# A Journey into A CNN Filter Size and Its Effects, Limitations and Challenges Exploratory Study

Mohamed Aboukhair[a], Abdelrahim Koura[a], and Mohammed Kayed[a]

[a] Computer Science Department, Faculty of Computers and Artificial Intelligence, Beni-Suef University, Beni Suef 62511, Egypt

## Abstract

This research delves into the investigation of the impact of filter size, deviating from the conventional notion of preferring smaller filter sizes (3x3), with the aim of shifting the focus towards the potential of larger filters and encouraging researchers to invest more effort in exploring their capabilities. The effectiveness of larger filters will increase as computational power increases, however, currently, there are too few pure CNN models that achieve optimal performance with large filters showing how far bigger filter size topic is neglected. Therefore, it is imperative to consider the necessity of integrating larger filter sizes in our models, as computing power continues to advance in the future. As advancements in computer power continue to occur and larger image sizes become more prevalent, the obstacles that have traditionally hindered researchers from utilizing larger filter sizes or not preferring to work with them will diminish. This will lead to increased exploration and study of larger filter sizes in the future. Our role is to provide guidance by elucidating the previous architectures, the effects, limitations, and challenges that must be addressed in order to capitalize on the unique opportunities and potential offered by larger filters. To the best of our knowledge, we have identified four such opportunities from utilizing bigger filters and conducted an analysis of filter size impact using all previous research that was related to filter size effects comparison showing bias toward smaller filter sizes and also suggesting test case design for future filter size effects studies.

**Keywords:** CNN, Classification, Filter Size Effects, CNN structures

## 1 Introduction

The advent of ample processing resources and vast data sets has cleared the path for the success of deep learning, elevating the importance of computer vision. This cutting-edge technology possesses immense value across various fields, including automated driving, object recognition, and image analysis. Through the study of computer vision, we can optimize manufacturing processes and spearhead innovative technological advancements[1]. There has been a noteworthy escalation in the number of papers presented at machine learning conferences wherein at least one author had affiliations with the industry. Specifically, the percentage of industry-affiliated papers at the International Conference on Machine Learning (ICML) increased from 20-25% in 2017 to 45% in 2018. Moreover, the number of publications by researchers from Google's DeepMind alone grew twofold from 6% in ICML'17 to 13% in ICML'18. Although the precise ratio of industry-sponsored research in computer vision remains unspecified, the immense success of deep learning has facilitated the growth of computer vision applications, resulting in significant investments by numerous companies in research and development in this domain. Industry-sponsored researchers have emerged as a prominent component of the computer vision community and have made significant contributions to all major conferences and journals in this field[2].

The proliferation of deep learning in computer vision has supplanted conventional machine learning algorithms and engenders pronounced benefits in feature extraction, rendering it ubiquitous in computer vision, as well as in other domains such as physics, biology, and manufacturing. Notably, the paper identifies several prevalent deep learning algorithms in computer vision, including dropout, convolutional neural networks (CNN), full convolutional networks (FCN), and generative adversarial networks (GAN). Consequently, deep learning has assumed a critical role in computer vision and is a highly sought-after research area in both academic and industrial settings[3]. The application of deep learning has yielded remarkable successes in various domains, including computer vision, natural language processing, and video/speech recognition. In the field of computer vision, deep learning has facilitated notable advancements in tasks such as image classification, object detection, and segmentation. In the realm of natural language processing, deep learning has been utilized for tasks such as sentiment analysis, machine translation, and text generation. In the domain of video/speech recognition, deep learning has enabled significant progress in tasks such as speech recognition, speaker identification, and emotion recognition. It is noteworthy that the significance of deep learning in these areas is contingent upon the specific context and field in which it is employed[4].

Kernels or filters are essential components of deep learning architectures, especially Convolutional Neural Networks (CNNs), which use a variety of kernels (filters) in convolutional layers to convolve the entire picture and intermediate feature maps, yielding different feature maps. The filters are used to extract features from input pictures, and their size might affect the network's computational efficiency. Szegedy et al. (2016), for example, advocated replacing filters with a size of 5x5 (7x7) with two stacked 3x3 filters to minimize the number of parameters and computational costs while keeping the receptive field size the same. As a result, filters play a critical role in the performance and efficiency of CNNs[5, 4]. This presents a thought-provoking yet insufficiently examined matter: the facility to utilize larger filters within our deep learning models. To the best of our knowledge, it appears that the use of 3x3 filters in convolutional neural network (CNN) designs is preferred. This preference extends to original models such as Inception, which incorporate larger filters within their modules. Subsequent studies have demonstrated that the substitution of bigger filters with two stacked 3x3 filters is more effective, further solidifying the dominance of 3x3 filters in all CNN models.

The emphasis on 3x3 filters in convolutional neural network (CNN) designs is not without trade-offs. While the use of larger filter sizes may lead to better learners, the associated increase in the number of parameters presents challenges for the optimizer. Overfitting is also a concern with larger filters, which are known to be powerful learners. Moreover, the exponential growth in parameters necessitates increased processing capacity, further complicating matters.

The present research endeavors to elucidate the challenges and drawbacks entailed in the application of larger filter sizes in convolutional neural network (CNN) architectures by conducting a comprehensive review of relevant literature. In addition, this study seeks to assess the benefits and limitations of larger filters and identify exceptional prospects by addressing the impediments and constraints associated with their use.

The contributions of this paper are summarized as follows:

- An exploration analysis of filter size effects on CNN
- Bigger filters challenges and opportunities

The remaining sections of this paper are organized as follows: Section 2 presents a comprehensive review of the previous literature relevant to our work. In Section 3, we discuss the output from the research that we targeted and show the rare opportunities that a bigger filter can offer. Finally, in Section 4, we offer concluding remarks summarizing the key findings and contributions.

# 2 Literature Review

As far as our understanding goes, no prior studies have comprehensively explored all previous work about the impact of filter size on CNN models, also no prior studies have highlighted both the limitations of utilizing larger filters and the potential opportunities that may arise from such exploration. Despite the lack of extensive research on this topic, it remains one of the most crucial aspects of the field of deep learning. Our selection criteria for relevant research include a focus on two key factors: comparison of various filter size impacts which is the main focus, and inclusion of modules and structures that incorporate larger filters in pure CNN architecture.

## 2.1 Preliminaries of CNN filter

Convolutional Neural Network (CNN) is a type of neural network that is commonly used in machine learning, especially in vision-related applications. It is well-known for learning representations from grid-like data, such as images, and has demonstrated significant performance improvement in a variety of machine-learning tasks. CNNs are made up of several layers, including convolutional layers, nonlinear processing units, and subsampling layers, which aid in extracting and categorizing valuable features from input.[6]. CNNs are inspired by the structure of neurons in human and animal brains and are designed to discover key aspects of input data without the need for human intervention. They include several layers, including convolutional layers that extract features from input images using convolutional filters.[7]. Since the late 1980s, CNNs have been used to do visual tasks. LeCuN et al. introduced the first multilayered CNN, ConvNet, in 1989, based on Fukushima's Neocognitron. This laid the groundwork for modern 2D CNNs. ConvNet demonstrated successful results in optical character and fingerprint recognition tasks in LeCuN's work, which incorporated supervised training using the backpropagation technique. It was further enhanced with the development of LeNet-5, which played an important role in character classification in document recognition applications.[6]. Deep CNNs had difficulties in the early 2000s due to their sophisticated architecture and limited hardware resources. However, the introduction of activation functions such as RelU, as well as developments in technology such as GPUs, have rekindled interest in CNN research.[6].

The availability of massive image databases like ImageNet, as well as platforms such as Kaggle and Colab, has expedited CNN research. A CNN is made up of several essential components that allow it to learn representations from grid-like data and perform tasks like feature generation and classification. Convolutional layers, nonlinear processing units, and subsampling layers are among the components.[6].

