IEEE *Access*
Multidisciplinary : Rapid Review : Open Access Journal

Bibliometric Analysis of the Application of Convolutional Neural Network in Computer Vision

HUIE CHEN AND **ZHENJIE DENG**
School of Internet Finance and Information Engineering, Guangdong University of Finance, Guangzhou 510521, China
Corresponding author: Zhenjie Deng (47-047@gduf.edu.cn)

ABSTRACT This article analyzes the research progress in field of Convolutional Neural Networks (CNNs) using the bibliometric method. Literature samples of CNNs are analyzed by a basic statistic and co-citation network. Experimental results show that CNNs are being utilized in many computer vision applications, such as fault and image recognition diagnosis, seismic detection, positioning, and automatic detection of cracks and signals, image classification and image segmentation. In addition, there is systematic research on unbalanced problems in CNNs. Quantitative experimental research, extensive application fields, and market research informatization will be the three vital research tendencies in the future. The ideas and conclusions of this article provide insights to the academic research of CNNs and their practical application in the corporate world.

INDEX TERMS Convolutional neural networks, computer vision, bibliometric analysis.

I. INTRODUCTION

In recent years, convolutional neural networks have become an important field of computer research. Computer vision enables artificial systems to "perceive" science from images or multidimensional data. In 1950s, the neural network appeared with the study of perceptrons. A multi-layer perceptron was designed to recognize two-dimensional shapes. This network structure is highly invariant to translation, scaling, oblique, or other forms of deformation. In 1962, Hubel and Wiesel proposed the concept of the receptive field through the study of cat visual cortical cells [1]. In 1984, Fukushima proposed a neocognitron model based on the concept of the receptive field. It was also the first application of the receptive field in artificial neural networks. C. Lawrence proposed a hybrid neural network for facial recognition, which has advantages over other existing methods [2]. Nebauer used CNNs to compare handwritten digit recognition and facial recognition against visual recognition [3].

CNNs were widely used in 1990s, but with the introduction of the support vector machine, CNNs decreased in

The associate editor coordinating the review of this manuscript and approving it for publication was Zhihan Lv.

popularity. In 2012, Krizhevsky *et al.* [4] demonstrated higher image classification accuracy on ILSVRC, reigniting interest in CNNS. Their success stemmed from training a large CNN with 1.2 million tag images and adding some twists on LeCun's CNN (for example, Max[X, 0] corrected nonlinearity and "drop-out" regularization).

In this article, the development of the CNNs is analyzed by the bibliometric method. The statistics on the total growth trend in CNN literature, research fields, scientific research institutions and core authors are used to construct a co-citation network for citation clustering. This article also focuses on analyzing the evolution of CNNs. We can observe that the research on CNN has experienced exponential growth in recent years. Computer vision is applied to the control process of industrial robots, the navigation of autonomous vehicles, the detection events of video surveillance, the information organization of the index database of images and image sequences, the modeling objects or environments such as medical image analysis systems or terrain models, and the automation of manufacturing industries. An example system for testing applications and computer human-computer interaction. However, the research field has also become more diversified. CNNs can be now used for seismic detection,

positioning, automatic detection, and fault and image recognition diagnosis. Another focus is on systematic research into the unbalanced problems in convolutional neural networks.

II. DATA AND METHODS

The search scope of the literature samples collected in this study is the dataset with the keyword ''Convolutional Neural Network'' in the title of the Web of Science website's core database. The search focuses on the title of the article, rather than the article's body, because there are so many CNN-related documents in many other fields. The body of literature that mentions the keyword ''convolutional neural network'' in the text is not the focus of our research. Retrieving documents containing the keyword convolutional neural networks in the title can ensure that the literature samples are more relevant. On this basis, manual preliminary screening was conducted. The final sample was 4,598 articles as of February 15, 2020. This study took 20 years of international literature as the research object. It used CiteSpace V software and a combination of quantitative and qualitative methods for analysis.

In this article, the CNNs-relevant papers published in the past 20 years were used as the research object, using Excel software and CiteSpace software. Descriptive statistics were employed with the cited literature and citations in the field of CNNs, bolstered by data mining. An in-depth analysis of the hotspots in research content was also conducted.

## III. STATISTICAL ANALYSIS OF CNN LITERATURE

The publication status of papers is usually regarded as an important indicator of measuring the development of a discipline and the level of scientific research achievements and contributions. CNN is the most important algorithm in the field of computer vision. It uses deep neural network to simulate the process of human pattern recognition and cognition of the outside world, and extracts, recognizes and classifies features from images, which is widely used in the field of computer vision. Compared with traditional computer vision methods of digital image processing and geometric optics, and statistical machine learning methods, it has more powerful feature learning and representation capabilities. The growth trend of scientific knowledge is closely related to the growth trend of the published literature. In this section, we analyze the trends of CNN research through the statistics of journal publications and cited frequencies over a period of time.

A. ANALYSIS OF THE OVERALL GROWTH TREND

Figure 1 shows the number of papers published and the total number of citations in various years. It shows a large increase in the CNN literature after 2014. There were 18 articles published between 2002 and 2010. About 1350 articles were published in 2018. The number of published articles reached a peak in 2019 with a total of 2561 articles. This means there were 142 times more articles published in a span of about 10 years.
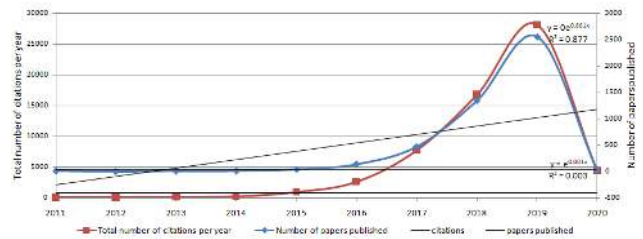
FIGURE 1. Number of publications and the total number of citations by year.
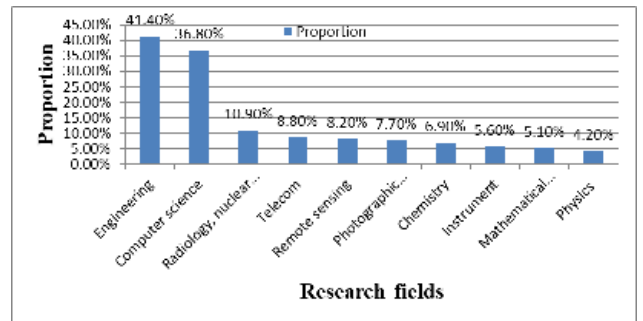
FIGURE 2. Main research areas and their proportions.

Figure 1 reflects that the number of citations related to CNN is also growing rapidly. The related literature had only 78 citations in 2010 but 60,739 in 2020. Among the literature samples, the most cited one is the study conducted by Krizhevsky *et al.* In order to reduce overfitting in the fully connected layer, the author adopts a newly developed regularization method. Deep CNNs are applied to classify ImageNet. This method was highly effective. As of 2020, the article has been cited 14,768 times, a groundbreaking achievement in the field of CNN research. An exponential function is fitted to the curve of the number of articles published and the number of citations per year. The R-square of the curve fitting is 0.877. This indicates that the research magnitude curve of the CNN is an exponential function.

