**ASSIGNMENT-2**

**DEEP LEARNING**

Comparative Analysis of Convolutional Neural Network Architectures: LeNet, AlexNet, GoogLeNet, ResNet, and SENet

**Abstract**

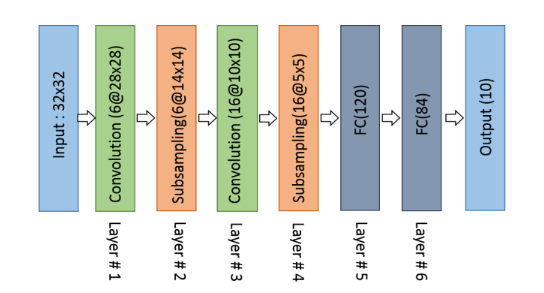
This paper presents a detailed comparative analysis of five prominent Convolutional Neural Network (CNN) architectures: LeNet, AlexNet, GoogLeNet, ResNet, and SENet. Each architecture is inspected in terms of its unique plan, methodology, error rates, and standout features. By investigating these aspects, the study aims to provide a comprehensive understanding of the advancement and headways in CNN architectures for image recognition tasks. The discoveries offer profitable insights into the qualities and characteristics of each network, contributing to the broader knowledge of deep learning models in computer vision applications.

**1. Introduction**

The evolution of Convolutional Neural Networks (CNNs) has revolutionized computer vision, with architectures like LeNet, AlexNet, GoogLeNet, ResNet, and SENet driving progress. These models introduce unique design principles, methodologies, and performance characteristics, shaping the neural network landscape. Researchers analyze these architectures to understand their architectural changes, methodological advancements, error rates, and standout features, unveiling their contributions to image recognition. CNNs have substantially reduced error rates in competitions like the ILSVRC ImageNet challenge, dropping the top-5 error rate from over 26% to less than 2.3% in six years. Analyzing winning entries offers insights into CNN operation and evolution. We'll examine LeNet-5 (1998), followed by ILSVRC winners: AlexNet (2012), GoogLeNet (2014), ResNet (2015), and SENet (2017).

**2. The Architecture**

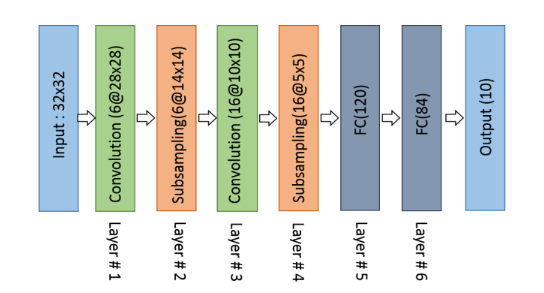
**2.1 LeNet**

****

LeNet is a classic convolutional neural network structure proposed by LeCun et al. in 1998, containing fundamental modules of deep learning like convolutional layers, pooling layers, and fully connected layers and broadly used for hand‐ written digit recognition (MNIST).Although LeNet was proposed in the 1990s, limited computation capability and memory capacity made the algorithm difficult to actualize until about 2010.

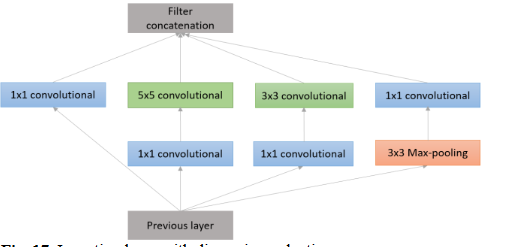
LeCun, however, proposed CNNs with the back-propagation algorithm and experimented on handwritten digits dataset to accomplish state-of-the-art accuracies.His architecture is well known as LeNet-5. The essential configuration of LeNet-5 is convolution (conv) layers, 2 sub-sampling layers,2 fully connected layers, and an output layer with Gaussian connection.

**2.2 AlexNet**

****

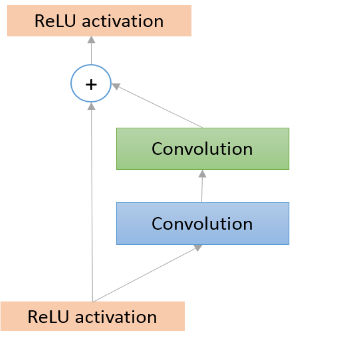
In 2012, Alex Krizhevsky and collaborators introduced AlexNet, a deeper CNN model that outperformed LeNet and won the challenging ImageNet Large Scale Visual Recognition Challenge (ILSVRC). It consists of multiple convolutional layers, with the first layer employing 96 11×11 receptive filters with Local Response Normalization (LRN), followed by max pooling using 3×3 filters with a stride size of 2. Similar operations are applied in subsequent layers, utilizing 5×5 filters in the second layer and 3×3 filters in the third, fourth, and fifth layers, generating 384, 384, and 256 feature maps, respectively. The model includes two fully connected (FC) layers with dropout and a Softmax layer. Two networks with identical structures and feature map counts are trained concurrently.

