Seminar II: A Survey on Convolutional Neural Networks: Innovations and Applications in Computer Vision

Bizuhan Abate

June 21, 2024

Introduction

- CNN is one of the major branches of neural networks and have revolutionized modern computer vision.
  - Initially designed to mimic the human visual cortex.
  - Evolved to handle tasks like object detection, image segmentation, and image classification etc.
- Comparable representations, parameter sharing and sparse interactions are significant advantages of CNNs.
- The CNN architecture is made up of many levels. CNNs consist of various layers, each with a different purpose.

Continued...

- The convolutional layer is one of the crucial components in the design of a CNN. It is constructed using several convolutional filters also known as kernels. The output feature map is created by convolving the input picture, spatial representation of learned features or response of a particular filter to the input image.
- The pooling layer is another important component with a primary function to subsample the feature maps for which convolutional procedures are used (Santos et al., 2020). In other words, this approach condenses large-scale feature maps into smaller feature maps. Further, it keeps the bulk of the dominating information in every stage of the pooling process.

Continued…

- Next is the Fully Connected Layer which is usually found near the conclusion of each CNN architecture. Each neuron in this layer is linked to all previous layer neurons, resulting in the so-called Fully Connected (FC) method. It serves as a CNN classifier.
- The data is in the format of a vector that is generated after flattening the feature map. Furthermore, the final classification is obtained from the output layer, the last layer of the CNN design.

Background

- Evolution of CNNs:
  - Inspired by the biological visual cortex.
  - Early work: LeNet by Yann LeCun.
  - Breakthrough: AlexNet in 2012, leveraging deeper architectures and GPU power.
- Milestones:
  - LeNet for handwritten digit recognition.
  - AlexNet for large-scale image classification.

AlexNet

- Put a landmark in the history of deep learning and computer vision.
- Consisted of eight layers, including five convolutional layers followed by three fully connected layers.
- Key innovations included the use of Rectified Linear Units (ReLU) as activation functions, local response normalization (LRN) for regularization, dropout for preventing overfitting, and efficient GPU implementation, which significantly accelerated training.
- By reducing the top-5 error rate on ImageNet from over 25 to below 17, AlexNet demonstrated the power of deep CNNs in handling large-scale image classification.

(Krizhevsky et al. (2012))

VGGNet

- VGGNet focused on exploring the impact of deeper architectures on image recognition tasks. It introduced several variants, including VGG16 and VGG19, which were characterized by their uniform architecture consisting primarily of 3x3 convolutional layers and max-pooling layers.
- Despite its simplicity in design, VGGNet demonstrated that increasing network depth could lead to improved performance on the ImageNet dataset. By systematically increasing the number of layers while maintaining a small receptive field size, VGGNet achieved competitive results in image classification tasks. However, its computational cost in terms of memory and parameter size paved the way for subsequent research to explore more efficient architectures.

(Simonyan and Zisserman (2014)).

ResNet

- ResNet introduced the concept of residual learning, which addressed the challenge of training very deep neural networks. The key innovation of ResNet was the use of residual blocks, where each block contained skip connections (or shortcuts) that bypassed one or more layers.
- These skip connections allowed gradients to flow more directly through the network during training, alleviating the vanishing gradient problem and enabling the training of exceptionally deep networks with hundreds of layers.
- As a result, ResNet achieved state-of-the-art performance on various benchmarks, including ImageNet, and significantly surpassed previous accuracy levels. ResNet's success demonstrated that increasing depth could lead to better representation learning and paved the way for the development of even deeper architectures.

(He et al. (2015))

GoogLeNet

- Inception, also known as GoogLeNet, introduced the inception module, which revolutionized the design of CNN architectures by incorporating multiple parallel convolutional operations within the same layer.
- This design allowed the network to capture features at multiple scales efficiently while reducing computational costs compared to traditional architectures.
- Inception achieved significant improvements in both accuracy and efficiency and won the ILSVRC 2014 competition.

(Szegedy et al. (2015)

Advanced CNN Architectures

MobileNets: Efficient for mobile and embedded vision applications.

- Based on depthwise separable convolutions, which split the standard convolution into two simpler operations: depthwise convolution and pointwise convolution (1x1 convolution).
- This significantly reduces the number of parameters and computations compared to traditional convolutions.
- Prioritize reducing the model size and computational complexity.
- Ideal for deployment on devices with limited computational power and memory, such as smartphones.

(Howard et al. (2017))

Applications of CNNs

- Medical Imaging: Early disease detection and diagnosis.
  - Autonomous Vehicles: Object detection and navigation.
  - Agriculture: Crop monitoring and yield prediction.
  - Surveillance: Real-time monitoring and threat detection.
  - Artistic Applications: Style transfer and creative designs.

Challenges in CNNs

- Computational Complexity: High resource requirements.
  - Overfitting: Needs regularization and large datasets.
  - Interpretability: Understanding decision processes in critical applications
  - Bias mitigation and domain adaptation.

Conclusion

- Summary:
  - CNNs have transformed computer vision tasks.
  - Innovations like ResNet and Inception have pushed boundaries.
  - MobileNets and EfficientNets addressed efficiency.

Thank you !!