Research on Video Quality Diagnosis System Based on Convolutional Neural Network

Yi Hu [1], Xiaodong Zhan [2]

[1,2] Training Center, Beijing Polytechnic, Beijing, China, 100176
[1] dhuyi@bpi.edu.cn*
* corresponding author

(Received: June 19, 2023; Revised: July 17, 2023; Accepted: August 31, 2023; Available online: September 10, 2023)

Abstract

In the era of rapid development in modern society, there is an escalating demand for high-performance products. However, this quest for excellence often encounters persistent quality issues during practical applications. Hence, to enhance the user experience and rectify this situation, this paper proposes a Convolutional Neural Network (CNN)-based Video Quality Diagnosis System. The system's design encompasses a myriad of construction methodologies, primary framework structures, and associated databases. This research primarily focuses on video quality during video conferencing as the subject of investigation, with the aim of constructing a Video Quality Diagnosis System grounded in CNN theory. The objective is to provide real-time identification, analysis, and enhancement of video quality, thereby offering timely solutions to issues that arise in the video conferencing experience. In this endeavor, the research amalgamates cutting-edge technology and meticulous study to create a smoother and more immersive video conferencing experience for individuals and organizations. By addressing the frequently encountered video quality issues, we hope to facilitate more effective and engaging communication on a global scale, bridging the gap between user expectations and practical implementation and paving the way for a future where video quality problems are a thing of the past.

Keywords: Convolutional Neural Network, Video Quality Diagnosis System, Internet

1. Introduction

Convolutional Neural Networks (CNNs) have made indelible strides in the field of computer vision, representing a profound shift in the way we approach visual data analysis. Their remarkable success has eclipsed traditional computer vision algorithms in numerous applications, enabling machines to grasp intricate visual patterns and features with unparalleled accuracy. Simultaneously, as the fabric of modern society continues to weave itself with the threads of advanced network technologies, video conferencing systems have evolved into a ubiquitous and vital conduit for disseminating multimedia information. These systems transcend geographic boundaries and serve as the lifeblood of real-time, remote communication across diverse domains, from business and education to healthcare and social interaction. In this scholarly endeavor, we embark on a mission to harness the transformative potential of Convolutional Neural Networks to address the perennial challenges that beset video quality within the context of video conferencing.

Our overarching objective is to conceive, design, and implement a Video Quality Diagnosis System rooted firmly in the principles of CNNs. By doing so, we aspire to transcend the boundaries of conventional video conferencing limitations and propel it into an era characterized by superior quality, reliability, and efficiency. This pursuit resonates profoundly with the modern digital landscape, where video conferencing has transitioned from being a mere convenience to an indispensable tool that underpins global communications, empowers remote workforces, and facilitates seamless educational experiences.

The intersection of CNNs and video conferencing represents a convergence of cutting-edge technology and contemporary communication imperatives. CNNs, initially inspired by the human visual system, have proved themselves as versatile, deep learning frameworks capable of decoding complex visual information. They excel in tasks ranging from image classification and object detection to facial recognition, pushing the boundaries of what machines can achieve in understanding the visual world. Harnessing this deep learning paradigm within the domain of video conferencing presents an auspicious opportunity to confront the multifaceted challenges that users frequently grapple with. These challenges span the gamut from image resolution and video compression artifacts to bandwidth constraints and unpredictable network conditions, all of which can adversely impact the quality and reliability of video communication.

Video conferencing, once considered a luxury, has metamorphosed into an indispensable tool for modern personal and professional interactions. The seismic shift toward remote work, online education, and virtual socialization, accentuated by the global events of recent years, underscores the critical importance of optimizing video conferencing quality. Effective communication, the exchange of ideas, and the preservation of engagement, comprehension, and overall user satisfaction all hinge on the quality of the video conferencing experience.

In the pages that follow, we shall embark on a comprehensive exploration of our proposed Video Quality Diagnosis System, grounded firmly in the bedrock of CNN technology. Our research journey will navigate the intricate architecture of CNNs, traverse the practical terrain of constructing this system, scrutinize the pertinent databases essential to our endeavor, and illuminate the methodologies that underpin the real-time diagnosis and enhancement of video quality during video conferencing sessions. The insights garnered from this scholarly pursuit aspire to make a significant contribution to the ongoing quest for a more seamless and immersive video conferencing experience. We anticipate that our findings will resonate not only with professionals and educators but also with individuals who increasingly rely on this mode of communication in today's interconnected and rapidly evolving world. Ultimately, our aim is to empower a global audience with the tools to transcend the constraints of distance and elevate the quality of their interactions through the power of Convolutional Neural Networks.

