Deep Learning Final Project: Analytical Paper on

A Survey of the Recent Architectures of Deep Convolutional Neural Networks

Asifullah Khan1, Anabia Sohail1, Umme Zahoora, and Aqsa Saeed Qureshi1

Professor Curro

Hyomin Seo

ECE 472 Deep Learning

2020 December

Table of Content

Introduction

The assignment is to write an analytical paper and a quiz for a choice of the recent Deep Learning paper. I have chosen the paper that discusses the most prominent CNN architectures, as a detailed and extensive response to a question that sums up the semester of learning: - 'final quiz: Discuss and analyze the building design of CNN'.

Building Design for CNNs

The components of a convolutional neural organization, for example, convolutional and pooling layers, are generally clear to comprehend.

The problematic piece of utilizing convolutional neural organizations is how to configuration model designs that best use these essential components.

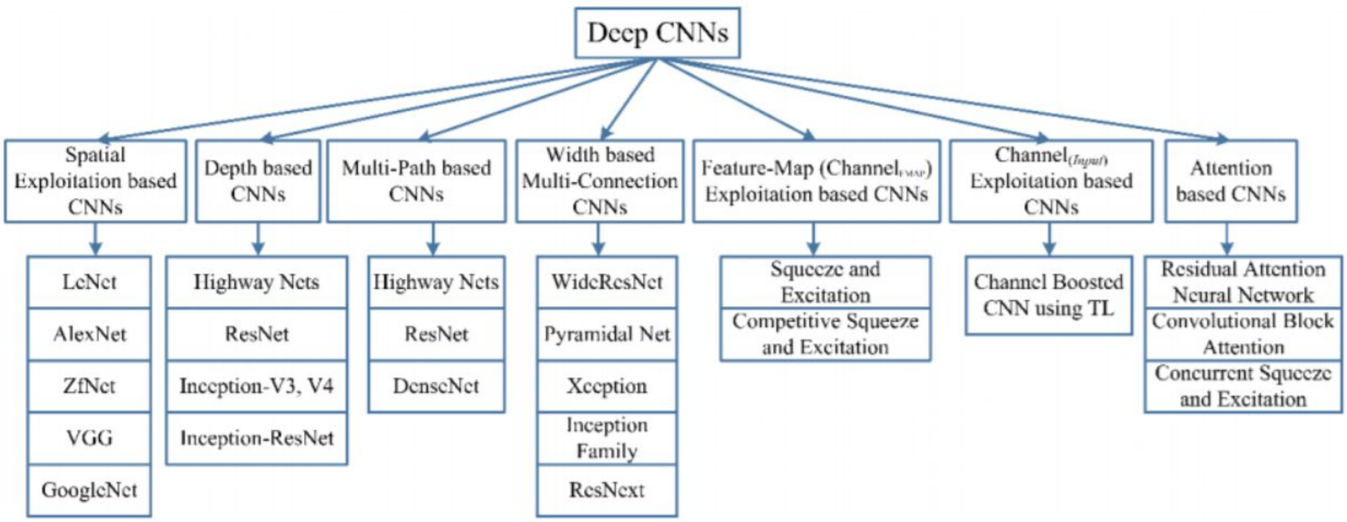
A helpful way to deal with figuring out how to plan compelling convolutional neural organization models is to consider fruitful applications. This is especially direct due to the superb examination and utilization of CNNs from 2012 to 2016 for the ImageNet Large Scale Visual Recognition Challenge, or ILSVRC. This test brought about both the quick progression in the cutting edge for troublesome PC vision assignments and the advancement of general developments in the design of convolutional neural organization models.

We will start with the LeNet-5 that is frequently portrayed as the principal effective and significant use of CNNs before the ILSVRC. At that point, take a gander at four diverse winning compositional advancements for the convolutional neural organization created for the ILSVRC, specifically, AlexNet, VGG, Inception, and ResNet.

By understanding these achievement models and their engineering or structural advancements from a significant level, you will create both a thankfulness for the utilization of these building components in present-day uses of CNN in PC vision and have the option to distinguish and pick engineering components that might be valuable in the plan of your models.

Architectural innovations in CNN

The diagram shows the overall content of the paper.



