How to Cite:
Kumar, T. V., & Latha, R. (2022). A generalized approach of convolutional and pooling layer in image processing using wavelet CNN. *International Journal of Health Sciences*, *6*(S2), 4568–4574. https://doi.org/10.53730/ijhs.v6nS2.6080

A generalized approach of convolutional and pooling layer in image processing using wavelet CNN

T. Vinoth Kumar
Research Scholar, Computer Science, St. Peter's Institute of Higher Education & Research, Chennai, Tamil Nadu, India
Email: t.lourdevinothkumar@gmail.com

R. Latha
Prof & Head, Department of Computer Science and Applications, St. Peter's Institute of Higher Education & Research, Chennai, Tamil Nadu, India

***Abstract*---**Spatial and spectral approaches area unit two major approaches for image processing tasks like and beholding. Among several such algorithms, convolutional neural networks (CNNs) have recently achieved significant performance improvement in several difficult tasks. CNNs enable the nation to utilize spectral data that is usually lost in typical CNNs however helpful in most image processing tasks. We tend to evaluate the sensitivity performance of Wavelet CNNs on texture classification and image annotation. The experiments show that Wavelet CNNs can do higher accuracy in each task than existing models, whereas having significantly fewer parameters than typical CNNs.

***Keywords*---**CNNs, pooling layer, image processing, using wavelet.

Introduction

CNNs primarily replaced standard hand woven descriptors like the Bag of Visual Words (BoVW) thanks to the superior performance in numerous tasks. The first specification has been extended to deeper architectures since then. One such design is Residual Networks (ResNets) that build deeper networks easier to coach by introducing crosscut connections; connections that skip a number of layers and perform identity mappings. Galvanized by ResNets, numerous networks with crosscut connections are projects.

***Corresponding author**: Kumar, T. V.; Email: t.lourdevinothkumar@gmail.com*
Manuscript submitted: 27 Feb 2022, Manuscript revised: 18 March 2022, Accepted for publication: 05 April 2022

Literature Review

Cimpoi et.al (2015) in contestable that a CNN (Cimpoi 2015) together with Fisher Vectors (FV-CNN) are able to do far better accuracy than employing a CNN alone. Their model uses a pre-trained CNN to extract image options and this CNN half isn't trained with existing datasets. Lin et.al (2017) achieved a motivating improvement in one-grained visual recognition (Lin 2017)by substitution the totally connected layers with additive pooling. The dimension of the encoded options within the additive model is usually higher than 250,000 and it is difficult to catch.

Gao et.al (2016) projected compact by linear pooling (Gao 2016) that reduces the quantity of parameters of additive pooling by ninetieth whereas maintaining its performance. Despite this need reduction within the range of parameters, genetic from standard CNNs, these models still have an outsizes range of trainable parameters that creates that creates to coach in follow. Andrearczyk et al.(2016) projected texture CNN (T-CNN) that could be a CNN (Whelan 2016)specialized for texture classification. T-CNN uses a completely unique energy layer within which every feature map is solely pooled by shrewd the typical of its activated output

Wavelet Convolutional Neural Network

Even with crosscut connections, however, deeper networks still have a retardant that data regarding the input and gradient will quickly vanish as they propagate through the networks. To deal with this drawback, Dense Convolutional Network (DenseNet) more adds crosscut connections that connect every layer with all its previous layers. Whereas these networks have achieved spectacular results on several PC vision tasks, they need machine resources as a result they need an outsized range of parameters. We have a tendency to design our network once DenseNet, however the mix with a multi-resolution analysis permits USA to USA scale back the quantity of parameters compared to standard networks. Many recent works use CNNs as a feature extractor and a BoVW approach as pooling and coding rather than the totally connected layers.

Spectral approaches rework pictures into the frequency domain employing a set of spacialspacial. The statistics of the spectral data at completely different scales and orientations image options. This approach has been well studied in image process and achieved sensible results. Feature extraction within the frequency domain has a plus. A space will be simply created selective by enhancing bound frequencies, whereas suppressing the others. This express choice of band frequencies is difficult to regulate in CNNs. Whereas CNNs square measure, understand to be universal approximates, in follow, it's unclear whether or not CNNs will learn to perform spectral analyses with accessible datasets. Instead of relying CNNs to be told activity spectral analysis, we have a tendency to propose to directly integrate spectral approaches into CNNs, significantly supported a multi-resolution analysis mistreatment Wavelet rework. Our experiments show that a CNN with a lot of parameters cannot be trained to become cherished our model with accessible datasets in follow.