Convolutional layers conduct the convolution operation, which aids in the extraction of relevant features from locally connected data points. To scan the input data and build feature maps, convolutional kernels (filters) are used. Nonlinear Processing Units: The output of the convolution process is passed through a nonlinear activation function. This function adds nonlinearity to the feature space and aids in the learning of abstractions and semantic differences in images. Subsampling Layers: Subsampling layers summarise the findings of previous levels and make the input geometrically invariant. They aid in reducing the dimensionality of feature maps and capturing the most critical data. These components collaborate hierarchically, allowing CNN to learn hierarchical representations of the input data. Automatic feature extraction hierarchical learning, and weight-sharing capabilities of CNNs contribute to their effectiveness in various machine learning applications, especially in vision-related tasks[6].

In a Convolutional Neural Network (CNN), a filter or kernel is a tiny matrix of weights that is convolved with the input image to extract features. Convolutional Neural Networks (CNNs) rely heavily on filters or kernels to extract features from input data. These filters are slid over the input image and conduct a dot product operation to obtain an output feature map. During the training phase, the weights of the filters are modified, helping them to learn to extract significant features. The usage of filters in CNNs has various advantages like filters enable weight sharing, reducing the amount of trainable network parameters. This weight-sharing function improves generalization while avoiding overfitting. Second, filters enable the CNN model to recognize relevant features automatically without human supervision. This automatic feature identification is a key advantage of CNNs compared to their predecessors[7]. CNN filters are intended to recognize specific patterns or features in input data. A filter, for example, could be built to identify edges, corners, or textures. CNNs may learn to recognize complicated patterns and features by applying many filters to the input data. The convolutional layer produces a set of feature maps, each reflecting the activation of a different filter. These feature maps are then subjected to non-linear activation functions, which introduce non-linearity and improve the network's ability to learn complicated relationships.[8, 7]. CNN Weights are filters or kernel values, and CNNs offer Weight sharing, which is an important aspect of CNNs. It is the practice of applying the same set of weights to different spatial places in the input data. CNNs achieve translation invariance by sharing weights, which means they can identify and recognize features regardless of where they are in the image. This is very beneficial in computer vision jobs where feature locations may vary.[8]. The selection of filter size is a critical design element in CNN architecture since it directly influences the network's capacity to extract meaningful information and achieve accurate results. Filters of varying sizes can be used to capture data at various levels of granularity. Small-size filters retrieve fine-grained features, and large-size filters capture coarse-grained information. CNNs can perform effectively on both coarse and fine-grained details by modifying the filters, boosting their overall performance[6].

Larger filter sizes in convolutional neural networks (CNNs) can have a number of effects. For starters, they broaden the network's receptive area, allowing it to detect greater spatial patterns in the incoming data. This can be useful for tasks that need a comprehension of the greater environment or structures. Larger filters can broaden the network's receptive field. The receptive field is the portion of the input that determines the activity of a certain neuron. By employing larger filters, the network can consider a broader context while making predictions, perhaps improving its overall performance. Larger filter sizes can increase model complexity and computational demands. With greater filters, the number of parameters in the network grows, which can make training and inference more computationally expensive [7].

## 2.2 Filter Size Key Challenges

The only significant challenges to filter size existed in the usage of larger filters, while smaller filters are advised for a variety of reasons. This section illustrates the primary issues that models encounter when using larger filter sizes, clarifying the restrictions while describing the challenges.

**Output Dimension Challenge**

One of the problems that may arise in using bigger filters is Dimension reduction leading to limiting the depth model that can reach with usage of bigger filter size That happens because of filter size. filter size factor can be removed by using padding yet this will lead to overpadding the higher we go in-depth which puts Dimension Challenge as one of the limitations that lash back the usage of bigger filters leading to prefer 3x3 filter size that allows for going deeper in depth which benefits the accuracy of the model [9, 10, 11].

**Computational Power Challenge**

One issue that arises with the use of larger filter sizes as time and model complexity increase is processing capacity. Because the filter is based on numbers, time complexity grows as the filter size increases. The issue of time complexity not only affects the time required to train the model, but it also leads to larger filters yielding bad results as it needs more time than smaller filter sizes to learn.[12, 13, 14].

**Overfitting Challenge**

Overfitting is more likely to occur in complex models with greater flexibility because larger filters increase model complexity; they can readily aid in overfitting due to the rise in model complexity. [15, 16]. This can be handled by a variety of methods, including dropout and regularisation, but it is underutilized because few model architectures include larger filters as part of their structure. As far as we know, there is no technique against overfitting that targets the use of larger filters; all techniques are generalized to target overfitting; however, one way to reduce the chances of overfitting is by lowering the complexity; in other words, even if larger filters are implemented, overfitting may occur, making this one of the most difficult challenges.

### 2.3 Collection of papers

After we collect the related papers to filter size effects comparison on pure CNN or pure CNN architecture that use the bigger filter as part of their design, we collect a sample of other papers that are Simi-related to compare with the previous and papers that are not related yet will give information about our deep learning CNN architectures like an estimated number of architectures that have a bigger filter in it and not, we will talk about them in section Non related to bigger filter. As far as we know we collected just a sample as there are too many papers that can be Simi-related and the same goes with architectures that do not apply bigger filter sizes we collected as far as we could to show just a bigger picture of our problem not to collect all custom architectures or all paper that simi-related. Also, we focus on the pure CNN architectures paper that uses pure CNN as a feature extractor but some non-pure CNN architectures employ bigger filters like (GCN, LR-et, Swin Transformers, ETC.), and these architectures have increased their performance by using bigger filters yet it is out of our research focus as we targeted pure CNN itself and its use of bigger filters.

### 2.4 Non-related to bigger filter

We collected the architectures that related to larger filter size in the process we have collected seven architecture that does not use the bigger filter size and noticed that a couple of in-processing papers that have continuation papers led to some architecture to improve by using two stacking layers of 3x3 filter size instead of bigger filter or by simply replace bigger filter sizes by smaller filter size[10]. For example, Yolo model version one included a bigger filter size but by versions two and three bigger filters were replaced by smaller filter sizes leading to a more efficient model in terms of model speed[21, 30, 31]. The same goes with Mobilenet which was specifically made for mobile applications leading to making baseline that depended on a smaller filter size to maintain its use on mobile applications[32]. The other architectures used the small filter (3x3) as its original design due to speed performance limitations or as recommended form multiple researchers' papers. and some architectures use 1x1 filter sizes like SqueezeNet that replace the bigger filter and get the same accuracy as in Alexnet with lower parameters[33, 34, 35, 36].

### 2.5 Simi-related to bigger filter

After we collected the architectures that related to larger filter size we found a problem, there are architectures that employ bigger filter size just in the beginning of the first couple layers in the architecture. we decided to exclude these architectures from our focus because they use bigger filters to deal with the image when it comes to the architecture as unprocessed big input leading to prefer a bigger filter to deal with this big input instead of a small filter almost in all architectures that use the bigger filter it uses in the first layers. for example, models like YOLO, Desnenet, Resnet, and ZFnet use filter 7x7 in the first layer then the rest of the layers employ 3x3 normally and this is commonly used in most of newer model architectures[20, 17, 21]. The only model that did not do that was the old model Lenet which employs a 5x5 filter size replacing 3x3 by 5x5 in the model which is the second smallest filter size[19]. Also, we decided to include a study titled "(input) size matters for CNN classifiers" which applied models that we use bigger filters in the first layers of CNN yet it is not the only reason to pick it, it has a relation between image size and model performance showing that bigger image is better even with no extra details from bigger images still model performs well on those bigger images leading us to see the bigger images as better input and smaller filter sizes will have a disadvantage with higher resolution images as it will fail to catch big pattern which will lead the research community in future to abandon smaller filter sizes if big resolution images are preferred with deep learning models in future[37].

