Effect of Kernel Size and Number of Filters in CNN Convolution Layers

Introduction

Convolutional Neural Networks (CNNs) have become a cornerstone of modern computer vision due to their ability to automatically learn spatial features from raw image data. The performance of CNNs is strongly influenced by architectural choices made within convolutional layers, particularly the kernel size and the number of filters. Kernel size determines the receptive field of each neuron and controls how much spatial context is captured from the input. In contrast, the number of filters governs the diversity and richness of feature representations learned at each layer. Selecting inappropriate values for these parameters can lead to underfitting, overfitting, or unnecessary computational cost. Despite their importance, kernel size and filter count are often chosen heuristically without systematic analysis. This research aims to investigate how varying kernel sizes and filter numbers affect feature extraction and classification performance. By conducting controlled experiments, the study highlights the trade-offs between accuracy, model complexity and efficiency. The findings provide practical guidance for designing effective CNN architectures. Ultimately, this work helps practitioners make informed decisions when tuning convolutional layers for real-world applications.

Convolutional Neural Network (CNN)

- Convolutional Neural Networks are a class of deep learning models specifically designed for processing grid-like data such as images.
- CNNs automatically learn hierarchical features, starting from simple patterns like edges to complex objects.
- They use convolutional layers to preserve spatial relationships in the input data.
- Pooling layers reduce dimensionality while retaining important information.
- CNNs are widely used in image classification, object detection and medical imaging tasks.

Kernels in CNN

- A kernel is a small matrix that slides over the input image to extract local spatial features.
- It performs element-wise multiplication and summation to produce a feature map.
- Kernel size (e.g., 3×3 or 5×5) determines the receptive field of the convolution operation.

- Smaller kernels capture fine details, while larger kernels capture broader context.
- Kernels help CNNs detect patterns such as edges, textures and shapes.

Filters in CNN

- A filter consists of a kernel and its associated learned weights applied across the input.
- Each filter produces one feature map representing a specific learned pattern.
- Increasing the number of filters allows the model to learn a wider variety of features.
- More filters improve representational power but increase computational cost.
- Proper selection of filter count balances accuracy, efficiency and overfitting risk.

Dataset

The **CIFAR-10** dataset is a widely used benchmark dataset for image classification tasks in deep learning. It consists of **60,000** color images with a fixed resolution of 32×32 pixels. The dataset is divided into **50,000** training images and **10,000** test images. CIFAR-10 contains images belonging to ten distinct object classes, including animals and vehicles. Each class is evenly represented which helps prevent class imbalance issues during training. The small image size makes the dataset computationally efficient for experimentation with convolutional neural networks. Despite its simplicity, CIFAR-10 presents challenges due to low resolution and background clutter. It is well suited for analysing architectural choices such as kernel size and number of filters. The dataset allows controlled comparison of CNN configurations under identical conditions.

Experimental Design

To isolate the effects of kernel size and filter count, we fix all other hyperparameters.

- Experimental Variables

Experiment	**Kernel Size**	**No. of Filter**
Baseline	3 × 3	32
Exp-1	5 × 5	32
Exp-2	3 × 3	64
Exp-3	5 × 5	64

- Fixed Parameters
    - Optimizer: Adam
    - Loss: Categorical Cross-Entropy
    - Epochs: 15
    - Batch size: 64
    - Activation: ReLU

Experimental Results and Analysis

This section presents a detailed analysis of the experimental results obtained by varying the kernel size (3×3 and 5×5) and the number of filters (32 and 64) in the convolutional layers of a CNN trained on the CIFAR-10 dataset. The evaluation focuses on training behaviour, generalisation performance and overfitting tendencies, as reflected through accuracy and loss curves.