The exponential growth trend of CNN research shows that CNN research is gradually transitioning from a fundamental stage to a mature development stage. With the boom in deep learning, there is an urgent need to understand and apply CNNs in various fields. Scholars have come to realize that CNN innovation requires many experiments. In addition, the rapid development of CNN research is being driven by real-world problems in various industries.

B. ANALYSIS OF RESEARCH FIELDS

This section analyzes the research on CNNs through the distribution of research fields and the distribution of statistical literature samples in journals.

Figure 2 shows the classification of the research directions for CNN documents and their proportion in the total literature samples. The research direction is given by the Web of Science database. Since the research directions of most of the

literature overlap, the proportions of all research directions add up to more than 100%. It can be seen from the figure that the main research fields of the CNN literature are engineering, computer science, radiology, nuclear medicine, imaging, and telecommunications. The more detailed research directions also include remote sensing, photographic technology of image science, chemistry, instruments, mathematical calculation, biology, and physics. The major reason is that with the development and application of shallow machine learning algorithms, including SVM, the application of CNNs is largely concentrated in engineering and computer science.

Statistics on CNN literature can help further analyze the direction of research hotspots and the quality of scientific research results. This article uses two indicators to evaluate the impact of a journal: the impact factor and the h index. Impact factor is an internationally used index for evaluating journals. It refers to the total number of citations in papers published by a journal in the previous two years divided by the total number of papers published by the journal in those two years. The h index is a mixed quantitative index, which means that at least h papers published in a certain period of time are cited at least h times. It measures the output level of high-quality documents.

C. ANALYSIS OF HIGH-LEVEL SCIENTIFIC RESEARCH INSTITUTIONS

This article offers statistics on research institutes with a high output of CNN research by their key research directions and cooperative relationships. Table 1 shows that the most published papers in all literature samples are from Chinese Academy of Sciences, which published a total of 246 papers. It is followed by Shanghai Jiao Tong University and Natl Tsing Hua University, each with 78 published papers. Moreover, Wuhan University and Stanford University each published 66 papers. Among the 10 research institutions with the most published papers, nine are in China and one is in the United States. The results suggest that the level of CNN research in China has improved significantly in recent years.

From Table 2, it can be seen that the cooperative research is more extensive in China ACAD SCI. The cooperative fields include CNN, deep learning, system, and classification issues. In the field of computer vision, we focus on the latest research progress of convolution neural network in image classification and location, target detection, target segmentation, target tracking, behavior recognition and image super-resolution reconstruction. The neural network architecture is improved in convolution layer, pooling layer and activation function. Stanford University, the University of California, San Francisco, and Columbia University have also conducted cooperative research. The cooperative fields are segmentation and CNN. Wuhan University in China has conducted cooperative research with University of Tokyo, Harvard Med School, and Kings College London. The gap between the scale and degree of cooperation between Chinese scientific research institutions in CNN research and the top

TABLE 1. Frequency analysis of highly cited papers.

Cited frequency	Number of Articles Published	Journal Title	Impact Factor(2018)	H in-dex(2015 to date)
35	320	IEEE Access	4.098	52
1402	134	Sensors	3.031	61
882	126	Remote Sensing	4.118	30
497	123	Neurocomputing	4.072	34
491	86	Applied Sciences Basel	2.217	88
442	82	Multimedia Tools and Applications	3.807	74
415	56	IEEE Geoscience and Remote Sensing Letters	3.534	69
408	55	Scientific Reports	4.011	76
364	48	Medical Physics	2.101	57
334	47	IEEE Transactions on Geoscience and Remote Sensing	5.63	96

The impact factors of all journals in table 1 are greater than 1 and the h index is greater than 30. all journals are either sci or ssci-included journals. in addition, the most influential achievements in the research of cnns are concentrated in computer science journals such as ieee access, sensors,remote sensing,and neurocomputing.

TABLE 2. Research institutions and their partnerships.

Sort	Organization Name	Research Focus	Number of Published Papers	Other Universities with Partnerships
1	Chinese Academy of Sciences	Convolutional neural network, Deep learning	246	Wuhan Univ, Shanghai Jiao Tong Univ, TianJin Univ, Dalian Univ Technol
2	Shanghai Jiao Tong Univ	System, Classification	78	Chinese Academy of Sciences
3	Tsdnghua Univ	Convolutional neural network	78	Shenzhen Univ, City Univ Hong Kong
4	Wuhan Univ	Model, Feature	66	Univ Tokyo, Harvard Med Sch, Kings Coll London
5	Stanford Univ	Segmentation, Convolutional neural network	66	Univ Calif San Francisco, Columbia Univ
6	Xidian Univ	Feature learning	66	Imperial Coll London, Sun Yat Sen Univ
7	Zhejiang Univ	segmentation, Image	63	Fudan Univ, Xi An Jiao Tong Univ
8	Beihang Univ	Object recognition, 3d mesh labeling	60	Beijing Jiaotong Univ, Shenzhen Univ
9	Harbin Institute Tech	object, feature Convolutional neural network	58	Nanyang Tech Univ, Shenzhen Univ
10	Huazhong Univ Academy of Sciences	Deep learning, Model	57	Tongji Univ, Adobe Res

foreign institutions is getting smaller. Cooperation between the institutions in China is also a growing trend.

There is a close cooperative relationship between the high-level institutions conducting research on CNNs.

TABLE 3. Core author statistics.

Serial number	Authors	Number	Number of Articles Published	Percentage
1	Zhang Y, Sun X	2	9	4.76%
2	Sun H	1	8	4.23%
3	Qiao Y	1	7	3.70%
4	Fu K, Li F	2	6	3.18%
5	Li B, Li W, Li Y, Liu Cl, Wang Q, Wang Y, Xiang Sm, Yan Ml	8	5	2.65%
6	Pan Ch, Waang L, Xie Yq, Zhang Zx	4	4	2.12%
7	Chen S, Feng Sl, Gao K, Hu Y, Li Bq, Li X, Liu S,Liu W, Liu Y, Qu Gx, Tan Tn, Wang D, Wang Hz, Wang P, Wang W, Wang Z, Wang ZF, Xing L, Xu B, Xu Gl, Yu Sd, Yuan Y, Zhang J, Zhang Xl, Zhang Yd, Zhang Yh, Zhang H	27	3	1.59%
8	Total	45	42	22.22%

Collaboration between researchers with different research backgrounds is conducive to producing high-quality scientific research. It also fosters innovation in CNNs. By cooperating with world-class scientific research institutions, researchers of CNNs in China can strengthen the exchange of scientific research resources and results.

D. ANALYSIS OF CORE AUTHOR

Core authors refer to the researchers who produce more published literature and have more influence in a certain research field. Price's law in bibliometrics can be used to determine the core author in a research field. In the literature sample searched in this article, the most papers published by the same author is nine. Place's law can be used to identify core authors in a field of study. Place's law is famous among scientists and historians of Science, Harvard, and the University of Place in his famous book, Small Science, big science, the proposed method of estimating the core authors can be expressed as follows:

$$M = 0.749\sqrt{N_{max}} \tag{1}$$

$\sqrt{N_{max}}$ is the number of papers with the most published papers among all authors, M is the minimum number of papers published to determine the core author, that is, authors who have published more than M papers are the core authors. Price's law comes out with the result M = 3. Therefore, authors who have published more than three papers are the core authors in the field of CNNs. Going by the number of published papers, the top 45 can be considered the core authors in the field of CNNs. The results are shown in Table 3.

