UPES

Technical Seminar
"Deep Learning For the Computer Vision"

SUBMITTED TO:-
Mrs.Dhiviya Rose

SUBMITTED BY:

JYOTI SINGH

SAP ID- 500106507

ROLL NO.-R271222006 , 2023-2024

MASTER OF COMPUTER APPLICATION
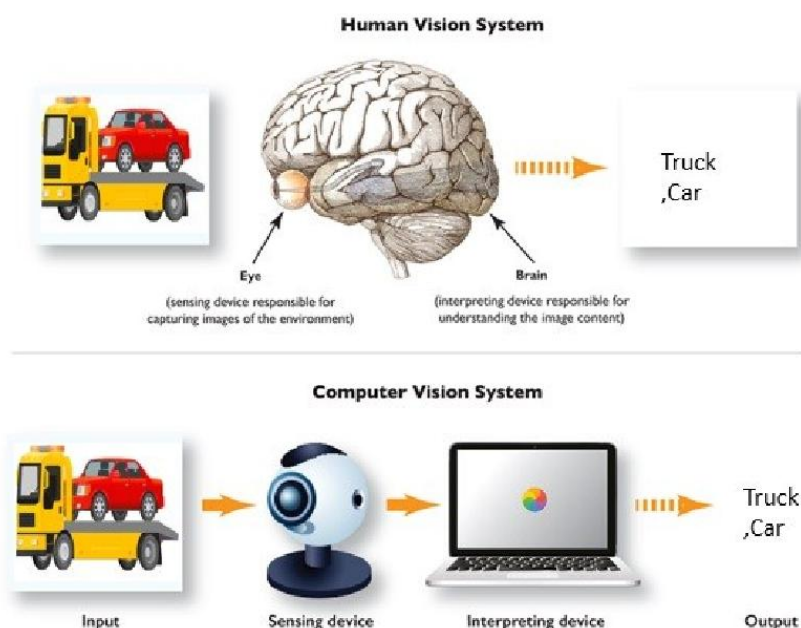
SCHOOL OF COMPUTER SCIENCE

UPES

CONTENT

ABSTRACT

Deep learning is a sub-field of machine learning that has shown remarkable success in solving complex problems in computer vision. It involves training artificial neural networks with multiple layers to learn representations of data, enabling them to automatically extract features and patterns from large data sets.

This seminar will provide an introduction to deep learning and its applications in computer vision. We will cover fundamental concepts like convolutional neural networks and recurrent neural networks, as well as advanced topics such as generative adversarial networks.

Through real-world examples and case studies, attendees will learn how deep learning can be used to tackle a variety of computer vision problems, including image classification, object detection, and semantic segmentation. We will also explore challenges related to training and deploying deep learning models, such as interpretability, bias, and ethical considerations.

By the end of the seminar, participants will have a solid understanding of deep learning and its applications in computer vision, and will be well-equipped to start applying these techniques to their own projects.

Human Vision System

Eye
(sensing device responsible for capturing images of the environment)

Brain
(interpreting device responsible for understanding the image content)

Truck ,Car

Computer Vision System

Input   Sensing device   Interpreting device   Output

Truck ,Car

<u>INTRODUCTION</u>

Deep Learning is a subset of Machine Learning. Deep Learning and Computer Vision have led to significant advancements in areas such as image recognition, object detection, and image generation. Applications of Deep Learning for Computer Vision include self-driving cars, facial recognition, and medical imaging.

- Convolutional Neural Networks (CNNs) are a popular type of deep learning architecture used in Computer Vision tasks such as image classification, object detection, and segmentation.

- Recurrent Neural Networks (RNNs) are another type of deep learning architecture used in tasks such as image captioning and video analysis.

- Generative Adversarial Networks (GANs) are a type of deep learning architecture used for image and video synthesis.

