DeepVision Odyssey

Navigating the Depths of Visual Learning

Saarthak Krishan (22B3959)

Week 1 :
CNNs -

- In the first week of the bootcamp, we were introduced to neural networks, specially focussing on convolutional neural networks.
- The assignment was based on two special CNNs AlexNet and ResNet
- The assignment included implementation of both the models on CIFAR10 dataset along with hyperparameter tuning and regularisation techniques
- We were also introduced to the concept of transfer learning
- AlexNet -
    - Classic CNN with eight layers (Five convolutional and three fully connected) followed by max-pooling
    - Dropout is used as a regularisation technique to prevent overfitting
    - Two versions - one with pre trained weights and one with untrained weights were implemented
- ResNet -
    - Uses residual blocks or skip connections which enable shortcut of information from one layer to another by adding the input of a layer to its output
    - The connections help in mitigating the vanishing gradient problem, which make it easier to train very deep networks
    - Makes it easier for gradient to flow in back propagation

Week 2 -
Transformers for Vision -

- The ViT architecture was introduced as an alternative to CNNs
- The ViT architecture adapts the transformers architecture (initially proposed for sequential data) to process images
- Divides the input image into patches which are linearly embedded into a sequence of vectors

- Uses multi head self attention mechanism to capture dependencies between different patches
- Comprises multiple transformer encoder blocks stacked on top of each other with each block containing a multi head self attention layer and a feed forward neural network layer
- Studied about the attention maps, a visualisation technique used to interpret the output of attention mechanisms in neural networks
- The ViT architecture along with visualisation of attention maps were also implemented on CIFAR 10 dataset

Week 3 -
Generative Models -

- Introduced to various generative models like Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs) and Diffusion Models
- GANs are a class of artificial intelligence algorithms used in unsupervised machine learning, consisting of two neural networks - generator and discriminator
- The generator is responsible for creating new data samples with random noise as the input while the discriminator acts as a binary classifier distinguishing between the real data from the dataset and the fake data generated by the generator
- GANs achieve the state of the art performance because of the competition between generator and the discriminator
- The generator aims to generate realistic data to fool the discriminator, while the discriminator tries to become more accurate in distinguishing real from fake data
- Implemented Conditional GAN (cGAN) on Fashion MNIST dataset and generated conditional samples for different fashion categories

Week 4-5 -
Final Project -

- Went through the three research papers thoroughly
- Densely Connected Convolutional Networks -

    - DenseNet is a neural network architecture with the dense connectivity pattern where each layer receives feature maps from all preceding layers and passes its feature to all subsequent layers
    - Transition layers control the growth of the network
    - 1x1 convolutions reduce the no. of input feature maps
    - Achieves competitive performance with significantly fewer parameters
- Unsupervised Representation Learning with DCGANs -
    - Deep Convolutional GAN is a special variant of GAN which emphasises use of convolutional layers
    - The avoidance of fully connected layer is highlighted and instead focus is on convolutional and convolutional transpose layer
    - Recommends weight initialisation techniques such as zero centred Gaussian initialisation
    - They are shown to learn a hierarchical feature representation of the input data, enabling generation of high quality synthetic samples
- Swin Transformers -
    - Introduces a hierarchical design that divides the input image into non overlapping patches at lowest resolution level
    - Uses shifted windows for self attention computation instead of using traditional self attention mechanisms
    - Employs local self attention within each window enabling local dependencies
    - Employs patch embedding and positional encoding similar to other ViT architectures
- Implemented the research paper Densely Connected Convolutional Networks for classification of the CIFAR 10 dataset with an accuracy of about 80% .