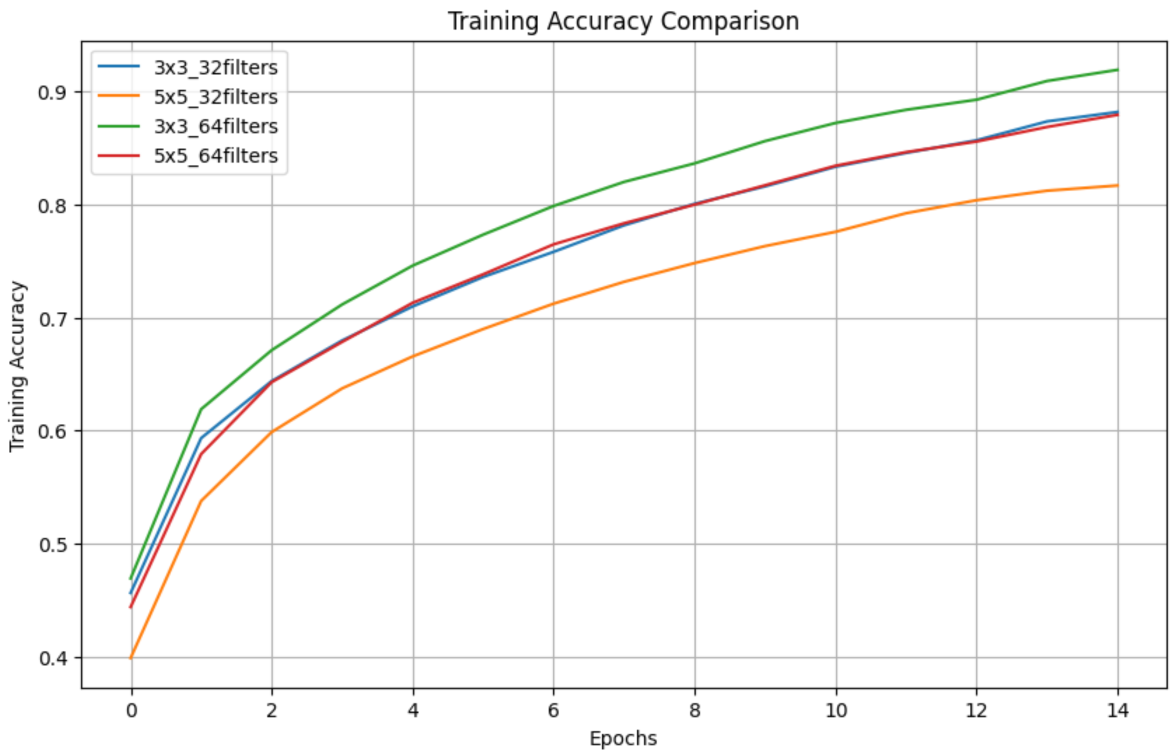

Figure 1: Comparison of Training Accuracy Across Different Kernel Sizes and Filter Counts.

The training accuracy results indicate that all CNN configurations successfully learn meaningful representations from the data, as evidenced by a steady increase in accuracy across epochs. However, clear performance differences emerge based on architectural choices. Models using 64 filters consistently outperform their 32-filter counterparts, achieving higher training accuracy and faster convergence. This demonstrates that increasing the number of filters enhances the network's capacity to learn a richer set of feature maps. For a fixed number of filters, models employing 3×3 kernels outperform those using 5×5 kernels throughout training. Smaller kernels enable more effective local feature extraction and allow deeper feature composition when stacked across layers. In contrast, larger kernels introduce more parameters early in the network, which slows learning and reduces training efficiency. These observations align with established CNN design principles adopted in architectures such as VGG and ResNet.

The training loss curves further support these findings. Models with 3×3 kernels and 64 filters exhibit the fastest reduction in loss, reaching the lowest final loss values. Conversely, the 5×5 kernel with 32 filters maintains the highest loss throughout training, indicating limited representational power and slower optimisation. While lower training loss indicates effective fitting, it also signals the need to assess generalisation through validation metrics.