2. Literature Review

2.1. Convolutional Neural Networks in Computer Vision

The emergence of Convolutional Neural Networks (CNNs) has reshaped the landscape of computer vision, representing a paradigm shift in visual data analysis. These deep learning architectures, inspired by the human visual system, have garnered remarkable success in various applications. CNNs excel in tasks such as image classification, object detection, and facial recognition, enabling machines to discern complex visual patterns and features with unparalleled accuracy. The transition from traditional computer vision algorithms to CNNs has been marked by improved performance and efficiency, making them a cornerstone of modern computer vision research [1].

2.2. Video Conferencing as a Communication Mainstay

In parallel, the development of network technologies has catapulted video conferencing into the mainstream as a critical means for disseminating multimedia information and fostering real-time communication. Video conferencing transcends geographical constraints, bridging the divide between individuals and organizations across the globe. This mode of communication has evolved beyond being a mere convenience, with its pivotal role highlighted in diverse domains, including business, education, healthcare, and social interaction. Recent global events have further underscored its significance as a linchpin of remote work, virtual education, and remote healthcare delivery [2].

2.3. The Challenge of Video Quality in Video Conferencing

However, despite the ubiquity and indispensability of video conferencing, it is not without its challenges. Users frequently encounter issues related to video quality during video conferencing sessions. These challenges encompass a spectrum of factors, including image resolution, video compression artifacts, bandwidth constraints, and the volatility of network conditions. Poor video quality can lead to decreased user satisfaction, diminished engagement, and hindered communication effectiveness. It is imperative to address these challenges comprehensively to unlock the full potential of video conferencing in the digital age [3].

2.4. Bridging the Gap with CNN-Based Video Quality Diagnosis

This paper seeks to bridge the gap between the transformative capabilities of CNNs in computer vision and the imperative to enhance video quality in video conferencing. By leveraging the deep learning capabilities of CNNs, we aim to develop a Video Quality Diagnosis System that operates in real-time, diagnosing and rectifying video quality issues during video conferencing sessions. Our research aligns with the broader objective of enhancing the user experience, enabling seamless and immersive communication across various domains. In the subsequent sections, we will delve into the theoretical foundations, practical implementation, and methodologies underpinning our proposed system, with the aspiration that our findings will contribute significantly to the advancement of video conferencing technology and its application in the modern digital landscape [4].

3. Convolutional Neural Networks in Brief

In recent years, convolutional neural networks have become more and more complex as the number of researchers in the field related to convolutional neural networks has increased and the technology is changing day by day. From the initial 5-layer, 16-layer, to the 152-layer ResNet proposed by MSRA [5,6] or even thousands of layers networks have become commonplace by a wide range of researchers and engineering practitioners.

A simple convolutional neural network is composed of various layers arranged in a sequence, each layer in the network uses a differentiable function to pass data from one layer to the next. Convolutional neural networks consist of three main types of layers: convolutional layers, pooling layers, fully connected layers. By stacking these layers together, a complete convolutional neural network can be constructed. As shown in Figure 1.

