ECG782: MULTIDIMENSIONAL DIGITAL SIGNAL PROCESSING

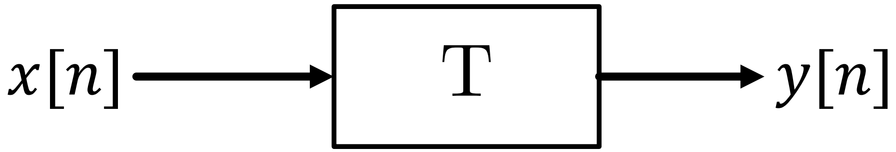
COURSE INTRODUCTION

OVERVIEW

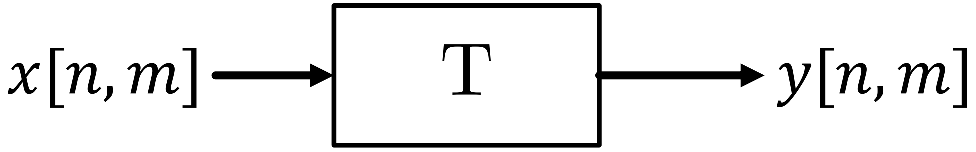
- Multidimensional signal processing
- Computer vision overview

MULTIDIMENSIONAL SIGNAL PROCESSING

- Standard DSP focuses on signals of a single independent variable (often time)
  - Use a system approach



- M-D signal uses more than one dimension for sampling
  - Similar ideas as 1D case
  - Images are popular concrete signal → computer vision



WHAT IS COMPUTER VISION? I

- Given an image, want to answer questions about what we see
- What type of scene is this?
- Is it a sunny day?
- Where was this picture taken?

WHAT IS COMPUTER VISION? I

- Given an image, want to answer questions about what we see
- What type of scene is this?
  - Beach, mountain+water, ...
- Is it a sunny day?
  - Yes
- Where was this picture taken?
  - Hanauma Bay, Hawaii

WHAT IS COMPUTER VISION? II

- Goal is to develop algorithms and programs that can interpret and understand images
  - Image can be a single image or come from a video
- Must bridge the gap between what we see and what a computer "sees"

WHY IS COMPUTER VISION DIFFICULT I

- Humans are very skilled with vision
  - We are designed with vision as our primary sensory input
  - It comes naturally
- Computers operate on numbers and do not have contextual clues we have wired in our brains

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

What a computer sees

WHY IS COMPUTER VISION DIFFICULT II

- Loss of information in 3D → 2D
  - The world is 3D but an image is only 2D → Loss of information from perspective imaging
- Interpretation
  - Many different interpretations of the same image
  - `interpretation:`
    `image data → model`
  - How can we develop a meaningful model?
- Noise – in data and labels
- Big data
  - High resolution imagery, HD video, lots of training data
- Brightness measurement
  - Complicated physical process that is hard to determine from an image
- Local window vs. need for global view
  - Processing done locally but must make inference globally

HUMANS VS. COMPUTERS

- Computers can't currently "beat" humans
  - Humans are much better at "hard" things
  - Computers can be better at "easy" things

- Computers are computational device so must be given memory and learn → machine learning (ML) and deep learning (DL)
- If the task requires lots of attention it may be better suited for a computer
  - Surveillance
  - Automotive blind spot detection
  - Searching for a face in a crowd

CV AS INTELLIGENT SYSTEMS

- Intelligence
  - The capacity to acquire knowledge
  - The faculty of thought and reason
- System
  - A group of interacting, interrelated or interdependent elements forming a complex whole

- This class uses computer vision to give a system intelligence
- The systems should perceive, reason, learn, and act intelligently