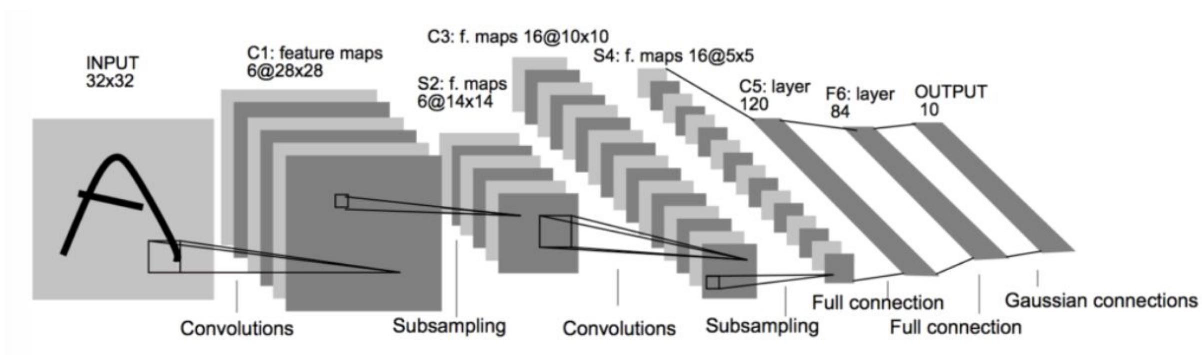
Spatial Exploitation based CNNs

In early 2000, the role of the various size filter in neural network was getting attention, as different sizes of filters encapsulate different levels of granularity. Consequently, researchers exploited spatial filters to improve performance and explored the relation of a spatial filter with the learning of the network. The two prominent example of such CNN are AlexNet and Inception.

AlexNet

These models incorporate AlexNet, the primary colossal scope network sent to beat traditional PC vision strategies for a vast scope vision challenge; the VGG network, which utilizes various rehashing squares of components; the organization in organization (NiN) which convolves entire neural organizations fix astute over sources of info; GoogLeNet, which uses networks with equal links; leftover organizations (ResNet), which remain the most famous off-the-rack design in PC vision; and thickly associated networks (DenseNet), which are costly to process yet have set some ongoing benchmarks.

Since we comprehend the nuts and bolts of wiring together CNNs, we will take you through a visit through present-day CNN models. In this part, each segment relates to a critical CNN design that was eventually (or as of now) the base model. After that, many explorations extended, and conveyed frameworks were constructed. Every one of these organizations was quickly a predominant design. Many layers), execution can fluctuate uncontrollably across structures and hyperparameter decisions. The neural organizations portrayed in this part are the. The square's yield is joined with the contribution to the square, for example, the alternate route association. An extended form of the info is utilized through 1×1 if the contribution to the square is diverse to the yield of the square supposed 1×1 convolutions. These are alluded to as extended alternate route associations, contrasted with the unweighted or character alternate way associations.



The creators start with what they call a plain organization, which is a VGG-motivated profound convolutional neural organization with narrow channels (3×3), gathered convolutional layers followed with no pooling in the middle of, and a regular pooling toward the finish of the component identifier part of the model before the completely associated yield layer with a softmax enactment work.

The plain organization is adjusted to turn into a remaining organization by adding easy route associations to characterize leftover squares. Regularly, the alternate way association's contribution is a similar size as the yield of the remaining square.

The picture presented from the paper and from left to right looks at the engineering of a VGG model, a plain convolutional model, and a variant of the plain convolutional with leftover modules called a remaining organization