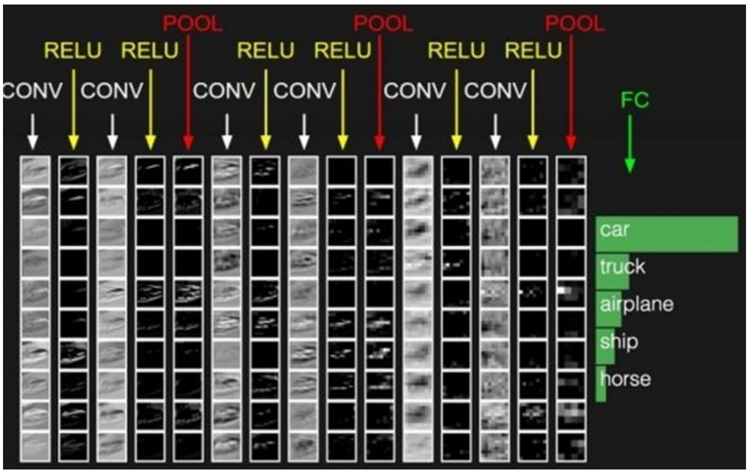

Figure 1. Complete convolutional neural networks

4. Video Quality Diagnosis System

In recent years, deep learning convolutional neural networks have developed rapidly in the field of machine vision, which can build very complex models by simulating the human nervous system to analyze and interpret data, have powerful expression capabilities to handle complex practical application scenarios, especially convolutional neural networks, which are now widely used in the field of pattern recognition, have shown superior performance in various tasks of computer vision, their unique deep The unique deep structure can effectively learn the complex mapping between input and output. Therefore, the video quality diagnosis algorithm based on deep learning convolutional neural network can extract more features of abnormal video, adapt to more complex rules, detect and classify the abnormal types more accurately is completely feasible [7,8]. Deep learning convolutional neural network algorithm is based on a large amount of data, so except deep learning convolutional neural network algorithm applied to the field of video quality diagnosis, the following points need to be done.

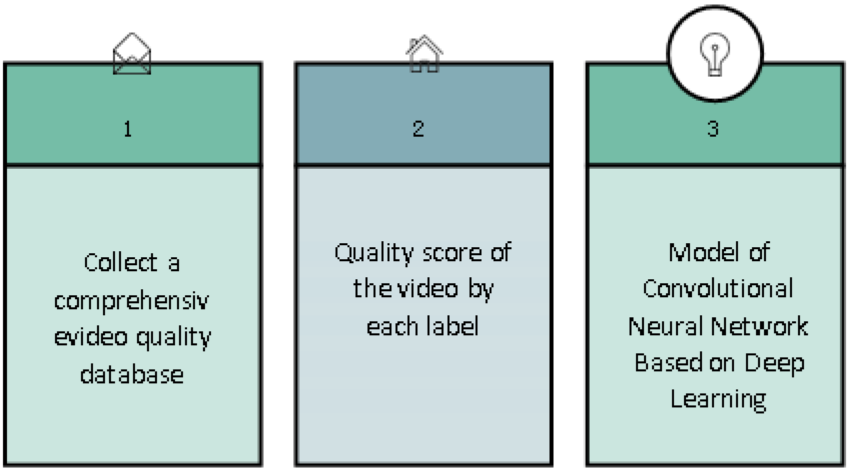
1
Collect a comprehensiv evideo quality database
2
Quality score of the video by each label
3
Model of Convolutional Neural Network Based on Deep Learning

Figure 2. Focus of building a convolutional neural network video diagnosis system

First, collect a comprehensive video quality database to form a deep learning convolutional neural network dataset, which is a prerequisite to ensure the generalization performance of the algorithm; second, the video database is organized into a dataset that conforms to the deep learning convolutional neural network model, then a set of schemes is developed, then each video in the video quality database is labeled with abnormal types, then the video is scored for quality according to each type of label, the process can be be called subjective evaluation of video; third, establish a deep learning convolutional neural network model based on which can accurately predict the video abnormality types and their video quality scores, the predicted results should be consistent with the manually labeled results, the process can be called an objective evaluation method of video quality [9,10].

5. Methodology

5.1. Pool layer of Convolutional neural network in video quality diagnosis

TUsually, a pooling layer is inserted periodically between successive convolutional layers. It serves to gradually reduce the spatial size of the data body, which in turn reduces the number of parameters in the network, making it less computationally resource intensive and also effective in controlling overfitting, as shown in Figure 3. The pooling layer usually uses the MAX operation, which operates independently on each slice of the input data body to change its spatial dimensions. The most common form is to use a filter of size 2x2 to downsample each depth slice in steps of 2, discarding 75% of all activation information in it. Each MAX operation takes the maximum value from 4 numbers (i.e., in some 2x2 region of the depth slice) [11,12]. Note that the number of channels in the data body remains constant during the pooling process.