### 2.6 History of Bigger Filter Size Architectures

To the best of our knowledge, we collected all models that are related to both pure CNN architecture and employ a bigger filter size than 3x3 which has been shown in Table 2 where the type column refers to the architecture name. Alexnet was the first CNN model that employed a bigger size of 11x11 in the first layer then lowered the size of the filter to 5x5 by the second layer and then 3x3 by third, which is one of CNN's old models that employed the bigger filter size which was developed in 2012, later in 2014 model named the Inception, proposed in the work of "Going Deeper with Convolutions,"[11] introduced a novel structure incorporating multiple filters including bigger filter sizes. The Inception module faced challenges related to computational limitations[38] and managing output size growth associated with 5x5 filter size in Convolutional layers, which were addressed through input reduction strategies. Furthermore, the Inception module was recommended to be used only in the upper layers of a model. In 2015 research of"Rethinking the Inception Architecture for Computer Vision"[10] aimed to enhance the module by replacing 5x5 Convolutional layers with two consecutive 3x3 Convolutional layers. This led to the death of the first model that used a bigger filter size as part of its structure not only used in a couple of first layers. This led to the idea of employing bigger filters to die for a couple of years until the year 2020 when the Efficientnet model was introduced as a model that has both performance and speed, its structure has multiple 5x5 filter sizes in the middle stages of the model

Table 1: Papers: Simi-Related to Bigger Filter

| Info / ID | Name | Author | Year | Type |
|---|---|---|---|---|
| paper8 | *Densely Connected Convolutional Networks*[17] | Huang et al. | 2018 | Densenet |
| paper9 | *Deep Residual Learning for Image Recognition*[18] | He et al. | 2015 | Resnet |
| paper10 | "Backpropagation applied to handwritten zip code recognition"[19] | LeCun et al. | 1989 | Lenet |
| paper11 | *Visualizing and Understanding Convolutional Networks*[20] | Zeiler and Fergus | 2013 | ZFnet |
| paper12 | *You Only Look Once: Unified, Real-Time Object Detection*[21] | Redmon et al. | 2016 | YOLO |
| paper13 | *OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks*[22] | Sermanet et al. | 2014 | OverFeat |

Table 2: Papers: Related to Bigger Filter

| Info / ID | Name | Author | Year | Type |
|---|---|---|---|---|
| paper14 | "ImageNet Classification with Deep Convolutional Neural Networks"[23] | Krizhevsky, Sutskever, and Hinton | 2012 | Alexnet |
| paper15 | *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*[24] | Tan and Le | 2020 | Efficientnet |
| paper16 | *Scaling Up Your Kernels to 31x31: Revisiting Large Kernel Design in CNNs*[25] | Ding et al. | 2022 | RepLKNet |
| paper17 | "Going Deeper with Convolutions"[11] | Szegedy et al. | 2014 | Inception |
| paper18 | "Analysis of Filter Size Effect In Deep Learning"[13] | Camgözlü and Kutlu | 2021 | comparison |
| paper19 | "Effect of filter sizes on image classification in CNN: a case study on CFIR10 and Fashion-MNIST datasets"[26] | Khanday, Dadvandipour, and Lone | 2021 | comparison |
| paper20 | "The Impact of Filter Size and Number of Filters on Classification Accuracy in CNN"[14] | Ahmed and Karim | 2020 | comparison |
| paper21 | "Convolution Kernel Size Effect on Convolutional Neural Network in Histopathological Image Processing Applications"[27] | Ozturk et al. | 2018 | comparison |
| paper22 | "Convolution Neural Networks and Impact of Filter Sizes on Image Classification"[12] | Khanday and Dadvandipour | 2020 | comparison |
| paper23 | "Filter size optimization on a convolutional neural network using FGSA"[28] | Poma et al. | 2020 | comparison |
| paper24 | *More ConvNets in the 2020s: Scaling up Kernels Beyond 51x51 using Sparsity*[29] | Liu et al. | 2023 | SLAKNet |

leading to good results showing that 3x3 filter size is not center of use to get results and speed performance[24]. The 5x5 filter is still one of the smallest filters so even if Efficientnet relays on it, it does not prove that an even bigger filter is better until 2022, the unnoticed RepLKNet architecture comes to light with a model that employs pure CNN architecture with a filter size of 31x31 with lower depth solving the problem of time complexity which highlights the efficiency of very large kernels. Although the 31x31 filter size is indeed big yet the research continuation of RepLKNet titled "*More ConvNets in the 2020s: Scaling up Kernels Beyond 51x51 using Sparsity*" increased the kernel size to 51x51 achieving even better results than RepLKNet by a novel model called SLAKNet which has broken the limit of kernel size into another new level suggesting that the scale can even increase more. The 51x51 filter size which is the biggest filter size known to be employed in CNN architectures is still an achievement that proves the need to study the bigger filter sizes more and make more efforts to conclude the right combinations to achieve optimal soulations[25, 29].

## 2.7 Dimensions of Filter Size Comparison in CNN

The present study aims to examine the efficacy of various filter sizes in two distinct dimensions, namely their application in architecture and their comparison for the purpose of investigating the impact of filter sizes on our comprehension of the same and subsequently enhancing the existing architectures.

**Architecture Comparison criteria**

In the realm of CNN architecture, numerous factors can be employed to compare and contrast different CNN architectures. However, the primary focus of this study is on filter size. To this end, we have chosen a dataset on which the architecture was trained and evaluated, while also considering whether the task was in the classification or object detection domain. In addition, we have selected the maximum parameters used in the architecture study, which may not necessarily be small, as researchers often aim to increase the number of layers and filters in their models, resulting in suboptimal parameters for the architecture. This consideration allows us to gauge the importance placed on computational power by researchers. Furthermore, we have examined the standard image size used in each architecture study to assess how well our architectures can handle image sizes, which could be related to larger filter sizes. The number of layers in the model is also taken into account as a criterion, as it can reveal the relationship between depth and filter size. Finally, we have also included the estimated number of published papers related to a study or architecture, sorted by year, as an indicator of the level of interest and engagement from the research community on the topic of larger filter-size architecture.

**Filter Size Effects Comparison criteria**

The present study aims to investigate the effects of filter size using all previous studies. While previous studies have examined the effects of filter size, simply showcasing the findings of each study is insufficient as it may imply agreement with their respective conclusions. Therefore, this study will present the results of each study in conjunction with our own perspective, while also identifying any potential biases or abnormalities. The reliability of each study will be evaluated by selecting a set of criteria, which will be used to determine their respective degrees of reliability. To begin, a specific year will be chosen to evaluate the progress made in the field of study. Additionally, a dataset and image size will be selected to examine the materials used in the creation of these studies. The image size will be ranked from largest to smallest to determine the degree to which it influences the results. Furthermore, the researchers' opinions on the use of larger filter sizes will be examined, and the extent to which each study demonstrates a bias towards smaller filter sizes will be assessed. This will be accomplished by analyzing the degree to which small images were utilized or if the study limited the range of filter sizes used, which may indicate a preference for smaller filter sizes.

## 2.8 Filter size effects papers

We showed previously bigger filter CNN architecture, in this section we will show the main target of this study which is filter size effects. As stated in earlier sections, the research issue of filter size effects is underexplored, with only six studies evaluating filter size effects as their main target that we are aware of, and practically almost all of them assume that using small filters is better which have been weakened with both RepLKNet and SLAKNet architecture. Because we do not have access to their outcomes data, we will portray the wider picture using their analysis graphs.

**Paper18**

The paper18 is "Analysis of Filter Size Effect In Deep Learning" by Camgözlü and Kutlu in 2021. The research illustrates the effects of three distinct filter sizes (3x3;5x5;9x9) on the Mendeley dataset in two ways. The first is that each filter size is standalone in the model, which means that each model includes only one filter. The second method is far more fascinating since the baseline in this research is six convolutional layers, and the researcher increases or decreases the filter size every two convolutional layers, resulting in the use of 3x5x9 filter size in this order or 9x5x3 filter size in this order.[13].

