American Journal of Engineering Research (AJER)
e-ISSN: 2320-0847 p-ISSN : 2320-0936
Volume-8, Issue-12, pp-184-188
www.ajer.org
Research Paper
Open Access

Deep learning introduction and visualization

Hao Yuan, Ping Li, Shanwen Zhang
School of Electronic Information Engineering, ZhengzhouSIAS University 451150, China
Correspondence: Hao Yuan

ABSTRACT: *With the rapid development of deep learning in recent years, the accuracy of image classification tasks has been greatly improved, which can meet the requirements of accurate classification in large data sets and is higher than human level in many aspects. Deep learning has the advantage of extracting classification features directly, and it does not need to design classifiers specially. At the same time, the feature extraction method of deep learning is often suitable for the classification of various occasions and has strong generalization. We briefly introduce the basic structure and visualization process of Convolutional Neural Network (CNNs), and give an example.*
KEYWORDS: *Convolutional Neural Network (CNNs); Visualization; Deep learning.*

Date of Submission: 20-12-2019 Date of acceptance: 31-12-2019

I. INTRODUCTION

In recent years, deep learning is widely used in image recognition [1], the speech recognition [2] [3], natural language processing, and other fields, for the complex networks such as social networks, telecommunications networks, protein interaction networks, and the application of road network and so on, there's still a lot of research on space. At present, the researchers apply depth study of thought on the graph theory [4-5]. To adopt the method of deep learning process diagram structure data and mining complex network community structure is possible. If only one keeps the traditional complex network community detection method clustering method [6] label propagation method, and the module of optimization method [7] and so on, these methods are based on the complexity of the large data of scientific research on the effect is poorer. Xue et al. [8] combined with random block model and neural network, this paper puts forward a community discovery method based on neural network diagram, for the research of complex network community detection provides a new way of thinking. Most map neural network model has a generic architecture. These models are collectively referred to as figure convolution network. Call it a convolution, because filter parameters are typically shared all position [9] in the picture. In view of the figure structure data, at present many researchers put forward different figure convolution neural network. Someone by using asymmetric Laplace module as an effective mechanism of embedded figure figure convolution neural network model is put forward, but this method limits its approximate expression of the power spectrum, and limits the application in thin figure. MICHAEL etc. [10] according to the spectrum convolution using the first-order approximate simplified calculation method, Direct operation in figure data of network structure model, this paper proposes a simple and effective layer type transmission method, and verify the figure structure neural network model can process data nodes in a semi-supervised learning problems, but in training due to the expansion of cross layer recursive neighborhood, a lot of computation time and memory consumption. In addition, adaptive neural network, input can be a variety of graph structure of original data, the network without the use of Shared nuclear spectrum, but each sample in the bulk to a particular graph Laplacian, objectively describe the unique topology. According to the topological structure of the unique figure,the customized graph Laplace operator will lead to customized spectrum filters. However, the network is less efficient in training and cannot scale to large data sets.

Deep learning, represented by Convolutional Neural Network (CNNs) [11-14], has a profound influence on computer vision and machine learning. But to fully understand the internal working principle of deep learning model, the depth of the design of high performance network structure is very difficult, have been widely its inner workings as a "black box", this is because there are a mass parameters, depth of CNN update to generate multiple iterations quite discrete and nonlinear mapping function between input and output. As well as