It can be seen from Table 3 that the 45 core authors published a total of 42 papers, accounting for 22.22% of the total number of papers published. Of these authors, Sun X and Zhang Y each published nine articles. Zhang Y's co-authored article ''Multi-view Convolutional Neural Networks for Multi-document Extractive Summarization'' uses

word embedding to represent sentences. The article proposed a multi-view enhanced CNN. Moreover, it obtains the characteristics of the sentence and sorts the sentence at the same time [5]. Object recognition has always been a popular topic in the field of remote sensing image analysis. Zhang Y, Fu K and Sun H proposed a pixel-based target recognition method based on deep belief networks (DBNs) [6]. SUN X has been cited 19 times in the co-authored article ''Sentiment analysis for Chinese microblog based on deep neural networks with convolutional extension features,'' which was published two years ago.

A novel convolutional automatic encoder, which can extract contextual information from conversations on Weibo as the features of the posts, was adopted. A customized deep neural network (DNN) model was then implemented. The experiments showed that the performance of the proposed DNN in sentiment classification is superior to the latest surface learning models such as SVM or NB based on the appropriate structure and parameters [7]. Nonetheless, traffic sign recognition (TSR) is an important and challenging task in intelligent transportation systems. Fu K proposed a method of Hinge Loss Stochastic Gradient Descent (HLSGD) to train CNNs. HLSGD is evaluated according to the German traffic sign recognition benchmark. It provides faster and more stable convergence, with a recognition rate reaching 99.65% [8]. In addition, Li F used a deep CNN to automatically distinguish between the view of glaucoma and non-glaucoma [9].

Liu Y's co-authored article ''Sentiment analysis for Chinese microblog based on deep neural networks with convolutional extension features'' has been cited 148 times in the past three years. The article proposed a multi-focus image fusion method that offers advanced fusion performance in visual quality and objective evaluation. It is useful in solving image fusion problems such as multi-focus image fusion and multi-mode image fusion [10]. Most of the research results of CNNs are still concentrated among a narrow range of researchers, with most of them having published only a few articles. However, as research on CNNs attracts more attention from researchers, the number and scope of publications by the core author group will continue to grow.

IV. CO-CITATION ANALYSIS OF CNN LITERATURE

The statistical analysis of CNN literature focuses on identifying the authors and institutions with the greatest influence in the field of CNNs. The co-citation analysis focuses on analyzing the contributions of specific literature. In this section, CiteSpace is used to construct a citation network of CNNs. Citation clustering is utilized in pinpointing important nodes of the network. This section identifies the development process and hot issues of CNN research.

A. CITATION CLUSTER ANALYSIS

By analyzing co-citations between articles, we can better understand the evolutionary pattern between studies. A time interval of four years is selected to generate a co-citation visualization map of the literature in this research field,
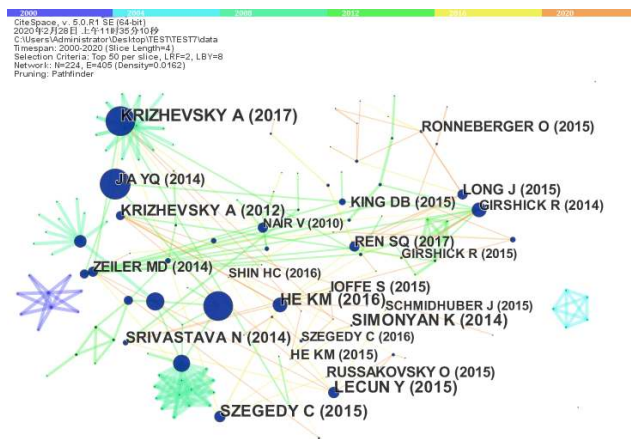
FIGURE 3. Literature co-citation network visualization map between 2000 and 2020.

as shown in Figure 3. The map contains a total of 224 nodes and 405 connections. Each node represents a cited article. The larger the radius, the higher the cited frequency of the article. The connection between nodes indicates the strength of co-citation. The key node is the literature with the greatest influence or significance in the field.

Table 4 shows that Krizhevsky A has continued his research during the period of 2012 to 2017 and there are 1347 citations published in the "ImageNet classification with deep CNNs." This study trained a large, deep CNN to classify 1.2 million high-resolution images into 1000 categories. To make the training faster, unsaturated neurons and a GPU are used to implement the operation of CNNs [11]. Ranking second was Lecun Y (2015), with 932 citations for his published "Deep learning." He also pointed out that deep CNNs have made breakthroughs in image, video, speech, and audio processing. In fact, the recursive network is fully reflected in text, speech, and other sequence data [12]. He KM (2016) proposed a residual learning framework to simplify more in-depth training than previous networks [13]. Szegedy C (2015) also pointed out that deep CNNs have recently achieved the most advanced performance in many image recognition benchmarks, including the ImageNet large-scale visual recognition challenge (ILSVRC-2012) [14]. Yangqing Jia suggested that Caffe provided impetus for ongoing research projects, large-scale industrial applications, and startup prototypes in vision, speech, and multimedia [15]. H. Cecotti and A. Graser (2011) proposed a P 300 wave detection method, a P 300 detection CNN and its application in brain-computer interface. The network topology is suitable for P 300 wave time domain detection. Seven CNN-based classifiers are proposed: four single-feature classifiers with different feature sets and three multiple classifiers [16].

According to the calculation of the occurrence value, Figure 3 shows that the most frequently cited articles from 2000 to 2007 are mainly about the application of CNNs in the fields of facial, handwritten character and visual image recognition, for example, "Gradient-based

TABLE 4. Information of key node documents from 2000 to 2020.

Order	Authors	Year of Publication	Cited References	Count
1	A. Krizhevsky	2017	A. Krizhevsky, I. Sutskever and G. Hinton, "ImageNet classification with deep convolutional neural networks," *ACM Commun*, vol. 60, no. 6, pp. 84-90, May 2017.	1347
2	K. He	2016	K. He, X. Zhang, S. Ren, et al., "Deep Residual Learning for Image Recognition," in *Proc.IEEE CVPR*, 2016, pp. 770-778.	821
3	K. Simonyan	2014	M. Jaderberg, K. Simonyan, A. Vedaldi, "Synthetic Data and Artificial Neural Networks for Natural Scene Text Recognition," *arXiv preprint*, arXiv:1406.2227, 2014.	699
4	N. Srivastava	2014	N. Srivastava, R.Salakhutdinov, "Multimodal Learning with Deep Boltzmann Machines," in *Proc. NIPS*, 2012, pp. 2222-2230.	650
5	A. Krizhevsky	2012	A. Krizhevsky, I. Sutskever, G. Hinton, ąřImageNet Classification with Deep Convolutional Neural Networks,ąś in *Proc. NIPS*, 2012, pp. 1097-1105	597
6	S. Yeung	2015	S. Yeung, O. Russakovsky, G. Mori, et al., "End-to-end Learning of Action Detection from Frame Glimpses in Videos," in *Proc. CVPR*, 2016, pp. 2678-2687.	383
7	J. Long	2015	J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE CVPR*, 2015, pp. 3431-3440.	364
8	Y. Jia	2014	Y. Jia, E. Shelhamer, J. Donahue, et al., "Caffe: Convolutional Architecture for Fast Feature Embedding," in *Proc. of MM ACM.*, New York, NY, USA, pp. 675-678	362

