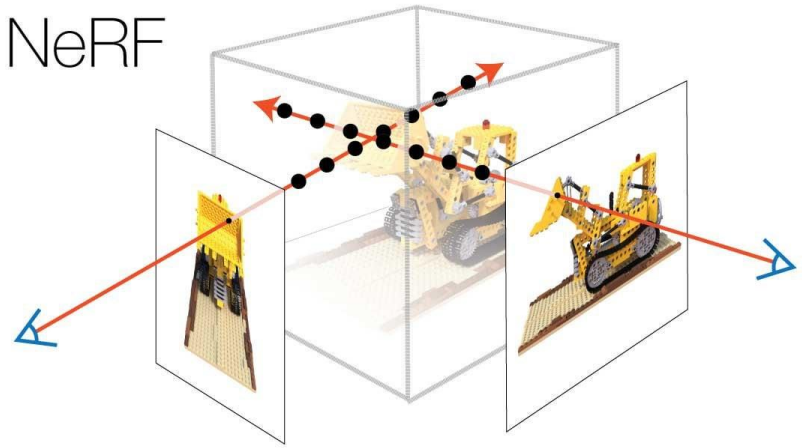
Camera calibration



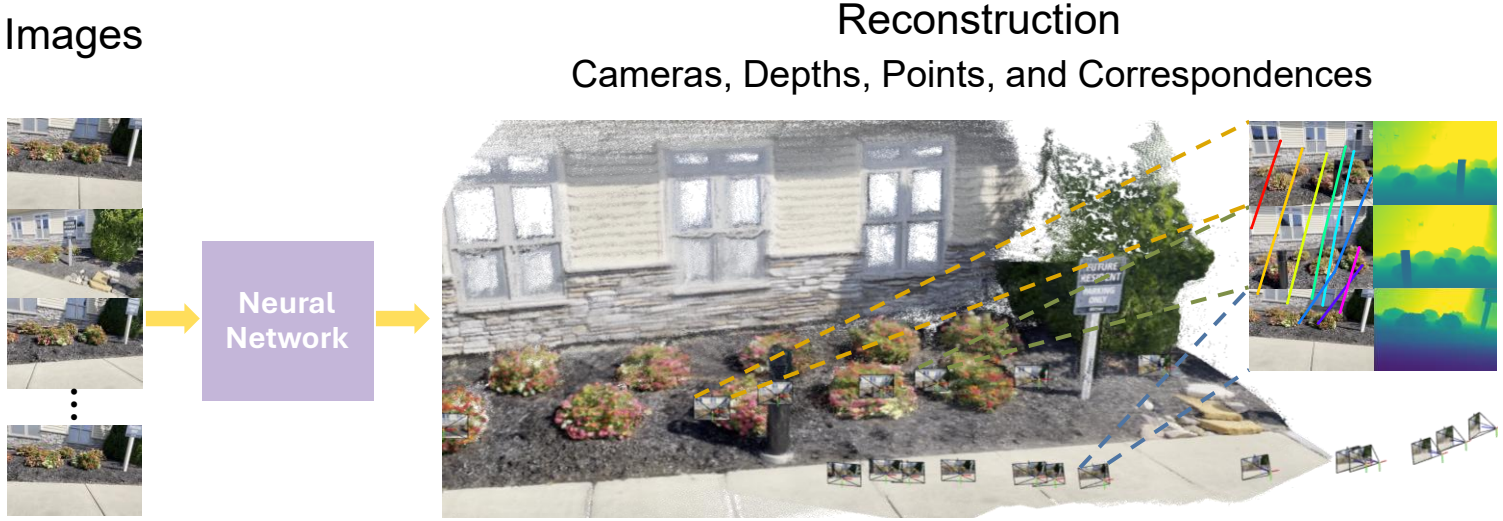
Two-view geometry, structure from motion



Stereo



Light fields and dense 3D reconstruction



Wang et al, VGGT: Visual Geometry Grounded Transformer, CVPR 2025.

Source: partly from Lazebnik

Segmentation

Segmentation Output (Semantic)

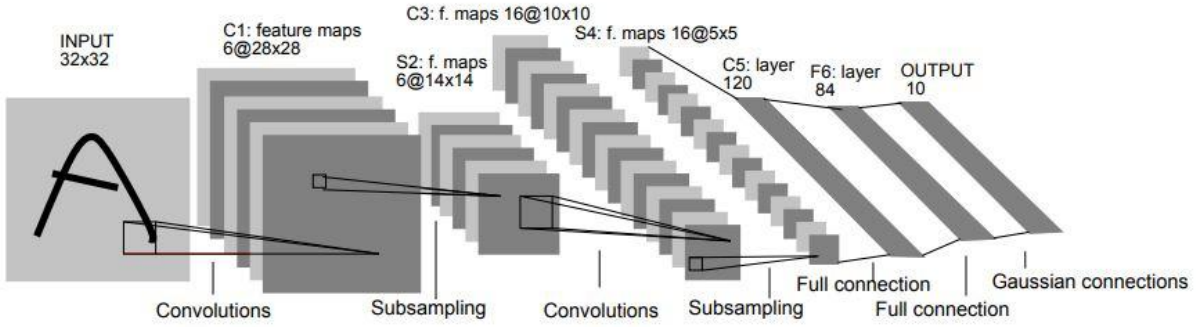
Input image → Segmentation → Labeled image

Binary Segmentation Output

Input image → Segmentation → Binary image (mask)

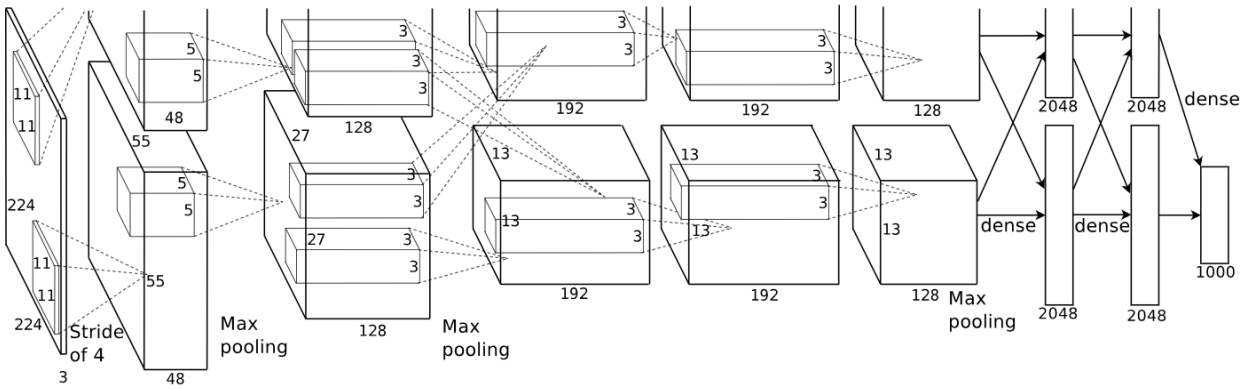
Ravi et al, SAM 2: Segment Anything in Images and Videos, ICLR 2025.

Deep Learning for Vision

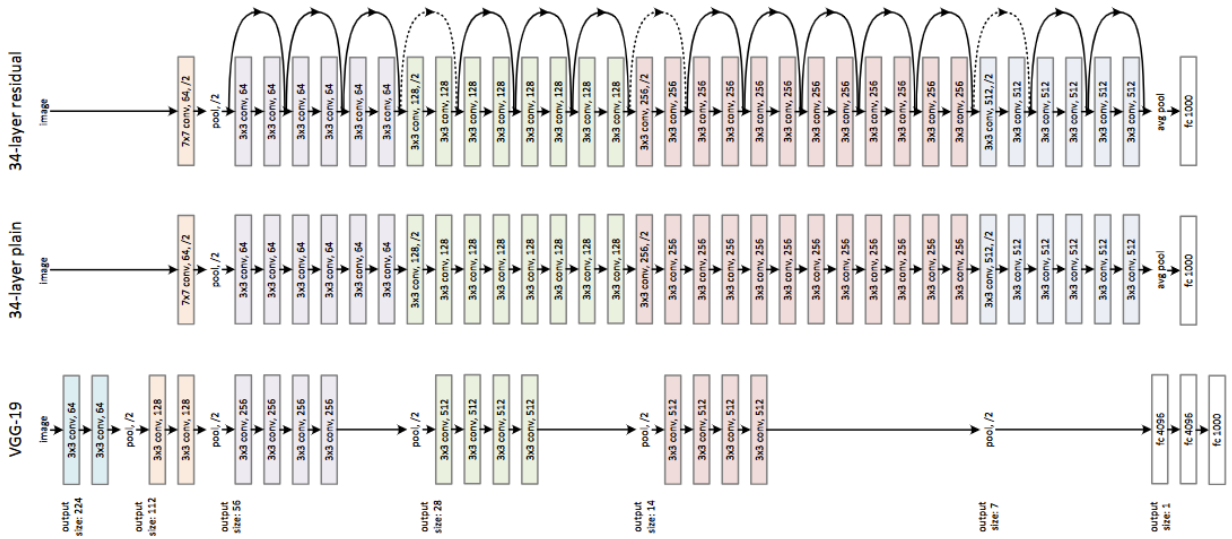


LeCun et al., Gradient-Based Learning Applied to Document Recognition, Proceedings of IEEE 1998.

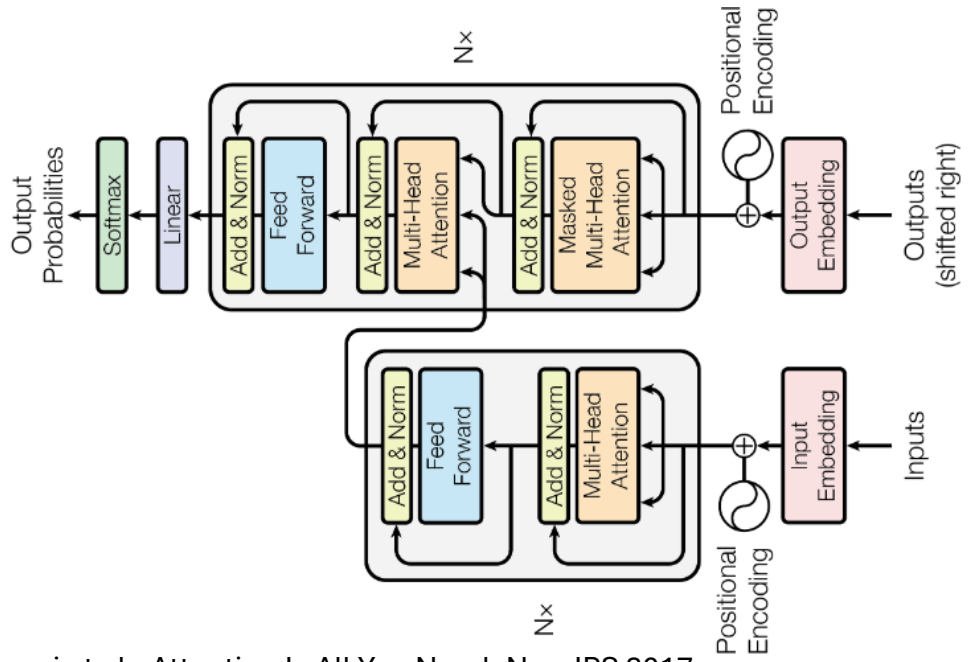
AlexNet



Krizhevsky et al., ImageNet Classification with Deep Convolutional Neural Networks, NeurIPS 2012.

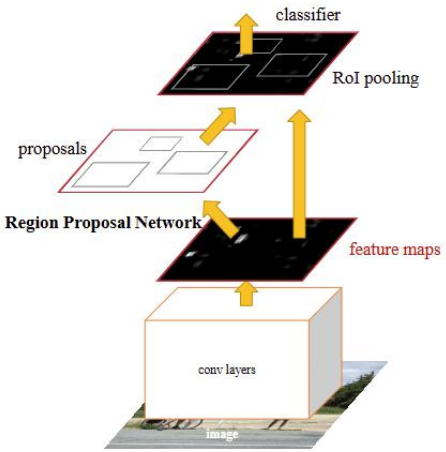


He et al., Deep Residual Learning for Image Recognition, CVPR 2016.

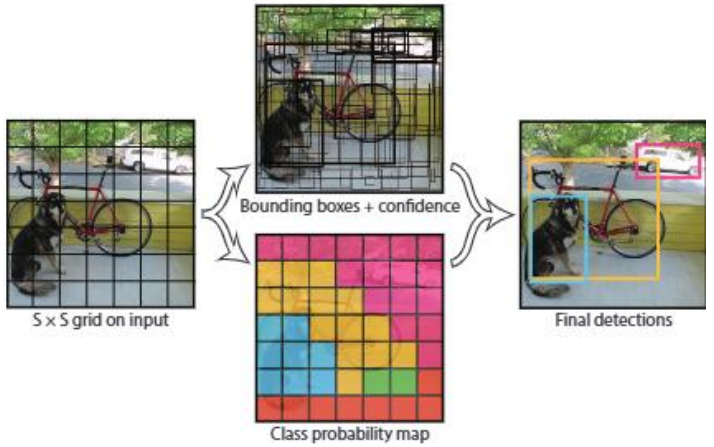


Vaswani et al., Attention Is All You Need, NeurIPS 2017.

