Computer Vision

SIAM ML

Vishak

Plan to cover Computer Vision

1. Basics of Image and Augmentation
2. Convolution process and the theory related to it
3. Image Classification and related research papers
4. Object Detection detailed discussion
5. Image segmentation detailed discussion

Theory + practical

How does it work?

- Extract features from images
- Process and analyze visual data
- Make predictions or generate new images

Where is Computer Vision Used??

Tasks We Can Do Using Computer Vision:

Image Classification – Identifying objects in images (*e.g., cat vs. dog classifier*)

Object Detection – Detecting multiple objects and their locations (*e.g., self-driving cars, security cameras*)

Image Segmentation – Separating objects from backgrounds (*e.g., medical image analysis*)

Face Recognition – Identifying and verifying faces (*e.g., phone unlock systems, surveillance*)

Optical Character Recognition (OCR) – Converting images of text into digital text (*e.g., Google Lens, license plate recognition*)

Pose Estimation – Detecting human body posture (*e.g., motion tracking in AR/VR, sports analytics*)

Image Generation (GANs & Diffusion Models) – Generating new images (*e.g., Deepfake, AI art*)

Anomaly Detection – Detecting defects in manufacturing and medical images (*e.g., cancer detection in X-rays*)

Scene Understanding – Analyzing complex environments (*e.g., self-driving cars understanding road signs*)
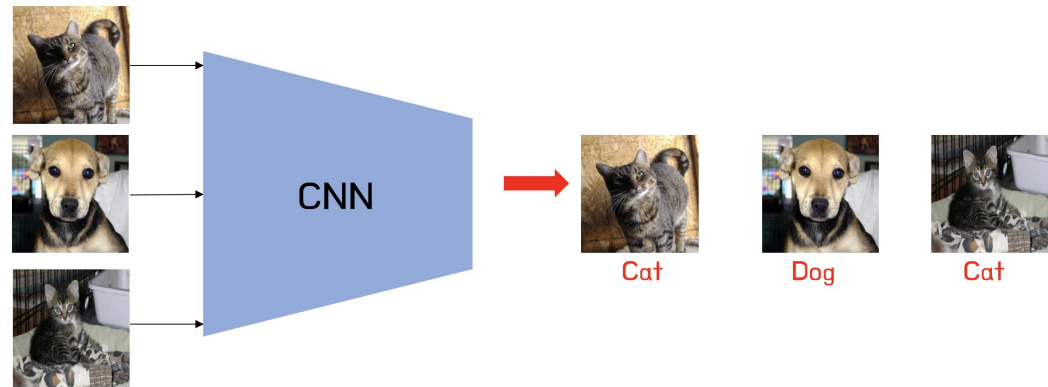
Why PyTorch for Computer Vision?

Why PyTorch?

- Easy to use and dynamic computation graphs
- Strong GPU acceleration and scalability
- Large community and support
- Pre-trained models in torchvision

Image Basics & Preprocessing

What is an Image?

Types of representing Images

Why do you think computers process images as numbers instead of visual representations?

Image Preprocessing for Deep Learning

Why Preprocess Images?

- Improves model performance
- Ensures consistent input size
- Normalizes pixel values

Common Preprocessing Steps:

Resizing – Ensures all images have the same dimensions
Normalization – Converts pixel values to a standard range
Augmentation – Helps improve model generalization

Overfitting?

What can we do ??

TTA ?

Lets just write the code quickly to load an image and make the transforms.

Soon we will even be writing codes for all the tasks seen before…

Torchvision ka transforms

Transformation	Description
`transforms.Resize(size)`	Resizes image to given size (single int for square, tuple for (H, W)).
`transforms.CenterCrop(size)`	Crops image at the center to given size.
`transforms.RandomCrop(size)`	Crops image randomly to given size.
`transforms.RandomResizedCrop(size)`	Randomly crops and resizes image (good for augmentation).
`transforms.Grayscale(num_output_channels=1)`	Converts image to grayscale.

Transformation	Description
`transforms.RandomHorizontalFlip(p=0.5)`	Flips image horizontally with probability p.
`transforms.RandomVerticalFlip(p=0.5)`	Flips image vertically with probability p.
`transforms.RandomRotation(degrees)`	Rotates image randomly within given degrees.
`transforms.ColorJitter(brightness, contrast, saturation, hue)`	Randomly changes brightness, contrast, etc.
`transforms.RandomAffine(degrees, translate, scale, shear)`	Applies affine transformations (rotation, scaling, translation, shearing).

Transformation	Description
`transforms.ToTensor()`	Converts image to a PyTorch tensor (C, H, W) and scales pixels to `[0,1]`.
`transforms.Normalize(mean, std)`	Normalizes pixel values using given mean & std.
`transforms.ToPILImage()`	Converts tensor back to PIL image.

What is CNN? And Why ??

What are the layers in a CNN ?

How does Convolution work?? (kernels, padding, stride etc)

What is dilated Conv?? (Hmmm many things will be there, hence read papers)

Hurraaayyyyy !!!! Done with the basics hehehehheehreheehe

Now comes the homework..

Homework !!

Read the different famous architectures for Image Classification and lets do Paper discussion Tomorrow !!!!!

- **LeNet-5**
- **AlexNet**
- **VGG**
- **GoogLeNet/Inception**
- **ResNet**
- **DenseNet**
- **EfficientNet**

Bonus:

      MobileNet

What we will do in next class?

Object Detection

Image Segmentation

Etc

Etc

Etc

Etc

Excited??

ANY QUESTIONS?