learning applied to document recognition" published by Y. Lecun (1998). Various methods of handwritten character recognition methods are reviewed and their performance compared on a task for recognizing standard handwritten numbers. CNNs are specifically designed to deal with the variability of two-dimensional shapes [17]. "Face recognition: a convolutional neural-network approach" published by S. Lawrence (1997) was cited most frequently in the period of 2000 to 2005. The researcher proposed a hybrid neural network for face recognition that is more effective than other methods. The method combines local image sampling, a self-organizing mapping (SOM) neural network and CNN. SOM quantizes image samples into a topological space, where the input near the original space is also near the output space. As a result, dimension reduction and invariance for small changes in image samples are achieved. CNNs provide a partial invariance in translation, rotation, scale, and deformation. "Evaluation of convolutional neural networks

for visual recognition," published by C. Nebauer (1998) was cited most frequently in the period of 2004 to 2007. The visual recognition performance of CNN was compared in two tests: comparing the improvement of the new cognitive network, and comparing the classifier based on the fully connected feedforward layer.

The most frequently cited articles from 2012 to 2016 are mainly about the complex functions that represent high levels of abstraction, for instance, in vision, language, and other tasks at the level of artificial intelligence. Yoshua Bengio (2009) discusses the motivation and principles of learning algorithms for deep architectures in "Learning Deep Architectures for AI," especially those that use single-layer models, such as restricted Boltzmann machines, as building blocks for unsupervised learning. The aim is to construct a deeper model, for example, deep trust networks [18]. In "Representation Learning: A Review and New Perspectives," Yoshua Bengio (2013) reviewed recent research in unsupervised feature learning and deep learning, including probabilistic models, autoencoders, manifold learning, and deep networks. This sought to find answers for long-standing unanswered questions about selecting the best learner, computational representation (i.e., inference), and the geometric relationship between representation learning, density estimation, and manifold learning [19].

"Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups" was the most frequently cited paper in the period of 2014 to 2017. The researcher, Geoffrey Hinton, pointed out that DNNs have multiple hidden layers, which are trained using new methods and have been shown to outperform Gaussian mixture models (GMMs) in a variety of speech recognition benchmarks [20].

This article uses CiteSpace to perform clustering in order to generate an automatic clustering label view, as shown in Figure 4. The automatic clustering label view is based on the default view. The knowledge clustering is generated through a spectral clustering algorithm. Moreover, an algorithm is used to extract the label words from the relevant literature that cites the cluster. It is used to characterize the basic advanced research corresponding to certain knowledge. As shown in Figure 3, among the seven subgroups, cluster labels 0, 1, 2, 6, and 7 are related to each other. Cluster labels 4, 5, and 6 are independent subgroups. Clustering labels include deep learning, face alignment, visual tracking, brain-computer interface, shunting inhibitory neuron, face detection, and steady-state visual evoked potential (SSVEP) electroencephalogram (EEG).

Cluster analysis based on a co-citation network is a specific application of cluster analysis technology. It mainly refers to the use of co-citation strength as the basic unit of measurement to quantify the classification and aggregation of a given citation or collection of cited documents. This technique can aggregate papers with close content into individual document clusters. It also quantitatively gives the degree of connection between clusters according to the relevant network indicators.
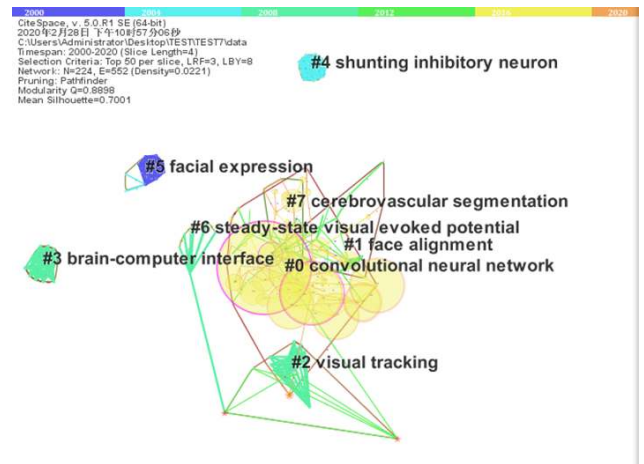
FIGURE 4. Automatic clustering label view of document co-citation network from 2000 to 2020.

A cluster analysis network graph of professional papers in a certain discipline will therefore be generated. CiteSpace first clusters the documents in different time divisions and then merges the sub-clusters to form a unified view. Figure 4 displays the knowledge network and clustering results of CNN literature citations. The node size in the graph represents the citation rate. The higher the cited frequency, the larger the node in the graph. The connection between two nodes represents that two documents have been cited together. The colors of the nodes and lines correspond to the time axis at the top of Figure 3 [21], where the right side of the time axis represents 2000 and the left side represents 2020. According to the color of the original image, the cluster co-citation time for cluster labels 0 and 1 is after 2014, for cluster label 2 is after 2007, for cluster labels 3 and 6 is after 2006, for cluster label 4 is after 2001, for cluster label 5 is after 1998 and for cluster label 3 is after 2016. Accordingly, Figure 3 can be said to represent the advanced knowledge of CNN research in different time periods.

Table 4 summarizes the basic situation of each cluster in the citation network shown in Figure 4. The clustering label is based on the abstract word segmentation of each highly cited document in each cluster. The mutual information (MI) method is used to extract the feature words, as shown in Table 5. The research results in Table 5 are based on the keywords and citations that frequently appear in each cluster. The results of citation clustering show that the citation network of CNN mainly forms seven clusters, which correspond to the research highlights in different time periods.

Table 4 summarizes the basic situation of each cluster in the citation network shown in Figure 4. The clustering label is based on the abstract word segmentation of each highly cited document in each cluster. The mutual information (MI) method is used to extract the feature words, as shown in Table 5. The research results in Table 5 are based on the keywords and citations that frequently appear in each cluster. The results of citation clustering show that

Cluster number	Cluster label	Main research points	Number of cluster members
0	Deep Learning	Gradient method deep learning, convolutional neural network classification, acoustic detection	49
1	Face Alignment	Facial recognition, convolutional neural network dimensions	24
2	Visual Tracking	Visual tracking, machine learning, effective position coding	19
3	Bci Competition	Brain-computer interface	15
4	Shunting Inhibitory Neuron	Second-order training method, visual pattern recognition shunting inhibitory neurons	12
5	Face Detection	Facial recognition	11
6	Steady-state Visual Evoked Potential	Neural network EEG application	10
7	Level Set	Neural network detection level set, automatic diagnosis	10

the citation network of CNN mainly forms seven clusters, which correspond to the research highlights in different time periods.

The clusters labeled 0 and 1 correspond to the main cluster in the middle part of Figure 4. This cluster has a high number of members, with 49 and 25 documents, respectively. These documents were all cited after 2014. The research direction is focused on the performance and impact of CNNs for image classification and the accuracy of deep learning detection methods based on CNNs. Chen *et al.* (2020) [22] proposed a deep learning framework for the analysis of histopathological images by using a CNN with a visualization scheme. She proposed a deep learning framework for histopathological image analysis and evaluated its use in the automatic and interpretable diagnosis of cervical cancer. S.H. Shabbeer Basha (2020) pointed out that the selection of the unique CNN architecture of the dataset is a time-consuming and error-prone process, because it is mainly based on human experience or expertise. In a bid to automate the process of learning the CNN architecture, an attempt was made to find the relationship between fully connected (FC) layers with certain dataset characteristics [23]. Chi-Hsuan Tseng proposed a method to automatically measure the length of fish in complex images based on CNNs [24].