Deep Learning for Vision

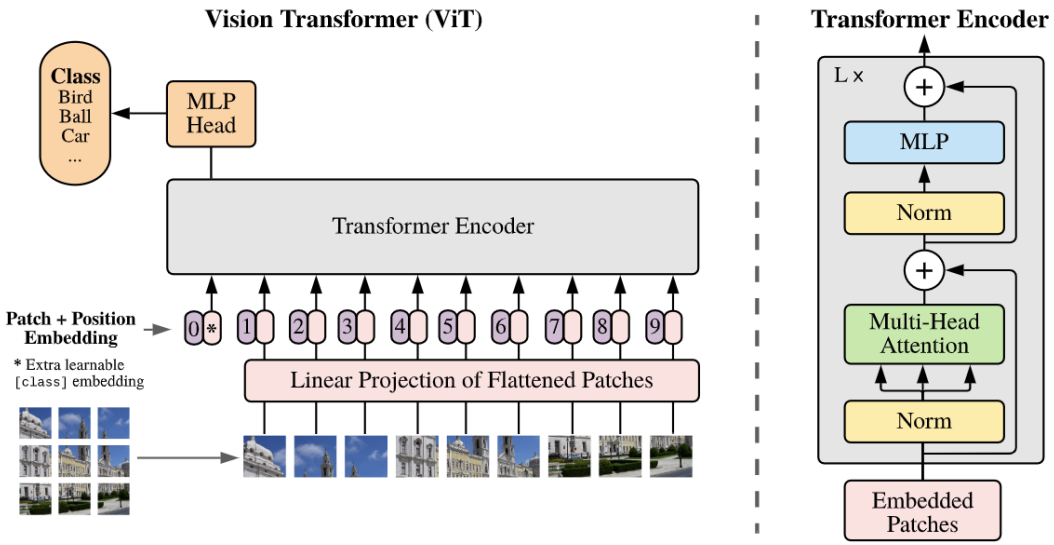


Ren et al, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, NeurIPS 2015.

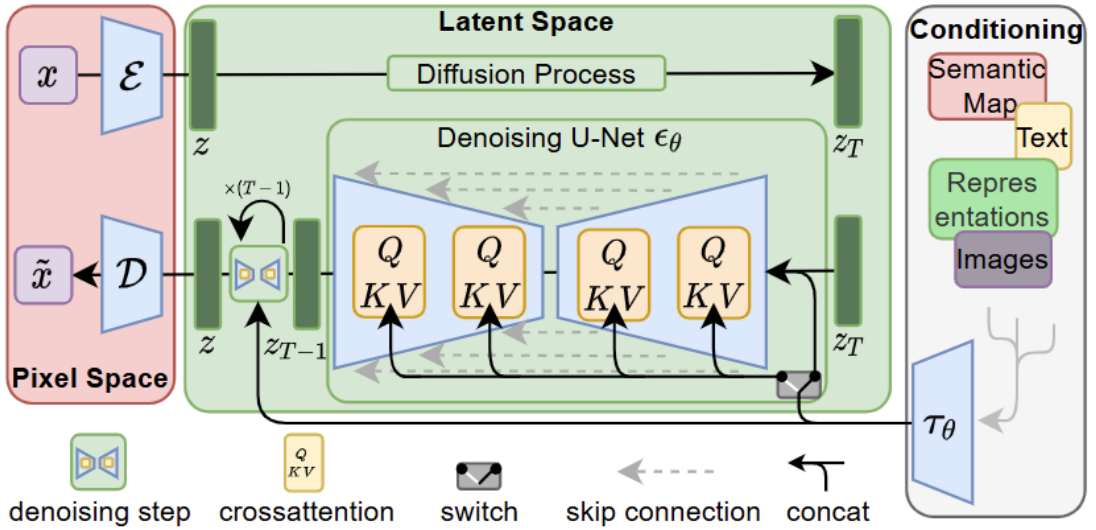


YOLO

Redmon et al., You Only Look Once:Unified, Real-Time Object Detection, CVPY 2016.



Dosovitskiy et al., An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, ICLR 2021.



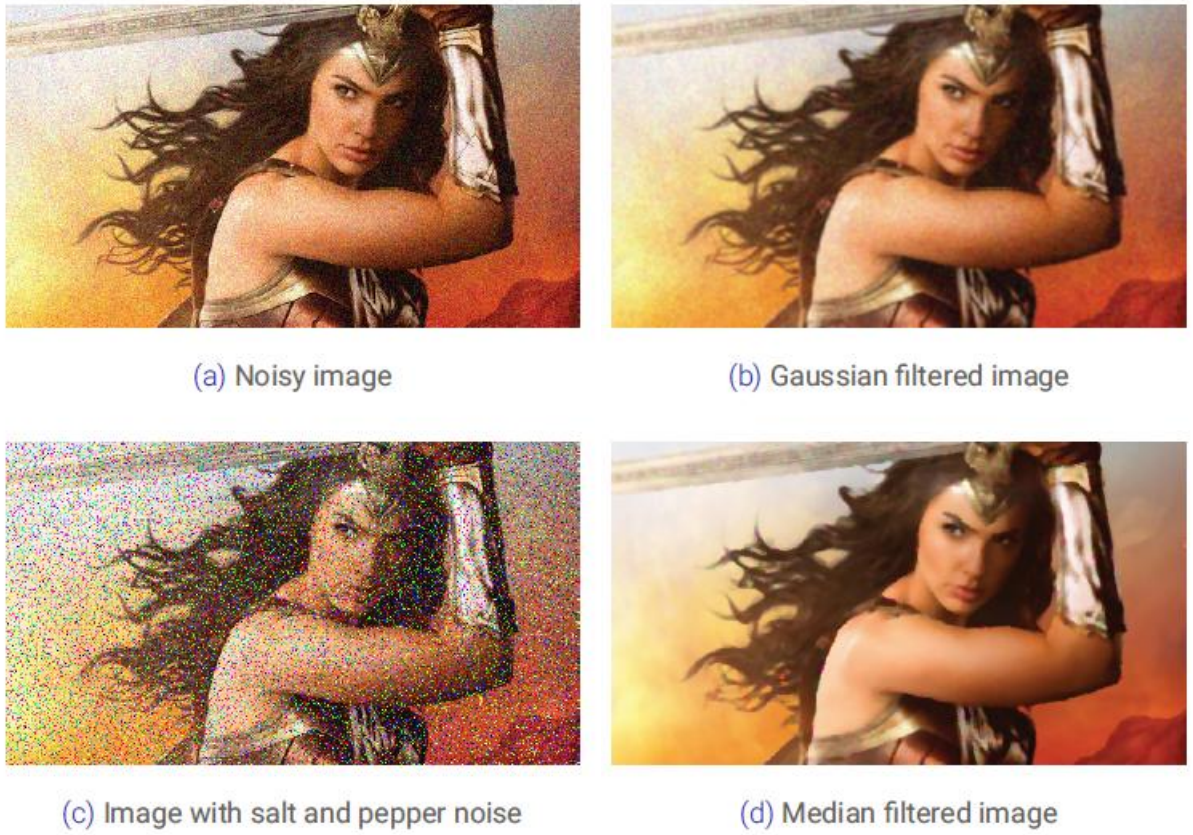
Rombach et al., High-Resolution Image Synthesis with Latent Diffusion Models, CVPR 2022.

Agenda

1	Image processing and computer vision	
2	Introduction to AI, ML, and Deep Learning, AI,	What is machine learning, neural networks, and deep learning? Convolutional Neural Networks (CNNs) Attention and transformers
3	Introductions to Generative AI	Introduction to generative AI Introduction to self-supervision Generative AI applications
4	ENTC Vision Projects	Projects done at the department

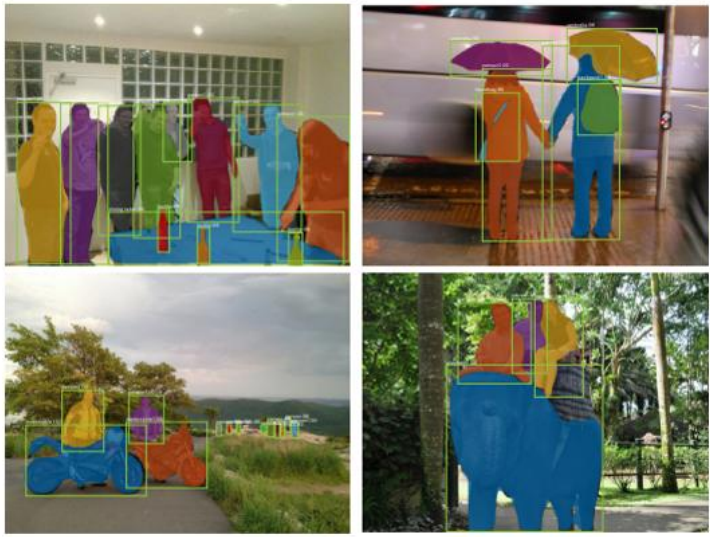
What Is Image Processing?

- 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 (?).

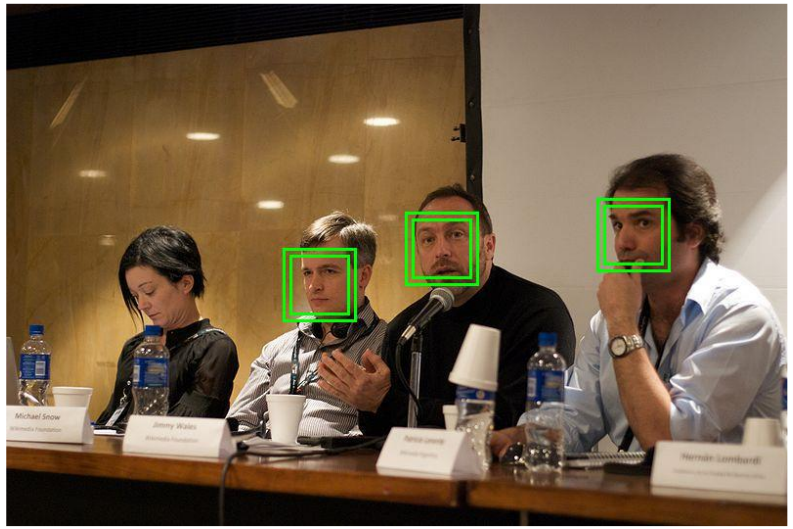


What is Computer Vision?

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



Instance segmentation in Mask R-CNN

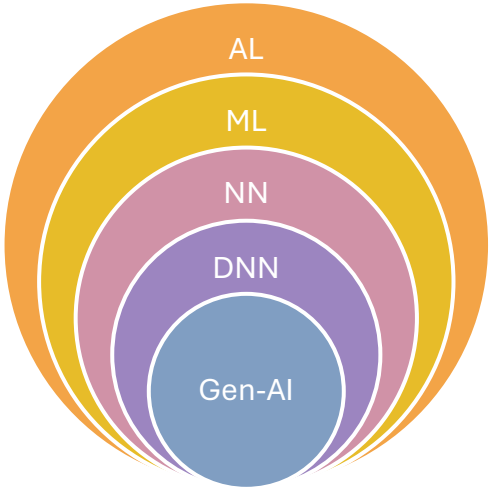


Face detection (Wikepedia)

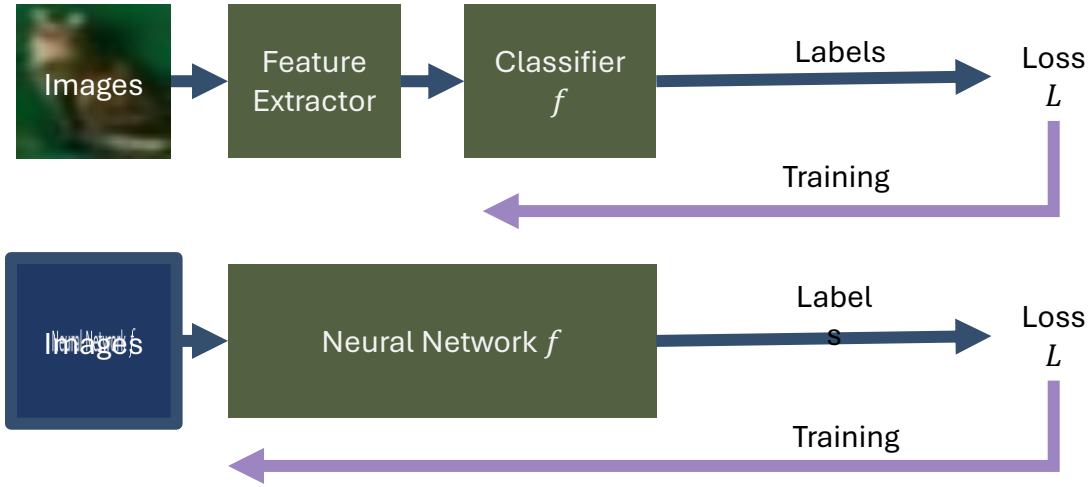
Google Whisk: Generate an image of a beautiful and serene scene of a young man asking for the hand of a beautiful girl.