Future directions for Deep Learning for Computer Vision include improving the interpretability and explain-ability of models, developing more efficient architectures, and integrating deep learning with other AI technologies such as robotics and natural language processing.

deep Learning for Computer Vision has the potential to revolutionize various industries and improve the quality of life for people worldwide.

deep learning have also resulted in the development of various architectures and techniques, such as transfer learning, data augmentation, and attention mechanisms, which have further improved the performance of computer vision models. One of the key advantages of deep learning for computer vision is its ability to learn complex features directly from raw data, without the need for handcrafted feature engineering. This has led to significant improvements in accuracy and efficiency, making deep learning a crucial tool in many real-world applications. However, deep learning also poses challenges, such as the need for large amounts of data and computing resources, as well as the potential for over-fitting.

FUNDAMENTALS OF COMPUTER VISION

Computer Vision is a field of study that focuses on enabling computers to interpret, analyze, and understand visual data from the world.

Computer Vision has applications in various industries such as healthcare, automotive, security, and entertainment. One of the fundamental aspects of Computer Vision is image processing, which involves manipulating and analyzing images to extract useful information.

- Computer Vision also involves analyzing patterns in visual data, which is typically done using machine learning algorithms such as clustering and classification.
- Computer Vision algorithms can be supervised, unsupervised, or semi-supervised, depending on whether they require labeled data for training.
- Computer Vision applications often involve processing large amounts of data, which can be challenging due to the complexity and variability of visual data.

The accuracy and efficiency of Computer Vision algorithms can be improved by using specialized hardware such as graphics processing units (GPUs) or field-programmable gate arrays (FPGAs).
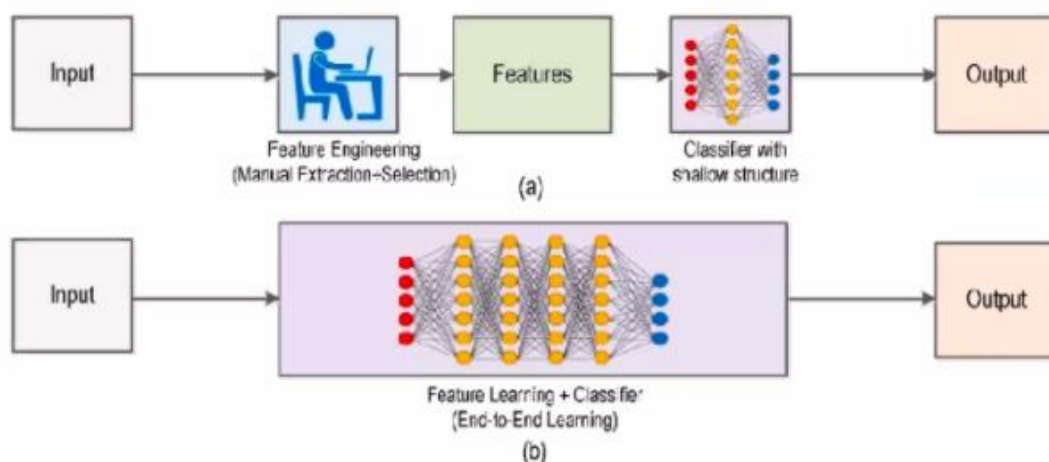
Computer vision; old (a) vs new (b) (Source: Angshuman Gosh | DLDC 2021)

CONVOLUTIONAL NEURAL NETWORKS (CNN)

CNNs are a type of Deep Learning architecture that are primarily used for image and video analysis tasks. They are designed to automatically learn hierarchical representations of visual data through the use of convolutional layers. CNNs use filters or kernels to extract local features from images, and these filters are learned during the training process.

The convolutional layers in CNNs are typically followed by pooling layers, which reduce the spatial dimensions of the feature maps. CNNs can have multiple layers, and the depth of the network contributes to its ability to learn increasingly abstract representations of visual data.

The final layer of a CNN is typically a fully connected layer, which is used for classification or regression tasks. CNNs can be trained using backpropagation, a gradient-based optimization algorithm that adjusts the weights and biases of the network to minimize the LOSS function.

Transfer Learning is commonly used with CNNs to leverage pre-trained models on large datasets, such as ImageNet.

Fine-tuning is another technique used with CNNs that involves training only the final layers of a pre-trained model on a new dataset.

Dropout regularization is often used in CNNs to prevent overfitting and improve generalization. Other techniques used with CNNs include data augmentation, which artificially increases the size of the training dataset, and batch normalization, which normalizes the output of each layer.