Implementation

Generalized approach of the Convolutional Layer – Wavelet CNN

Given an input vector with n components x = $(x_0, x_1, . . . , x_{n-1}) € R^n$, R be the Convolutional Layer.

A convolution layer outputs a vector of the same number of components

$$y = (y_0, y_1, . . . , y_{n-1}) € R^{n.}$$

$$y_i = \sum_{j \in N_i} w_j x_j \qquad \rightarrow (1)$$

Where N_i is a set of indices of neighbors at x_j,
w_j is a weight,
Notational Convention in CNN, Consider that w_j be the bias by having constant input from equation (1)
By Sharing the parameters, each layer defines the weight of $w_j$over i, equation 1 can be rewritten as y_i is a weighted sum of neighbors $\sum_{j \in N_i} w_j x_j$ plus constant.

$$y = X * W \qquad \rightarrow (2)$$

Where w = $(w_0, w_1, . . . , w_{o-1}) € R^o$

Generalized approach of the Pooling Layer – Wavelet CNN

Pooling layer will be followed by the convolutional Layer. This layer to allow the connection with multiresolution analysis. The Average pooling output vector, referred as

Input x € R^n, average pooling outputs a vector of a fewer components y € R^m as

$$y_j = \frac{1}{p} \sum_{k=0}^{p-1} x_{pj+k} \qquad \rightarrow (3)$$

Where p defines the support of pooling and m=n/p.

$$y = (X * P) \downarrow p \qquad \rightarrow (4)$$

p = (1/p, . . . , 1/p) €R^p represents the average filter. Average pooling mathematically involves convolution p followed by down sampling.

Generalized approach of the Convolution Layer and Pooling Layer – Wavelet CNN

Equation (2) and equation (4) can be combined into the generalized form of a down sampling and convolution neural network as

$$y = (\mathrm{X} * k) \downarrow p \qquad \rightarrow (5)$$

Where k = w with p = 1 (convolution in Equation 2),
k = p with p > 1 (pooling in Equation 4),
k = w*p with p > 1 (convolution followed by pooling).

Equation 5 referred to multiresolution analysis.

Let us consider convolution followed by pooling with p = 2 and a pair of convolution kernels k_l and k_h:

$$x_l = (\mathrm{X} * k_l) \downarrow 2 \qquad \rightarrow (6)$$
$$x_h = (\mathrm{X} * k_h) \downarrow 2$$

By defining $x_{l,0}$ = x, a multiresolution analysis performs a hierarchical decomposition of $x_{l,t}$ into $x_{l,t+1}$ and $x_{h,t+1}$ by repeatedly applying Equation 6 with different $k_{l,t}$ and $k_{h,t}$ at each t:

$$\mathrm{X}_{l.t+1} = (\mathrm{X}_{l,t} * k_{l,t}) \downarrow 2 \qquad \rightarrow (7)$$
$$\mathrm{X}_{h.t+1} = (\mathrm{X}_{l,t} * k_{h,t}) \downarrow 2$$

The number of applications t is called a level in a multiresolution analysis.
CNNs essentially discard $x_{h,t}$ entirely and use only one set of kernels k_t:

$$\mathrm{X}_{l.t+1} = (\mathrm{X}_{l,t} * k_t) \downarrow 2 \qquad \rightarrow (8)$$

Therefore, CNNs can seen as a limited form of a multiresolution analysis.

Experiments

Provide the details of 2 applications of CNNs. We have a tendency to applyWavelet CNNs to texture classification to confirm that Wavelet CNNs will capture little options of pictures. We have a tendency to additionally investigate, however CNNs perform on nature pictures in a picture annotation task. Texture category could be a difficult drawback since texture soften vary tons among an equivalent class, due to changes in viewpoints, scales, lighting configurations, etc. Additionally, textures sometimes don't contain enough data concerning the form of objects that square measure informative |to completely differentiate to tell apart, different objects in image classification tasks. Thanks to, even the newest approaches supported convolution neural networks achieved a restricted success, in comparison to the r^th tasks like image classification. This ends up in one price for every feature map, the same as Associate in nursing energy response to a bank. This approach doesn't accurately, however, its straightforward design reduces the quantity of parameters.