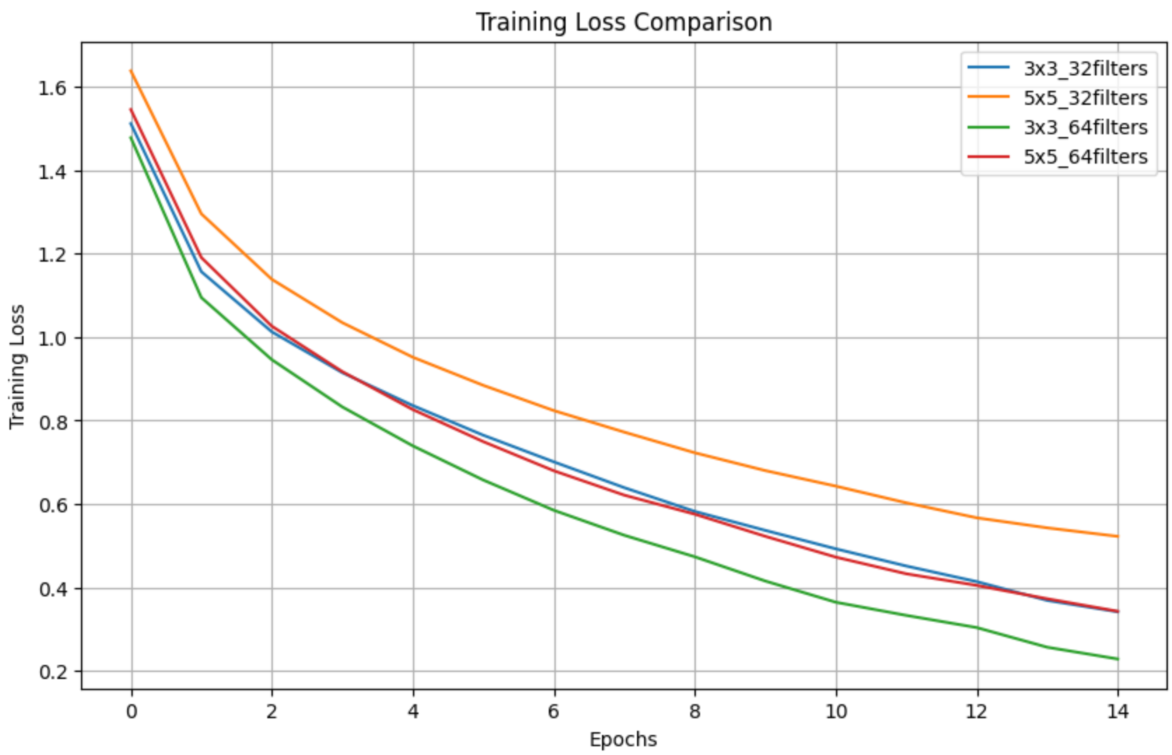

Figure 2: Comparison of Training Loss for CNN Models with Varying Convolutional Configurations.

Validation accuracy provides critical insight into how well each model generalises to unseen data. Across all configurations, validation accuracy increases rapidly during the early epochs, peaking around epochs 5–7 before stabilising or fluctuating. The 3×3 kernel with 64 filters achieves the highest peak validation accuracy, indicating superior generalisation compared to other configurations.

However, beyond the peak point, validation accuracy for higher-capacity models begins to plateau or slightly decline, suggesting the onset of overfitting. Models with 5×5 kernels consistently show marginally lower validation accuracy, despite reasonable training performance. This indicates that larger kernels do not necessarily improve generalisation and may instead introduce redundant parameters that fail to capture additional useful information. The validation accuracy results demonstrate that smaller kernels combined with a sufficient number of filters provide the best balance between learning capacity and generalisation performance.