AI, ML, NN, and DEEP LEARNING

Deep learning is a machine learning advancement that allows the system to automatically discover (learn) **representations** (representation learning) using a stack of multiple (deep) layers (of neurons, a neural network).



- Artificial intelligence
- Machine learning
- Neural networks
- Deep neural networks
- Generative AI



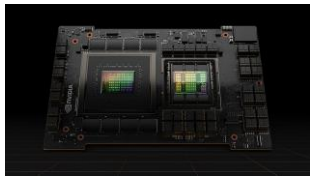
https://cs231n.github.io/linear-classify/
https://www.nvidia.com/en-us/data-center/a100/

WHY NOW

Research → industry
Computation power

Intel i9

Nvidia RTX 4090



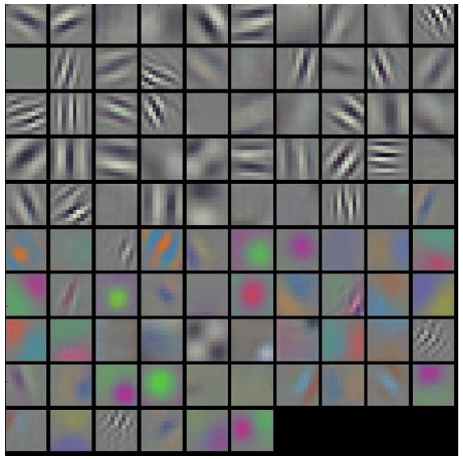
Nvidia H100

TensorFlow

PyTorch

ONNX

TensorRT



AlexNet layer 1, 11 × 11 weights

DISRUPTIVE ADVANCEMENTS

Perceptron 1958
Convolutional neural networks 2012
GPU based parallelization
Transformers 2017
Generative pre-trained models 2022

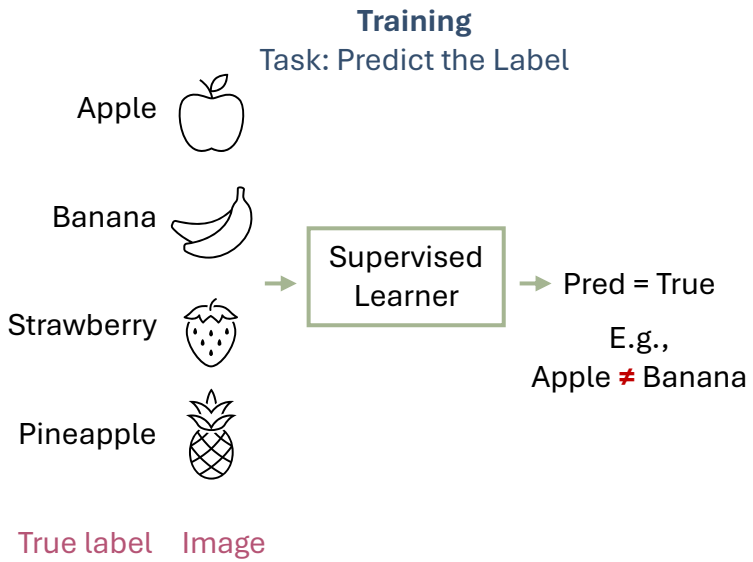
https://www.nvidia.com/en-us/geforce/graphics-cards/30-series/rtx-3090/
https://www.intel.com/content/www/us/en/products/details/processors/core/x.html

Major Areas of Application of AI

Area	Examples	Key Technologies	Example Use Cases
Computer Vision	Tesla's self-driving Avatar 2 production	CNNs, Transformers	Defect detection Volume application VR displaying Self driving
Natural Language Processing (NLP)	Google search, Google's Bard OpenAI's ChatGPT Meta's Llama2	Transformers	Public chatbots Negotiators
Data Analytics	Amazon's product recommender system YouTube's video recommended system Quantitative finance models of a hedge fund	Machine learning, big data, NLP, cloud, visualizations	Recommender systems Customer profiling Business analytics

Supervised learning

Learns from labeled examples, like a student studying with answers provided. Needs pre-labeled data (think flashcards).



Self-supervised learning

Trains on unlabeled data by creating its own supervisory signals. Discovers patterns by itself. Two common approaches: reconstruction, contrastive learning.

Reinforcement learning

Learns through trial and error, like an animal receiving rewards for desired actions. Interacts with an environment and gets feedback (rewards or penalties) to improve.

Supervised learning

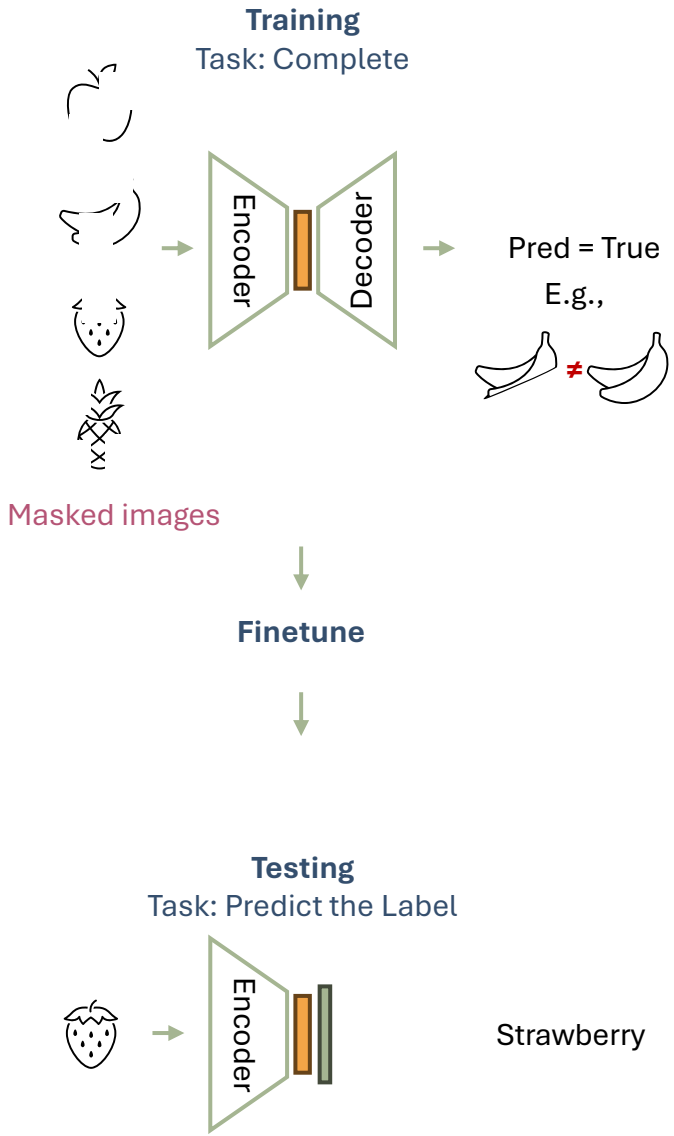
Learns from labeled examples, like a student studying with answers provided. Needs pre-labeled data (think flashcards).

Self-supervised learning

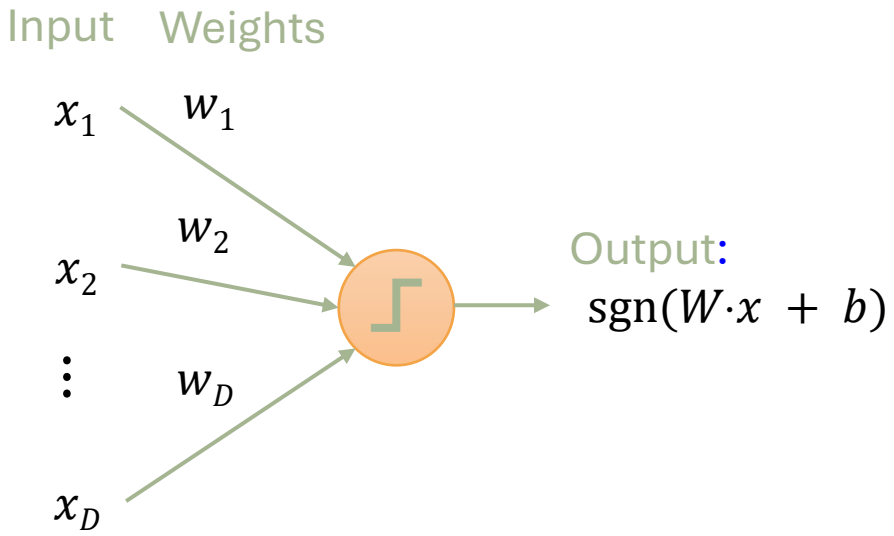
Trains on unlabeled data by creating its own supervisory signals. Discovers patterns by itself. Two common approaches: reconstruction, contrastive learning.

Reinforcement learning

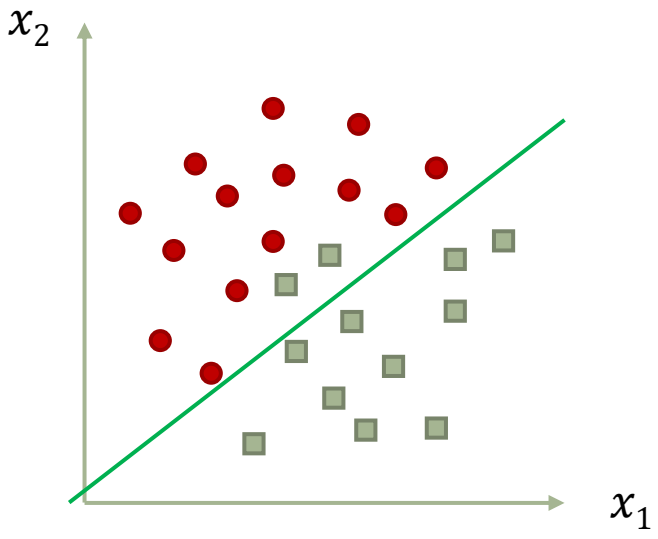
Learns through trial and error, like an animal receiving rewards for desired actions. Interacts with an environment and gets feedback (rewards or penalties) to improve.



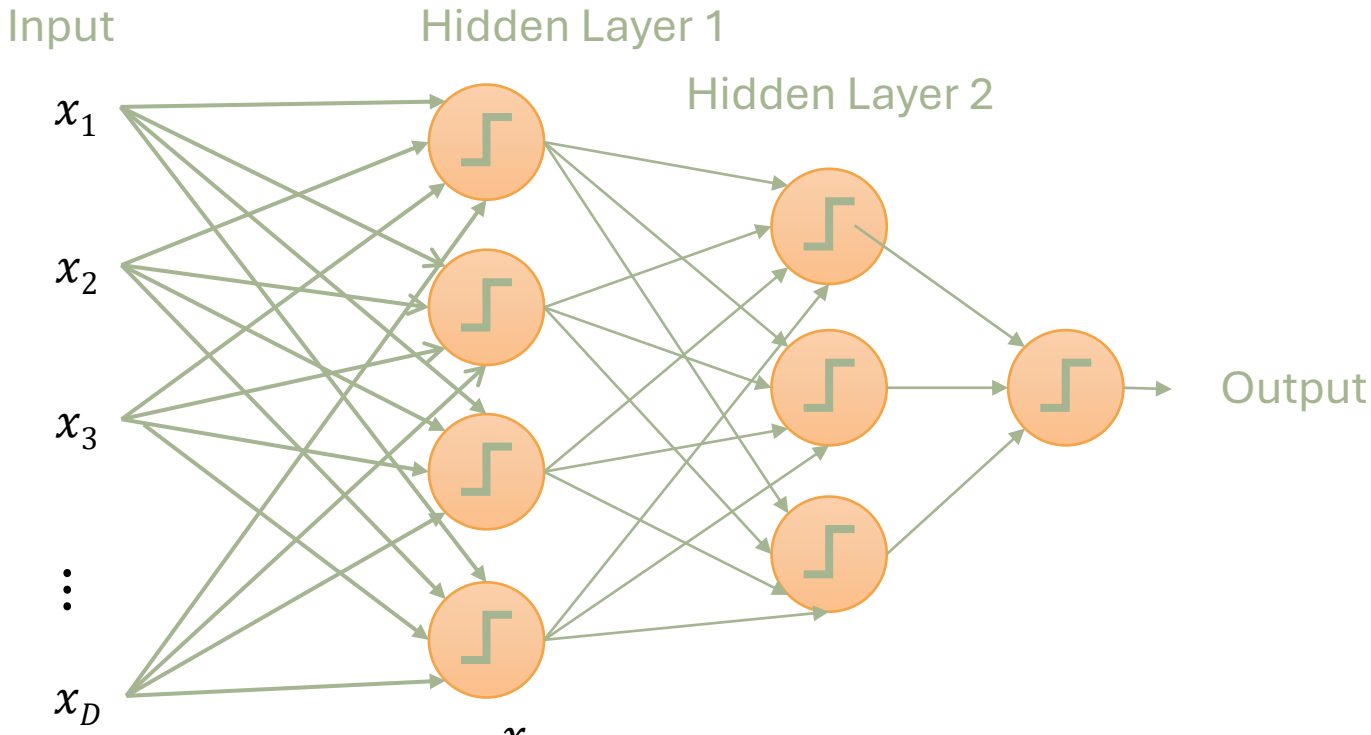
The Perceptron



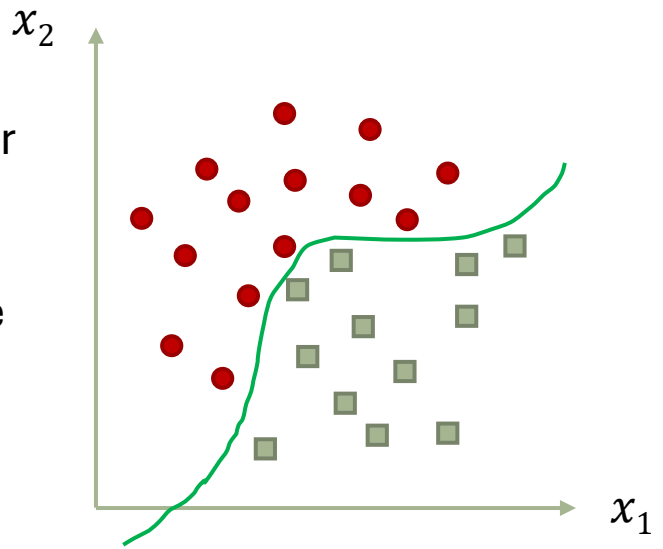
Perceptron is a linear classifier.