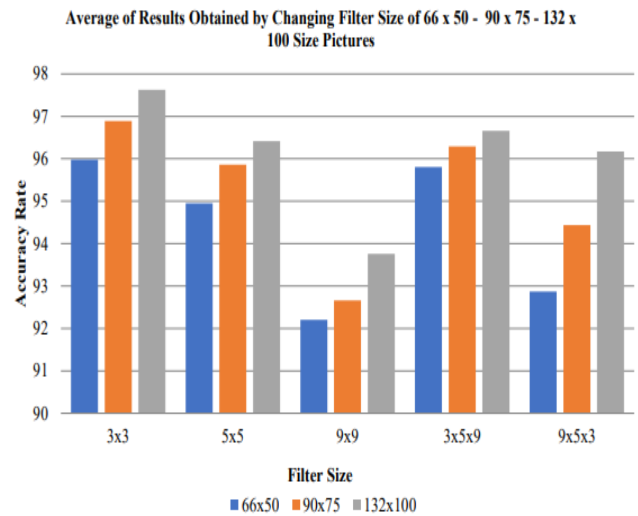


Figure 1: Accuracy Over Filter Size with Images Sizes

From Figure 1, The impact of filter size on CNN models has been observed to affect accuracy in various ways. Notably, the efficacy of 3x3 filter size and larger filter sizes have been observed to have contrasting effects on model accuracy. Further, the use of a combination of incremental filter sizes has been observed to produce more favorable results than larger filter sizes alone. Interestingly, the use of decremental filter sizes has been observed to have a significant impact on model accuracy in relation to image size, highlighting the potential benefits of utilizing larger filter sizes.

From Figure 2, The influence of filter size and image size on time complexity has been observed to have varying effects on model performance. Specifically, the efficacy of 3x3 filter size and smaller image sizes have been observed to positively impact time complexity. Conversely, the order in which multi-filter sizes are utilized has been noted to have a significant impact on time complexity, understandably due to the increased processing time required for larger filter sizes. Notably, decremental filter sizes have been observed to result in the second-worst time complexity performance, highlighting the importance of carefully selecting filter sizes to optimize deep learning models.

From Figure 3, The pars represent the parameter number

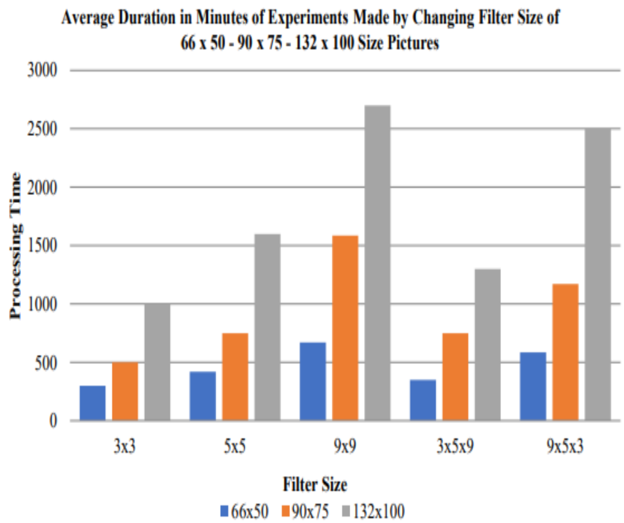


Figure 2: Time over Over Filter Size with Images Sizes

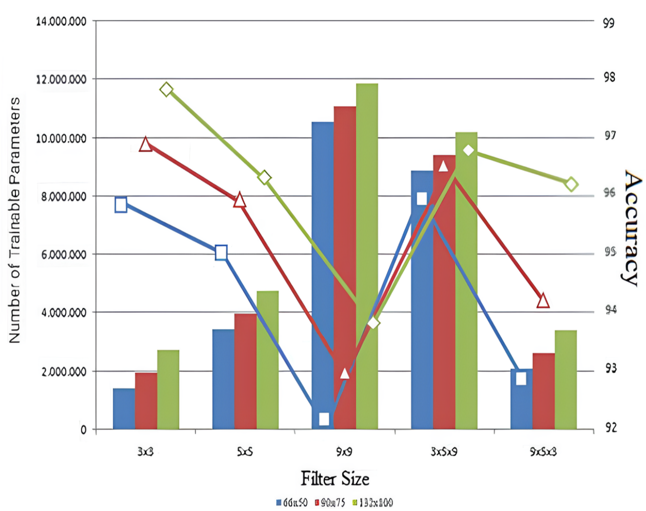


Figure 3: Accuracy Over Parameters Over Filter Size

while the line represents the accuracy of each approach. The impact of filter size on model performance can be seen in the number of parameters, as well as the accuracy of the approach. It has been observed that the use of smaller filter sizes, such as 3x3, can positively affect time complexity. Additionally, the order in which multi-filter sizes are utilized has been found to significantly impact time complexity. However, it should be noted that the illusion of varying parameter numbers for different image sizes is due to the fully connected layer parameters depending on the dimension of input that reaches it. The second illusion is fascinating as 3x5x9 has lower time than 9x5x3 from Figure 2 yet 3x5x9 has way higher parameters than 9x5x3. the use of many filter numbers the more we go in-depth can result in the last layers being dense in parameters, causing a larger filter to have more parameters yet work with smaller dimensions in parallel, resulting in faster processing time. The structure used by the researcher is not supplied, but this explanation is the most plausible.

**Paper 19**

The paper 19 is "Effect of filter sizes on image classification in CNN: a case study on CFIR10 and Fashion-MNIST datasets" by Khanday, Dadvandipour, and Lone in 2021 [26]. The research presents the effects of different filter sizes (3x3;5x5;7x7) in two datasets( CFIR10 and Fashion-MNIST).

| Filter Size | Training Data | Validation Data | Test Data |
|---|---|---|---|
| $3 \times 3$ | 0.942625 | 0.7275 | 0.7304 |
| $5 \times 5$ | 0.923275 | 0.7261 | 0.7297 |
| $7 \times 7$ | 0.87725 | 0.7067 | 0.635 |

Table 3: dataset-Cifar10-acc

| Filter size | Training data | Validation data | Test data |
|---|---|---|---|
| $3 \times 3$ | 0.929 | 0.9235 | 0.9268 |
| $5 \times 5$ | 0.926 | 0.9196 | 0.9264 |
| $7 \times 7$ | 0.918 | 0.910 | 0.911 |

Table 4: dataset-FashionMNIST-acc

From Table [3;4], We only notice that the larger the filter size, the worse the results get, and filter 5x5 is the closest to 3x3 filter size from an accuracy standpoint, leading to the common misconception that 3x3 filter size is better than 5x5 filter size, but this is not the case because the image size in Fashion-MNIST is 28x28 and the image size in CFIR10 is 32x32, which means that the effects of bigger filters were tested on the smaller image, which gives a huge advantage.

**Paper 20**

The paper 20 is "The Impact of Filter Size and Number of Filters on Classification Accuracy in CNN" by Ahmed and Karim in 2020 [14]. The research presents the effects of different filter sizes (3x3;5x5) with different numbers of each filter being applied in the convolutional layer on KTH dataset.