Other successful CNN training methods have appeared in recent years. Erdal Basaran (2020) proposed a diagnosis method that combined a rapid R-CNN and a pre-trained CNN. This method could be applied to future otology clinical decision support systems to improve the diagnostic accuracy of physicians and reduce the overall misdiagnosis rate [25]. J. Jin (2014) described the architecture of a TSR model and proposed to use the HLSGD method to train CNNs [8].

The cluster members labeled 2 and 3 contain only a total of 34 documents. The cited time is concentrated after 2007. The research focuses on the application of CNNs for object

tracking, object feature analysis and object motion recognition in computer vision. By analyzing the cluster members, we can see that there are some insightful results in the early neural network research. J. Fan used CNNs to estimate the scale through the precise positioning of some key points. She designed a displacement variable CNN architecture to alleviate the drift problem when the distracting object is similar to the target object in a chaotic environment. This method can also be used to track other types of objects [26]. Lin Hui proposed a method for vehicle traversability analysis based on CNN, which can extract implicit features of vehicles [27]. S. Ji developed a novel 3D CNN model for motion recognition. The model extracts features from the spatial and temporal dimensions by performing 3D convolution. As a result, it captures motion information encoded in multiple adjacent frames and achieves higher performance than the benchmark method [28].

B. ANALYSIS OF LANDMARK AND PIVOT NODES

By Landmark and pivot nodes are the key research objects in the analysis of citation co-occurrence networks. Identifying landmark and pivot nodes in the citation network can help in the discovery of the important research results of the discipline. Landmark nodes are the nodes with a large radius in the citation network. The outer circles of keyword nodes with large landmark nodes are highlighted in purple, which implies a higher frequency of citations. Moreover, the larger the landmark node, the more information flow with other nodes. Nevertheless, pivot node is the only connecting point that connects two clusters at the same time. It is represented by a red circle. The appearance of the pivot node often leads to a change in the research focus of a field in a specific discipline because it emphasizes the emerging trends in the discipline.

In Figure 5, the points indicated by the purple circles and stars are the landmark nodes and the pivot nodes of the citation network, respectively. Table 6 summarizes the basic information of landmark nodes, pivot nodes and most cited documents in each cluster in the citation network.

The work done by S. Lawrence *et al.* is one of the most important landmark nodes in the citation network. It provides an in-depth discussion of the relationship and role of CNNs in terms of company strategies and innovation. It has been cited a myriad of times in the past five years since its publication. It can be considered as a milestone of the development and evolution of network research.

The two pivot nodes in the study of CNNs are Geoffrey E. Hinton and Hubert Cecotti. Hinton showed how to use "complementary priors" to eliminate the explaining-away effects that lead to difficult reasoning in densely connected belief networks with many hidden layers [29]. Cecotti pointed out that a brain-computer interface (BCI) is a special kind of human-machine interface. It analyzes the measurement data of the brain in order to achieve the direct communication between human and computer. It also provides a new method for analyzing the receiving field of the CNN model [16]. Their literature as a pivot node document in the co-citation
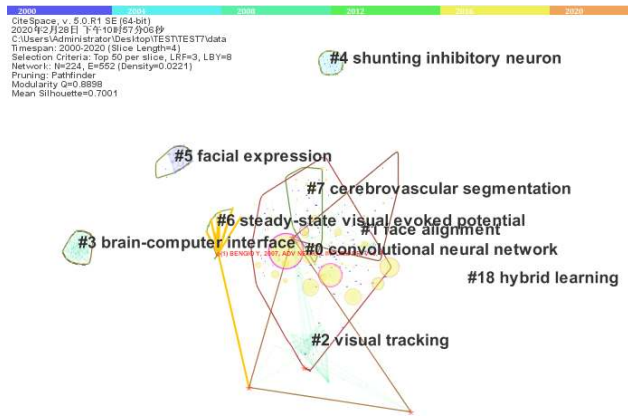
FIGURE 5. Study of landmarks and turning points.

TABLE 6. Landmarks and turning points in citation networks.

Author	Monograph \ Thesis Title	Published journal or conference ("\" Means monograph)	Citation frequency	Cluster label
A. Krizhevsky (2017)	ImageNet Classification with Deep Convolutional Neural Networks	*ACM Commun.*	1347	0
Y.Lecun (2015)	Deep Learning	*Nature*	932	0 and 1
K. He (2016)	Deep Residual Learning for Image Recognition	*IEEE CVPR*	821	0 and 2
G. Hinton (2006)	A fast learning algorithm for deep belief nets	*Neural Computation*	3	2 and 6
H. Cecotti	Convolutional Neural Networks for P300 Detection with Application to Brain-Compute Interfaces	*IEEE Trans. PAMI*	2	1 and 2

network further confirms their critical position in the field of CNN research. The cooperation in the literature node also reflects the research field and direction of institutional cooperation discussed in Section 2.3.

C. DYNAMIC TREND ANALYSIS OF HOT SPOTS

The development of CNNs has been deeply influenced by advances in artificial intelligence (AI). CNN applications will continue to expand into the fields of medicine, physics, chemistry, industry, telecommunications, and the military.

This article divides the entire sample period into sub-sample sets of five or six years. The evolution of CNN research will be analyzed by citation clustering. Figure 6 shows the citation clustering results in the sample periods of 2000-2004, 2005-2009, 2010-2014, and 2015-2020. Most clustering centers have only appeared for one period. It can be seen from Figure 6 that the research focus has continuously changed over the past 20 years.

In the R&D of CNNs prior to 2000, more attention was paid to documents about face detection and visual pattern recognition. From 2000 to 2004, the CNN was still in a

relatively independent position, one that was far away from other clustering centers. More noteworthy is the literature on image processing and layered spiking coupled with CNNs. For example, Matsugu M proposed a convolutional spike neural network model with an explicit pulse sequence (pulse grouping) timing structure for encoding and decoding local visual features [30]. J. Schemmel proposed an ANS VLSI architecture based on the combination of digital signaling and analog computing [31]. More than 200 of Egmont-Petersen M neural networks are applied in image processing, especially feedforward neural networks, Kohonen feature mapping and Hopfield neural networks [32].

After 2005, the literature of CNNs became more pivotal. The frequency of citations increased significantly. Important literature includes motion or visual recognition, visual tracking, machine learning, gradient learning, BCIs, convolutional codes, and artificial neural networks. Hinton uses "complementary priors" to eradicate the explanation of leaving effect that leads to difficult reasoning in a densely connected belief network with many hidden layers. A fast and greedy algorithm is therefore derived. It can learn deep directed belief networks layer by layer, but the first two layers have to be an undirected associative memory [29]. Larochelle H pointed out that the deep-level empirical evaluation of multi-factor problems has good prospects in solving difficult learning problems [33]. Ahmed A proposed a training framework for a visual recognition layered feedforward model based on pseudo-task transfer learning. These pseudo-tasks generate information inverse Wisart priors to the network's functional behavior. They provide an effective way to incorporate useful prior knowledge into network training [34]. This research is an important landmark in the citation network, offering an in-depth discussion of the role of CNN in visual recognition and action recognition, and is frequently cited.