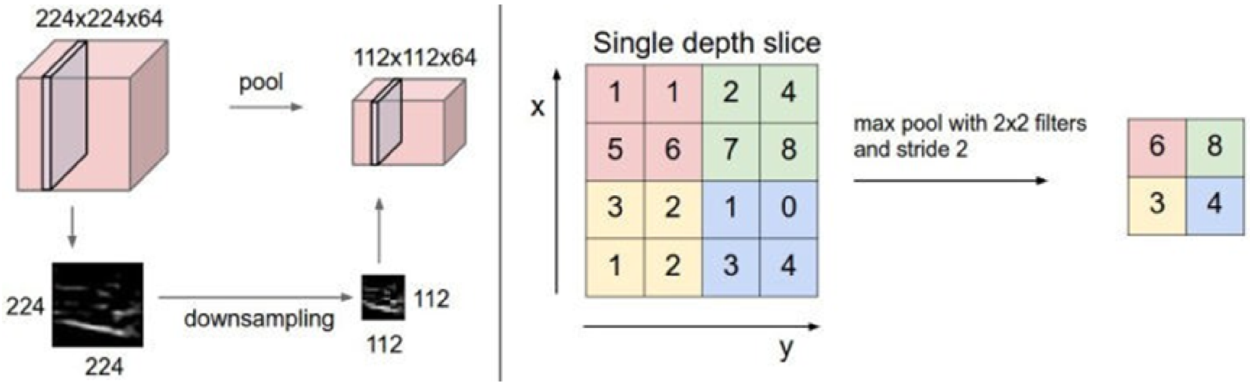

Figure 3. Role of the pooling layer

5.2. Convolutional neural network video diagnosis algorithm in this paper

The algorithm in this paper is a multi-task multi-label deep learning network model implemented based on the VGG-16 convolutional neural network as a prototype, using the convolutional network for feature map extraction of multi-frame images, which are then connected together to do anomaly type classification and quality scoring regression tasks. To extend the training set and testing environment, pyramid pooling (SPP) is introduced, the size of the input video images can be unrestricted. Different size images will get different size feature maps after convolutional layers. Since the number of parameters of the fully connected layer is fixed, it cannot be connected with the fully connected layer, the different size feature maps can be converted into feature vectors of the same size by means of pyramid pooling [13,14]. The network structure block diagram is shown in Figure 4.

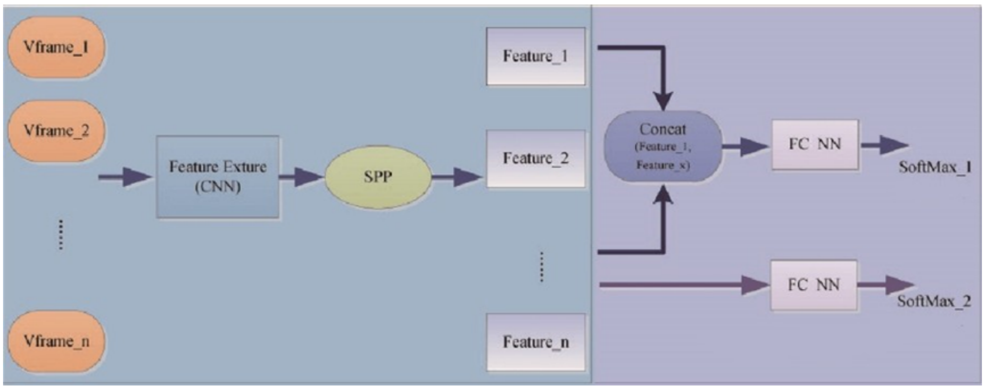

Figure 4. Network structure block diagram

During the training process, the anomaly type classification is trained first, on this basis the quality score is trained based on the classification results. When using the trained model for prediction, video sequences of different sizes can be input, the abnormality type of the video can be obtained after model processing. If multiple consecutive frames have the same abnormality, it can be assumed that there is some kind of fault in the system, the quality score corresponding to each frame can be obtained by statistical analysis of the relative quality of each frame of the video.

6. Overall System Design Scheme

6.1. Overall System Architecture

The video quality diagnosis system adopts a modular design, including five modules: video frame interception module, OpenCV image processing module, image abnormality detection module, abnormality recording and display module, abnormality alarm module. The workflow of the video quality diagnosis system is mainly as follows: firstly, the required video frames are obtained from the stored surveillance video and saved as Mat entities: secondly. With the help of 0penCV image processing technology. Get the spatial domain structure information that can represent the image content: Finally, give the OpenCV processed image spatial domain structure information to the designed image abnormality detection algorithm for different faults to realize the automatic detection of image abnormality[15,16], the main functions of each module of the video quality diagnosis system are as follows.