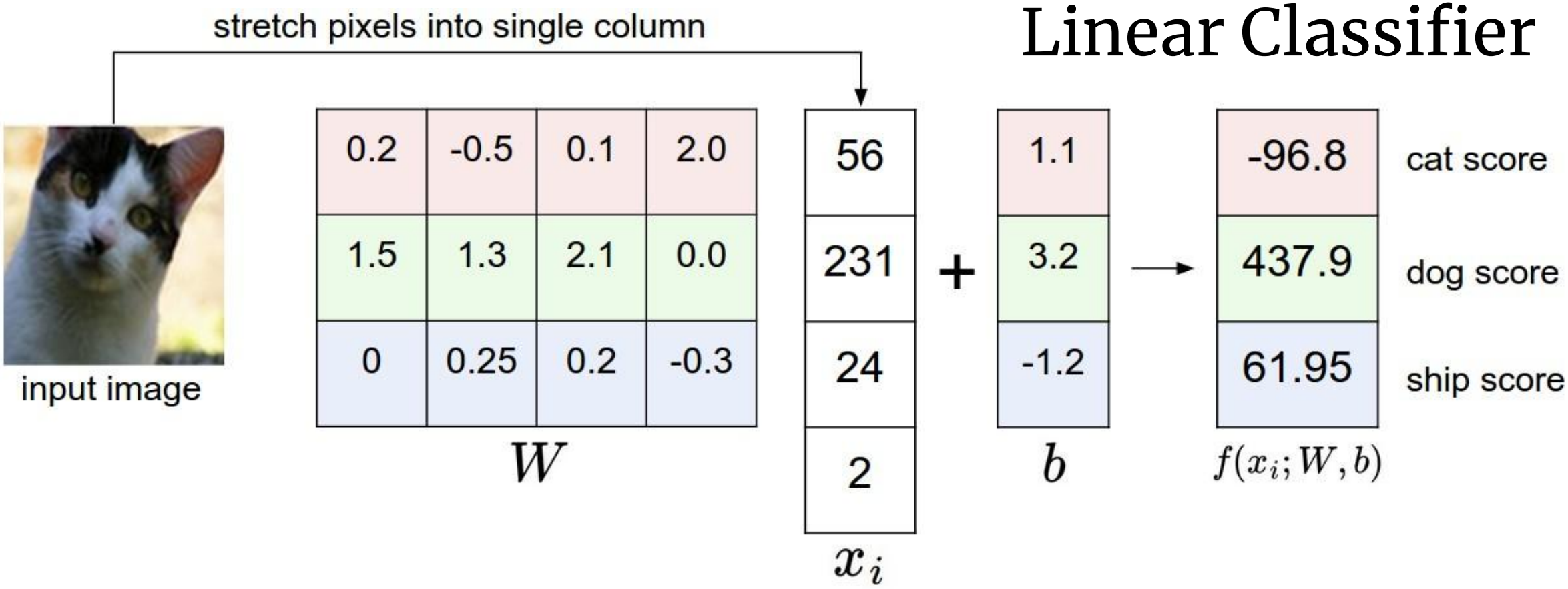
A Multi-Layer Perceptron



Can learn nonlinear functions provided each perceptron has a differentiable nonlinearity.



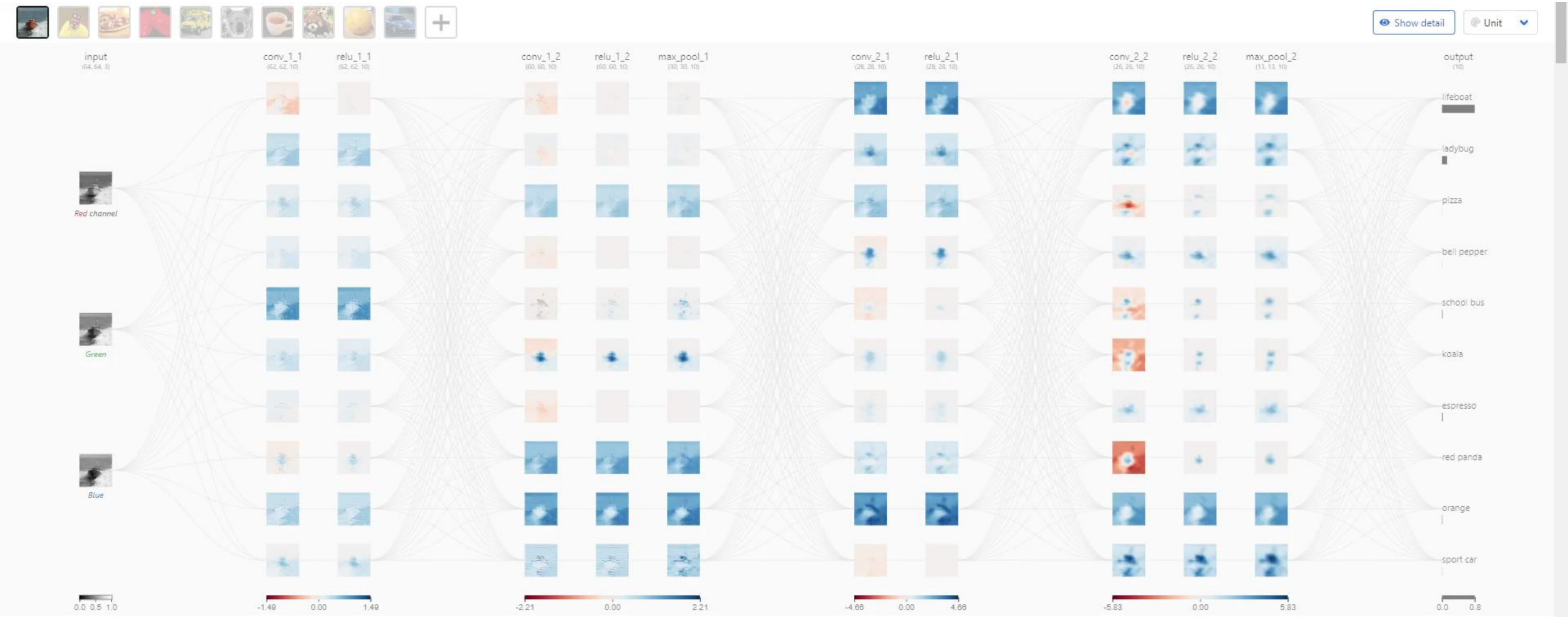
Linear Classifier



Example learned weights at the end of learning for CIFAR-10. Note that, for example, the ship template contains a lot of blue pixels as expected. This template will therefore give a high score once it is matched against images of ships on the ocean with an inner product.

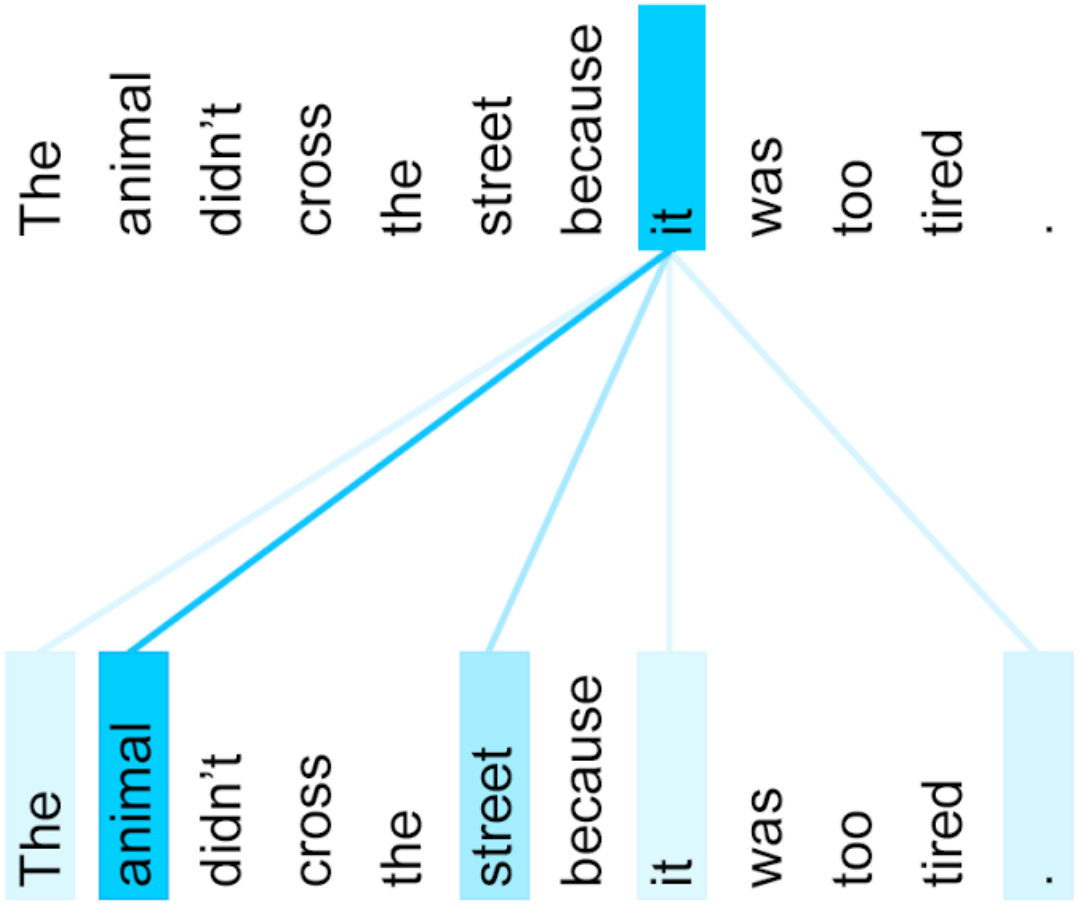
https://cs231n.github.io/linear-classify/