sensitivity to the initial state of the parameters, there are many local optima. To explore the operation mechanism of CNN, the core lies in what features it automatically extracts. After the convolutional layer and pooling layer, the features are all distributed and expressed, and each feature will overlap on the original image. Therefore, we hope to establish the connection between the feature map and the original image, that is, depth visualization. This technique tries to find a better qualitative explanation for the feature of each layer extracted by depth model and plays an important role in designing and developing new network structure. At present, researches on CNN visualization mainly focus on how to understand the expression of layered features that CNN automatically learns from massive data and can reflect the nature of images, that is, to obtain the connection between hidden layer neurons in the network and human interpretable concepts. The most direct method is to show the convolution kernel and the corresponding feature map obtained by learning, but except the first layer convolution kernel and feature map have intuitive interpretation, the other layers are not interpretable. From the perspective of signal processing, the classifier based on CNN high level features needs a large perceptual field in the input domain, so as to carry out multi-layer nonlinear response to the input image dominated by low frequency, and produce smooth and constant output for small input changes. At the same time, due to the nonlinear activation function transformation and pooling, introducing spatial invariance to obtain better recognition performance also brings new challenges to visualization. Depth visualization technology can be divided into three categories: gradient update method; Method based on feature reconstruction; A method based on correlation. The idea of net-based gradient update was introduced, where fixed model parameters changed input values through gradient update to maximize the probability of activating a single neuron or label category. The unnatural images generated by activation maximization can also be antagonistic samples of the network model [13]. Jaromiretc. [6] iteratively searched for the optimal image to maximize the activation of one or some specific neurons of CNN through the gradient ascent method, and assumed that the neuron gradient to the pixel described the change of the current pixel could affect the intensity of the classification result. Literature [2] introduced L2 regularization prior (or weight attenuation) to improve the visualization effect. FEDERICO et al. [9] further proposed gaussian fuzzy regularization and gradient clipping, among which gradient clipping refers to updating only the most favorable part of the gradient for classification each time to improve the image quality. Literatures [3,6] considers the multifaceted nature of neurons and USES the generation network as a priori for natural images to synthesize more natural images. Zeiler et al. [7] proposed the use of deconvolution network and the use of back propagation to reconstruct the mapping of features of each layer to pixel space, and used it to guide the design and optimization of network structure to improve classification and recognition accuracy. In the process of deconvolution, the inverse convolution kernel approximation is used as deconvolution kernel. Jaromiret al. [2] proposed to reconstruct the features of each layer of CNN by learning the 'up' convolutional network, and pointed out that, combined with strong priors, even the high-level activation features used for classification also contained color and contour information. Jaromiret al. [3] proposed to use full variational regularization and natural image priors, and extended L2 norm regularization to p norm regularization through anti-coding reconstruction of the feature expression of each layer learned, so as to obtain a better visualization effect.

II. VISUALIZATION OF MATHEMATICAL MODELS

Activation maximization and feature expression anti-coding reconstruction are both for the trained model, and for the given input $x_i \in R^{C\times H\times W}$, where C is the number of color channels, H and W are the height and width of the image, respectively. The CNN model can be abstracted as the function: $R^{C\times H\times W} \rightarrow R^d$, whose activation value of the i-th neuron is$\varphi_i(x)$, encodes$\varphi_0 = (x_0)$ for the features of the given image x_0, defines the regularization term R(x) of the parameter, and looks for the initial input x* that minimizes the energy functional, and its mathematical model is

$X^* = \arg_x \min(l(\varphi(x),\varphi_0) + \lambda R_\theta(x)\ )$ (1)

wherelis to compare the loss and the difference between$\varphi(x)$ and targetφ_0, choosing different loss functions to define different visualizations. However, this optimization is usually a non-convex optimization problem, and the gradient descent method is usually used to find the local optimal value,

$$x \leftarrow x + \alpha \frac{\partial \phi_i(x)}{\partial x} \quad (2)$$

The activation maximization method is to find the input mode that maximizes the response value$\varphi_0 \in R^d$ of a given hidden layer unit based on the features extracted by any neuron in any layer of the deep architecture, which can be defined by the inner product form,

$$\text{謂}(\varphi(x), \varphi_0) = -\langle \varphi(x)\ ,\ \varphi_0 \rangle \quad (3)$$

whereφ_0 should be manually specified, the target to maximize activation can be the feature vector of the full connection layer or the activation value of a neuron in a channel in the convolutional layer.In the reverse coding reconstruction of feature expression, by minimizing the loss between the given feature vector and the

feature vector of the reconstructed target image, the Euclidean distance is generally used to measure the loss error, which is defined as follows:

$$l(\phi(x),\phi_0)=\frac{\|\phi(x)-\phi_0\|}{\|\phi_0\|^2} \quad (4)$$

But other distance measures can also be used to evaluate the loss.

III. THE ARCHITECTURE OF CONVOLUTIONAL NEURAL NETWORKS (CNN)

Convolutional Neural Networks (CNNs) take advantage of the fact that the input consists of images and they constrain the architecture in a more sensible way. In particular, unlike a regular Neural Network, the layers of a ConvNet have neurons arranged in 3 dimensions: width, height, depth. (Note that the word depth here refers to the third dimension of an activation volume, not to the depth of a full Neural Network, which can refer to the total number of layers in a network.) For example, the input images in CIFAR-10 are an input volume of activations, and the volume has dimensions 32x32x3 (width, height, depth respectively). As we will soon see, the neurons in a layer will only be connected to a small region of the layer before it, instead of all of the neurons in a fully-connected manner. Moreover, the final output layer would for CIFAR-10 have dimensions 1x1x10, because by the end of the ConvNet architecture we will reduce the full image into a single vector of class scores, arranged along the depth dimension.The architecture of CNNs isshowed in Fig.1 and introduced as follows.

