**Computer Vision Fundamentals Writeup**

Computer Vision is a field where Algorithms are designed to help computers see and interpret the physical world as humans do with their eyes (internally brain neurons). Here come the deep neural networks which is more similar to the human brain neurons.

After training the model, given an image, sequence of images or frames from a video a deep neural network would be able to interpret the image as humans does.

Computer Vision (CV) = Digital Signal Processing (DSP) + Neuroscience + Artificial Intelligence

**CV pipeline:**

Image Acquisition

(Capture images as Digital Signals)

Digital Signal Processing

Image-to-Image Processing

(Image compression

Image restoration

Image enhancement)

Pre-Processing

Feature Extraction (using CNN)

Detection & Segmentation

High-level processing

(mimicking the human brains)

Applications

(medical, entertainment, automation, video surveillance, robots, facial recognition)

**Types of Cameras:** It matters!

1. Pinhole cameras – old style, poor lens
2. Digital Cameras – better lens – light after passing through lens hits imaging sensor – Photoelectric effect principle

**Digital Camera Pipeline:**

**Optics**

Light travels through the optics

**Aperture**

The Optics step is controlled by Aperture

Size of Aperture decides how much light enters image sensor

**Shutter**

It decides the exposure time

Time at which Image Sensor is exposed to light

**Sensor**

Finally, light reaches the Image Sensor

**Color Theory**: 2 types

1. **RGB** – 3 intensity values – Requires pixel values
   1. **Additive mixing** of colors: dealing with light sources, replicating colors using active light, using RGB mixtures, we can create spectrum of colors.
   2. **Subtractive mixing** of colors: dealing with colored pigments, replicating colors using passive light, [cyan, yellow, magenta] creates all the spectrum of colors.
2. **Gray Scale –** no need of pixel values

**Low-Mid-High-level Vision:**

**Low-level:** Basically, it works more with pixels

* The filters you use in your mobile phone camera/apps is the best example for this.
* Identifies edges, corners, performing filtering, morphology.

**Mid-level:** Scene Geometry, Camera and Object motion

**High-level**: Objective is to infer semantics

* Video and image understanding
* Image/video frames classification which requires detection and recognition.
* Model based (we already know what we are trying to recognize) and Learning based (requires image as the input and deep learning model learns the representation of the object) recognition