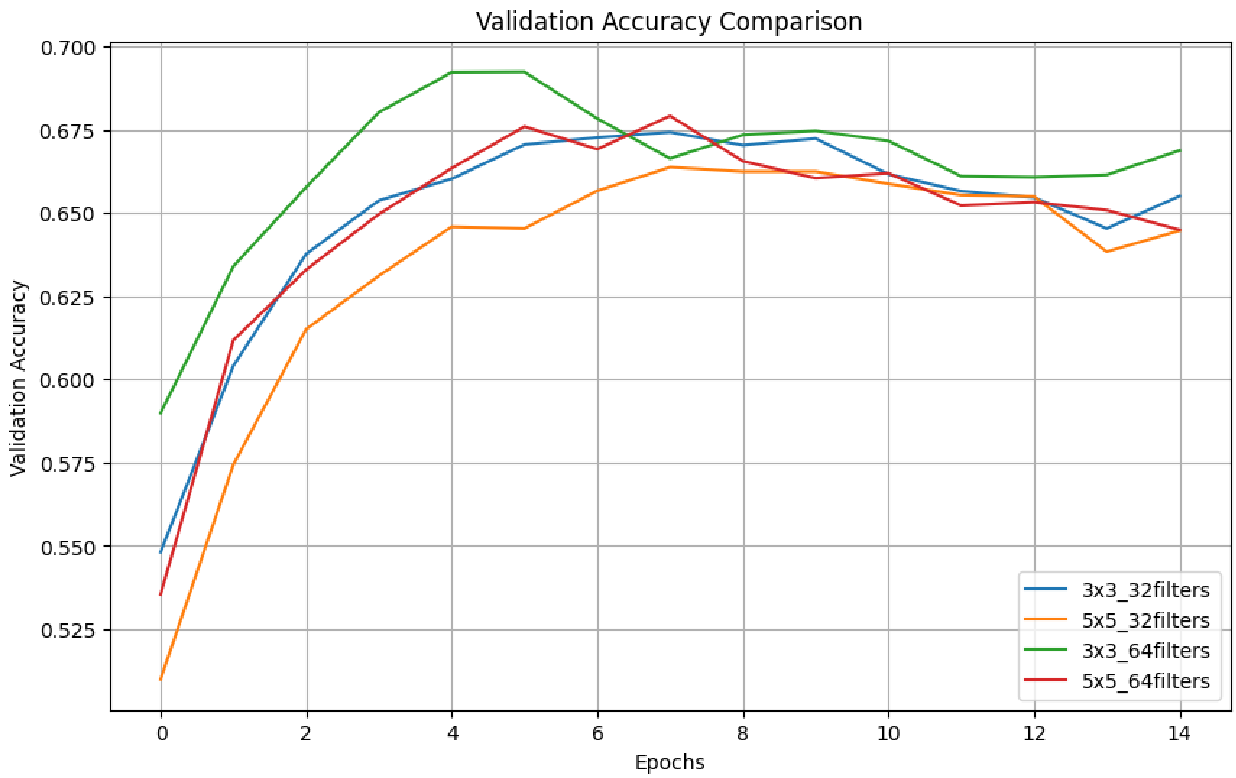

Figure 3: Validation Performance Analysis of CNN Architectures Using Different Kernel Sizes and Number of Filters.

The validation loss curves provide further evidence of overfitting behaviour. For all models, validation loss decreases during the initial training phase, reaching a minimum between epochs 4 and 6. After this point, validation loss begins to increase, even as training loss continues to decrease. This divergence clearly indicates overfitting. The increase in validation loss is most pronounced for models with 64 filters, particularly when combined with a 3×3 kernel, reflecting their higher model capacity. While these models achieve the highest training and validation accuracy, they are also more prone to memorising training data. In contrast, the 5×5 kernel with 32 filters shows a slower rise in validation loss, suggesting slightly better inherent regularisation but at the cost of reduced accuracy. These results highlight the importance of incorporating techniques such as early stopping, dropout or data augmentation when deploying higher-capacity CNN models.

