A Comparative Survey of Recent Deep Learning Architectures for Object Recognition

The goal of Object Recognition (OR) models is to identify what objects appear in an image and where they are located. OR combines the principles of image classification and object detection. While all OR systems aim to determine the present objects, their implementation varies widely. Traditional approaches relied on features such as edges, corners, and textures, failing to generalize across varying lighting conditions, viewing angles, and object occlusions. As a result, their performance was often limited by the designer's ability to manually engineer suitable feature representations.

The emergence of deep learning has allowed models to automatically learn hierarchical feature representations directly from data. Convolutional Neural Networks (CNNs) extract spatial features at multiple levels of abstraction, leading to significant improvements in accuracy and robustness. This shift from manually designed features to data-driven learning has changed how object recognition systems are developed and deployed.

Despite these advances, selecting an appropriate deep learning architecture remains a challenge. Numerous models have been proposed offering trade-offs between accuracy, computational efficiency, and suitability for deployment. Understanding these trade-offs is essential for designing systems that balance performance with resource constraints.

The objective of this work is to present a comparative survey of recent deep learning architectures used for object recognition, such as: ResNet-50, EfficientNet-B0, MobileNetV3 and the Vision Transformer. The study aims to analyze their design principles, performance characteristics, and application contexts, highlighting the strengths and limitations.

References:

1. Ajit, A., Acharya, K., Samanta, A., 2020. A review of convolutional neural networks, in: 2020 international conference on emerging trends in information technology and engineering (ic-ETITE), IEEE. pp. 1–5.
2. Ankile, L.L., Heggland, M.F., Krange, K., 2020. Deep convolutional neural networks: A survey of the foundations, selected improvements, and some current applications. arXiv preprint arXiv:2011.12960 .
3. He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.
4. Howard, A., Sandler, M., Chu, G., Chen, L.C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., et al., 2019. Searching for mobilenetv3, in: Proceedings of the IEEE/CVF international conference on computer vision, pp. 1314–1324.
5. Khan, A.A., Laghari, A.A., Awan, S.A., 2021. Machine learning in computer vision: A review. EAI Endorsed Transactions on Scalable Information Systems 8.
6. Liang, J., 2020. Image classification based on resnet, in: Journal of Physics: Conference Series, IOP Publishing. p. 012110.
7. Pinaya, W.H.L., Vieira, S., Garcia-Dias, R., Mechelli, A., 2020. Convolutional neural networks, in: Machine learning. Elsevier, pp. 173–191.
8. Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks, in: International conference on machine learning, PMLR. pp. 6105–6114.