VGG

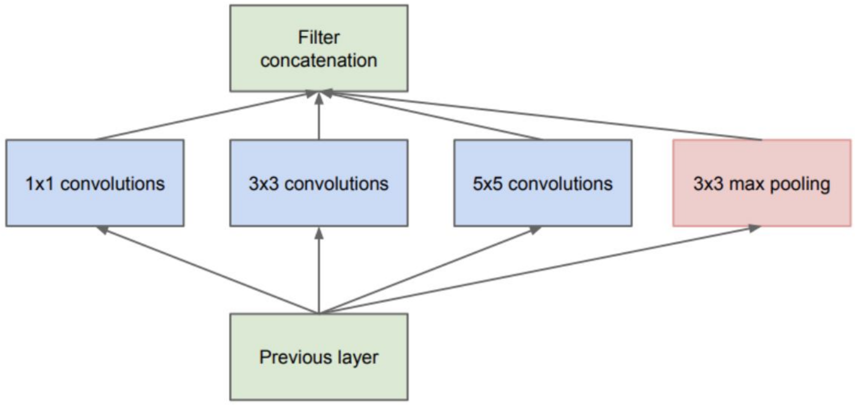
Max pooling layers are utilized after most, yet not all, convolutional layers, gaining from the model in AlexNet. Yet, all pooling is performed with the size 2×2 and the very step, which also has gotten a valid norm. In particular, the VGG networks use instances of two, three, and even four convolutional layers stacked together before a maximum pooling layer is utilized. The reasoning was that stacked convolutional layers with more modest channels estimated the impact of one convolutional layer with a bigger measured channel; for example, three stacked convolutional layers with 3×3 channels approximates one convolutional layer with a 7×7 channel.

Inception / GoogLeNet

The critical development of the commencement models is known as the origin module. This is a square of equal convolutional layers with various measured channels (for example, 1×1, 3×3, 5×5) and a 3×3 max-pooling layer, the aftereffects of which are then linked. The following is an illustration of the origin module taken from the paper.

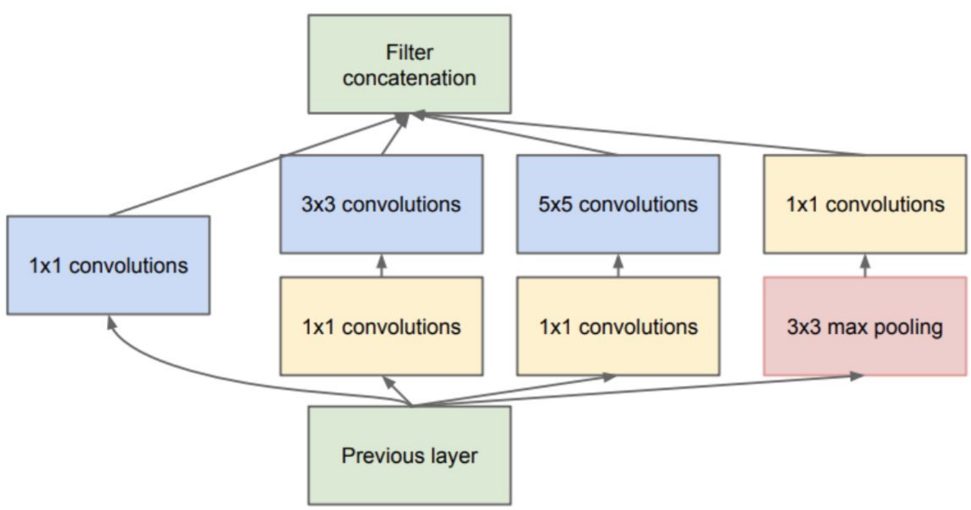
An issue with an innocent usage of the initiation model is that the quantity of channels (profundity or channels) starts to develop quickly, particularly when commencement modules are stacked.

Performing convolutions with bigger channel sizes (for example, 3 and 5) can be computationally costly on an enormous number of channels. To address this, 1×1 convolutional layers are utilized to diminish the number of channels in the original model, explicitly before the 3×3 and 5×5 convolutional layers and after the pooling layer. The picture underneath taken from the paper shows this change to the origin module.



A second significant plan choice in the commencement model interfaced the yield at various focuses in the model. This was accomplished by making little off-shoot yield networks from the fundamental organization prepared to make a forecast. The goal was to give an extra mistake signal from the characterization task at various purposes of the profound model to address the evaporating inclinations issue. These little yield networks were then eliminated in the wake of preparing. There do exists other limitations; the main drawback of the GoogleNet was its heterogeneous topology that needs to be customized from module to module. Another limitation was a representation bottleneck that drastically reduces the feature space in the next layer and thus sometimes may lead to loss of useful information.

Beneath shows a pivoted adaptation (left-to-appropriate for contribution to-yield) of the GoogLeNet model's design, taken from the paper utilizing the Inception modules from the contribution on the left to the yield characterization on the privilege and the two extra yield networks that were used just during preparing.