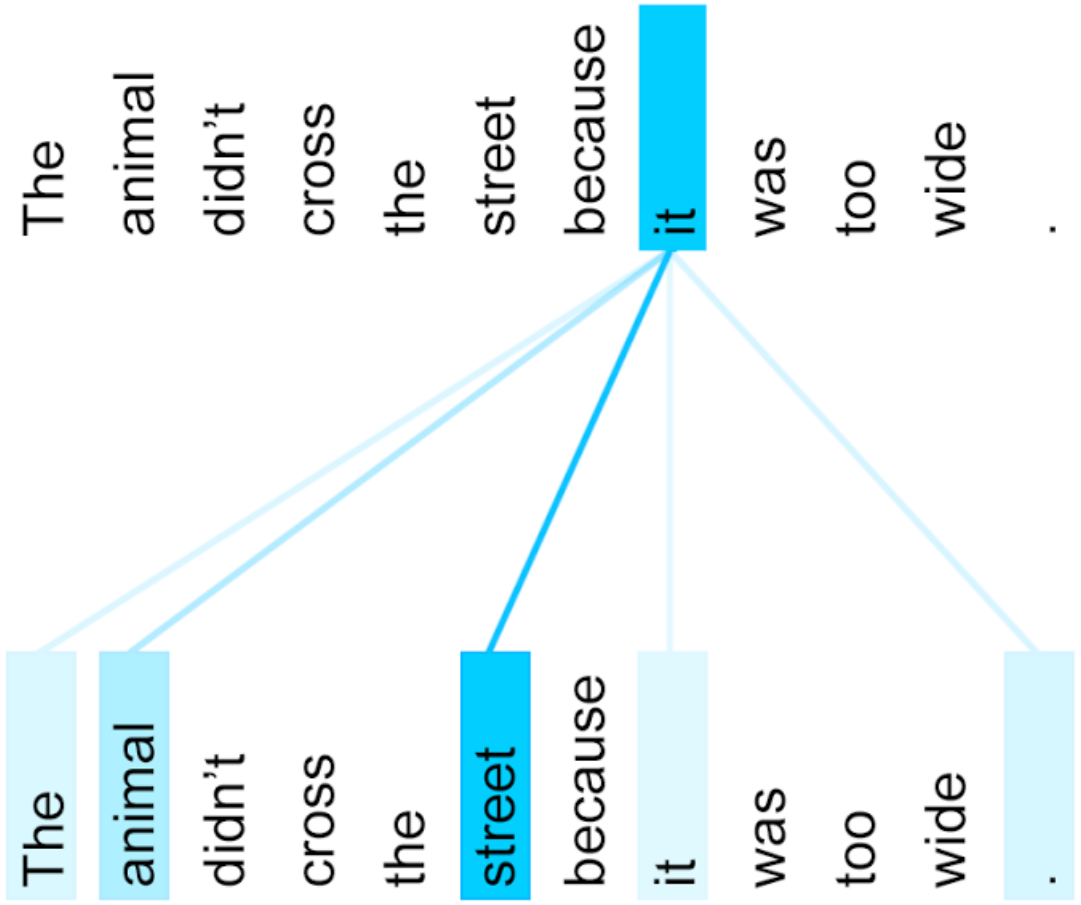
CNNs in Image Classification



https://poloclub.github.io/cnn-explainer/

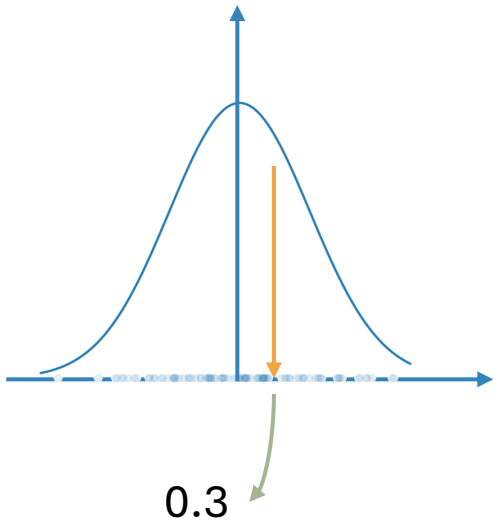
Attention (in Transformers) in NLP

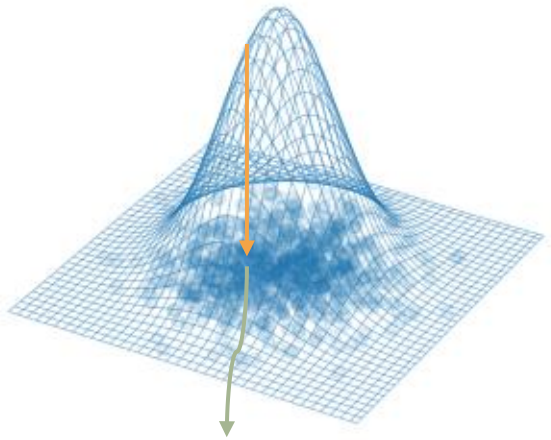




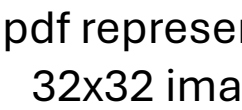
https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html

Generative Models

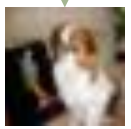




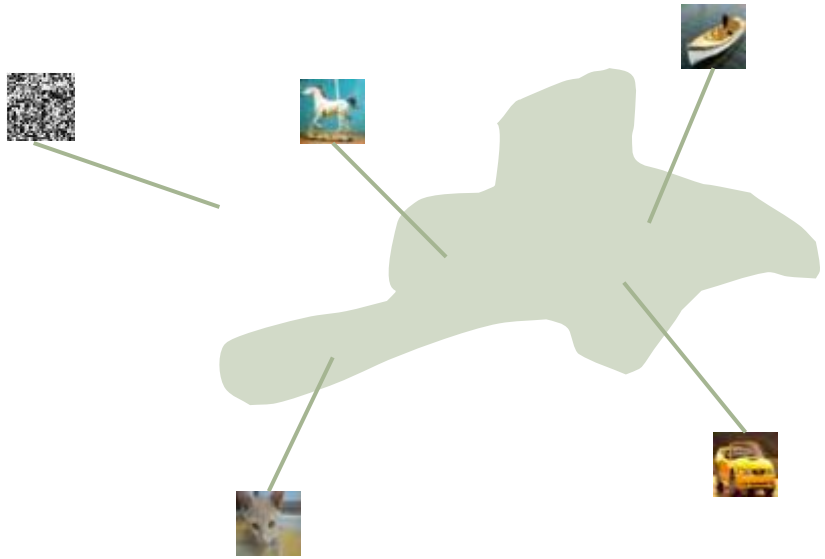
pdf representing
32x32 images



sample a point



- Generative models learn the underling distribution of data.
- Once learned, they can generate new data samples that resemble the training data.



Generative AI

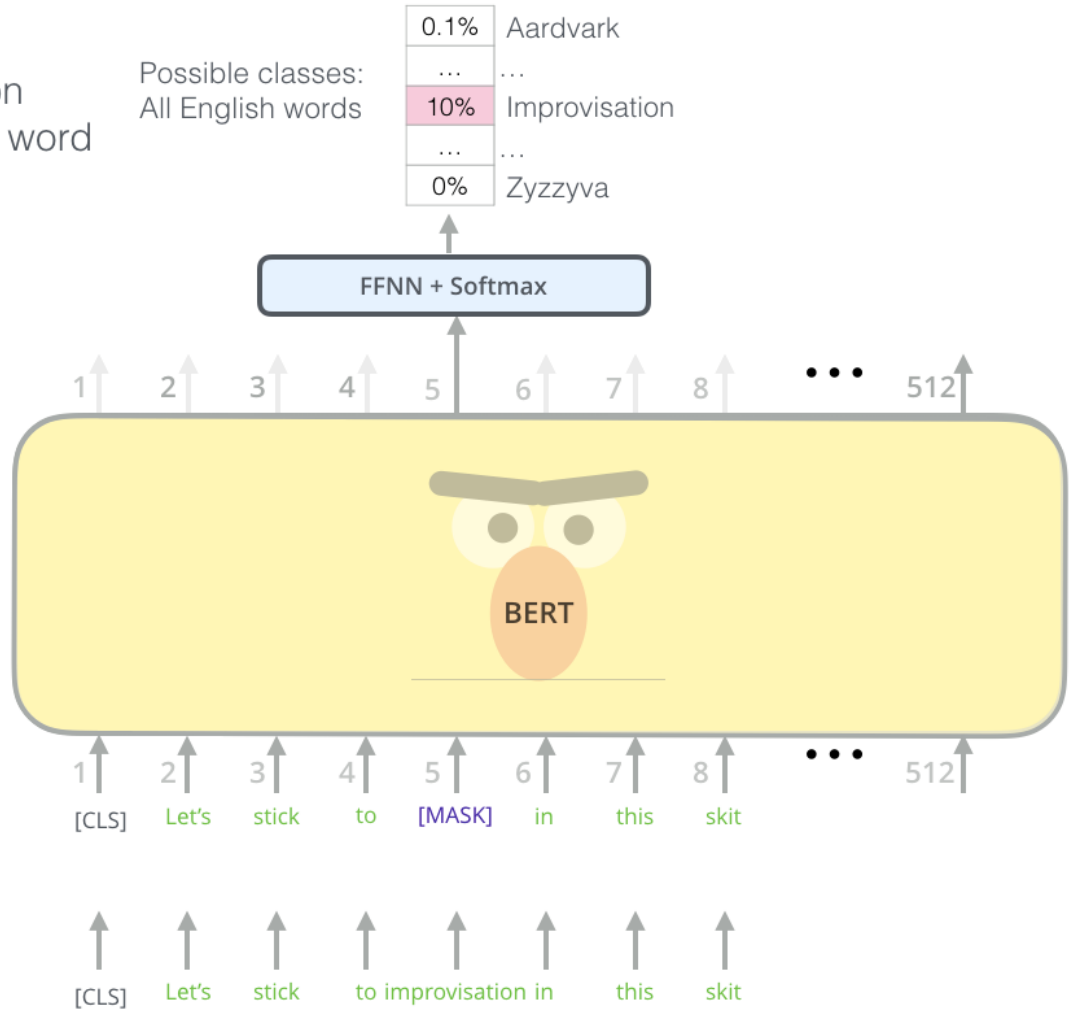
Generates new content, data, or information that is similar to, or mimics, human-created content.

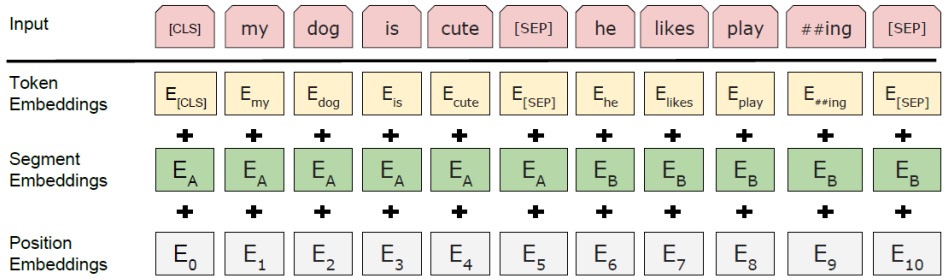
Text Generation	Generative AI models like Generative Pre-trained Transformers (GPT) can generate human-like text	Chatbots, and content creation. E.g., ChatGPT, Gemini.
Image and Video Generation	Stable diffusion generates realistic images, art, and visual content based on a textual description	Advertising and entertainment industry. E.g., MidJourney
Data Augmentation	Generative AI can generate additional data samples to augment datasets for training machine learning models	Training models with less actual data
Music Composition	AI systems can generate music compositions in various styles and genres, based on patterns learned from existing music.	Entertainment industry

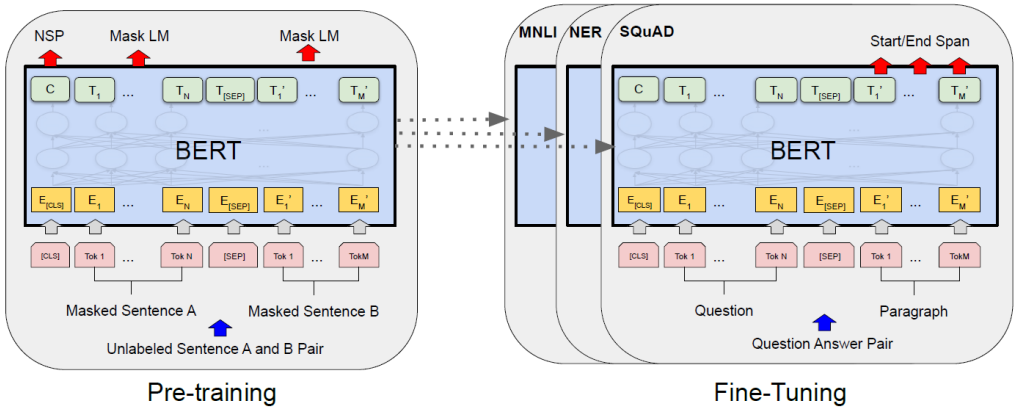
BERT

Devlin et al., BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL 2019.

