Deep_Learning

Deep Learning Architectures and Their Impact on Computer Vision

Abstract:
Deep learning has revolutionized computer vision by enabling machines to learn hierarchical representations of images. This paper explores convolutional neural networks (CNNs), residual networks, EfficientNet, Vision Transformers, GANs, and their applications in image classification, segmentation, object detection, and generation.

Keywords: deep learning, computer vision, CNN, ResNet, ViT, GAN, image recognition

1. Introduction

Deep learning architectures allow automatic feature extraction, surpassing traditional computer vision methods. The hierarchical structure captures spatial patterns from low-level edges to high-level semantic features.

2. Convolutional Neural Networks

CNNs, using convolution, pooling, and fully connected layers, are fundamental in image analysis. They reduce parameters while maintaining spatial hierarchies.

3. Advanced Architectures

Residual networks address vanishing gradients, enabling very deep architectures. EfficientNet optimizes accuracy and efficiency through compound scaling. Vision Transformers leverage self-attention for global context understanding.

4. Generative Models

GANs generate realistic images by training a generator and discriminator in adversarial settings. Variants enable image-to-image translation, super-resolution, and style transfer.

5. Applications

Applications span autonomous vehicles, medical imaging, security surveillance, and augmented reality. Integration with NLP enables multi-modal understanding.

6. Challenges

Challenges include high computational cost, interpretability, adversarial robustness, and dataset bias. Research focuses on lightweight models, explainability, and transfer learning.

References:

[1] LeCun et al., 2015, "Deep Learning"

[2] He et al., 2016, "Deep Residual Learning for Image Recognition"

[3] Dosovitskiy et al., 2021, "An Image is Worth 16x16 Words: Transformers for Image Recognition"