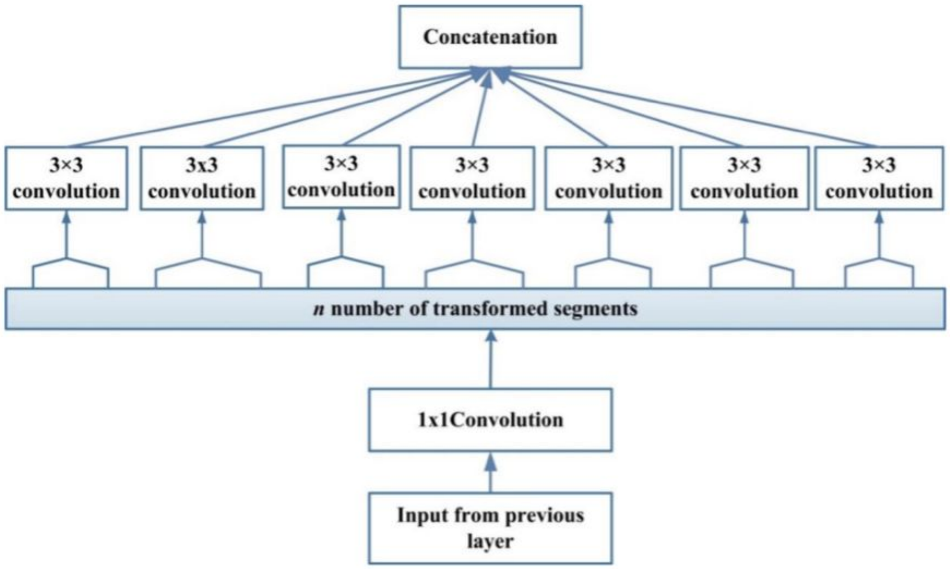
Width based Multi-Connection CNNs

During 2012-2015, the attention was chiefly on abusing the intensity of profundity, alongside the viability of multi-pass administrative associations in organization regularization. Notwithstanding, Kawaguchi et al. revealed that the organization's width is significant (Kawaguchi et al., 2019). Multilayer perceptron picked up the benefit of planning complex capacities over perceptron by utilizing numerous preparing units inside a layer. This recommends that width is a fundamental boundary in characterizing standards of learning along with profundity. Lu et al. (2017) and Hanin and Sellke (2017) have, as of late, indicated that NNs with ReLU initiation work must be wide enough to hold general estimate property alongside an expansion top to bottom (Hanin and Sellke 2017). Besides, a class of persistent capacities on a smaller set can't be subjectively approximated by a discretionarily profound network if the organization's greatest width isn't bigger than the information measurement (Lu et al. 2017b; Nguyen et al. 2018). Even though. One significant issue connected with profound structures is that a few layers or handling learn valuable highlights. To handle this issue, the focal point of examination moved from profound and limited engineering towards dainty and wide designs.

Xception

Xception, as you can imagine, is the extreme version on inception architecture. Xception modified the original inception block by making it wider and replacing the different spatial dimensions (1x1, 5x5, 3x3) with a single dimension (3x3) followed by a 1x1 convolution to regulate computational complexity.

Xception involves each feature-map across spatial axes, perform cross-channel correlation by pointwise 1by 1 convolution, 1x1 convolutions. Although the transformation strategy adopted by Xception does not reduce the number of parameters, it makes learning more efficient and results in improved performance.



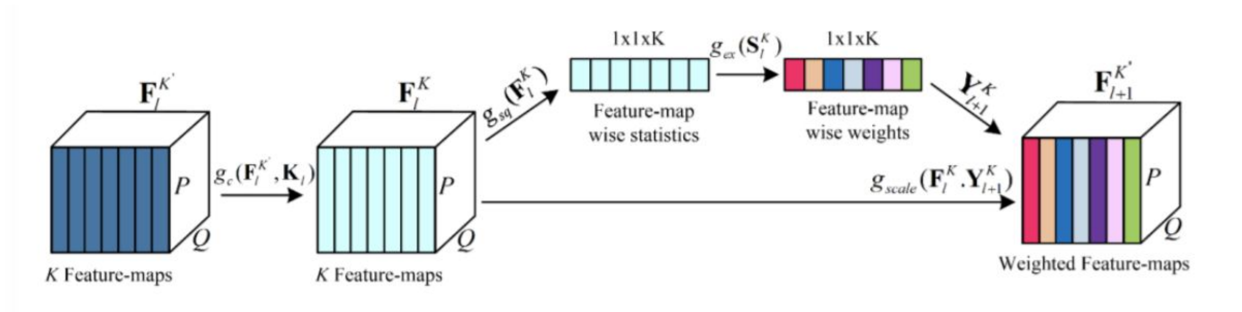
Xception building block and its *n* sets of transformation.