VISION

- Signal to symbol transformation

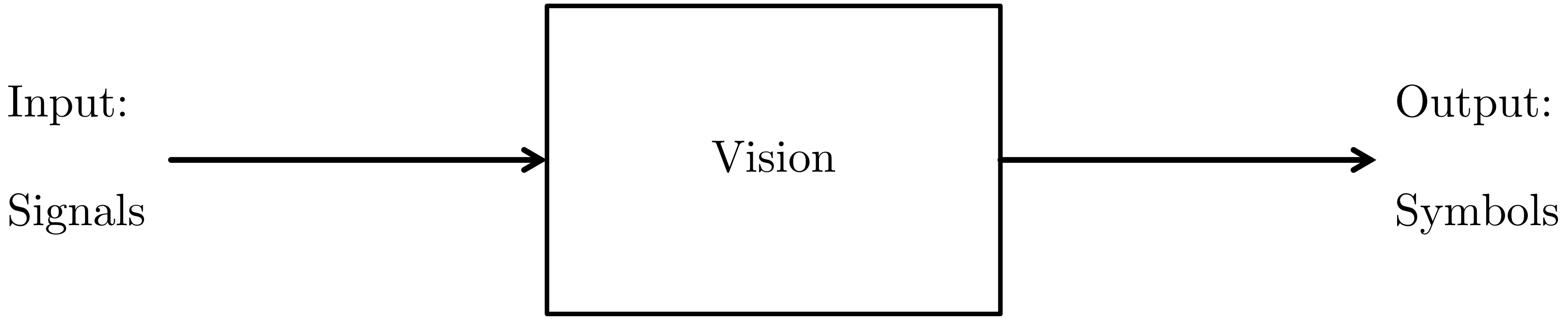
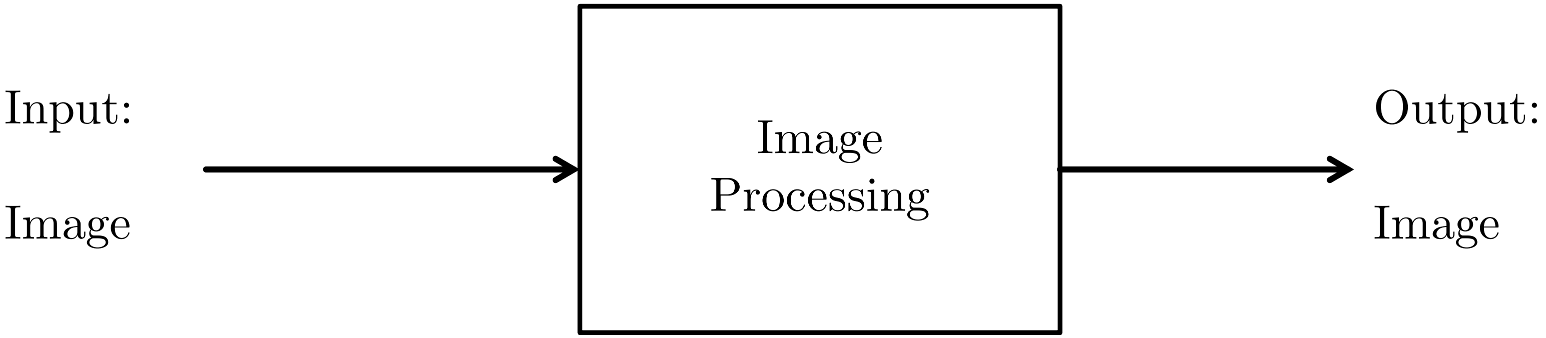