**2.3 GoogLeNet**

****

Google's GoogLeNet, winner of the ILSVRC 2014, aimed to reduce computing complexity compared to traditional CNNs. Created by Christian Szegedy, it boasted 22 layers, making it significantly more intricate than predecessors, yet using fewer parameters than AlexNet or VGG. It introduced Inception modules, combining convolutional layers of 1x1, 3x3, and 5x5 filters to create a single output vector, with parallel pooling pathways for speed enhancement. Deeper layers emphasize higher abstraction, balancing computing efficiency with network depth for optimal image classification and detection.

**2.4 ResNet**

****

In 2015, Microsoft Research introduced ResNet to address the vanishing/exploding gradient problem in deep neural networks. Residual blocks are pivotal in ResNet architecture, resolving gradient issues with skip connections. These connections aid in learning residual functions with respect to layer inputs, facilitating optimization and training of very deep networks. As network depth increases, skip connections enable learning of residual mappings rather than direct functions, preserving accuracy. The architecture of ResNet closely resembles GoogLeNet, except for the dropout layer. ResNet consists mainly of a stack of basic residual units, each comprising two convolutional layers using 3 × 3 kernels with Batch Normalization (BN) and ReLU activation, maintaining spatial dimensions without pooling layers (stride 1, SAME padding).

**Different Types of ResNet Architectures**

ResNet, has various architectures with different numbers of layers, each designed to address specific challenges and optimize performance. Here are some key types of ResNet architectures:

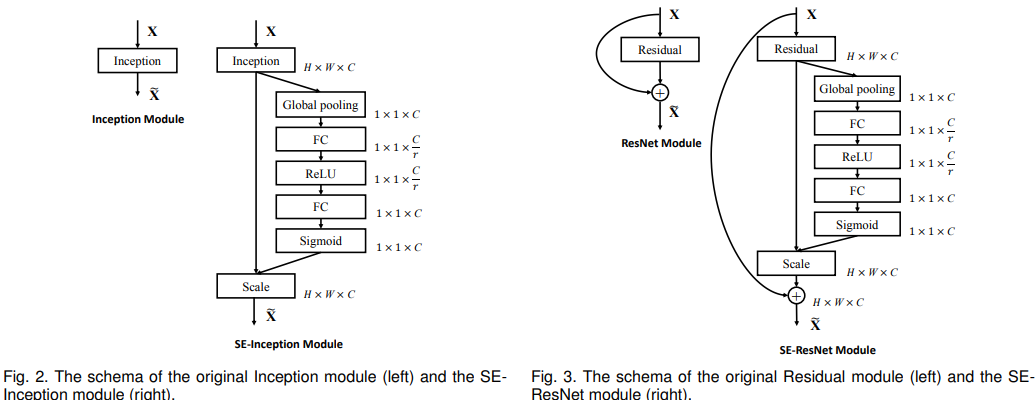
1. ResNet-34: The first ResNet model to be introduced was the ResNet-34 architecture. It involves inserting shortcut connections to change a plain network into a residual network, and it has 34 weighted layers. This architecture delivers 3.6 billion FLOPs [via the use of 3×3 filters in convolutional networks].

2. ResNet-50: Expanding upon the ResNet-34 paradigm, ResNet-50 adds a bottleneck design by substituting 3-layer bottleneck blocks for the 2-layer blocks seen in ResNet-34. With this update, accuracy is improved and the 50-layer ResNet performs at 3.8 billion FLOPs.

3. ResNet-101 and ResNet-152: Using more 3-layer blocks, these architectures build larger residual networks. In comparison to previous networks such as VGG-16 or VGG-19 nets, the 152-layer ResNet maintains a reduced complexity despite its greater depth.

4. ResNet-V2: ResNet-V2 emphasises the usage of weight layers before to activation rather than after activation. It emphasises identity mapping and ensures smoother signal propagation over the network by eliminating the final non-linearity left over from the addition operation.