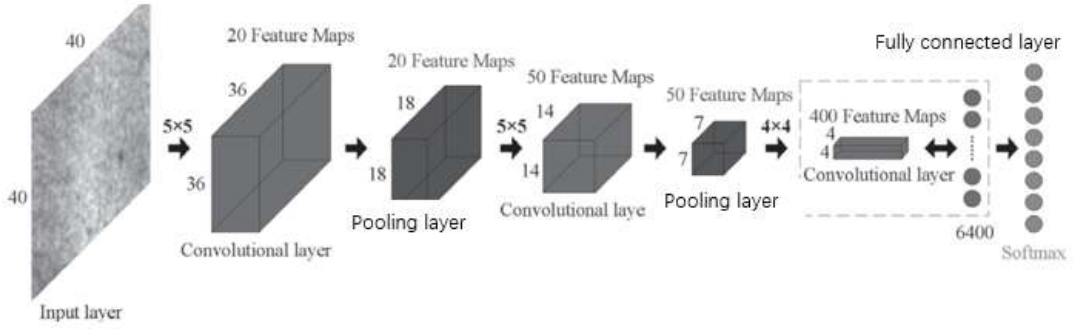

Fig.1 The architecture of CNNs.

Input layer. It will hold the raw pixel values of the image, in this case an image of width 32, height 32, and with three color channels R,G,B.

Convolution layer.The convolutional layer is a feature extraction layer in the network, and the links between each layer are local links and weight sharing. Under the action of multiple convolutional layers, an image is input to obtain multiple feature maps, and then the feature map is extracted through a convolution kernel. If multiple feature graphs are operated by convolution again, then the combination of multiple feature graphs in the upper layer will be the lower neurons obtained after the operation, as shown in figure 4.2. Compared with the fully connected mode, the weighted parameters of the network are greatly reduced, mainly because the local receptive field is used in the convolutional neural network, and the feature of weight sharing makes the scale not deform to some extent.

Pooling layer.If all the extracted features are trained in the process of processing, the whole calculation process will be very tedious and prone to overfitting. You can calculate the maximum number of a particular feature in the image, to avoid the above problems in the process of calculation, this process is called "the biggest pooling, as shown in figure 4.3: use the size of a 2 * 2 templates through the largest pool of don't overlap each other, the size of 7 * 7 of the original image into the image size is 3 * 3, pixel changed from 4 to 1. Compared with the dimension obtained by convolution, the characteristic dimension after pooling is relatively low, and it is not easy to overfit. The whole process is conducive to reducing the complexity of calculation and enhancing the robustness of the network. In different directions, the input image will be shifted, but the features extracted through maximum pooling will remain unchanged. This point is called the maximum pooling process and has the characteristic of translation invariability.

RELU layer will apply an elementwise activation function, such as the max(0,x)max(0,x) thresholding at zero. This leaves the size of the volume unchanged ([32x32x12]).

Fully connected layer.The fully connected network refers to the common neural network in which every neuron in the upper layer is connected to every node in the next layer, and a large number of training parameters are included. The fully connected network is adopted between each layer.

Normalization Layer.Many types of normalization layers have been proposed for use in ConvNet architectures, sometimes with the intentions of implementing inhibition schemes observed in the biological brain. However, these layers have since fallen out of favor because in practice their contribution has been shown to be

minimal, if any.

Softmaxclassification layer.During the experiment, the Softmax classification layer is mainly used for two types of classification. The output is a conditional probability of a size between 0 and 1. The sigmoid transfer function is used in the network to convert the output value of the neuron node to between 0 and 1.

A ConvNet is made up of Layers. Every Layer has a simple API: It transforms an input 3D volume to an output 3D volume with some differentiable function that may or may not have parameters.The architecture ofConvNet model is in the simplest case a list of Layers that transform the image volume into an output volume: There are a few distinct types of Layers (e.g. CONV/FC/RELU/POOL are by far the most popular); Each Layer accepts an input 3D volume and transforms it to an output 3D volume through a differentiable function; Each Layer may or may not have parameters (e.g. CONV/FC do, RELU/POOL don't); Each Layer may or may not have additional hyperparameters (e.g. CONV/FC/POOL do, RELU doesn't). Fig.1 shows the convolutional features and convolutional kernels.