Use the output of the masked word's position to predict the masked word

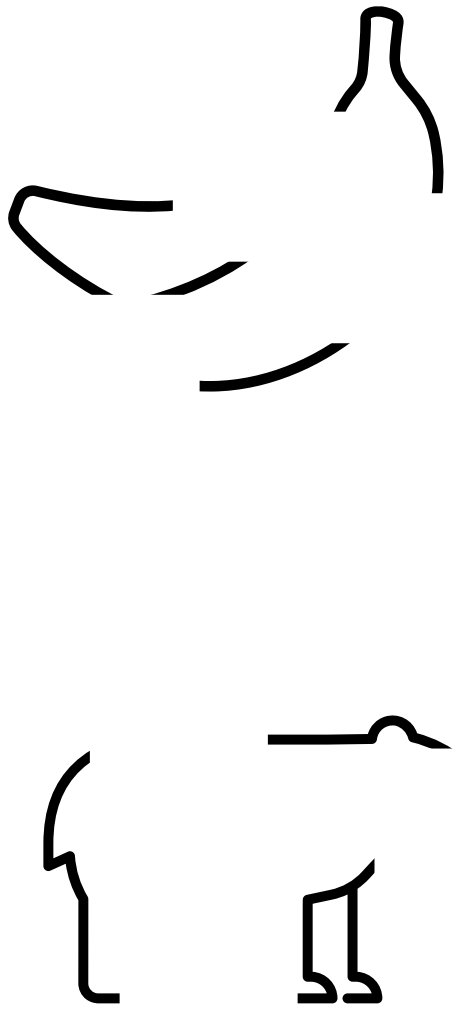


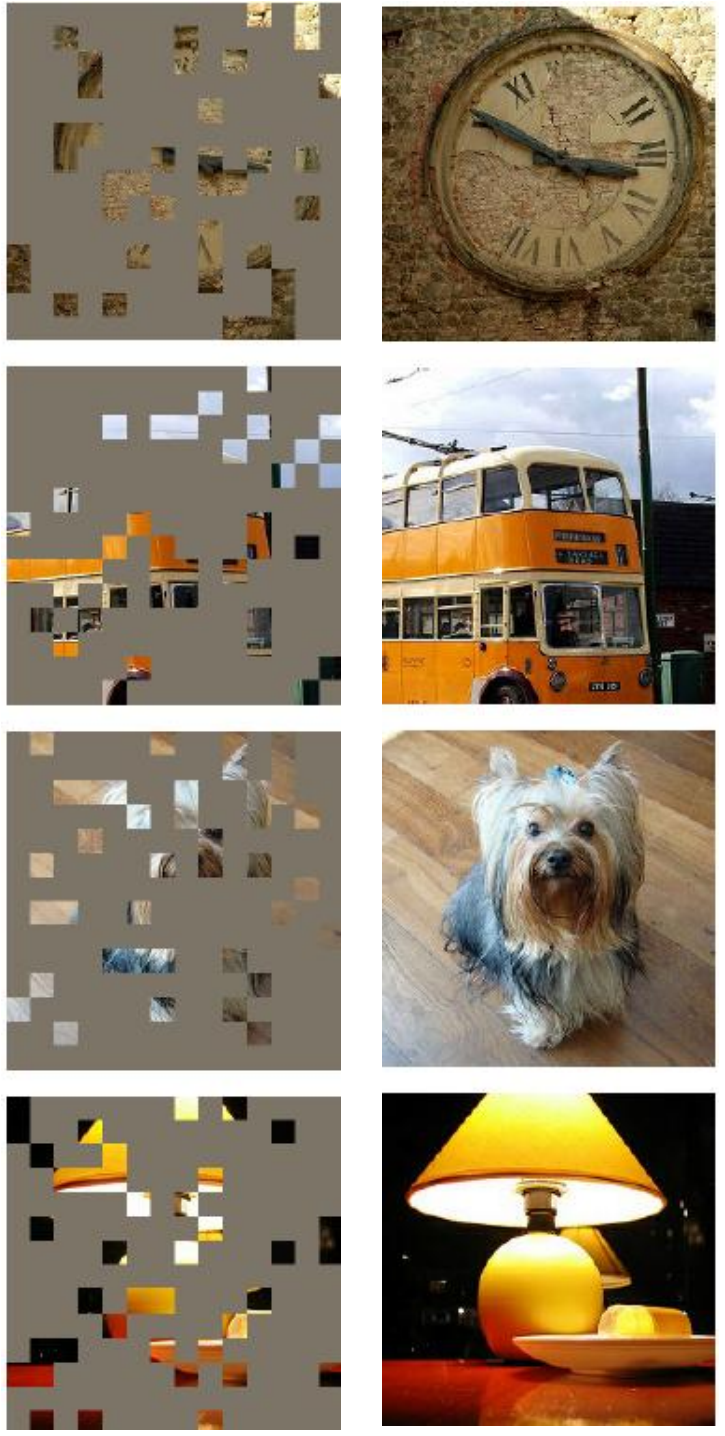




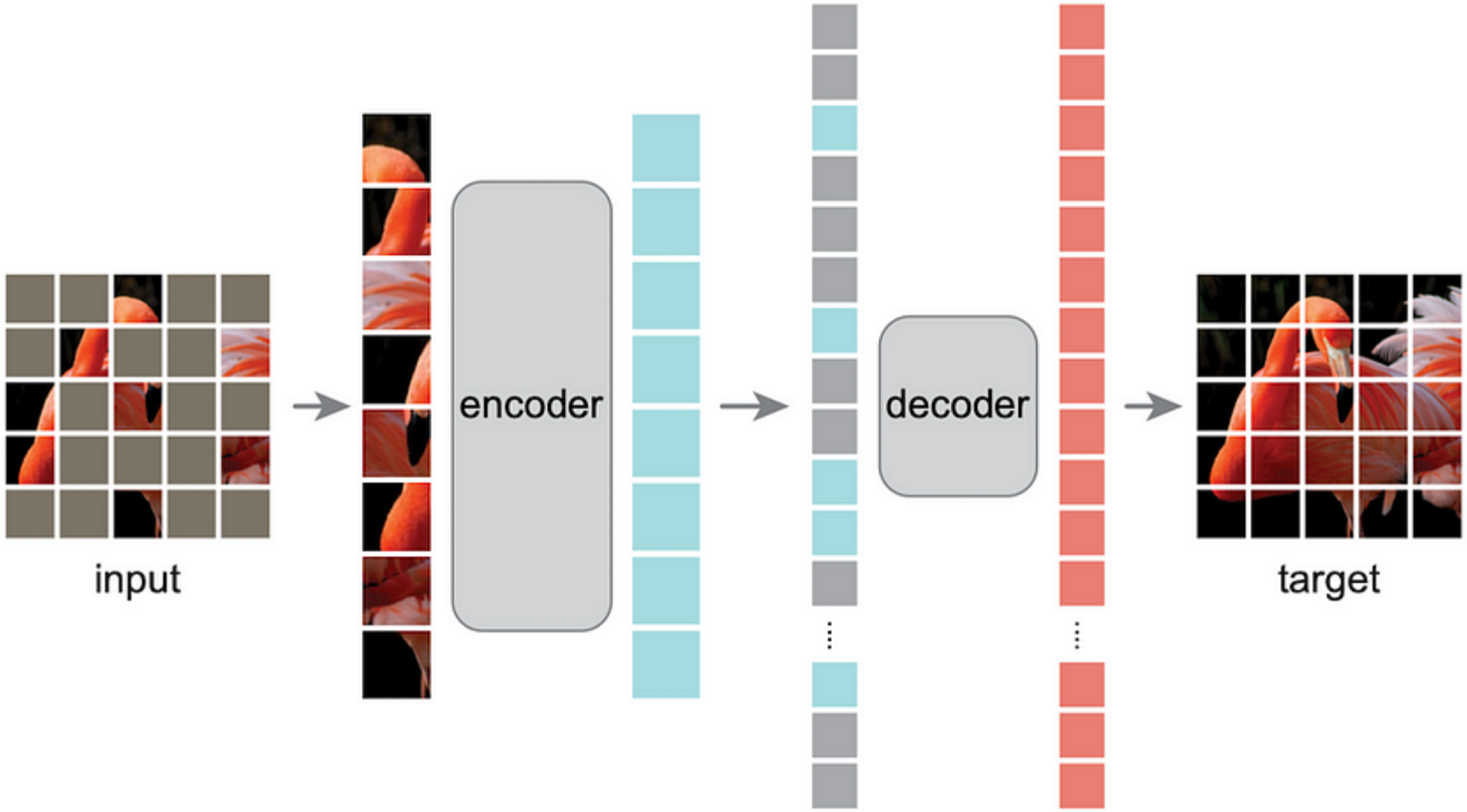
https://jalammar.github.io/illustrated-bert/

Fun Exercise: What is Hidden?





Self-Supervised Pre-Training: Masked Auto Encoder

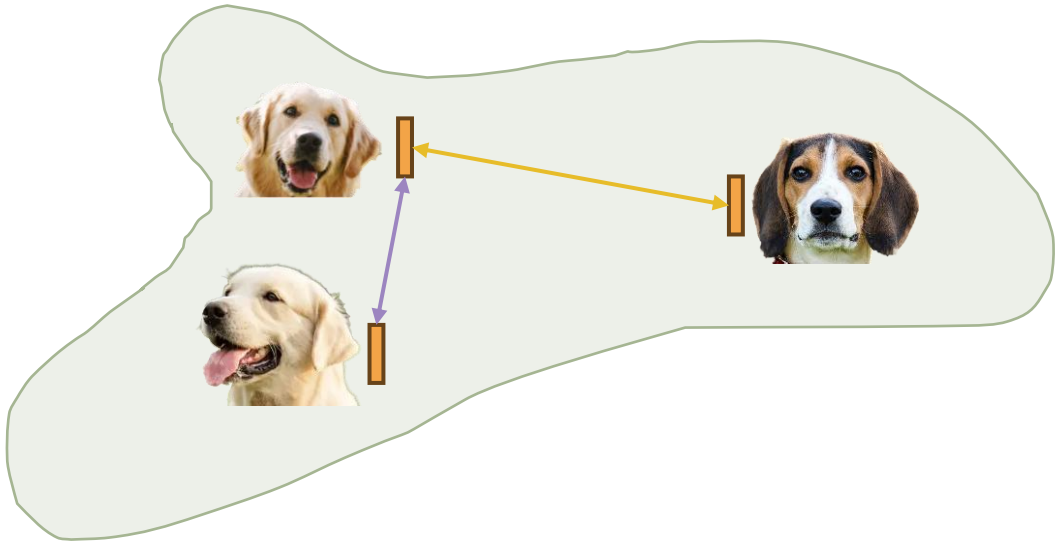


He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2022). Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 16000-16009).

Contrastive Learning

Encoder

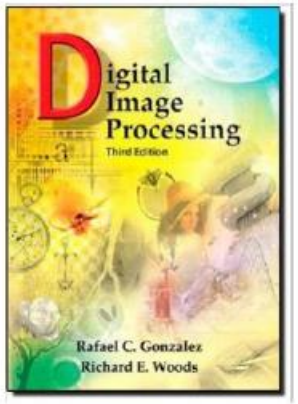
Bring together representations of similar examples.
Push apart representations of deferring examples.

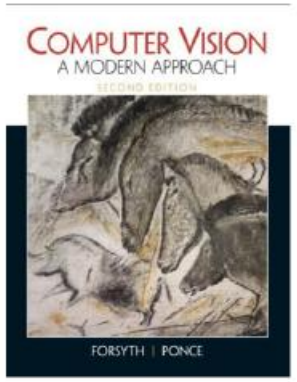


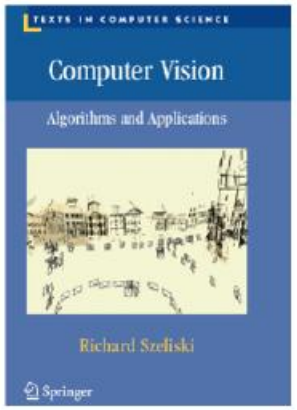
Encoder learns how to “Represent” dogs

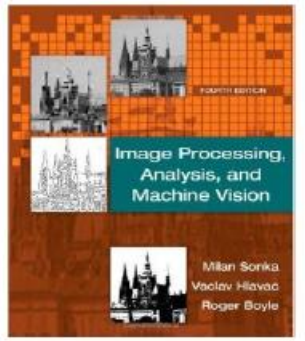
Text Books, Journals, and Conferences

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









Conferences: CVPR, ICCV, ECCV, WACV, BMVC, ICIP
Journals: PAMI, IJCV, Pattern Recognition, TIP, Cybernetics, Computer Vision and Image Understanding

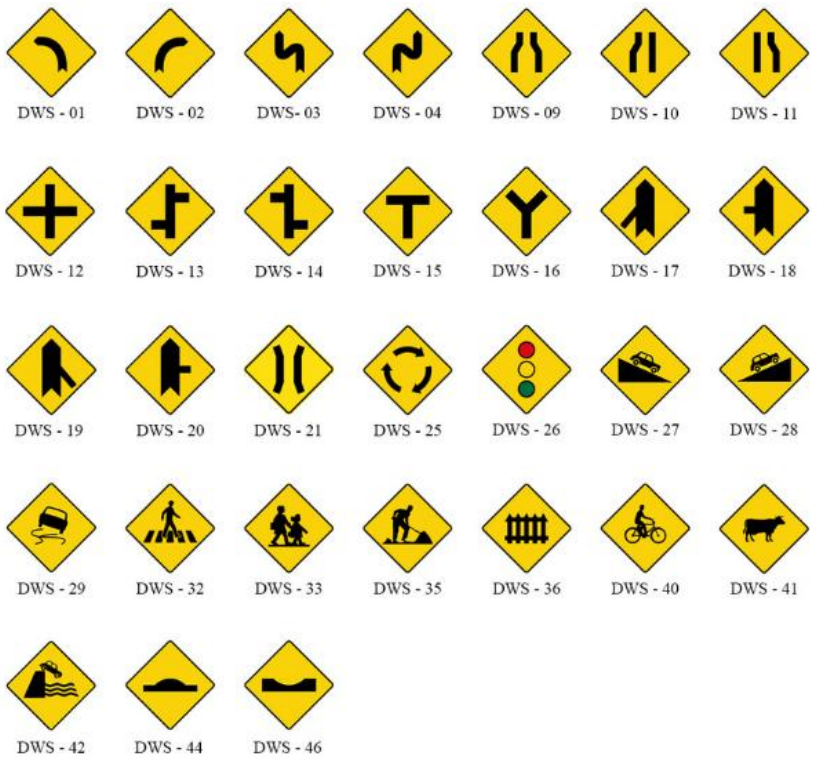
Tips for Successful Completion

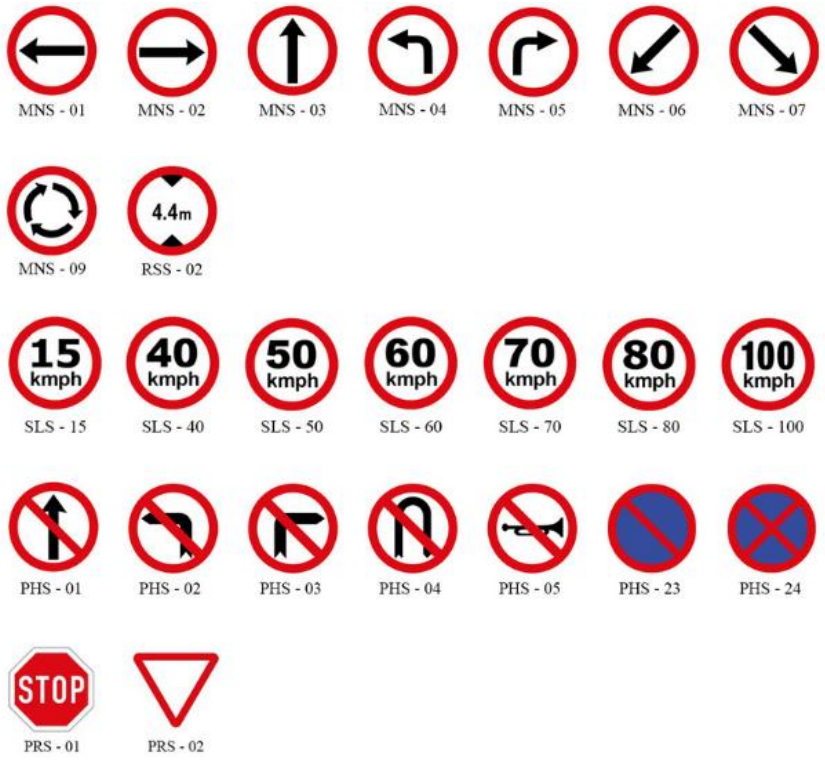
1. Mindset: “I want to solve this vision problem. What are the tools available? How have others solved it? How should I go about solving it?”
2. Attend every single lecture.
3. Go through the material the night before and engage in a good discussion in class.
4. Implement at least one algorithm discussed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