**2.5 SENet**

****

A convolutional neural network architecture called a Squeeze-and-Excitation Network (SENet) uses SE blocks to dynamically recalibrate features by changing channel-wise weights. By extending current models, such as inception networks or ResNets, this architecture improves their functionality. Three layers make up a SE Block: a dense output layer with sigmoid activation, a hidden dense layer with ReLU activation, and a global average pooling layer. After calculating the mean activations for every feature map, the global average pooling layer compresses the data to provide a low-dimensional vector representation. For every feature map, the output layer creates a recalibration vector, minimizing the size of unnecessary features while maintaining the same size of useful ones.

**3.Error Rate**

Two error rates are typically reported on ImageNet: top-1 and top-5. The top-5 error rate is the percentage of test images for which the correct label is not one of the five labels the model deems most likely.

**3.1 LeNet**

After about ten iterations through the training set, LeNet-5's error rate stabilises at 0.95%. After 19 passes, the training error approaches 0.35%. Since the learning curves did not display this behaviour, the study was unable to examine the usual phenomena of overtraining, which is characterised by an increase in test error after a minimum.Furthermore, it has been demonstrated that expanding the size of the training set improves accuracy; even with specialised architectures like LeNet-5, more training data improves performance.

**3.2 AlexNet**

It exhibited significantly superior top-1 and top-5 error rates—37.5% and 17.0%, respectively—than the prior state-of-the-art on the test data.Additionally, they entered a modified version of this model in the ILSVRC-2012 competition, which it won with a top-5 test error rate of 15.3%, better than the second-best entry's 26.2%.

* AlexNet introduced ReLU activation to CNNs, addressing gradient vanishing in deep networks.
* Dropout in AlexNet prevents overfitting by randomly ignoring neurons, primarily in the last few fully-connected layers.
* AlexNet uses overlapping max pooling to improve feature richness compared to average pooling. In comparison to the non-overlapping scheme s = 2, z = 2, which yields output of equal dimensions, this strategy reduces the top-1 and top-5 error rates by 0.4% and 0.3%, respectively.
* LRN in AlexNet simulates biological lateral inhibition, enhancing generalization by suppressing neurons with small values.Response normalization reduces our top-1 and top-5 error rates by 1.4% and 1.2%, respectively
* AlexNet utilizes two GPUs for group convolutions, combining feature maps from distinct GPUs for final output.This scheme reduces our top-1 and top-5 error rates by 1.7% and 1.2%, respectively, in contrast to a net trained on a single GPU that has half as many kernels in each convolutional layer.
* AlexNet employs data augmentation by extracting random patches and using PCA to alter RGB values, mitigating overfitting and enhancing generalization.This scheme reduces the top-1 error rate by over 1%.

**3.3 GoogLeNet**

GoogLeNet showcased remarkable performance in image classification, achieving a top-5 error rate of 6.67% on the ImageNet dataset. Noteworthy for its innovative design, the architecture employed methods like global average pooling and 1x1 convolutions. To address the vanishing gradients problem and regularize the network, the loss was augmented by 70% throughout training, although their impact was found to be minimal. Subsequent variations of GoogLeNet, such as Inception-v3 and Inception-v4, further improved performance through slightly different Inception modules.

**3.4 ResNet**

As the architecture with the lowest top-5 error rate (3.57%) on the ImageNet test set, ResNet is considered a leader in image classification tasks.Kaiming He et al.'s Residual Network (or ResNet), which produced an amazing top-5 error rate of less than 3.6%, was used to win the ILSVRC 2015 challenge.With ResNet-152 achieving the top results at an astounding 3% error rate, ResNet's mistake rate on the ImageNet dataset is remarkably low—beyond even that of human judges. Shortcut connections were added to ResNets, which is why they performed so well.

**3.5 SENet**

SENet-154 performed well in picture classification tasks on the ImageNet dataset, with a top-5 error rate of 4.47% using a 224 x 224 centre crop evaluation. Furthermore, a top-5 error rate of 6.00% and a top-1 error rate of 22.20% for SE-ResNet-50 were found, indicating the efficacy of SENet in enhancing accuracy and performance in contrast to other architectures such as ResNet.