From Table 5, We see the same pattern of favoring smaller filter sizes due to the input image size of 96x96, but in this study, we can see a different perspective, which is the number of filters, the results improve with a higher filter number in the case of smaller filter 3x3 size, but in the 5x5 filter size it drops-down then improves but falls short of its best score. The impacts of a larger filter size are undoubtedly overfitting, and increasing the number of filters will just make the output worse than its original form with fewer filters. The number of filters of 128,256 is still greater than the number of filters of

| Dataset | Convolution layer | Number of filters | Filter Size | Train accuracy | Test accuracy | Loss value |
|---|---|---|---|---|---|---|
| KTH | 1, 2 | 32, 64 | 3×3 | 0.999% | 97.64% | 0.002 |
| | 1, 2 | 64, 128 | | 0.999% | 97.96% | 0.003 |
| | 1, 2 | 128, 256 | | 0.998% | 98.09% | 0.005 |
| | 1, 2 | 32, 64 | 5×5 | 0.997% | 97.58% | 0.007 |
| | 1, 2 | 64, 128 | | 0.999% | 97.33% | 0.002 |
| | 1, 2 | 128, 256 | | 0.998% | 97.46% | 0.005 |

Table 5: accuracy over filter size

64,128 and so reduces the effect of overfitting yet it is lower than the score of the number of filters of 32,64.

| Dataset | Number of filters | Filter Size | Training Time in second per epoch |
|---|---|---|---|
| KTH | 32 | 3×3 | 38 s |
| | 64 | | 96 s |
| | 128 | | 259 s |
| | 32 | 5×5 | 67 s |
| | 64 | | 171 s |
| | 128 | | 516 s |
| | 32 | 7×7 | 160 s |
| | 64 | | 428 s |
| | 128 | | 1400 s |

Table 6: time over filter size

From Table 6, we notice that the number of filters affects the time complexity exponentially, and the ratio of time increment increases with higher filter number increase even more with higher filter size.

**Paper 21**

The paper 21 is "Convolution Kernel Size Effect on Convolutional Neural Network in Histopathological Image Processing Applications" by Ozturk et al. in 2018 [27]. The research presents the effects of different filter sizes (3x3;5x5;7x7;9x9) in the training process on The CAMELYON17 challenge dataset which has the second biggest image size (128x128) in all previous papers.

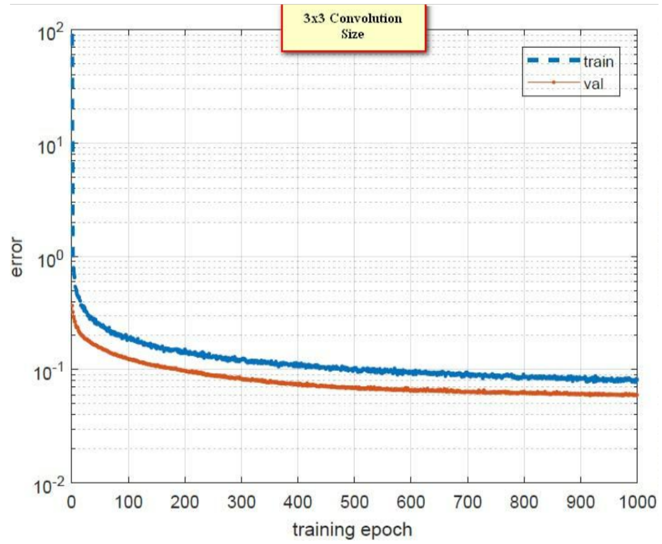


Figure 4: 3x3 train vs val loss

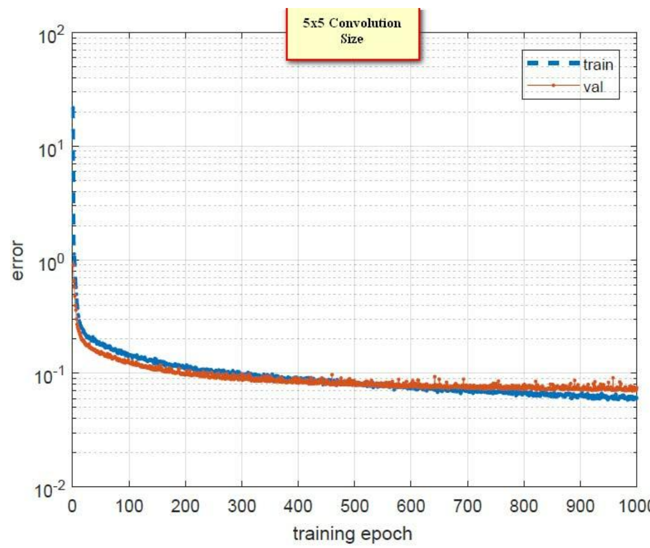


Figure 5: 5x5 train vs val loss

From Figure [4;5], We notice that the 3x3 filter has a stable curve where the line of validation loss is closing to train loss, indicating stable results (6.1%) error. On the other hand, the 5x5 filter had stable results before overfitting took over after 500 epoch, resulting in a decrease in train loss lower than validation loss, which is the effect of overfitting, resulting in 7.5% error, as expected from an overfitting model.

From Figure [6;7], Based on our observations, it seems that the 7x7 filter displays a stable curve with respect to total epochs. However, there are some dropouts that cause slight disturbances in the line. Interestingly, the curve of training and validation loss for this filter converges better than the 3x3 filter in Figure 4, resulting in a lower error rate of 6%. Suggesting that larger filter sizes may have untapped potential,

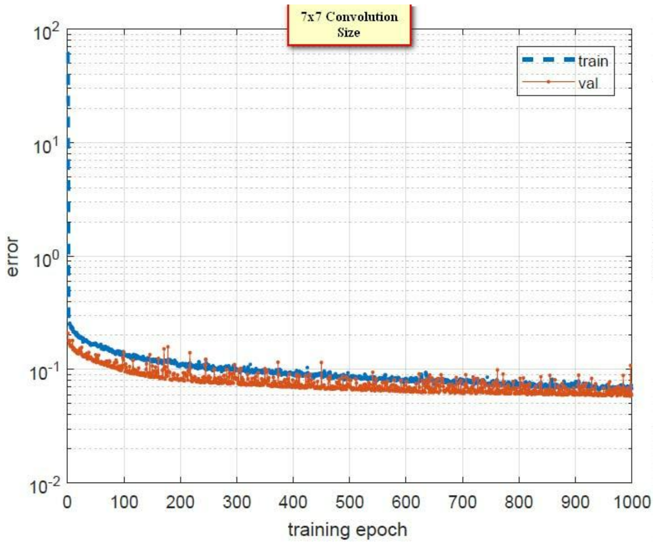


Figure 6: 7x7 train vs val loss

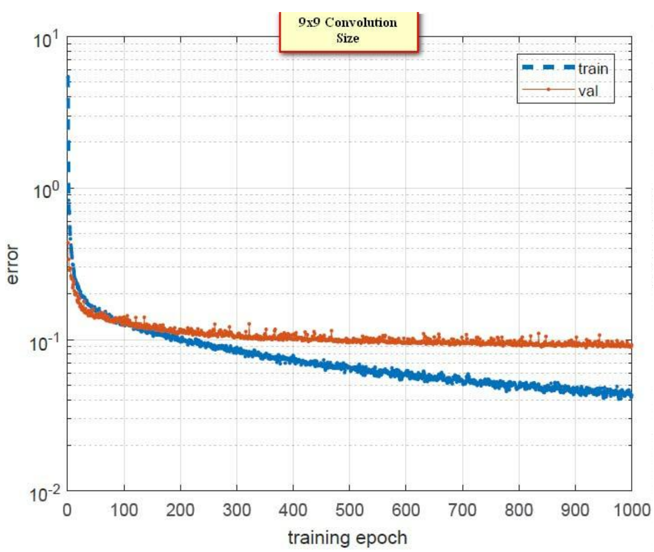


Figure 7: 9x9 train vs val loss

but also pose a risk of overfitting. For instance, the 9x9 filter tended to overfit after just 100 epochs, resulting in an error rate of 9.1%.

### Paper 22

The paper 22 is "Convolution Neural Networks and Impact of Filter Sizes on Image Classification" by Khanday and Dadvandipour in 2020 [12]. The research presents the effects of different filter sizes (3x3;5x5;7x7) in the digits MNIST dataset and gets similar results to paper 19 favoring a smaller filter size with 1% improvement while using an image size of 28x28.