CNNs have been applied to a wide range of applications, including image classification, object detection, segmentation, and style transfer.

One of the challenges of CNNs is the need for large amounts of labeled data to train the network effectively. Another challenge is the need for specialized hardware, such as GPUs, to train and run large CNNs.

The future of CNNs includes developing more efficient and accurate architectures, improving the interpretability and explainability of models, and integrating with other AI technologies.

Convolutional neural networks (ConvNets) are a type of neural network that are commonly used in image recognition and processing tasks. They are composed of multiple layers of artificial neurons that are designed to recognize patterns in images. Each layer generates several activation functions that are passed on to the next layer. The first layer usually extracts basic features such as horizontal or diagonal edges. This output is passed on to the next layer which detects more complex features such as corners or combinational edges. As we move deeper into the network it can identify even more complex features such as objects, faces, etc.

Based on the activation map of the final convolution layer, the classification layer outputs a set of confidence scores (values between 0 and 1) that specify how likely the image is to belong to a "class." For instance, if you have a ConvNet that detects cats, dogs, and horses, the output of the final layer is the possibility that the input image contains any of those animals.
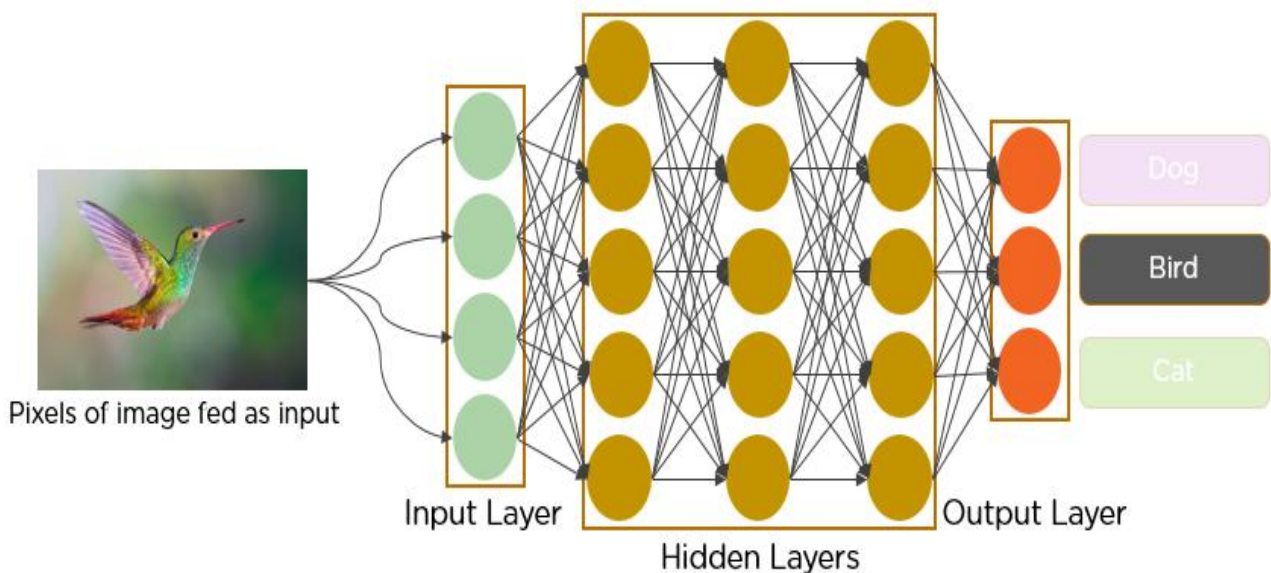
RECURRENT NEURAL NETWORKS (RNN)

RNNs are a type of neural network that can process sequences of varying lengths, making them well-suited for tasks such as speech recognition, natural language processing, and time series analysis.

They have feedback loops that allow information to be passed from one step of the sequence to the next, allowing them to maintain a "memory" of previous inputs. RNNs can be trained using gradient-based optimization methods such as backpropagation through time (BPTT), which is similar to standard backpropagation but takes into account the recurrence of the network.