**4. Unveiling the Distinctive Essence**

**4.1 LeNet**

In 1998, LeCun proposed LeNet-5, an effective convolutional neural network trained for handwritten character recognition using the backpropagation algorithm.The ground-breaking convolutional neural network LeNet-5 combines shared weights, local receptive fields, and spatial or temporal subsampling to provide some degree of shift, scale, and distortion invariance. It serves as the basis for the current CNN. LeNet-5 falls short of the conventional support vector machine (SVM) and boosting techniques, despite being helpful at identifying handwritten characters and reading bank checks. LeNet-5 did not receive enough attention at that time as a result.

.

**4.2 AlexNet**

AlexNet, a groundbreaking convolutional neural network (CNN), introduced several fundamental concepts. It was among the first CNNs to incorporate dropout, ReLU activation, and local response normalization (LRN). GPUs were leveraged for accelerated computation. Key innovations include using ReLU for activation, dropout to ignore neurons during training, and overlapping max pooling for richer features. LRN mimics biological lateral inhibition, enhancing neuron activity based on value magnitude. AlexNet utilizes two powerful GPUs and implements data augmentation techniques such as extracting random patches and applying PCA to RGB values, reducing overfitting and enhancing generalization.

**4.3 GoogLeNet**

The primary innovation of this architecture lies in its efficient use of computing resources, achieved through strategic planning of greater depth and width within a fixed computational budget. GoogLeNet, a 22-layer network, improved classification and detection through the Hebbian principle and multi-scale processing intuition. Efficiency is crucial in deep learning, as shown by GoogLeNet's enhanced accuracy with reduced parameters. Considering the rise of mobile and embedded systems, computational economy was a key design consideration alongside accuracy. The architecture aims to capture characteristics with multiple abstraction levels, featuring parallel pooling pathways and balanced filter sizes.

**4.4 ResNet**

Residual Blocks, a novel design used by ResNet, include skip connections that link a layer's activations to subsequent levels by omitting some layers in between. As a result, the network can fit the residual mapping rather than directly acquiring the underlying mapping. ResNets allow training incredibly deep neural networks without running into problems with vanishing or expanding gradients since they use skip connections.ResNets are able to train very deep networks more effectively than typical deep neural networks, with improved accuracy and practicality for implementation because to the use of skip connections and these block types.

The key components of the ResNet architecture include:

1. Residual Blocks: ResNet fits residual mappings by utilising residual blocks with skip connections, which enables effective training of very deep networks.

2. Identity and Convolutional Blocks: To ensure flexibility in addressing various scenarios, ResNet uses Identity Blocks when input and output activations match and Convolutional Blocks when dimensions diverge.

3. Skip Connections: By enabling smoother gradient flow, these connections help deep networks be trained more efficiently without requiring more processing power, protecting significant characteristics all the way through the network.

4. Customisation: Using pre-trained models from the ImageNet dataset and transfer learning, PyTorch allows users to modify the number of layers in ResNet designs.

5. Vanishing Gradient Problem: ResNets introduce skip connections that facilitate the smooth flow of gradients, hence solving the vanishing gradient problem that arises when training very deep neural networks.

**4.5 SENet**

Due to its distinct architecture, the network may more effectively modify how important particular features are in the input data, which improves accuracy and performance. Unlike typical CNNs, which maintain constant weights during training and produce less-than-ideal outcomes, SENet introduces a novel method of recalibrating features by dynamically altering weights at each level of the network.

Key points about the SENet architecture include:

* Integration of SE Blocks: By integrating SE blocks, SENet improves architectures such as Inception-ResNet-v2 and ResNeXt, leading to notable performance improvements with examples such as SE-Inception-ResNet-v2 and SE-ResNeXt15.
* Flexible Integration: Because SE blocks are flexible, different integration techniques can be incorporated into architectures such as ResNet, allowing for the effective testing of various inclusion designs to evaluate integration strategies1.
* Computational Efficiency: As network depth increases, SE blocks maintain computational efficiency while delivering notable performance enhancements, striking a balance between better outcomes and model complexity.15.
* SE Block Structure: The SE block structure consists of an excitation operation that comes after a squeeze operation. This results in channel weights that, through element-wise multiplication with input features, dynamically adjust feature relevance throughout the network.

**References:**

1. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324.

2. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

3. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).

4. Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7132-7141).

5. Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 31, No. 1).

6. Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc.".

7. Lakshmanan, V., Görner, M., & Gillard, R. (2021). *Practical machine learning for computer vision*. " O'Reilly Media, Inc.".

8. Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, *33*(12), 6999-7019.