### Paper 23

The paper 23 is "Filter size optimization on a convolutional neural network using FGSA" by Poma et al. in 2020 [28]. The present study introduces the FGSA algorithm, namely the fuzzy gravitational search algorithm, which aims to optimize filter size for improved recognition rates in face recognition tasks. The algorithm was applied to the CROPPED YALE database, consisting of 380 records with an image size of 640 (w) x 480 (h), using a small convolution model to mitigate computational requirements. Results indicate that the optimal filter size is 9x9, suggesting that larger filter sizes may be advantageous for extracting relevant information from larger image sizes.

### Bias Degree

We present all the previous work related to comparing filter sizes showing the favoritism of 3x3 filter size while explaining why 3x3 filter size is tested on cases that favor its use in the first place. In this section, we will discuss these effects with the possibility of improving the test case design in some papers while showing both collective analysis and bias in the studies. Based on the findings presented in Figure 3, it is evident that the image size limitation to a maximum size of 132x100 results in a significant advantage for smaller filter sizes. Notably, the absence of the 7x7 filter size in the experiment can not be ignored because as shown in Figure 6, the application of the 7x7 filter size on slightly larger image sizes yields more promising results than the 5x5 and 3x3 filter size. Therefore, it can be concluded that the evaluation of bigger filter sizes should consider the image size as a crucial factor[37]. Additionally, the results presented in Table 6 show that the increment of time complexity exponentially affects the performance of models. This observation serves as an explanation for the second illusion of the 9x5x3 with high time complexity despite having a lower parameter number. These findings align with the discussion presented in the previous section of Paper 18[27, 13].

The dimensions of the images in papers 19 and 22 are between 28x28 and 32x32, respectively, indicating a preference for a 3x3 filter size. This finding is supported by papers 23 and 21, which have larger image sizes, and consequently, showcase better outcomes for larger filter sizes. However, due to the limitations imposed by the selected datasets, this test case design cannot be improved. The only potential avenue for improvement is to test larger filter sizes with a lack of bias using images with larger dimensions and a greater number of epochs, which would provide sufficient time for the optimizer to manage the parameter increment. Paper 21 is the only study that nearly satisfies both of these factors, with a reduced bias towards the 3x3 filter size. Nonetheless, the image size was not sufficiently large to determine whether a larger filter size is significantly superior or only marginally better[12, 27, 28, 26, 37].

The only Study that does not favor small filter size is a pa-

Table 7: Architecture Comparison

| Info / ID/year | Archit. | Dataset | Task | max-param. | Image-size | filter-size | Layers | Estimated papers-published |
|---|---|---|---|---|---|---|---|---|
| paper10 1989 | Lenet [19] | handwritten zip codes | clf. | 60k | 16x16 | 5x5/2x2 | 5 | 45 |
| paper14 2012 | Alexnet [23] | ImageNet | clf. | 60 M | 256×256 | 11x11/5x5/3x3 | 8 | 282 |
| paper11 2013 | ZFnet [20] | ImageNet | clf. | 62M | 224x224 | 7x7/5x5/3x3 | 8 | 5 |
| paper13 2014 | OverFeat [22] | ImageNet/ILSVRC13 | clf./OD | 145M | 221x221 | 11x11/5x5/3x3 | 8 | 7 |
| paper17 2014 | Inception [11] | ILSVRC 2014 | clf./OD | 5M | 224x224 | 3x3/5x5 | 22 | 131 |
| paper2 2015 | VGG16 [33] | ImageNet | clf. | 144M | 224x224 | 3x3 | 16 | 457 |
| paper9 2015 | Resnet [18] | CIFAR-10/ImageNet | clf. | 19.4M | 224x224 | 7x7/3x3 | 18-152 | 1933 |
| paper12 2016 | YOLO [21] | ImageNet/VOC 2012 | clf./OD | 24M | 224×224 | 7x7/3x3 | 9 | 11 |
| paper1 2017 | Mobilenet [32] | ImageNet | clf./OD | 4.9M | 128 to 224 | 3x3 | 28 | 65 |
| paper8 2018 | Densenet [17] | CIFAR-10/CIFAR-100/SVHN/ImageNet | clf. | 15.3M | 224x224 | 7x7/3x3 | 121-264 | 370 |
| paper15 2020 | Efficientnet[24] | ImageNet/CIFAR-10/CIFAR-100/etc. | clf. | 66M | 224/299 | 3x3/5x5 | 18-36 | 237 |
| paper16 2022 | RepLKNet [25] | ImageNet/COCO | clf./OD | 335M | 224/384 | 31x31-29x29 - 27x27-13x13 | 24 | 2 |
| paper24 2023 | SLAKNet [29] | ImageNetADE20K/ PASCAL VOC 2007/COCO | clf./OD | 95M | 224/384 | 51x51-49x49 - 47x47-13*13 | 24 | 1 |

per23 titled "Filter size optimization on a convolutional neural network using FGSA" which uses an optimization algorithm to find the best filter size that enhances the recognition rate, having the biggest image size with an optimizer that is not biased towards any filter size lead that study which is not a full comparison of filter size to be a simple comparison of filter size showing which filter size it better for the task leading bigger filter size to have chance to prove its potential[28].

In both paper16 and paper24 titled "*Scaling Up Your Kernels to 31x31: Revisiting Large Kernel Design in CNNs*" and "*More ConvNets in the 2020s: Scaling up Kernels Beyond 51x51 using Sparsity*" which created model RepLKNet and SLAKNet, the biggest models in terms of bigger filter size use, RepLKNet and SLAKNet showed better results on bigger image size than 224x224 which also supports that image size is indeed a crucial factor to use in filter sizes effects comparison which was not cared enough for in previous research[25, 29].

# 3 Discussion

## 3.1 Architectures Comparison

From Table 7, The growth in the field of computer vision has led to an increase in the maximum number of parameters that can be achieved in architecture, from 2012 to 2015. However, researchers began to lower their standard for parameters, seeking computationally efficient models. This resulted in a widespread belief in the research community that a 3x3 filter size was the best choice, leading to a heavy dependence on this filter size in architectures from 2013 to 2022. This belief persisted until a relatively unpopular research paper titled "*Scaling Up Your Kernels to 31x31: Revisiting Large Kernel Design in CNNs*" emerged, which challenged this common sense by utilizing a 31x31 filter size with its architecture, resulting in an efficient architecture that can scale up to a maximum of 335M parameters while prioritizing performance, scalability, and time complexity. A subsequent research paper titled "*More ConvNets in the 2020s: Scaling up Kernels Beyond 51x51 using Sparsity*" further expanded on this approach, scaling the filter size up to 51x51. These findings suggest that the research community is beginning to take note of the potential of larger filter sizes, though awareness of this phenomenon remains limited among some researchers. Addressing this lack of awareness can facilitate more rapid advancement in the computer vision field.

Also From Table 7, The standard image size of 224x224 has long been utilized by researchers until recent advancements in RepLKNet and SLAKNet have demonstrated the optimal results achieved through larger image sizes. As such, it is our recommendation to encourage the usage of larger image sizes accompanied by corresponding filter size increases to enhance model performance. The increase in image size will mitigate the loss of information resulting from downsampling. The depth of a model has also been a subject of much interest among researchers, with an increasing number of layers requiring a corresponding increase in the use of 3x3 filter sizes. However, the exponential increase in parameters associated with larger filters has previously limited utilization to larger filter sizes such as 5x5 and 7x7. Nonetheless, the recent successes of RepLKNet and SLAKNet have highlighted the potential efficiency of larger filter sizes with fewer layers, thereby demonstrating the promise of utilizing larger filters.

Although the potential for bigger filter size is increasing, few types of research have been conducted about bigger filters in CNN as even if we combined From Table 7 all bigger filter architectures estimated papers published on the same topic will not reach 5% of the number of yearly computer science published papers in 2017 showing how under-explored this topic is[39].