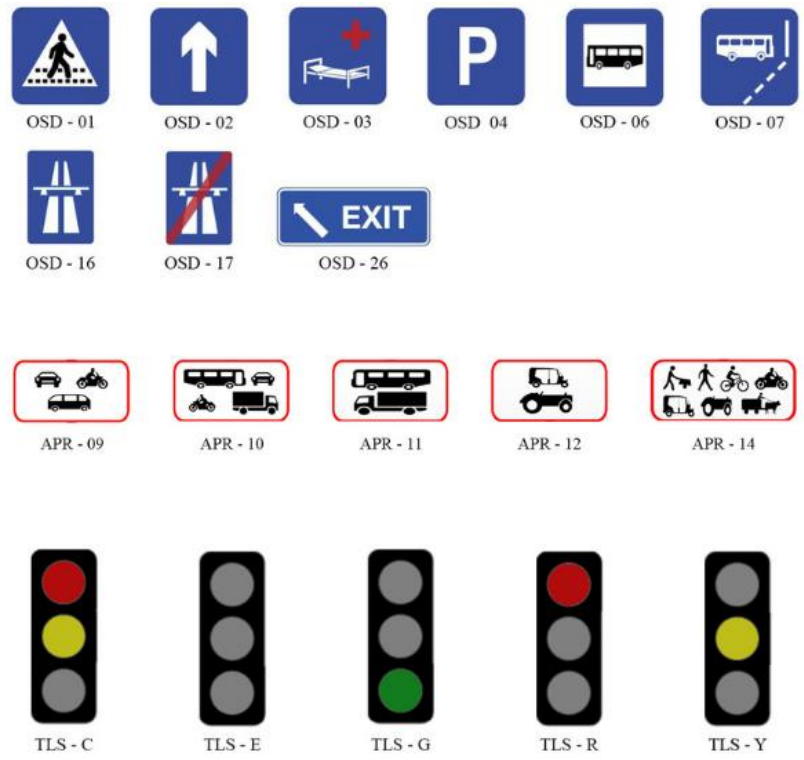
Software and Cameras for Industrial Vision

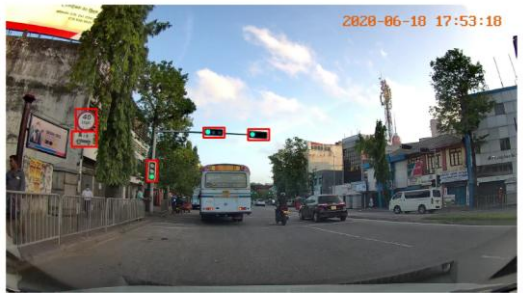
1. In EN3160:
   - We use Python, OpenCV and PyTorch.
2. Computer vision software companies
   - https://www.cognex.com/
   - https://www.mvtec.com/
3. Camera companies
   - https://www.baslerweb.com/en/
   - https://www.flir.com/iis/machine-vision/
4. Camera catalogue
   - https://www.edmundoptics.com/c/cameras/1012/

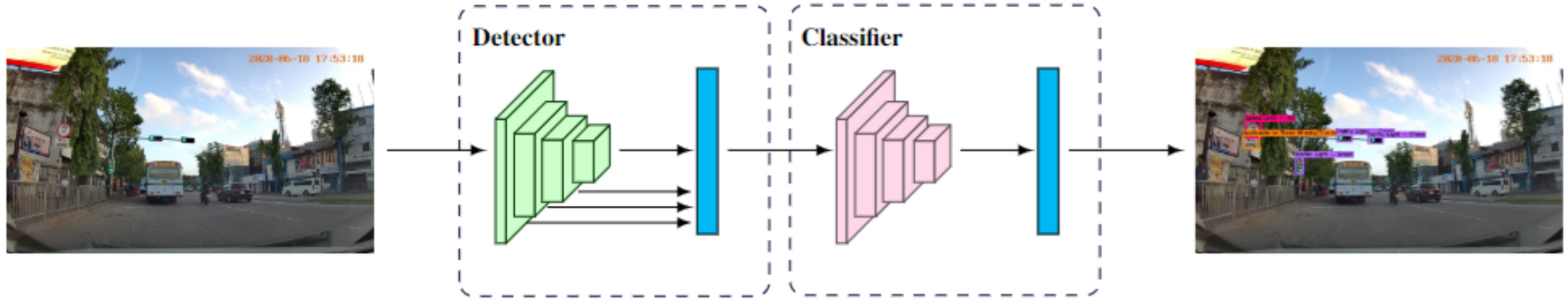
Traffic Sign and Traffic Light Detection on Embedded Systems





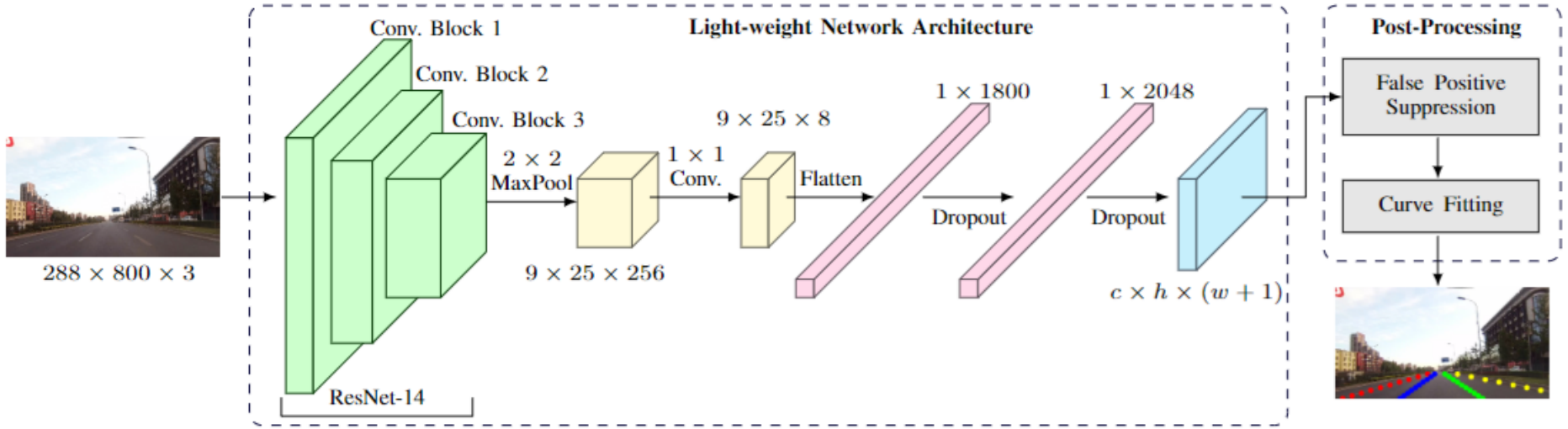






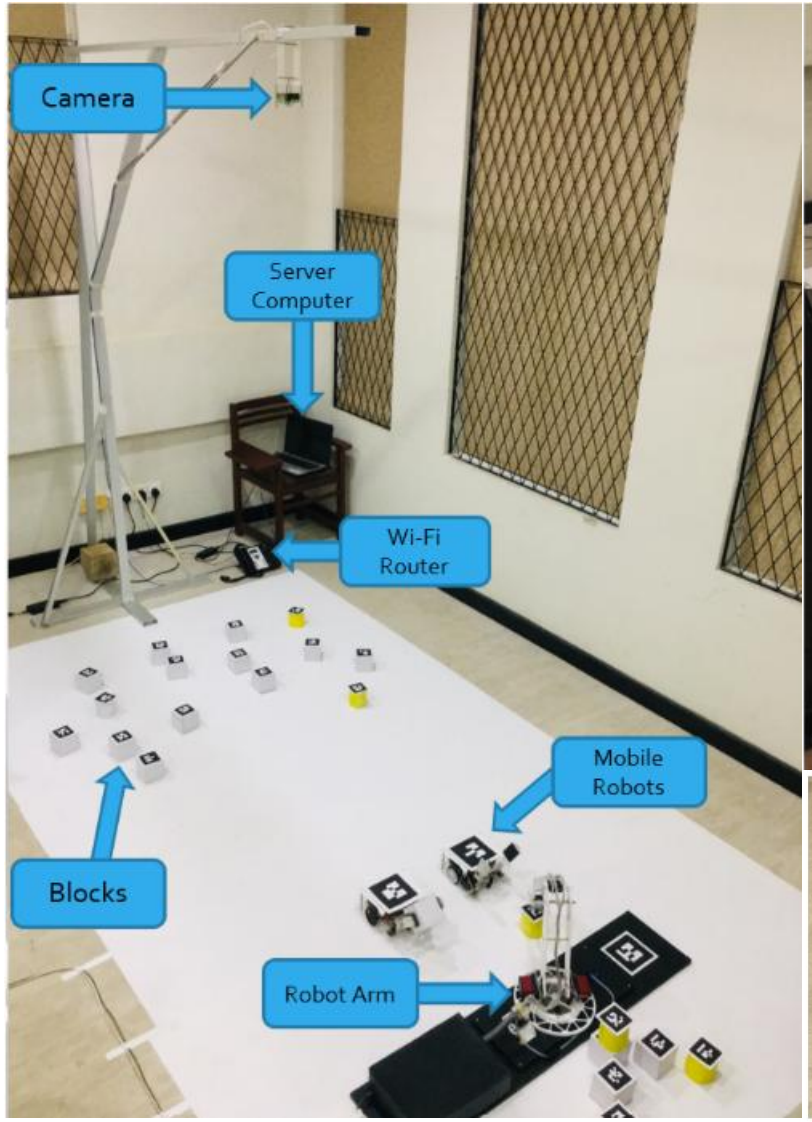
Jayasinghe, Oshada, Sahan Hemachandra, Damith Anhettigama, Shenali Kariyawasam, Tharindu Wickremasinghe, Chalani Ekanayake, Ranga Rodrigo, and Peshala Jayasekara. "Towards real-time traffic sign and traffic light detection on embedded systems." In *2022 IEEE Intelligent Vehicles Symposium (IV)*, pp. 723-728. 2022.