Result

Four experiments, we used two publicly available texture datasets: kth-tips2-b and DTD. The kth-tips2-b dataset contains 11 classes of 432 texture images. Each class consists of four samples and each sample has 108 images. Each sample is used for training once while the remaining three samples are used for testing. The results for kth-tips2-b are shown as the mean and the standard deviation over the four splits. The DTD dataset contains 47 classes of 120 images “in the wild” which means that the images are collected in uncontrolled conditions. This dataset includes 10 available annotated splits with 40 training images, 40 validation images, and 40 testing images for each class. The results for DTD are averaged over the 10 splits. We processed the images in each dataset by global contrast normalization. We calculated the accuracy as percentage of images that are correctly labeled which is a common metric in texture classification**.** Consider two tables, one with level and other with comparison as shown in Table 1 and table 2 below with appropriate graph

	AlexNet	T-CNN	Wavelet CNN
kth tip2-b	47.6	49.4	65.4
DTD	22.8	29.6	35.6

Table 1: Classification with levels

The appropriate graphs are generated from the above table as shown in Fig. 1

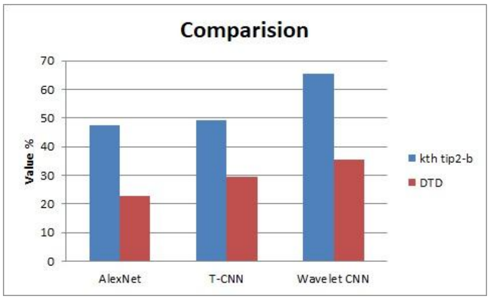

Fig.1 Comparison

Classification with different level as shown below

	2-level	3-level	4-level	5-level
kth tip2-b	60.12	61.23	62.35	65.89
DTD	34.2	35.6	35.71	35.75

Table 2. Classification at various levels

From the table 2, the following fig.2 has been generated as shown below

Fig 2. Classification Level

Conclusions

We concluded a unique CNN design which contains a spectral analysis into CNNs. We tend to show a way to develop convolution and pooling layers in CNNs into a generalized kind of kind of down sampling. This reformulation shows, however typical CNNs perform a restricted version of multi-resolution analysis that then permits North American nation to integrate multi-resolution analysis into CNNs as one model referred to as rippling CNNs. We tend to incontestable that our model achieves higher accuracy for texture classification and image annotation with a smaller variety of trainable parameters than existing models.

Reference

1. Cimpoi, S. Maji, and A. Vedaldi. "Deep filter banks for texture recognition and segmentation." In IEEE Conference on Computer Vision and Pattern Recognition , 2015.
2. Gao, O. Beijbom, N. Zhang, and T. Darrell. " Compact bilinear pooling." In IEEE Conference on Computer Vision and Pattern Recognition , 2016: 317-326.
3. Lin, A. RoyChowdhury, and S. Maji. "Bilinear cnns for fine-grained visual recognition." In Transactions of Pattern Analysis and Machine Intelligence, 2017.
4. Whelan, Andrearczyk and P. F. "Using filter banks in convolutional neural networks for texture classification." Pattern Recognition Letters, 2016: 63-69.
5. S.Arivazhagan,T.G.SubashKumar,andL.Ganesan. Texture classification using curvelet transform. International Journal of Wavelets, Multiresolution and Information Processing, 05(03):451–464, 2007.
6. Y. Boureau, J. Ponce, and Y. Lecun. A theoretical analysis of feature pooling in visual recognition. In International ConferenceonMachineLearning(ICML-10),pages111–118. Omnipress, 2010.
7. K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. Return of the devil in the details: Delving deep into convolutional nets. In British Machine Vision Conference, 2014.

8. M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, and A. Vedaldi. Describing textures in the wild. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.
9. G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Q. Weinberger. DeepNetworkswithStochasticDepth,pages646–661. Springer International Publishing, 2016.
10. S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. CoRR, abs/1502.03167, 2015.
11. T. B. J. T. Springenberg, A. Dosovitskiy and M. Riedmiller. Striving for simplicity: The all convolutional net. In International Conference on Learning Representations Workshop Track, 2015.
12. J.JinandH.Nakayama. Annotationordermatters: Recurrent image annotator for arbitrary length image tagging. In International Conference on Pattern Recognition (ICPR), pages 2452–2457, 2016.