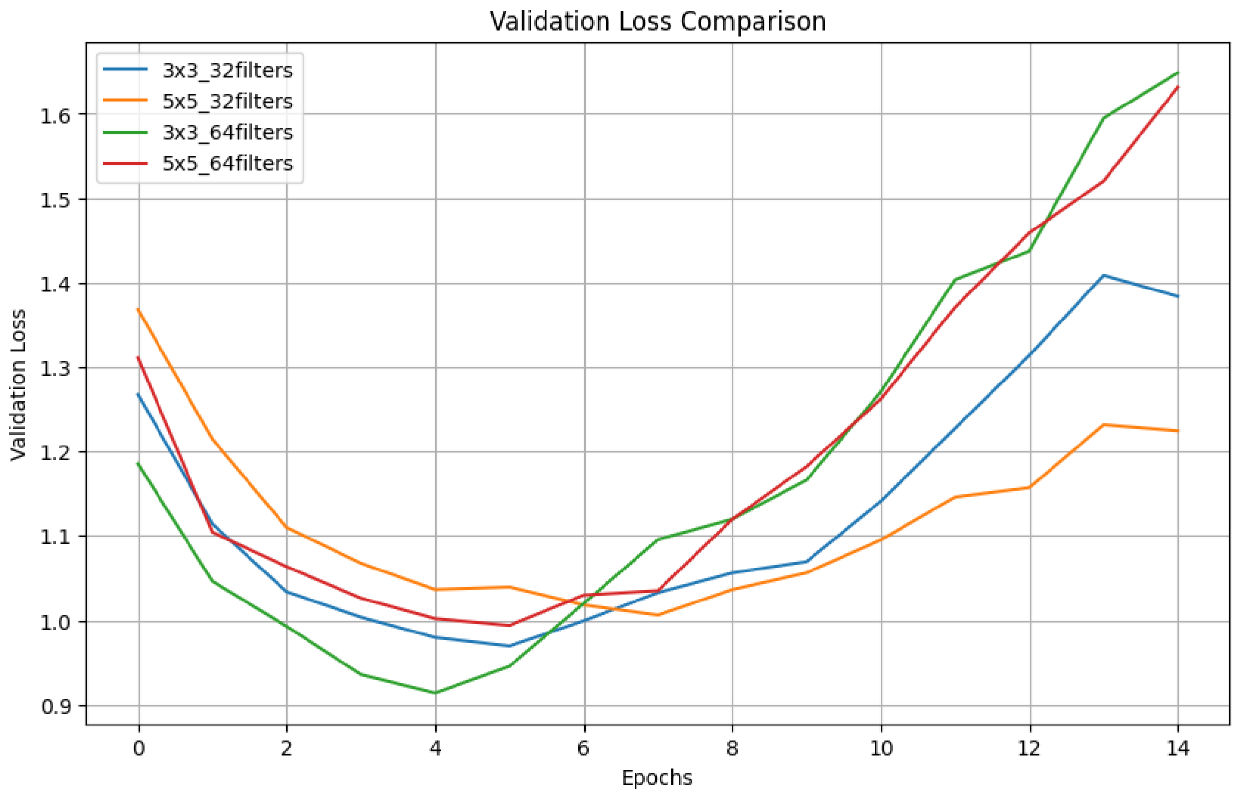

Figure 4: Comparison of Validation Loss for CNN Models with Varying Convolutional Configurations.

From a comparative perspective, the experimental results clearly show that:

- Kernel size influences feature extraction efficiency, with smaller kernels being more effective.
- Number of filters strongly affects model capacity, training speed accuracy.
- Increasing filters improves performance but increases overfitting and risk.
- Larger kernels increase computational cost without proportional gains in accuracy.

The 3×3 kernel with 64 filters emerges as the best-performing configuration overall, offering the highest accuracy and fastest learning. However, this configuration also requires careful regularisation to ensure robust generalisation.

These results align with modern deep learning practices, where deep architectures rely on smaller convolutional kernels and increased channel depth rather than large spatial filters. The experiment confirms that architectural hyperparameters significantly influence learning dynamics and must be chosen carefully based on dataset size, task complexity and computational constraints.

Conclusion

This research examined the effect of kernel size and number of filters on the performance of convolutional neural networks. The results show that smaller kernels, particularly 3×3, consistently provide more efficient feature extraction than larger kernels. Increasing the number of filters improves learning capacity and training accuracy by enabling richer feature representations. However, higher filter counts also increase the risk of overfitting, as indicated by rising validation loss. Models with 3×3 kernels and 64 filters achieved the best performance but required careful regularisation. Larger kernels offered no significant generalisation benefits despite higher computational cost. The study highlights the importance of balancing model complexity, accuracy and efficiency when designing CNN architectures.

References:

1. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (2002). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324.
2. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
3. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.
4. Goodfellow, I. (2016). Deep learning.
5. 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).
6. 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).

External Links:

1. Github : https://github.com/ms24ama/Effect-of-Kernel-Size-and-Number-of-Filters-in-CNN-Convolution-Layers