1) Intercept video frame module: The function of this module is mainly to intercept the stored video image frames, the acquired color image signal is color-separated, separately amplified and corrected to obtain RGB. and the basic image information and pixel data are encapsulated. as the image data to be detected.
2) OpenCV image processing module: This module makes full use of the image processing library provided by OpenCV to pre-process the image to be detected, such as converting grayscale images, segmenting image channels, color clustering, and image space conversion. and obtaining image pixel channel values and other required image structure information.
3) Image anomaly detection module: this module is the core module of the whole video quality diagnosis system, the main function is to detect the image data information after OpenCV processing, determine whether the detected video frames have abnormal abnormalities including video signal lack of clarity, brightness noise, snow, streaks, color bias, screen freeze, PTZ motion out of control and other abnormalities each image anomaly detection is done by (1) independent algorithm class to complete the input set of video frames for sequential detection[17,18], return a variety of anomaly detection results.
4) Anomaly logging and display module: this module accepts the return value of the algorithm class of the image anomaly detection module, specifies the description and stores it in the log, so that it can be called when querying the detection results.

5) Abnormal alarm module: this module provides feedback on the detection results to detect abnormal images of the camera according to the abnormal type of mark to remind remind maintenance personnel to solve the problem in a timely manner.

6.2. Functional Module

Video quality diagnosis system mainly includes 6 modules.

1) Each quality diagnosis algorithm uses a single frame or two frames of video frame images at close moments to complete various diagnoses, which do not depend on the background image, so there is no need for background modeling and the update of kenjing, reducing the false detection caused by unreasonable background model.

2) Video quality diagnosis is completed by multiple servers, since each server is equally configured, the diagnosis task is equally distributed to each device. With the corpse only need to set up the pre- program of polling detection, LOTUS will start the task according to the start time set in the pre-program, without the need for manual intervention.

3) Diagnostic results are stored for each recent detection, regardless of whether the camera is working properly or not. Users can query detection records by region[19,20], fault type.

4) Fault information stores all the historical fault information of the problematic camera, also contains screenshots of the video frames when the fault is detected, which is easy for the user to view visually. The corpse can query the fault information records of a certain time period by camera, region, fault type. At the same time can be based on the stored video screenshots to determine whether the system is misdetected, can allow misdetected cameras to learn to reduce the probability of misdetection.

5) In addition, users can count the number of failures and failure rate of cameras in different areas and brands in different time periods according to their own needs, which can be displayed in different forms to facilitate users to understand the operation status of cameras.

6) In the pre-program management part, the user can set the inspection start time, detection items. In terms of algorithm parameter setting [21,22], at the early stage of system operation, the algorithm parameters of each camera are set according to the unified threshold; after the system runs for a period of time, the threshold of each detection algorithm of each camera will be adjusted due to various reasons such as equipment quality, service life, site environment and power supply, transmission line., that is, each camera has its own optimal algorithm threshold. In addition, users can set the thresholds of algorithm parameters applicable to special weather such as rain and snow to cope with these bad weather.

6. Conclusion

In the ever-evolving landscape of technology, one area that has seen remarkable progress is video surveillance. Over the years, video surveillance technology has become an integral part of our daily lives, contributing significantly to various sectors, including security, transportation, and even personal monitoring. However, as the utilization of video surveillance has proliferated, so too has the need for efficient and accurate quality diagnosis and control.

Traditionally, the quality of video feeds required human operators to monitor them, a labor-intensive and often error-prone process. These operators would need to constantly sift through hours of footage to identify any anomalies or issues, making the process not only time-consuming but also susceptible to human error. Recognizing these challenges, this paper presents an innovative solution in the form of a video image detection system based on convolutional neural networks (CNNs).

The introduction of CNNs into video surveillance represents a groundbreaking advancement that aims to circumvent the limitations of traditional video diagnosis methods. By leveraging the power of artificial intelligence and deep learning, this system can automatically and efficiently analyze video feeds, detecting and diagnosing potential quality issues with a high degree of accuracy. It does so by mimicking the human brain's ability to recognize patterns, thereby making it an invaluable tool in the realm of video quality diagnosis.

This novel approach not only streamlines the quality control process but also opens up new avenues for enhancing video surveillance technology further. It offers a promising path towards making video surveillance systems more autonomous, reducing the burden on human operators, and ultimately improving the overall reliability and effectiveness of video surveillance across various applications. As we continue to witness the rapid evolution of technology, innovations like the CNN-based video image detection system showcased in this paper serve as a testament to our ongoing commitment to harnessing technology's potential for the betterment of society.

References

[1] MLeCun, Y, Boser, B, Denker, J. S, Henderson, D, Howard, R. E, Hubbard, W, Jackel, L. D. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, vol. 1, no. 4, pp. 541– 551, 1989.