Another training method for RNNs is the use of teacher forcing, where the network is fed the correct output at each step of the sequence during training, rather than the predicted output from the previous step. Various techniques have been developed to mitigate these problems, such as gradient clipping, which limits the magnitude of the gradients, and weight initialization methods, which can help to prevent the gradients from exploding or vanishing.

RNNs have been applied to a wide range of applications, including language modeling, machine translation, sentiment analysis, and speech recognition. RNNs have been used to generate text, music, and other types of creative content through the use of generative models.

Another area of research is the use of RNNs for unsupervised learning, where the network learns to represent the structure of the input data without explicit labels. The future of RNNs includes developing more efficient and scalable architectures, improving their interpretability and explainability, and integrating them with other AI technologies.

Another consideration is the potential for RNNs to be used for malicious purposes, such as generating fake news or deepfakes. Ensuring the responsible development and deployment of RNNs will be crucial to realizing their potential benefits while mitigating potential risks.

- **Advantages of Recurrent Neural Network**

    1. An RNN remembers each and every piece of information through time. It is useful in time series prediction only because of the feature to remember previous inputs as well. This is called Long Short Term Memory.
    2. Recurrent neural networks are even used with convolutional layers to extend the effective pixel neighborhood.

- **Disadvantages of Recurrent Neural Network**
    1. Gradient vanishing and exploding problems.
    2. Training an RNN is a very difficult task.
    3. It cannot process very long sequences if using tanh or relu as an activation function.

- **Applications of Recurrent Neural Network**
    1. Language Modelling and Generating Text
    2. Speech Recognition
    3. Machine Translation
    4. Image Recognition, Face detection
    5. Time series Forecasting
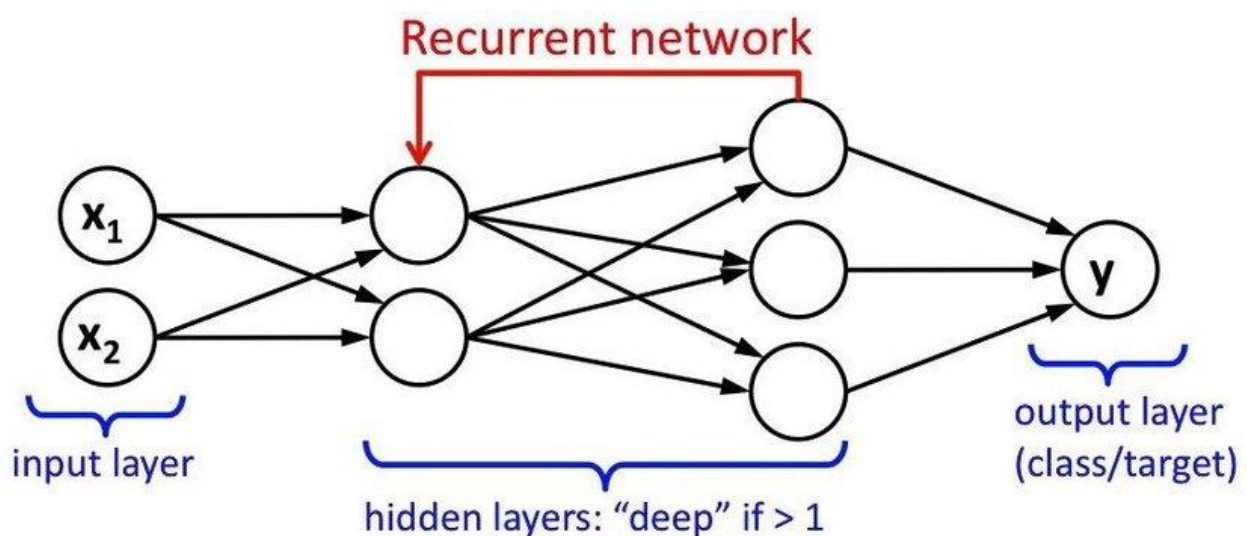
Generative Adversarial Networks (GAN)

GANs are a type of neural network architecture that consists of two components: a generator network and a discriminator network.

The generator network takes in a random noise vector as input and outputs a sample that is meant to resemble data from a particular distribution. The discriminator network takes in either a real sample from the distribution or a generated sample from the generator network, and outputs a probability estimate of whether the input is real or fake.