After 2010, the literature in the field of visual pattern recognition and motion recognition in CNNs began to gradually decrease. It was replaced by literature on deep CNNs. In the past five years, despite the widely used traditional disciplines of machine vision and natural language processing, the number of papers related to urban traffic sign recognition data representation, and cognitive science (emotional feature analysis, feature learning, speech emotion recognition, and acoustic modeling) have sharply increased. This reflects the trend that DNNs have become a new field of interest for researchers. Hau D used deep CNNs to explore hierarchical speech representations [35]. A.R. Mohamed used the reduced visualization of the relationship between the phone recognition performance on the Timit corpus and the feature vectors learned by the DBN to maintain the similarity structure of the feature vectors on multiple scales. Deep trust networks were also used for acoustic modeling [36].

It is worth noting that, after 2015, neural networks began to be applied in other fields. The number of papers related to weather radar, optical imaging canopy detection, CNN dimension crown detection, similarity detection, control door dimension scene classification, and network dimension frame
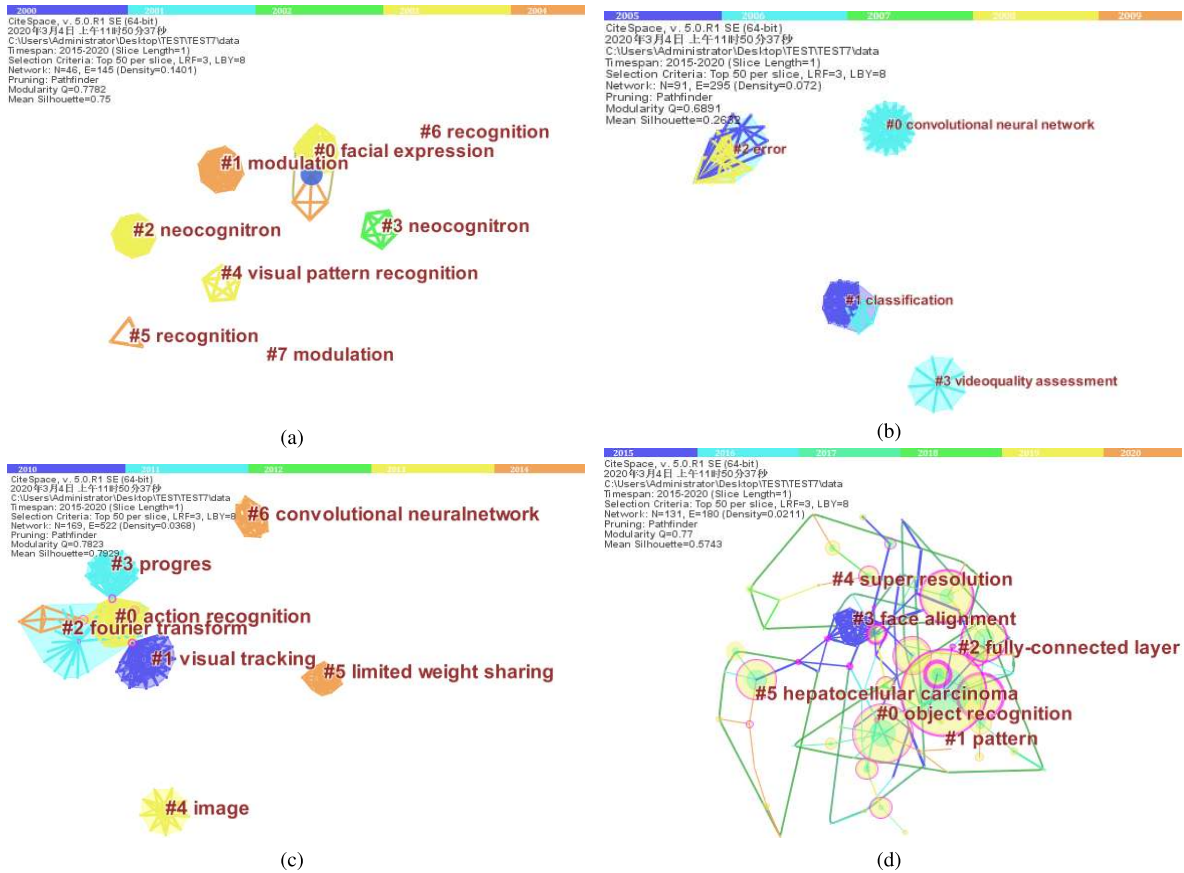
FIGURE 6. Dynamic evolution of the literature citation network of CNNs.

model increased. Sander Dieleman proposed a deep neural network model that uses translational symmetry and rotational symmetry to classify galaxy morphology. Applying these algorithms to larger training data sets is essential for analyzing the results of future surveys, such as large-scale weather measurement telescopes [37]. Since deeper neural networks are more difficult to train, K. He proposed a learning framework for residual networks. Experiments show that these residual networks are easier to optimize, and accuracy can be improved by greatly increasing the depth. This research is an important component in the citation network because it investigates the deep CNN and its application in image network classification thoroughly. It has been frequently cited since its publication five years ago.

The research focus on CNNs differs by country and region. Table 7 shows the high-frequency keywords and word frequencies for CNNs in China, the United States and Europe, in the four sub-sample periods. As can be seen from the table, prior to 2000, CNN research was still in its infancy. In China the literature on CNNs had not yet been studied, while in Europe and the United States, there was yet no obvious research focus. From 2000 to 2004, visual pattern recognition, motion recognition, and face detection became the areas of research focus. At the same time, early stage research related to CNNs appeared in European countries and

Japan. From 2005 to 2009, researchers in Australia, China and the United States also began to pay more attention to visual pattern recognition, video quality assessment, recurrent neural networks, decoding algorithms, image analysis, training methods, and convolutional codes. In recent years, deep CNNs have become the latest focus worldwide. Chinese researchers have focused on speech recognition, deep feature extraction and classification of hyperspectral images, Wi-Fi positioning and target tracking, land scene classification, image super-resolution algorithms and crack damage detection deep CNNs. Researchers in the United States have paid more attention to image network classification of deep CNNs, energy-efficient target recognition, human detection and activity classification, monitoring feature learning, fault diagnosis methods, video super-resolution, and crack damage detection. European researchers tend to focus on deep CNN pattern recognition, image recognition segmentation extraction, remote sensing architecture, and object detection.

Figure 7 shows a time-zone view of CNN citation clustering. The time-zone view arranges the citation network clusters in chronological order from left to right in order to list the documents in each cluster. It highlights evolution of the research field in terms of time periods. The words in the box as shown in the figure are the keywords of highly cited documents in each time period. As shown in Figure 4,

TABLE 7. Landmarks and turning points in citation networks.