IMAGE PROCESSING

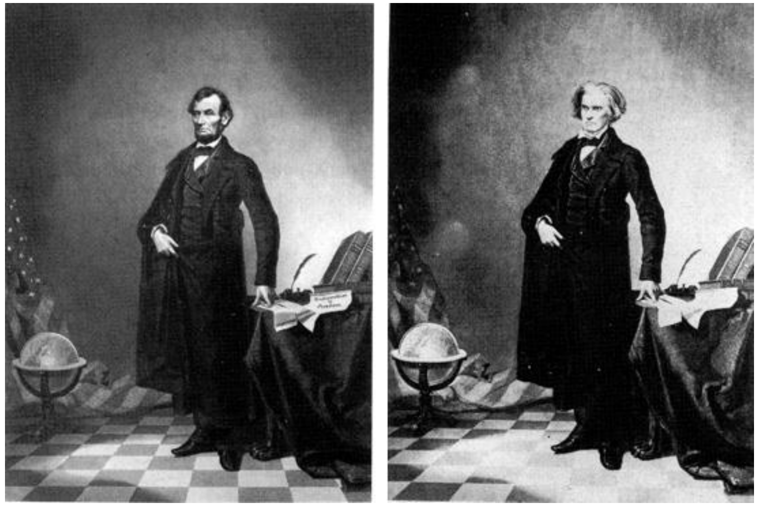
- Manipulation of images



Examples:

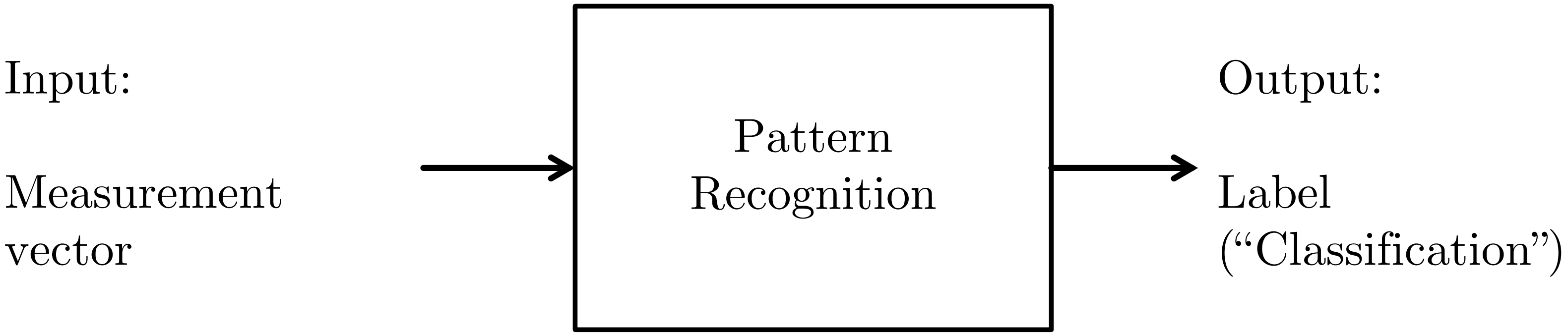
- “Photoshopping”
- Image enhancement
- Noise filtering
- Image compression