Feature-Map Exploitation based CNNs

Around 2015, CNN became popular for MultiView tasks because of its automatic feature extraction ability. However, researchers noticed that enormous feature sets may create an effect of noise and thus lead to overfitting of the network. This suggested that selection of feature-maps can play an important role in improving the generalization of the network, apart from network engineering itself. One exemplary network of this category is Squeeze and Excitation Network (SE Network)

Squeeze and Excitation Network.

Crush and Excitation Network (SE-Network) was accounted for by Hu et al. (Hu et al. 2018a). They proposed another square to determine highlight maps (regularly known as channels) important to protest segregation. This new square was named as SE-block, which smothers the less significant component maps, yet gives high weightage to the class indicating highlight maps. SE-Network announced a record decline in blunder on the ImageNet dataset. SE-block is a preparing unit planned conventionally and can be included in any CNN engineering before the convolution layer. The working of this square comprises two activities; crush and excitation.

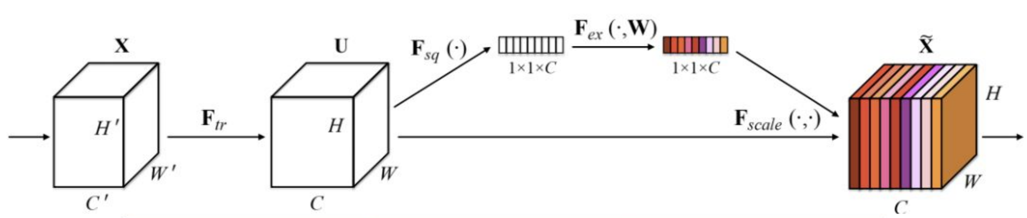


$$F_{tr}: \ Mapping\ X \in R^{H' * W' * C'}\ to\ U \ \in \ R^{H * W * C}$$
$$F_{sq}\ : \ Squeeze\ global\ spatial\ information\ into\ a\ channel\ descriptor$$
$$F_{es}\ : \ Fully\ capture\ channel - wise\ dependencies$$
$$F_{scale}\ : \ Channel - wise\ multiplication$$

The first is the Squeeze operation, which performs feature compression along the spatial dimension, transforming each two-dimensional feature channel into a real number. This real number has a global receptive field to some extent, and the output dimension and the input feature channel number match. It characterizes the global distribution of responses on feature channels, and allows layers close to the input to obtain global receptive fields,

This is followed by the Excitation operation. A weight is generated for each feature channel by the parameter w, where the parameter w is learned to explicitly model the correlation between the feature channels.

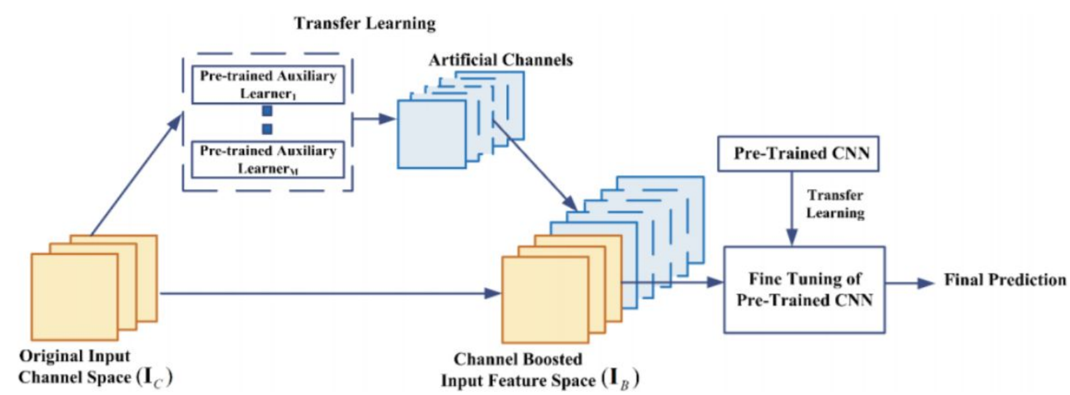
Finally, a Reweight operation treats the weight of the output of Excitation as the importance of each feature channel after the feature selection, and then weights the previous feature by multiplication, completing the original pair on the channel dimension.