Years	China		USA		Europe	
	Hot Keywords	freq-uency	Hot keywords	freq-uency	Hot keywords	freq-uency
2000-2004					recognition	6
					visual pattern recognition	2
2005-2009			classification	2	convolutional code	3
					convolutional neural network	2
2010-2014	convolutional neural network	5	Recognition communica-tion	4	feature, spa-tial filter	4
	neural network	4	classification	2	artificial neural network	2
2015-2020	convolutional neural network	1528	deep learning	1022	deep convo-lutional neu-ral network	121
	deep convo-lutional neu-ral network	121	classification	612	system, algorithm, feature, machine learning	705
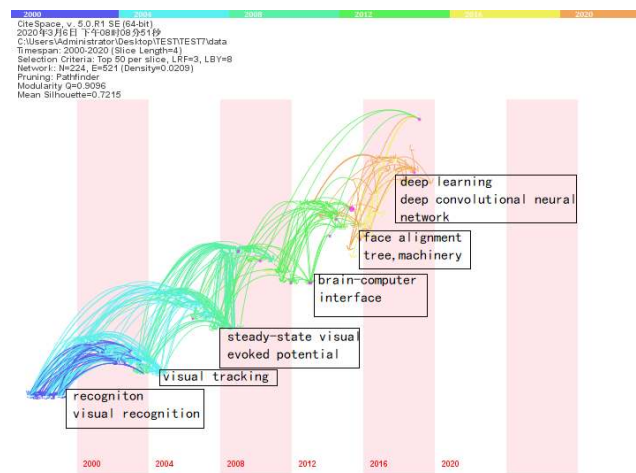
FIGURE 7. Time zone view of CNN citation clustering.

some early literature, for instance, Lecun Y (1995), Sahiner B (1996), Lawrence S (1997), Matsugu M (2002), Nair V (2010), Glorot X (2010), and Krizhevsky A (2012) has been cited in various periods, suggesting that the achieve-ments of these documents have laid the foundation for the long-term development of CNN research. The time zone view is a view that represents the evolution of knowl-edge from a time span. Can clearly show the update and mutual influence of the literature. Through the connection relationship of each time period, we can see the inher-itance relationship between the time periods. There are many connections between the nodes in the 2000 time zone and the nodes in the 2004 time zone, indicating that the convolutional neural network research in these two time periods has gradually passed on from target recognition and visual recognition to object positioning, and the inheritance relationship is strong. There are fewer connections between the nodes in the 2008 time zone and the

nodes in the 2018 time zone, indicating that the inheritance relationship between these two time periods is weak. In 2019, we focused on perfecting and improving the functions and forms of traditional neural networks. The keyword evolution of highly cited literature in various periods shows that the research of CNNs has gradually changed from target recog-nition, visual recognition, and object positioning tracking to the application of steady-state visual evoked potential, BCI, detection and diagnosis, and deep learning. This further validates the conclusions obtained by citation clustering.

V. CONCLUSION

Through citation network and cluster analysis, it can be shown that the hot areas of CNN research are target detec-tion and location, fault diagnosis, image recognition, and the systematic research of unbalanced problems in CNNs. Papers focusing on these topics were published after 2010 and formed the most important cluster in the citation network. They represent the frontier of CNN research.

A. OBJECT DETECTION AND LOCATION OF CNNs
In recent years, the use of CNNs for detection research has become a focus of academia. Shin H. C. and Roth H. R. studied two specific computer-aided detection (Cade) prob-lems: chest and abdominal lymph node (LN) detection and interstitial lung disease (ILD) classification. The researchers achieved mediastinal LN detection. They also reported the results of first five cross-validation classification when pre-dicting axial CT slices in ILD [38]. This article has been cited 905 times to date. Object detection based on machine learning is gaining more attention. Cheng G. proposed a novel method to learn the rotation-invariant CNN (RICNN) model to improve the performance of target detection. It is used for target detection in VHR optical remote sensing images [39]. Cha Y. J. proposed a vision-based method that uses a deep CNN architecture to detect concrete cracks without calculat-ing defect characteristics [40].

B. FAULT DIAGNOSIS AND IMAGE RECOGNITION OF CNNs
Fault diagnosis is critical to manufacturing systems, because it helps discover problems at an early stage. In recent years, intelligent fault diagnosis algorithms using machine learn-ing technology have achieved great success. Data-driven fault diagnosis has become a hot topic. However, traditional data-driven fault diagnosis methods rely on features extracted by experts. Wen L. proposed a new CNN based on LeNet-5 for fault diagnosis. The signal is converted into a 2-D image. This method can extract the features of the converted 2-D image and eliminate the influence of manual features [41]. Zhang proposed a new model based on deep learning to solve the problems of noisy environment and bearing intelli-gent fault diagnosis methods under different workloads [42]. Acharya UR uses a computer-aided diagnosis system and machine learning technology to automatically differentiate the types of the electroencephalogram (EEG) signals. This solution has conducted a CNN analysis of EEG signals, and

broke through the restrictions of recognition abnormalities that are brought about by direct visual inspection [43].

C. A SYSTEMATIC STUDY OF THE IMBALANCE PROBLEM IN CNNS

Classification of imbalance has become a common problem in research on classic machine learning. An unbalanced dataset contains an uneven distribution of data samples between various types, which makes it difficult to learn the concepts of a few types, and therefore poses a challenge to the learning algorithm. Mateusz B uses three benchmark datasets (MNIST, CIFAR-10, and ImageNet) to continuously reduce the complexity of and study the impact of imbalance on classification performance [44].

D. THE FUTURE OF CNNs

Over the past decade, the CNNs has demonstrated state of the art performance in AI missions. To speed up CNN testing and development, several software frameworks have been released, mainly for power-hungry CPUs and GPUs. In this case, Venieris Stylianos proposed that FPGA-based reconfigurable hardware constitutes a potential alternative platform that can be integrated into existing deep learning ecosystems for performance, while providing an adjustable balance between power consumption and programmability [45]. There is a close relationship between target detection and video analysis and image understanding, and so it has attracted much attention in recent years. Traditional object detection methods are based on hand-made features and shallow trainable architectures. With the rapid development of deep learning, Zhao introduced more powerful tools that can learn semantics, and more advanced and deeper functions, to solve problems in traditional architectures [46].

The use of CNNs in various industries around the world, particularly the construction and medical industries, is accelerating, as are the applications for them. CNNs are now used in object detection and positioning, fault diagnosis, image recognition, and optical remote sensing.

Based on the above analysis, it can be seen that CNN solves the problem of feature representation in the field of computer vision, in many research fields, such as image classification, object tracking and recognition, instance segmentation, image generation, human feature recognition and efficient data coding, it has become the most concerned research area. It also improves our quality of life in commercial applications such as face painting, traffic detection systems, decision support systems, and driverless cars. In recent years,The development of CNNs towards high performance, functionalization and ecology. Although there are many problems related to CNNs that need to be solved, this will not affect their further development and application in the fields of artificial intelligence in the future.

REFERENCES

[1] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol.*, vol. 160, no. 1, pp. 106–154, Feb. 1962.

[2] S. Lawrence, C. L. Giles, A. Chung Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE Trans. Neural Netw.*, vol. 8, no. 1, pp. 98–113, Jan. 1997.

[3] C. Nebauer, "Evaluation of convolutional neural networks for visual recognition," *IEEE Trans. Neural Netw.*, vol. 9, no. 4, pp. 685–696, Jul. 1998.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. NIPS*, 2012, pp. 1097–1105.