IP EXAMPLES



PATTERN RECOGNITION

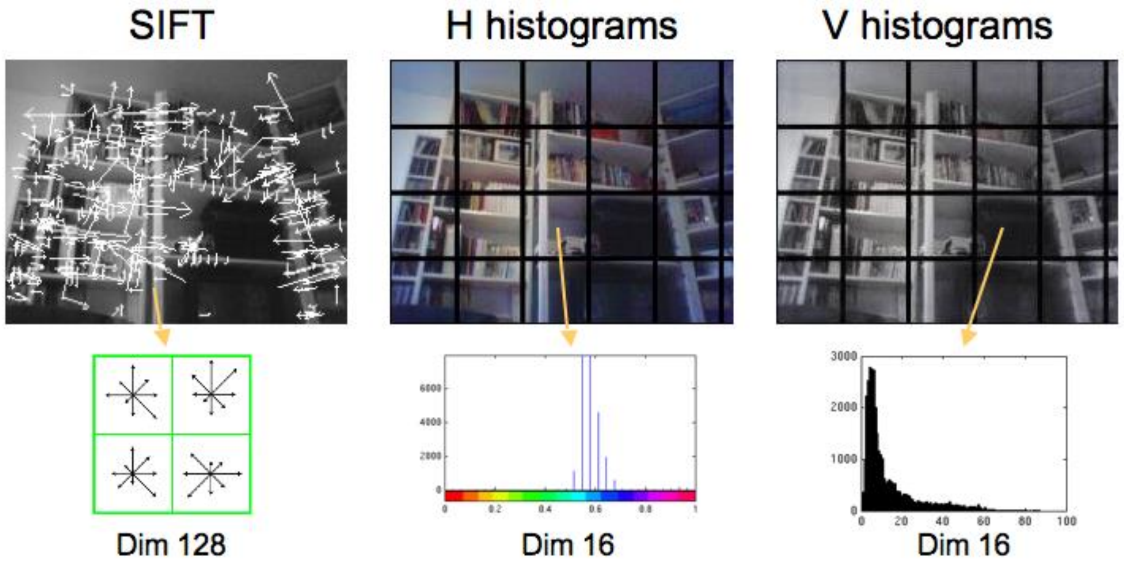
- Assignment of a label to input value

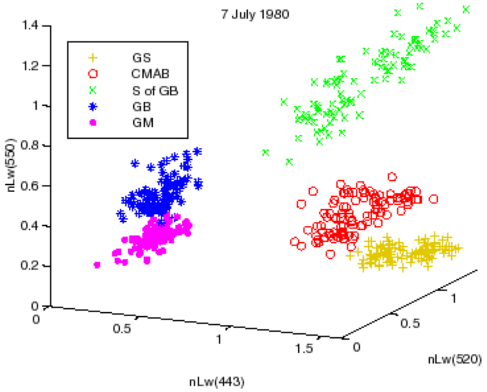


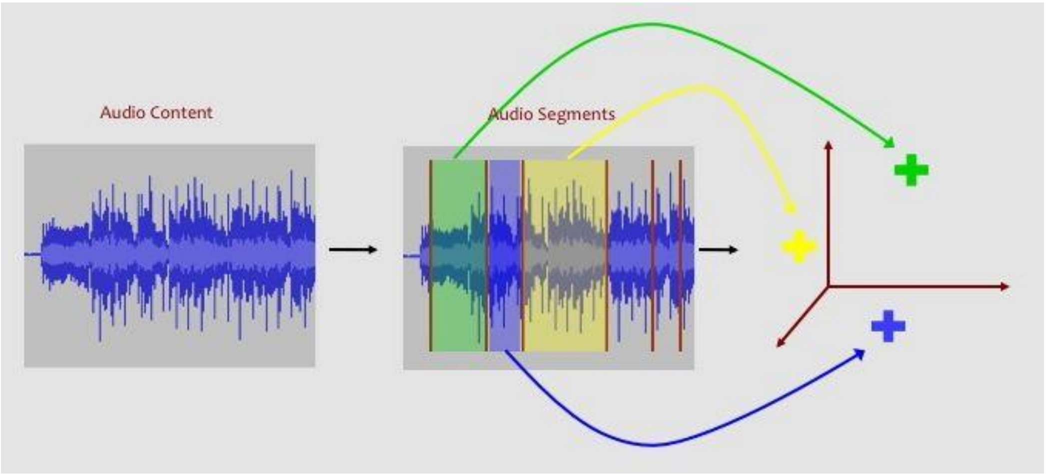
Examples:

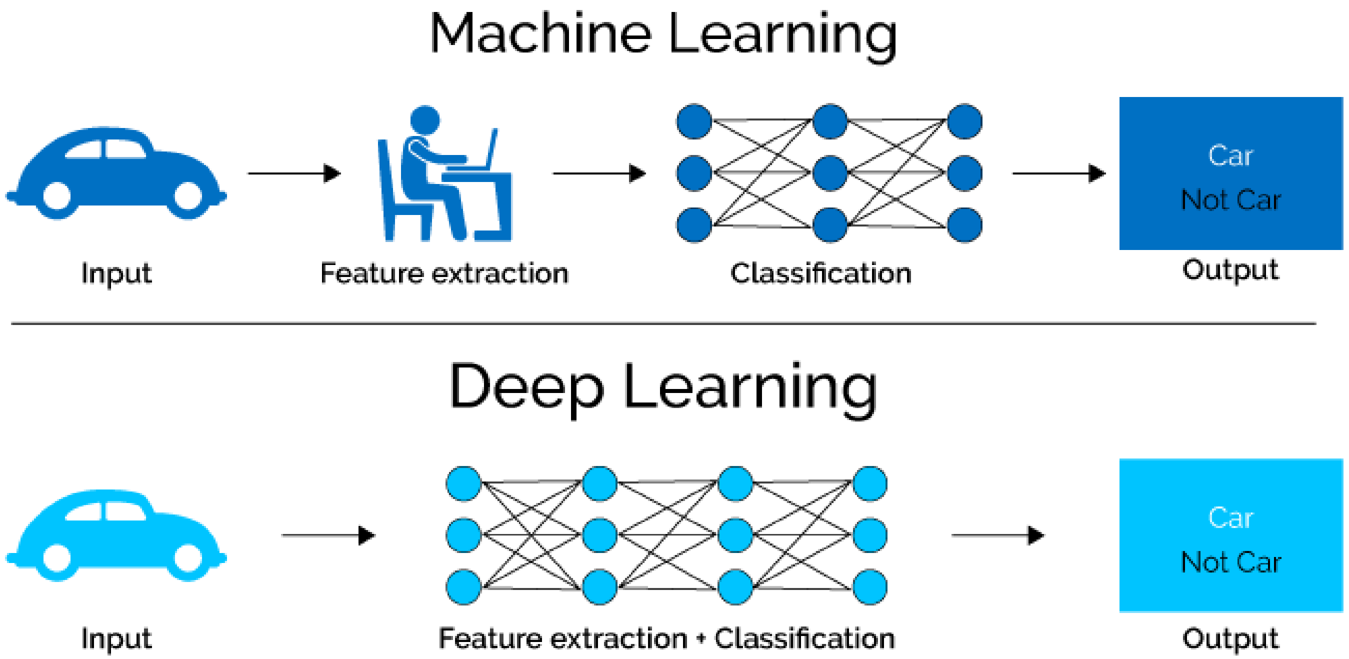
- Classification (1/0)
- Regression (real valued)
- Labeling (multi label)

PR EXAMPLES


SIFT
H histograms
V histograms
Dim 128
Dim 16
Dim 16


7 July 1980
GS
CMAB
S of GB
GB
GM
nLw(550)
nLw(443)
nLw(520)


Audio Content
Audio Segments


Machine Learning
Input
Feature extraction
Classification
Car
Not Car
Output
Deep Learning
Input
Feature extraction + Classification
Car
Not Car
Output

COMPUTER GRAPHICS

- Create realistic images (“forward problem”)

Input:

Mathematical model of objects and events → Computer Graphics → Output:

Images (“synthesized”)

Examples:

- Simulation (flight, driving)
- Virtual tours
- Video games
- Movies

CG EXAMPLES





COMPUTER VISION

- Interpretation and understanding of images

Input:

1. Image derived measurements
2. Models (prior knowledge)

Computer Vision

Output:

Recognition of objects and events embedded in images and video

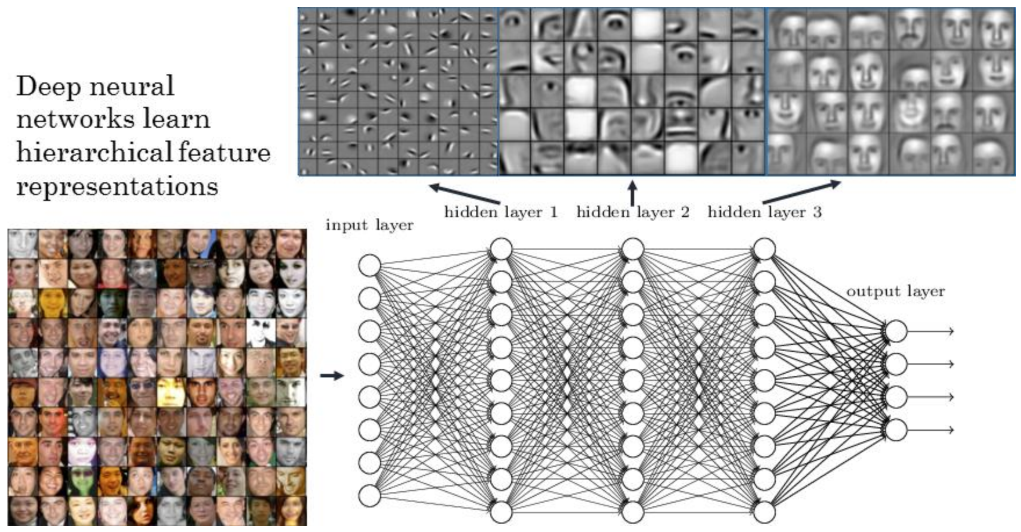
("Semantic" level classification)

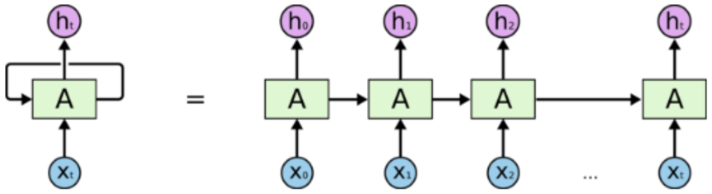
Examples:

- Object recognition
- Face recognition
- Lane detection
- Activity analysis

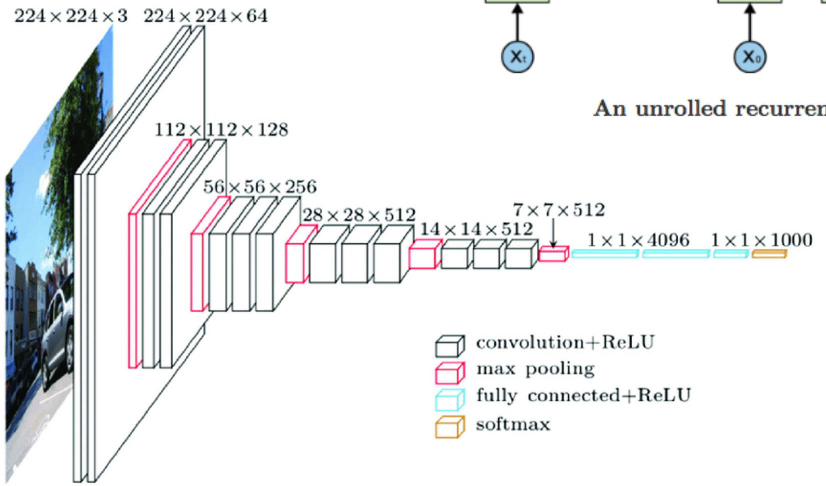
COMPUTER VISION EXAMPLES

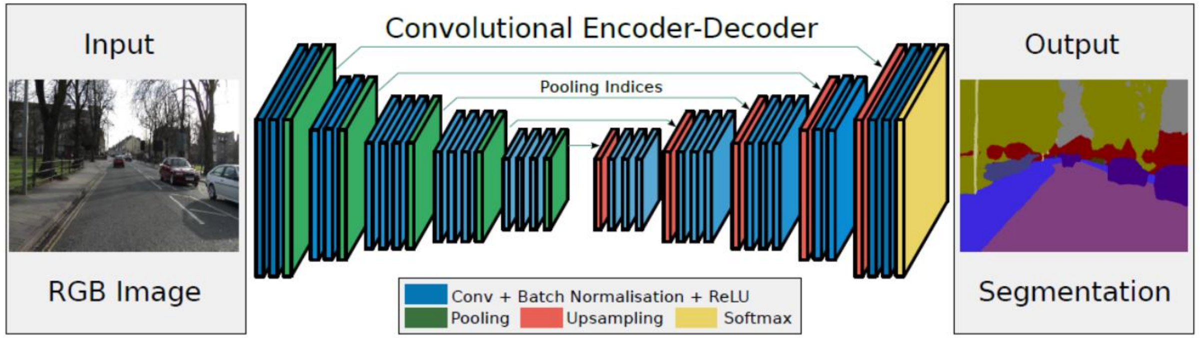
Deep neural networks learn hierarchical feature representations





An unrolled recurrent neural network.





SegNet: Encoder Decoder Architecture

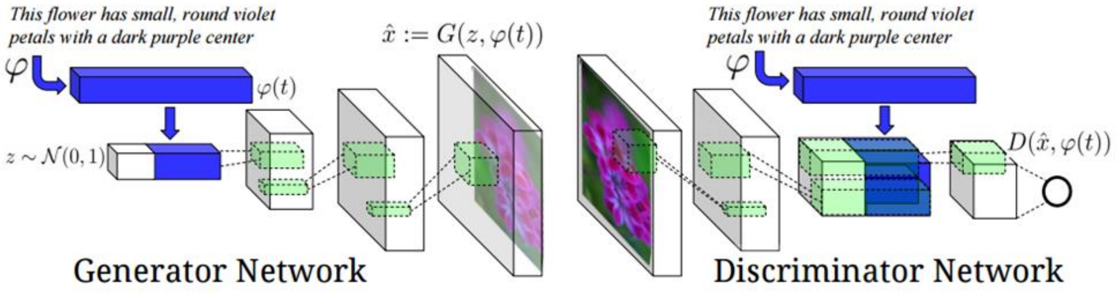


Figure 2. Our text-conditional convolutional GAN architecture. Text encoding $\varphi(t)$ is used by both generator and discriminator. It is projected to a lower-dimensions and depth concatenated with image feature maps for further stages of convolutional processing.

Network Architecture

SCOPE OF COMPUTER VISION

- Very broad: Cfp for Computer Vision and Pattern Recognition (CVPR) conference (pre deep learning – convoluational neural networks)

• Motion and Tracking • Stereo and Structure from Motion • Shape-from-X • Color and Texture • Segmentation and Grouping • Image-Based Modeling • Illumination and Reflectance Modeling • Shape Representation and Matching • Sensors • Early and Biologically-Inspired Vision • Computational Photography and Video	• Object Recognition • Object Detection and Categorization • Video Analysis and Event Recognition • Face and Gesture Analysis → FG • Statistical Methods and Learning → DL • Performance Evaluation • Medical Image Analysis • Image and Video Retrieval • Vision for Graphics • Vision for Robotics • Applications of Computer Vision

CVPR2021

- 2D object recognition
- 3D computer vision
- 3D object recognition
- Action and behavior recognition
- Adversarial learning, adversarial attack and defense methods
- Biometrics, face, gesture, body pose
- Computational photography
- Datasets and evaluation
- Efficient training and inference methods for networks
- Explainable AI, fairness, accountability, privacy, transparency and ethics in vision
- Image and video retrieval
- Image and video synthesis
- Image classification
- Low-level and physics-based vision
- Machine learning architectures and formulations
- Medical, biological and cell microscopy
- Motion and tracking
- Optimization and learning methods
- Pose estimation
- Representation learning, deep learning
- Scene analysis and understanding
- Transfer, low-shot, semi- and un- supervised learning
- Video analysis and understanding
- Vision + language, vision + other modalities
- Vision applications and systems, vision for robotics and autonomous vehicles