Table 8: Papers: Filter Size Effects Comparison

| ID \ Info | year | Dataset | Filter-size | Image-Size | Image-Size Rank | Use a bigger filter size | Bias-Degree |
|---|---|---|---|---|---|---|---|
| paper18 [13] | 2021 | Mendeley | 3x3/5x5/9x9 | 66x50/90x75/132/100 | 3 | Against | 2 |
| paper19 [26] | 2021 | CFIR10/Fashion-MNIST | 3x3/5x5/7x7 | 32x32/28x28 | 5 | Against | 4 |
| paper20 [14] | 2020 | KTH | 3x3/5x5/7x7 | 96x96 | 4 | Against | 3 |
| paper21 [27] | 2018 | CAMELYON17 | 3x3/5x5/7x7/9x9 | **128x128** | **2** | **Support** | 1 |
| paper22 [12] | 2020 | digits MNIST | 3x3/5x5/7x7 | 28x28 | 6 | Against | 4 |
| paper23 [28] | 2020 | CROPPED YALE | Auto-Selection(1-31) | **640x480** | **1** | **Support** | 0 |

### 3.2 Filter size effects Studies Comparison

We present the effects and results of all previous Filter size effects studies in the CNN domain, we showed that there is bias in the design of test cases. We showed which research is against and which is supporting the use of bigger filter sizes, we also used the factors that favor the 3x3 filter size as a bias degree criteria which has been added to Table 8.

From Table 8, Only two studies appear to support the utilization of larger filter sizes out of a total of six studies; however, upon further examination of the ranking of image sizes within these studies, it becomes apparent that the top two ranked image sizes, corresponding to the largest image sizes utilized in all studies, support the use of bigger filter sizes. This observation may be attributed to the fact that, in smaller image sizes, smaller filter sizes appear like larger filter sizes. Specifically, when the ratio of filter size to image size is calculated for the smallest image sizes (28x28 and 32x32), it is revealed that the smaller filter sizes ratio appears to be smaller than the ratio of largest filter size in paper21 being applied in image size 128x128 which shows that smallest filter size is too big for the image size to get smaller ratio. Notably, two studies with the smallest image sizes exhibit a notable bias towards the 3x3 filter size and did not include a diversity of image sizes in their analyses.

This leaves us with a balanced set of two corroborative studies using a bigger filter size and two contradictory studies. While it may appear that no definitive conclusion can be drawn, it is important to note that one of the non-supportive studies lacked a 7x7 filter size, which has been demonstrated to be superior to a 3x3 filter size in a corroborative study with an image size of 128x128. Moreover, the two non-supportive studies had smaller image sizes than the two corroborative studies, while recent architectures have employed both filter sizes of 31x31 and 51x51 proving the potential of a bigger filter size. Therefore, it can be argued that these two contradictory studies had a bias toward 3x3 filter size. The prevalent belief in the computer science community that favors 3x3 filter size without any empirical evidence is short-sighted and neglects the potential of larger filter sizes. This is evident from the two oldest studies which are closest to unbiased and confirm the importance of the bigger filter sizes showing that we must open that topic and encourage researchers to neglect one of the oldest beliefs in CNN and start researching bigger filter sizes effects.

### 3.3 Opportunities

**Capture larger spatial patterns**

The utilization of larger filter sizes has the potential to capture larger patterns, which is advantageous for tasks that require the perception of global context or larger objects. Larger filter sizes have the potential to yield significant improvements in results within the appropriate context. In order to fully capitalize on these benefits, it is necessary to solve the challenges and limitations associated with larger filters[7].

**Large ERFs**

the Effective Receptive Fields (ERF) is a new term used in deep learning to define the relative importance of each input pixel[40]."Large kernel design significantly increases the Effective Receptive Fields (ERFs)" is quoted from the study titled "*Scaling Up Your Kernels to 31x31: Revisiting Large Kernel Design in CNNs*" that created RepLKNet model which employs 31x31 filter size. As ERF increases, it will lead increase in the number of pixels that determine the activity of a certain neuron in the classification process which will enhance not only the ability of the model to identify bigger objects or classify them but also identify objects that exist in the pixels that are originally inside small ERF leading to improvement in classification, object detection, image segmentation, etc [25].

The receptive field is the portion of the input that

**Handling Bigger Image Better**

The issue of image input size limitations is often overlooked or disregarded by machine learning engineers. While past machine learning engineers' constraints may have forced this issue, new advancements in technology, such as TPUs, have increased image input size limits by over tenfold[41, 42]. However, the exploration and implementation of larger image sizes is still lacking due to limitations and challenges, including difficulty in handling larger image sizes with smaller filter sizes. Despite the benefits of utilizing larger images, such as increased information extraction and reduced pixelation and data loss from downsampling, deep learning users often resort to downsampling as a recommended solution to circumvent issues. Two significant challenges of utilizing larger images include limited vision with 3x3 filters as image size increases and exponential increases in time complexity for model processing[37].

**Stronger Learner**

As technology advances, the utilization of strong learners without the constraints of overfitting and computational power can greatly assist in enhancing deep learning models. The need for more powerful learners opens up vast potential for deep learning models in various applications. Even a slight increase in accuracy, such as 1%, can be appreciated in fields like medicine, where it can assist in diagnosing patients more accurately and potentially save more lives.

### 3.4 Test Case Design Conditions

As we explained in the previous sections, we had a bias in previous studies' experiments we would like to suggest a couple of criteria based on our collection of papers and our understanding that are needed for any filter size effects research in the future to use to give meaning to their experiment while adding over our knowledge and getting away from repeating previous mistakes or repeating work with standard test case design.

Test Case Design Conditions:

- Range of filter size should be between (3-11) at least while preferring bigger upper bound
- Image size must be considered as an important strong factor with a range of (128-512) at least while preferring bigger upper bound
- Model hyperparameters should also be considered as a strong factor including the most important hyperparameter (epoch)
- Test must be concluded on multiple datasets at least 2 datasets to show there is bias to certain data
- Model depth should also be considered as a normal factor yet needs to be tested and taken into account.
- Filter size effects on Effective Receptive Fields(ERFs) will be preferred with the rest of the conditions as it will complete the investigation of filter size effects from all factors.

Those conditions will surely lead to showing more reliable results because by comparing filter sizes as having a bigger range of filters with a higher range of image sizes with at least 2 datasets to test on it while taking into account the time (epoch) that the model needs to converge to better results, all these conditions will guarantee adding over the overall knowledge in filter size effects research in the future.

## 4 Conclusion

We are using 3x3 filters for our common belief of bigger filter limitations and this is also why we prefer to use smaller image sizes because we can not use bigger filter sizes as there is a limit for which resolution can get better results with 3x3 filter size. We present the challenges and limitations that come with the use of bigger filter sizes in the form of problems that need to be solved to access bigger filter sizes benefits. we also show most biggest models in filter size factor which could reach 51x51 filter size surprisingly. the analysis demonstrates the bias towards smaller filter sizes in test case design of filter size effects studies, which can lead to favoring the 3x3 filter size. Additionally, the potential benefits and opportunities that larger filter sizes offer are discussed, such as the ability to capture larger spatial patterns, handle bigger image sizes better, and have bigger ERFs, resulting in a stronger learner. Finally, we suggest some conditions and criteria for conducting the next filter size effects studies to lower the bias towards 3x3 filter size and add over our knowledge without the repetition of work. Solutions in various ways are needed to neutralize the challenges that the use of larger filters requires to facilitate their implementation in image-related tasks in deep learning and open a new era of bigger filter research.

## References


[1] Ahmed Fathi Bekhit. "Introduction to Computer Vision". In: *Computer Vision and Augmented Reality in iOS: OpenCV and ARKit Applications*. Berkeley, CA: Apress, 2022, pp. 1–20. ISBN: 978-1-4842-7462-0. DOI: 10.1007/978-1-4842-7462-0_1. URL: https://doi.org/10.1007/978-1-4842-7462-0_1.

[2] Iuliia Kotseruba, Manos Papagelis, and John K. Tsotsos. "Industry and Academic Research in Computer Vision". In: *CoRR* abs/2107.04902 (2021). arXiv: 2107.04902. URL: https://arxiv.org/abs/2107.04902.