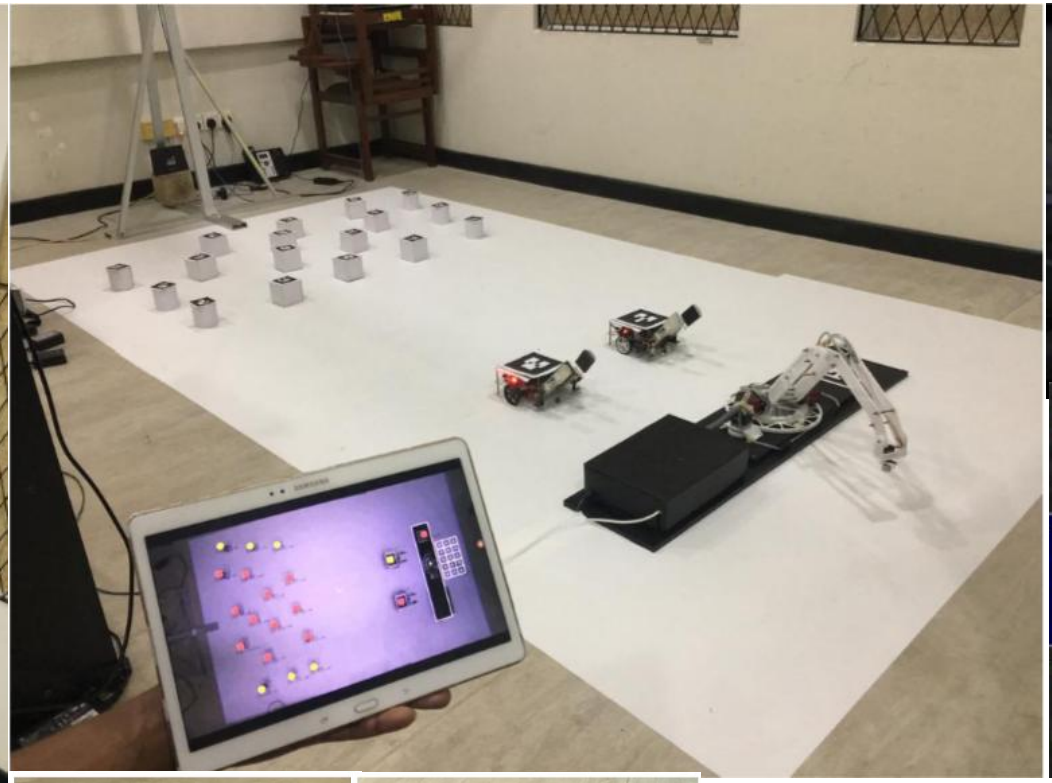
SwiftLane: Towards Fast and Efficient Lane Detection

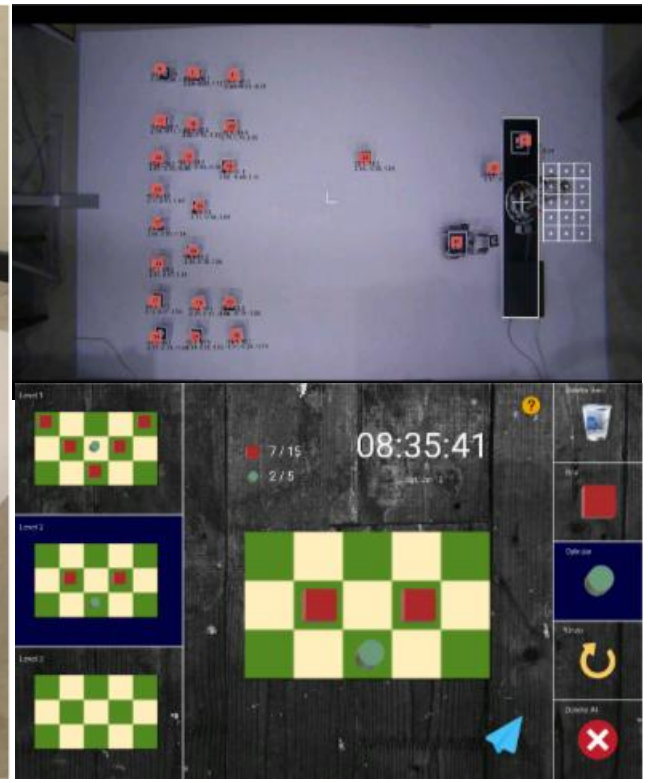


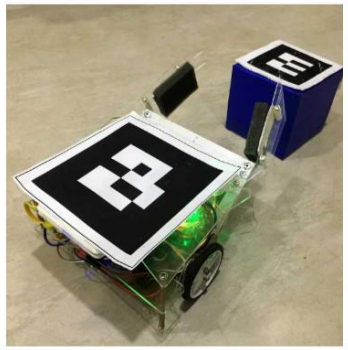
Jayasinghe, Oshada, Damith Anhettigama, Sahan Hemachandra, Shenali Kariyawasam, Ranga Rodrigo, and Peshala Jayasekara. "SwiftLane: towards fast and efficient lane detection." In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 859-864, 2021.

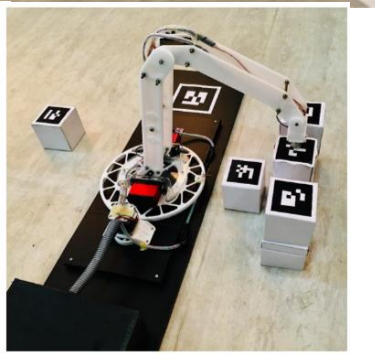
AR based Human Interactions for Automating a Robot Team





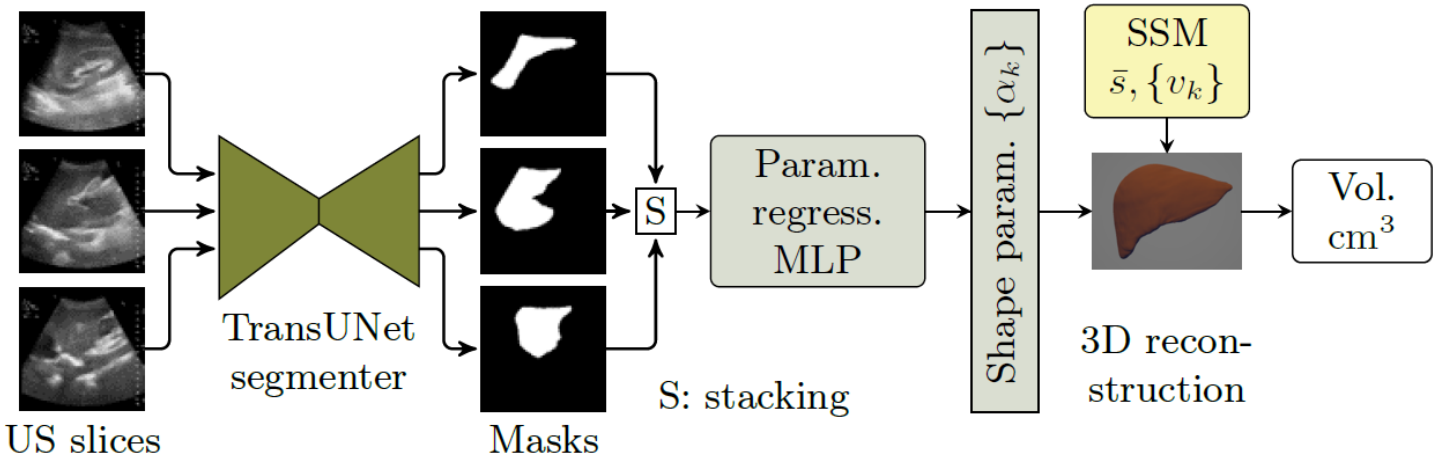


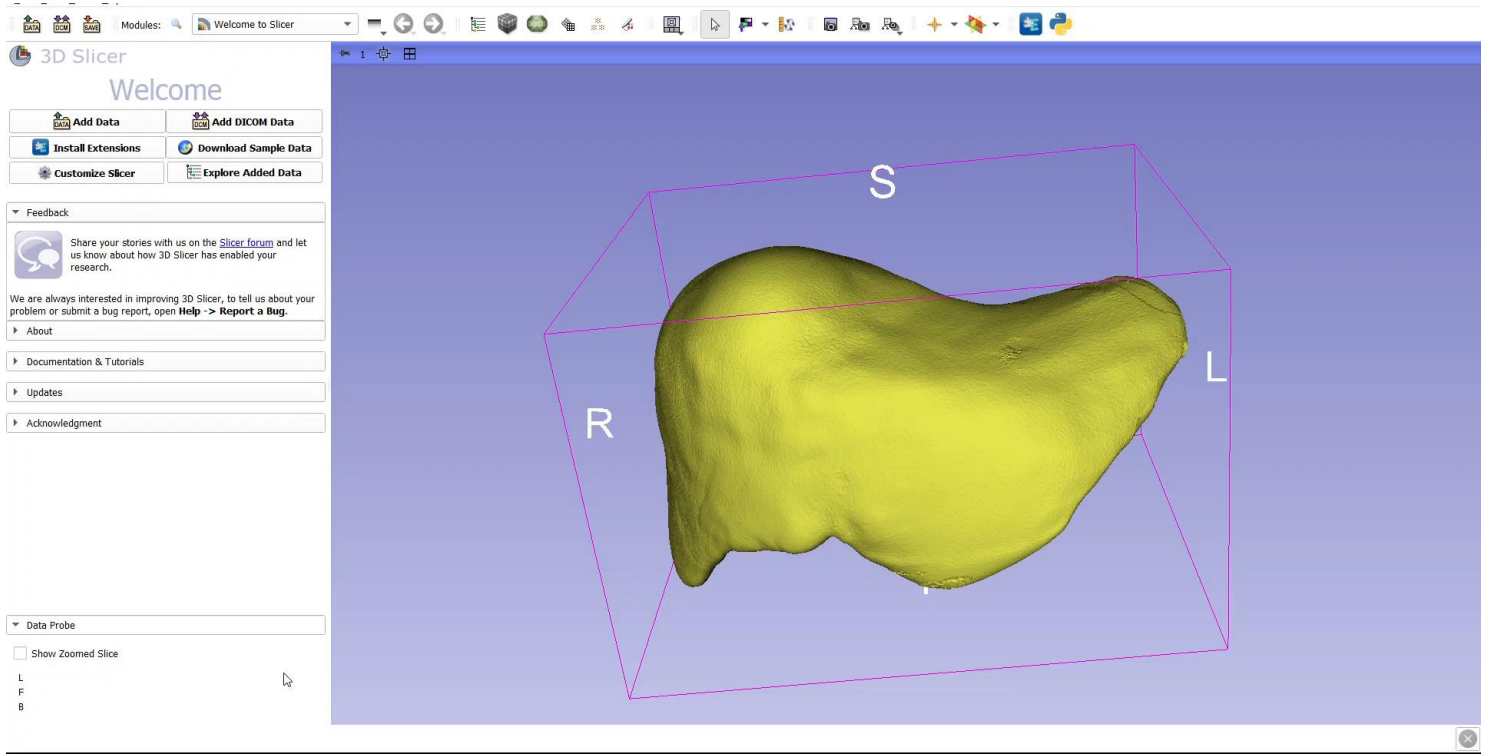




Dias, Adhitha, Hasitha Wellaboda, Yasod Rasanka, Menusha Munasinghe, Ranga Rodrigo, and Peshala Jayasekara. "Deep learning of augmented reality based human interactions for automating a robot team." In *2020 6th International Conference on Control, Automation and Robotics (ICCAR)*, pp. 175-182, 2020.

3D Reconstruction of the Liver with a Few US Scans





2 or 3 US scans
Volumetry possible

Kaushalay Sivayogaraj
Sahan Guruge
Jeevani Udupilille
Fariha Sitheeque
Saroj Jayasinghe
Gerard Fernando
Ranga Rodrigo
Rukshani Liyanaarachci

Patent Pending.

1. What is the primary difference between image processing and computer vision?
A. Image processing enhances images; computer vision interprets them
B. Image processing is used only in photography; computer vision is used in robotics
C. Computer vision converts analog signals to digital; image processing doesn't
D. Image processing needs machine learning; computer vision doesn't
E. There is no difference

Which of the following is an example of a computer vision task?
A. Histogram equalization
B. Gaussian blur
C. Semantic segmentation
D. Image sharpening
E. Padding

3. What is the goal of early vision in a computer vision pipeline?
A. Compressing video
B. Extracting high-level semantics from images
C. Creating 3D point clouds
D. Extracting low-level features like edges and textures
E. Rendering synthetic scenes

4. Which of the following best describes the structure of this course (EN3160)?

A. It focuses entirely on deep learning for image generation
B. It includes image processing, computer vision, and recent vision methods
C. It is limited to classical image processing
D. It covers only industrial inspection techniques
E. It is a hardware-oriented course

5. Which of the following is recommended for succeeding in the course?

A. Memorize all the equations before exams
B. B. Implement at least one algorithm before the next class
C. C. Avoid reading papers; just follow lecture notes
D. D. Focus only on coding assignments
E. E. Use MATLAB and ignore Python

6. What is an advantage of using PyTorch in EN3160?

A. It requires no coding knowledge
B. It is optimized only for image compression
C. It supports dynamic computation graphs and deep learning libraries
D. It is a hardware simulation tool
E. It only works with C++

7. Which of the following is a real-world application of computer vision?
A. Applying a median filter to smooth an image
B. Detecting faces in surveillance footage
C. Printing high-resolution images
D. Converting RGB to grayscale
E. Saving images in compressed formats

8. What is an example of a generative model in vision?
A. Histogram matching
B. Optical flow estimation
C. Diffusion models
D. Sobel edge detection
E. Bilinear interpolation

9. Which of the following is a typical use-case of segmentation in vision systems?
A. Enhance the brightness of an image
B. Estimate the motion of objects
C. Detect boundaries between different regions in an image
D. Convert images into JPEG format
E. Add color to grayscale images

10. What is meant by “self-supervised pre-training” in deep learning for vision?

A. Training a model on random images from the internet
B. Learning from labeled datasets with minimal supervision
C. Supervised training using multiple GPUs
D. Using reinforcement learning to train vision models
E. Learning meaningful representations from unlabeled data using surrogate tasks