The goal of GANs is for the generator network to learn to produce samples that are indistinguishable from real samples, while the discriminator network learns to distinguish between real and fake samples.

GANs use a minimax game framework, where the generator network tries to maximize the probability of the discriminator network making a mistake, while the discriminator network tries to minimize its error rate.

Training GANs involves alternating between updating the weights of the generator network and the discriminator network using gradient-based optimization methods. One of the challenges of GANs is mode collapse, where the generator network produces a limited set of outputs that do not fully capture the diversity of the true distribution.

Another challenge is instability during training, where the generator and discriminator networks can get stuck in a local minima and fail to converge to an optimal solution. Various techniques have been developed to mitigate these problems, such as adding noise to the input of the discriminator network, using different loss functions, and modifying the network architecture.

GANs have been applied to a wide range of applications, including image and video generation, text generation, and style transfer.

They have also been combined with other types of neural networks, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to improve their performance on certain tasks.

GANs have been used to create realistic images of people, animals, and landscapes, as

well as to generate novel artwork and music. Another area of research is the use of GANs for data augmentation and synthesis, where they can be used to create new training data for machine learning models. GANs have the potential to revolutionize many industries, including entertainment, advertising, and manufacturing, by enabling more efficient and creative content generation.

One of the ethical considerations of GANs is the potential for misuse, such as the creation of deepfakes or other types of fake content. The future of GANs includes developing more efficient and scalable architectures, improving their stability and performance, and exploring new applications and use cases. GANs have the potential to transform the way we create and interact with digital content, and will likely continue to be a major area of research and innovation in the field of deep learning.
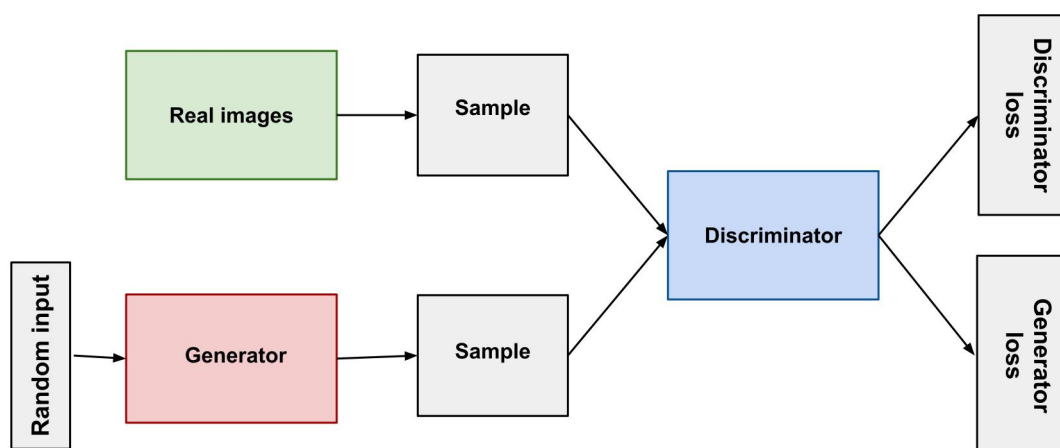
CONCLUSION

In conclusion, deep learning has emerged as a powerful and versatile set of techniques for solving a wide range of computer vision problems. Convolutional neural networks have revolutionized image classification and object detection, while recurrent neural networks have enabled the modeling of sequential data such as text and speech. Generative adversarial networks have opened up new possibilities for content generation and synthesis.

Despite the remarkable progress that has been made in deep learning, there are still many challenges that must be addressed. These include improving the robustness and interpretability of deep learning models, addressing issues related to bias and fairness, and developing more efficient and scalable algorithms for training and inference.

As the field of deep learning continues to evolve, there will be many exciting opportunities for research and innovation. From developing new neural network architectures to exploring new applications and use cases, the potential for deep learning to transform the way we solve problems in computer vision is truly limitless.

REFERENCE

- Deep Learning By Ian Goodfellow, Yoshua Bengio, Aaron Courville.
- Geeks for Geeks .
- Javatpoint.
- Medium.