Squeeze	Excitation	Scale
Shrinking feature maps $U \in R^{H * W * C}$ Through spatial dimension (w X h) Global distribution of channel-wise responses	Learning $W \in R^{C * C}$ to explicitly model channel association Gating mechanism to produce channel-wise weights.	Reweighting the feature maps $U \in R^{H * W * C}$

Channel Boosted CNN using TL

CNN design named as Channel supported CNN (CBCNN)based on boosting the quantity of info channels for improving the illustrative limit of the organization (Khan et al. 2018a). The Block outline of CB-CNN appears in Fig. 11. Channel boosting is performed by misleadingly making additional channels (known as helper channels) through assistant profound generative models and afterward abusing it through the profound discriminative models.



Basic architecture of CB-CNN showing the deep auxiliary learners for creating artificial channels.

Conclusion

CNN has gained momentous ground, particularly in picture preparing and vision-related assignments, and has along these lines resuscitated scientists' interest in ANNs. In this unique circumstance, a few exploration works have been completed to improve CNN's exhibition on such errands. The progressions in CNNs can be classified in an unexpected way, including enactment, misfortune work, advancement, regularization, learning calculations, and developments in engineering. This paper surveys headway in the CNN models, particularly dependent on the plan examples of the handling units, and subsequently has proposed the scientific categorization for ongoing CNN designs. Notwithstanding the classification of CNNs into various classes, this paper additionally covers the historical backdrop of CNNs, its applications, challenges, and future bearings.

Reference

1. Asifullah Khan, Anabia Sohail1, (2019) A Survey of the Recent Architectures of Deep Convolutional Neural Networks Published in Artificial Intelligence Review, DOI: https://doi.org/10.1007/s10462-020-09825
2. Abbas Q, Ibrahim MEA, Jaffar MA (2019) A comprehensive review of recent advances on deep vision systems. Artif Intell Rev 52:39–76.DOI: 10.1007/s10462-018-9633-3
3. Abdel-Hamid O, Deng L, Yu D (2013) Exploring convolutional neural network structures and optimization techniques for speech recognition. In: Interspeech. pp 1173–1175
4. Amer M, Maul T (2019) A review of modularization techniques in artificial neural networks. Artif Intell Rev 52:527–561. DOI: 10.1007/s10462-019-09706-7
5. Aurisano A, Radovic A, Rocco D, et al (2016) A convolutional neural network neutrino event classifier. J Instrum. doi: 10.1088/1748-0221/11/09/P09001
6. Batmaz Z, Yurekli A, Bilge A, Kaleli C (2019) A review on deep learning for recommender systems: challenges and remedies. Artif Intell Rev 52:1–37. DOI: 10.1007/s10462-018-9654- y
7. Bengio Y (2009) Learning Deep Architectures for AI. Found Trends® Mach Learn 2:1–127. DOI: 10.1561/2200000006
8. Berg A, Deng J, Fei-Fei L (2010) Large scale visual recognition challenge 2010 Bettoni M, Urgese G, Kobayashi Y, et al (2017) A Convolutional Neural Network Fully Implemented on FPGA for Embedded Platforms. 49–52. DOI: 10.1109/NGCAS.2017.16
9. Bhunia AK, Konwer A, Bhunia AK, et al (2019) Script identification in natural scene image and video frames using an attention-based Convolutional-LSTM network. Pattern Recognit 85:172–184
10. *SENet (Squeeze-and-Excitation Networks) Network Detailed - Programmer Sought, www.programmersought.com/article/5144760128/.*
11. *Inception(GoogLeNet) 리뷰, kangbk0120.github.io/articles/2018-01/inception-googlenet-review.*
12. *Fung, Vincent. “An Overview of ResNet and Its Variants.” Medium, Towards Data Science, 17 July 2017, towardsdatascience.com/an-overview-of-resnet-and-its-variants-5281e2f56035.*