Original image of dataset

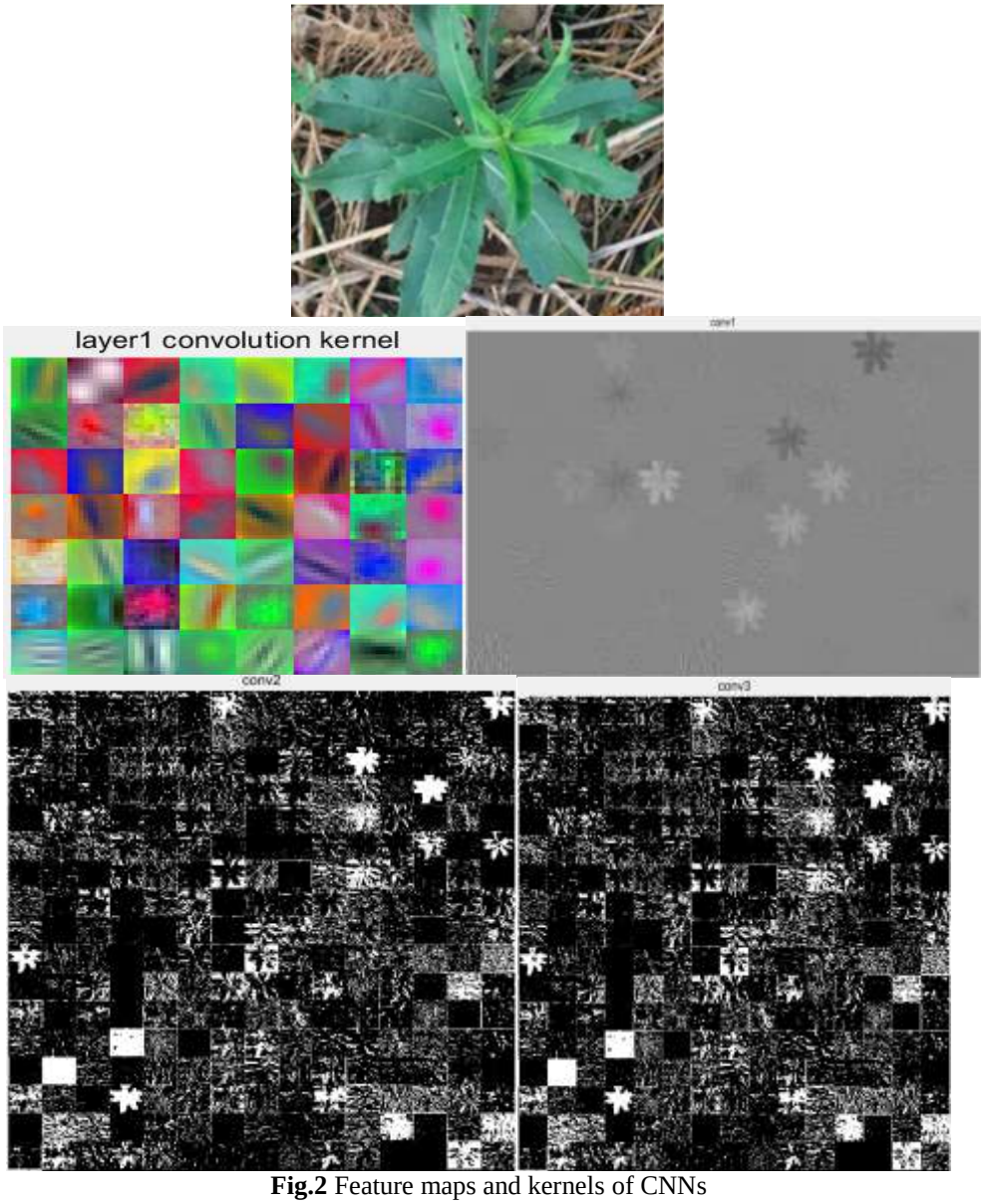

Fig.2 Feature maps and kernels of CNNs

From Fig.2, it is found that the designed CNN model can naturally integrate low/mid/highlevels features, and the multi-levels features can be enriched by the number of stacked layers including convolutional, pooling, global pooling and other layers. The multi-levels featuresrevealthatCNNsis effective to extract the features to describe the image features more detail.