[3] Qing Wu et al. "The application of deep learning in computer vision". In: *2017 Chinese Automation Congress (CAC)*. 2017, pp. 6522–6527. DOI: 10.1109/CAC.2017.8243952.

[4] Junyi Chai et al. "Deep learning in computer vision: A critical review of emerging techniques and application scenarios". In: *Machine Learning with Applications* 6 (2021), p. 100134. ISSN: 2666-8270. DOI: https://doi.org/10.1016/j.mlwa.2021.100134. URL: https://www.sciencedirect.com/science/article/pii/S2666827021000670.

[5] Christian Szegedy et al. *Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning*. 2016. arXiv: 1602.07261 [cs.CV].

[6] Asifullah Khan et al. "A survey of the recent architectures of deep convolutional neural networks". In: *Artificial Intelligence Review* 53.8 (Apr. 2020), pp. 5455–5516. DOI: 10.1007/s10462-020-09825-6. URL: https://doi.org/10.1007%2Fs10462-020-09825-6.

[7] Laith Alzubaidi et al. "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions". en. In: *J. Big Data* 8.1 (Mar. 2021), p. 53.

[8] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. "Understanding of a convolutional neural network". In: *2017 International Conference on Engineering and Technology (ICET)*. Antalya: IEEE, Aug. 2017.

[9] Vincent Dumoulin and Francesco Visin. *A guide to convolution arithmetic for deep learning*. 2018. arXiv: 1603.07285 [stat.ML].

[10] Christian Szegedy et al. "Rethinking the Inception Architecture for Computer Vision". In: *CoRR* abs/1512.00567 (2015). arXiv: 1512.00567. URL: http://arxiv.org/abs/1512.00567.

[11] Christian Szegedy et al. "Going Deeper with Convolutions". In: *CoRR* abs/1409.4842 (2014). arXiv: 1409.4842. URL: http://arxiv.org/abs/1409.4842.

[12] Owais Khanday and Samad Dadvandipour. "Convolution Neural Networks and Impact of Filter Sizes on Image Classification". In: *Multidiszciplináris Tudományok* 10 (Apr. 2020), pp. 55–60. DOI: 10.35925/j.multi.2020.1.7.

[13] Yunus Camgözlü and Yakup Kutlu. "Analysis of Filter Size Effect In Deep Learning". In: *CoRR* abs/2101.01115 (2021). arXiv: 2101.01115. URL: https://arxiv.org/abs/2101.01115.

[14] Wafaa Ahmed and Abdulamir Karim. "The Impact of Filter Size and Number of Filters on Classification Accuracy in CNN". In: (Apr. 2020), pp. 88–93. DOI: 10.1109/CSASE48920.2020.9142089.

[15] Osval Antonio Montesinos López, Abelardo Montesinos López, and Jose Crossa. "Overfitting, model tuning, and evaluation of prediction performance". In: Cham: Springer International Publishing, 2022, pp. 109–139.

[16] Matias Valdenegro-Toro and Matthia Sabatelli. "Machine Learning students overfit to overfitting". In: (2022).

[17] Gao Huang et al. *Densely Connected Convolutional Networks*. 2018. arXiv: 1608.06993 [cs.CV].

[18] Kaiming He et al. *Deep Residual Learning for Image Recognition*. 2015. arXiv: 1512.03385 [cs.CV].

[19] Y LeCun et al. "Backpropagation applied to handwritten zip code recognition". en. In: *Neural Comput.* 1.4 (Dec. 1989), pp. 541–551.

[20] Matthew D Zeiler and Rob Fergus. *Visualizing and Understanding Convolutional Networks*. 2013. arXiv: 1311.2901 [cs.CV].

[21] Joseph Redmon et al. *You Only Look Once: Unified, Real-Time Object Detection*. 2016. arXiv: 1506.02640 [cs.CV].

[22] Pierre Sermanet et al. *OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks*. 2014. arXiv: 1312.6229 [cs.CV].

[23] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems*. Ed. by F. Pereira et al. Vol. 25. Curran Associates, Inc., 2012. URL: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

[24] Mingxing Tan and Quoc V. Le. *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. 2020. arXiv: 1905.11946 [cs.LG].

[25] Xiaohan Ding et al. *Scaling Up Your Kernels to 31x31: Revisiting Large Kernel Design in CNNs*. 2022. arXiv: 2203.06717 [cs.CV].

[26] Owais Khanday, Samad Dadvandipour, and Mohd Aaqib Lone. "Effect of filter sizes on image classification in CNN: a case study on CFIR10 and Fashion-MNIST datasets". In: *IAES International Journal of Artificial Intelligence (IJ-AI)* 10 (Dec. 2021), p. 872. DOI: 10.11591/ijai.v10.i4.pp872-878.

[27] Saban Ozturk et al. "Convolution Kernel Size Effect on Convolutional Neural Network in Histopathological Image Processing Applications". In: *2018 International Symposium on Fundamentals of Electrical Engineering (ISFEE)*. IEEE, Nov. 2018. DOI: 10.1109/isfee.2018.8742484. URL: https://doi.org/10.1109/isfee.2018.8742484.

[28] Yutzil Poma et al. "Filter size optimization on a convolutional neural network using FGSA". In: *Intuitionistic and Type-2 Fuzzy Logic Enhancements in Neural and Optimization Algorithms: Theory and Applications*. Studies in computational intelligence. Cham: Springer International Publishing, 2020, pp. 391–403.

[29] Shiwei Liu et al. *More ConvNets in the 2020s: Scaling up Kernels Beyond 51x51 using Sparsity*. 2023. arXiv: 2207.03620 [cs.CV].

[30] Joseph Redmon and Ali Farhadi. *YOLO9000: Better, Faster, Stronger*. 2016. arXiv: 1612.08242 [cs.CV].

[31] Joseph Redmon and Ali Farhadi. *YOLOv3: An Incremental Improvement*. 2018. arXiv: 1804.02767 [cs.CV].

[32] Andrew G. Howard et al. *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. 2017. arXiv: 1704.04861 [cs.CV].

[33] Karen Simonyan and Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015. arXiv: 1409.1556 [cs.CV].

[34] Forrest N. Iandola et al. *SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and ¡0.5MB model size*. 2016. arXiv: 1602.07360 [cs.CV].

[35] Dongyoon Han, Jiwhan Kim, and Junmo Kim. *Deep Pyramidal Residual Networks*. 2017. arXiv: 1610.02915 [cs.CV].

[36] Zeming Li et al. *DetNet: A Backbone network for Object Detection*. 2018. arXiv: 1804.06215 [cs.CV].

[37] Mats L Richter et al. "(input) size matters for CNN classifiers". In: *Lecture Notes in Computer Science*. Lecture notes in computer science. Cham: Springer International Publishing, 2021, pp. 133–144.

[38] Sergey Ioffe and Christian Szegedy. "Batch Normalization: Accelerating deep network training by reducing internal covariate shift". In: (2015).

[39] Dalibor Fiala and Gabriel Tutoky. "Computer science papers in web of science: A bibliometric analysis". In: *Publications* 5.4 (Sept. 2017), p. 23.

[40] Wenjie Luo et al. *Understanding the Effective Receptive Field in Deep Convolutional Neural Networks*. 2017. arXiv: 1701.04128 [cs.CV].

[41] Norman P. Jouppi et al. *In-Datacenter Performance Analysis of a Tensor Processing Unit*. 2017. DOI: 10.48550/ARXIV.1704.04760. URL: https://arxiv.org/abs/1704.04760.

[42] David Paper. "Introduction to Tensor Processing Units". In: *State-of-the-Art Deep Learning Models in TensorFlow*. Apress, 2021, pp. 127–152. DOI: 10.1007/978-1-4842-7341-8_5. URL: https://doi.org/10.1007/978-1-4842-7341-8_5.