[2] A. Faizah, P. H. Saputro, A. J. Firdaus, and R. N. R. Dzakiyullah, “Implementation of the convolutional neural network method to detect the use of masks,” IJIIS: International Journal of Informatics and Information Systems, vol. 4, no. 1, pp. 30–37, 2021. doi:10.47738/ijiis.v4i1.75

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *In Advances in neural information processing systems*, vol. 1, no. 1, pp. 1097–1105, 2012.

[4] C.-W. Hung, “Application of Quality Function Development Method to Establish Application of New Product Development Information System”, Int. J. Appl. Inf. Manag., vol. 1, no. 1, pp. 23–27, Apr. 2021.

[5] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *In CVPR*, vol. 1, no.1, pp. 770-7778, 2016.

[6] N. Ky Vien, “Modelling The Relationship of Perceived Quality, Destination Image, and Tourist Satisfaction at The Destination Level”, Int. J. Appl. Inf. Manag., vol. 1, no. 4, pp. 165–172, Aug. 2021.

[7] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv*, vol. 1, no.1, pp 1409-1556, 2014.

[8] H. Y. Su, “Algorithm analysis of clothing classification based on Neural Network,” Journal of Applied Data Sciences, vol. 3, no. 2, pp. 82–88, 2022. doi:10.47738/jads.v3i2.61

[9] C.Szegedy, W.Liu, Y.Jia, P.Sermanet, S. Reed, D. Anguelov, D.Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, no. 1, pp. 1–9, 2015.

[10] F. Fang and X. Zhang, “Embedded image recognition system for lightweight convolutional Neural Networks,” Journal of Applied Data Sciences, vol. 3, no. 3, pp. 102–109, 2022. doi:10.47738/jads.v3i3.62

[11] S.Ioffe and C.Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift.*In ICML,* vol. 1, no.1, pp. 448-456, 2015.

[12] R. Xin, J. Zhang and Y. Shao, "Complex network classification with convolutional neural network," in Tsinghua Science and Technology, vol. 25, no. 4, pp. 447-457, Aug. 2020, doi: 10.26599/TST.2019.9010055.

[13] C.Szegedy, V.Vanhoucke, S.Ioffe, J.Shlens, and Z.Wojna. Rethinking the inception architecture for computer vision. *arXiv preprint arXiv*, vol. 1, no.1, pp. 2818-2826, 2015.

[14] J. Huang, S. Huang, Y. Zeng, H. Chen, S. Chang and Y. Zhang, "Hierarchical Digital Modulation Classification Using Cascaded Convolutional Neural Network," in Journal of Communications and Information Networks, vol. 6, no. 1, pp. 72-81, March 2021, doi: 10.23919/JCIN.2021.9387706.

[15] C.Szegedy, S.Ioffe, V. Vanhoucke, A. Alemi. Inceptionv4, inception-resnet and the impact of residual connections on learning. *arXiv*, vol. 31, no. 1, pp. 4278-4284, 2016.

[16] Y. Lou, R. Wu, J. Li, L. Wang, X. Li and G. Chen, "A Learning Convolutional Neural Network Approach for Network Robustness Prediction," in IEEE Transactions on Cybernetics, vol. 53, no. 7, pp. 4531-4544, July 2023, doi: 10.1109/TCYB.2022.3207878.

[17] G. Huang, Z. Liu, K. Q. Weinberger, L. Maaten. Densely connected convolutional networks. *In CVPR*, vol. 1, no.1, pp. 4700-4708, 2017.

[18] P. Zhang, X. Wang, W. Zhang and J. Chen, "Learning Spatial–Spectral–Temporal EEG Features With Recurrent 3D Convolutional Neural Networks for Cross-Task Mental Workload Assessment," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 27, no. 1, pp. 31-42, Jan. 2019, doi: 10.1109/TNSRE.2018.2884641.

[19] S. Xie, R. Girshick, P. Dollar, Z. Tu, K. He. Aggregated residual transformations for deep neural networks. *In CVPR*, vol. 1,

no. 1, pp. 1492-1500, 2017.

[20] Jie Hu, Li Shen, Gang Sun. Squeeze-and-Excitation Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* vol. 1, no.1, pp. 7132-7141, 2018.

[21] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv: 1704.04861,* vol. 1, no. 1, pp. 1-9, 2017.

[22] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, Jian Sun. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, vol. 1, no.1, pp. 6848-6856, 2018.