[5] Y. Zhang, M. J. Er, R. Zhao, and M. Pratama, "Multiview convolutional neural networks for multidocument extractive summarization," *IEEE Trans. Cybern.*, vol. 47, no. 10, pp. 3230–3242, Oct. 2017.

[6] Y. Zhang, K. Fu, H. Sun, X. Sun, X. Zheng, and H. Wang, "A multi-model ensemble method based on convolutional neural networks for aircraft detection in large remote sensing images," *Remote Sens. Lett.*, vol. 9, no. 1, pp. 11–20, Jan. 2018.

[7] X. Sun, C. Li, and F. Ren, "Sentiment analysis for Chinese microblog based on deep neural networks with convolutional extension features," *Neurocomputing*, vol. 210, pp. 227–236, Oct. 2016.

[8] J. Jin, K. Fu, and C. Zhang, "Traffic sign recognition with hinge loss trained convolutional neural networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 1991–2000, Oct. 2014.

[9] F. Li, Z. Wang, G. Qu, D. Song, Y. Yuan, Y. Xu, K. Gao, G. Luo, Z. Xiao, D. S. C. Lam, H. Zhong, Y. Qiao, and X. Zhang, "Automatic differentiation of glaucoma visual field from non-glaucoma visual filed using deep convolutional neural network," *BMC Med. Imag.*, vol. 18, no. 1, p. 35, Oct. 2018.

[10] Y. Liu, X. Chen, H. Peng, and Z. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Inf. Fusion*, vol. 36, pp. 191–207, Jul. 2017.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE CVPR*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[14] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE CVPR*, Boston, MA, USA, Jun. 2015, pp. 1–9.

[15] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. ACM MM*, New York, NY, USA, 2014, pp. 675–678.

[16] H. Cecotti and A. Graser, "Convolutional neural networks for P300 detection with application to brain-computer interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 433–445, Mar. 2011.

[17] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[18] Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.

[19] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[20] G. Hinton, L. Deng, D. Yu, G. Dahl, A.-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.

[21] C. Chen, "CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 57, no. 3, pp. 359–377, Feb. 2006.

[22] Y. Chen, X. Qin, J. Xiong, S. Xu, J. Shi, H. Lv, L. Li, H. Xing, and Q. Zhang, "Deep transfer learning for histopathological diagnosis of cervical cancer using convolutional neural networks with visualization schemes," *J. Med. Imag. Health Informat.*, vol. 10, no. 2, pp. 391–400, Feb. 2020.

[23] S. H. S. Basha, S. R. Dubey, V. Pulabaigari, and S. Mukherjee, "Impact of fully connected layers on performance of convolutional neural networks for image classification," *Neurocomputing*, vol. 378, pp. 112–119, Feb. 2020.

[24] C.-H. Tseng, C.-L. Hsieh, and Y.-F. Kuo, "Automatic measurement of the body length of harvested fish using convolutional neural networks," *Biosyst. Eng.*, vol. 189, pp. 36–47, Jan. 2020.

[25] E. Ba aran, Z. Cömert, and Y. Çelik, "Convolutional neural network approach for automatic tympanic membrane detection and classification," *Biomed. Signal Process. Control*, vol. 56, Feb. 2020, Art. no. 101734.

[26] J. Fan, W. Xu, Y. Wu, and Y. Gong, "Human tracking using convolutional neural networks," *IEEE Trans. Neural Netw.*, vol. 21, no. 10, pp. 1610–1623, Oct. 2010.

[27] L. Linhui, W. Mengmeng, D. Xinli, L. Jing, and Z. Yunpeng, "Convolutional neural network applied to traversability analysis of vehicles," *Adv. Mech. Eng.*, vol. 5, Jan. 2013, Art. no. 542832.

[28] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.

[29] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.

[30] M. Matsugu, "Hierarchical pulse-coupled neural network model with temporal coding and emergent feature binding mechanism," in *Proc. IJCNN*, Washington, DC, USA, 2001, pp. 802–807.

[31] J. Schemmel, F. Shurmann, S. Hohmann, and K. Meier, "An integrated mixed-mode neural network architecture for megasynapse ANNs," in *Proc. IJCNN*, Honolulu, HI, USA, 2002, pp. 2704–2709.

[32] M. Egmont-Petersen, D. de Ridder, and H. Handels, "Image processing with neural networks—A review," *Pattern Recognit.*, vol. 35, no. 10, pp. 2279–2301, 2002.

[33] H. Larochelle, D. Erhan, A. Courville, J. Bergstra, and Y. Bengio, "An empirical evaluation of deep architectures on problems with many factors of variation," in *Proc. ACM ICML*, Corvallis, OR, USA, Jun. 2007, pp. 473–480.

[34] A. Ahmed, K. Yu, W. Xu, Y. Gong, and E. Xing, "Training hierarchical feed-forward visual recognition models using transfer learning from pseudo-tasks," in *Proc. ECCV*, Berlin, Germany, Oct. 2008, pp. 69–82.

[35] H. Darren and K. Chen, "Exploring hierarchical speech representations with a deep convolutional neural network," in *Proc. UKCI*, 2011, p. 37.

[36] A.-R. Mohamed, G. Hinton, and G. Penn, "Understanding how deep belief networks perform acoustic modelling," in *Proc. IEEE ICASSP*, Kyoto, Japan, Mar. 2012, pp. 4273–4276.

[37] S. Dieleman, K. W. Willett, and J. Dambre, "Rotation-invariant convolutional neural networks for galaxy morphology prediction," *Monthly Notices Roy. Astronomical Soc.*, vol. 450, no. 2, pp. 1441–1459, Jun. 2015.

[38] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[39] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.

[40] Y.-J. Cha, W. Choi, and O. Büyüöztürk, "Deep learning-based crack damage detection using convolutional neural networks," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 32, no. 5, pp. 361–378, Mar. 2017.

[41] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5990–5998, Jul. 2018.

[42] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mech. Syst. Signal Process.*, vol. 100, pp. 439–453, Feb. 2018.

[43] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Comput. Biol. Med.*, vol. 100, pp. 270–278, Sep. 2018.

[44] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Netw.*, vol. 106, pp. 249–259, Oct. 2018.

[45] S. I. Venieris, A. Kouris, and C.-S. Bouganis, "Toolflows for mapping convolutional neural networks on FPGAs: A survey and future directions," *ACM Comput. Surv.*, vol. 51, no. 3, pp. 1–39, Jul. 2018.

[46] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019.
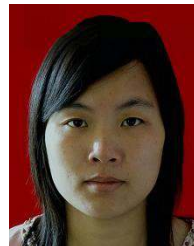
HUIE CHEN received the M.S. degree from Guangdong Polytechnic Normal University, Guangdong, in 2012. Since 2015, she has been a Computer Application Instructor with the Computer Engineering Department. She is the author of two books. Her research interest includes computer science.

ZHENJIE DENG received the B.S. degree in communication engineering from Jilin University, China, in 2011, and the Ph.D. degree in signal and information processing from the Institute of Information Engineering, Chinese Academy of Sciences, in 2016. He is currently pursuing the Ph.D. degree with the Institute of Acoustics, Chinese Academy of Sciences. He is also a Lecturer with the Guangdong University of Finance, Guangdong, China. In 2015, he joined the Institute of Information Engineering, Chinese Academy of Sciences. His research interests include multimedia communication and heterogeneous networks.

• • •