IV. VISUALIZATION OF GRADIENT UPDATE

Deep CNN, which is used for classification, extracts high-level semantic information while losing a lot of low-level structural information. Since the convolution kernel of the first layer is mostly similar to Gabor filter, the visualized image generated by gradient update contains a lot of high-frequency information. Although it can generate a large response activation value, the image generated is unnatural for visualization. Moreover, linear operations of the network model (such as convolution) lead to the existence of confrontation samples [13]. In order to obtain visualization results more similar to real natural images, regularization should be introduced into the optimization objective function as a prior. Deep CNNs, which is used for classification, extracts high-level semantic information while losing a lot of low-level structural information. Since the convolution kernel of the first layer is mostly similar to Gabor filter, the visualized image generated by gradient update contains a lot of high-frequency information. Although it can generate a large response activation value, the image generated is unnatural for visualization. Moreover, linear operations of the network model (such as convolution) lead to the existence of confrontation samples. In order to obtain visualization results more similar to real natural images, regularization should be introduced into the optimization objective function as a prior.

V. CONCLUSIONS

Convolutional neural networks (CNNs)arean important class of neural networks used to learn image representations that can be applied to numerous computer vision problems. Deep CNNs, in particular, consist of multiple layers of linear and non-linear operations that are learned simultaneously, in an end-to-end manner. To solve a particular task, the parameters of these layers are learned over several iterations. CNN based methods have become popular in the recent years for feature extraction from images and video data.In the paper, we simply introduce its architecture and visualization.

ACKNOWLEDGMENTS

This work was supported by the grant of Key Scientific Research Projects of Henan Higher Institutions in 2019 (Nos. 19B520028&19B520029), and the basic and frontier technology research projects of Henan Province (No.182102210544).

REFERENCES

[1] Schmidhuber, Jürgen. Deep learning in neural networks: An overview. Neural Networks, 2015, 61:85-117.
[2] JaromirPrzybyło, MirosławJabłoński. Using Deep Convolutional Neural Network for oak acorn viability recognition based on color images of their sections. Computers and Electronics in Agriculture, 156 (2019) 490–499.
[3] Yang Z, Jin L, Tao D. A comparative study of several feature extractionmethods for person re-identification. Chin. Conf. Biomet. Recog., 2013, 268-277
[4] Gómez-Ríos, Anabel, Tabik S, Luengo, Julián, et al. Towards Highly Accurate Coral Texture Images Classification Using Deep Convolutional Neural Networks and Data Augmentation. Expert Systems with Applications, 118 (2019) 315-328.
[5] Kim E J, Brunner R J. Star-galaxy Classification Using Deep Convolutional Neural Networks. Monthly Notices of the Royal Astronomical Society, 2016. DOI: 10.1093/mnras/stw2672.
[6] JaromirPrzybyło, MirosławJabłoński. Using Deep Convolutional Neural Network for oak acorn viability recognition based on color images of their sections. Computers and Electronics in Agriculture, 156 (2019) 490–499.
[7] SUN Y, WANG X, TANG X. Deeply learned facerepresentations are sparse, selective, and robust. Proceedings of the IEEE Conference on ComputerVision and Pattern Recognition. Boston: IEEE, 2015:2892–2900.
[8] XUE S, YAN Z. Improving latency-controlledBLSTM acoustic models for online speechrecognition. Proceedings of the IEEEInternational Conference on Acoustics, Speechand Signal Processing. New Orleans: IEEE, 2017:5340–5344.
[9] FEDERICO M, DAVIDE B, JONATHAN M, et al.Geometric deep learning on graphs and manifoldsusing mixture model CNNs. Proceedings of theIEEE Conference on Computer Vision andPattern Recognition. Honolulu: IEEE, 2017:5425–5434.
[10] MICHAEL M. B, JOAN B, YANN L, et al. Geometricdeep learning: going beyond euclidean data. IEEESignal Processing Magazine, 2017, 34(4): 18–42.
[11] SHIGA M, TAKIGAWA I, MAMITSUKA H. Aspectral approach to clustering numerical vectors asnodes in a networks. Pattern Recognition, 2011,44(2): 236–251.
[12] RAGHAVAN U N, ALBERT R, KUMARA S. Nearlinear-time algorithm to detect community structuresin large-scale networks. Physical Review E,2007, 76(3): 036106.
[13] WANG M, CAI X, ZENG Y, et al. A Communitydetection algorithm based on jaccard similarity labelpropagation. Intelligent Data Engineeringand Automated Learning. Guilin: IDEAL, 2017:45–52.
[14] MAHENDRAN A，VEDALDI A． Understanding deep image representations by inverting them. IEEE Conference on Computer Vision and Pattern Recognition． Boston: CVPR，2015: 5188-5196.

Hao Yuan "Deep learning introduction and visualization" American Journal of Engineering Research (AJER), vol. 8, no. 12, 2019, pp 184-188