ANALYSIS ON APPLICATIONS OF MACHINE LEARNING FOR THE COMPRESSION OF BRAIN VIDEO

Atif Ali Mohammed, Information technology, Charles Sturt University, aliphd59@gmail.com.

ABSTRACT

Due to advancements in multimedia technology, video transmission now necessitates greater capacity for storage, more bandwidth, and faster transmission speeds. Compression methods for videos are crucial since they allow for content to be transmitted over less network capacity and with less room for storage. In addition to its widespread use in fields such as image and video analysis, NLP, speech recognition, etc., machine learning has just lately begun to find applications in BCI. Machine learning, which expedites training and boosts performance by drawing on information and precedents from similar situations, might be very helpful in BCI for accommodating differences between users and between jobs. Integrating cutting-edge machine learning technologies is another way to reap the benefits of both fields. This paper provides a comprehensive literature review and bibliometric study of all Machine Learning (ML) techniques that have been used to video compression in the recent past. Popular academic databases include Scopus and Web of Science. This analysis makes use of the data retrieved from them. There are two sorts of analysis done on the extracted data. You'll find both quantitative and qualitative findings. Quantitative analysis is the examination of records according to criteria such as the number of citations they include, the relevance of their keywords, the journals they were published in, and the countries in which those journals are based. The benefits, drawbacks, and difficulties of utilising ML-based techniques to video compression are detailed in the qualitative analysis provided.

Key words: Video compression, Machine Learning methods, Brain,Magnetic resonance imaging (MRI).

1. INTRODUCTION

A video is made up of many moving pictures (called frames). Intra- and inter- [P- and B-] frames of a video sequence can be used to examine the large and small differences introduced by camera movement, respectively. For subsequent frames, I-frames serve as the primary reference frame. A P-frame can be anticipated from an I- or P-frame that came before it. Interpolation between forward and reverse frames is used to create the B-frames. It is the frame type that decides how much video can be compressed and what kind of quality is produced. The quality of a video appears to increase as the number of intra-frames increases. Increasing the amount of inter-frames, on the other hand, improves the compression ratio but degrades the video quality. The frames of a GoP are formatted similarly to IP, IBP, or IBBP packets. The data between frames has traditionally been compressed using prediction analysis for codecs like MPEG-4 and H.264. The method of motion estimation and motion correction makes this possible. A motion vector is a representation of the motional relationship between two frames.

Medical imaging, such as an MRI or a CT scan, produces digital pictures of the human body. Large amounts of data are generated by these imaging techniques, so compression is required for convenient storage and transmission [1]. While modern compression techniques can reduce file sizes, they typically do so at the expense of quality. In some fields of medicine, it may be necessary to preserve image quality just where it is most needed, namely, in

areas that are clinically significant. A standard X-ray for a medical facility will be 2048 pixels wide by 2560 pixels tall, and will have 12 bits of resolution. An equivalent file would be 10,485,760 bytes in size. A typical 16-bit mammography image could be 4500 pixels by 4500 pixels in dimension, or 40 MB in file size [2]. This impacts how much storage space is required for a given image and how long it takes to transfer. The amount of digital imaging produced by hospitals and their new filmless radiology departments has increased at an incredible rate, despite the fact that disc capacity has grown over time. The problem of sending the photographs would still be present even if storage were inexhaustible.
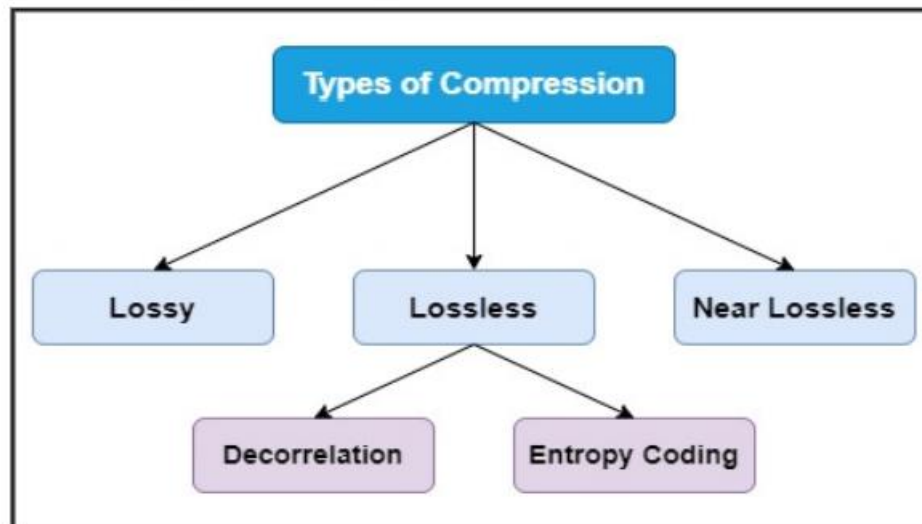

Figure 1. Types of compression.

In the last few decades, data compression has seen a dramatic transformation. Figure 1 illustrates how lossless and lossy data compression can be differentiated. Furthermore, new compression methods are now being developed that support a new compression type known as Near lossless compression. With lossless compression, you may decompress an image or video and get it back to its original size and quality without losing any quality. When a lossy compression method finds duplicate data in an image's pixels, it will attempt to get rid of them. Therefore, lossy compression is rarely used for text documents or software but is commonplace for multimedia content such as audio, video, and still photographs. Because of the visual processing limits of the human brain, lossy compression methods can effectively eliminate visible detail. Currently available software that uses lossy compression includes the JPEG (Joint Photographic Expert Group) [4], JPEG2000 [5], BPG (Better Portable Graphics), MPEG (Motion Picture Experts Group) [6], MPEG CDVS (MPEG Compact Descriptors for Visual Search) [7], MPEG CDVA (MPEG Compact Descriptors for Visual Analysis) [7], and MP3 [8].

2. LITERATURE REVIEW

Recent years have seen a meteoric rise in the academic and professional adoption of machine learning (ML) [1, 2, 3, 4, 5, 6]. Numerous ML approaches are available for this purpose [7,8,9], and ML is commonly employed in such contexts. representation learning (RL) is another name for ML (RL). Both the unexpected expansion of data collection capabilities and the astounding advances made in hardware technologies, such as High Performance Computing (HPC), have contributed to the continued creation of creative studies in the fields of deep and distributed learning [10].

To some extent, ML may be traced back to the classic neural network, but it is far more efficient. An additional advantage of ML is that it employs transformation and graph technologies concurrently to build multi-layer learning models. The most up-to-date ML methods have shown exceptional performance in many areas [11,12,13,14], including audio and speech processing, visual data processing, and NLP.

In most situations, an ML algorithm's success is highly dependent on the precision with which the input data is represented. It has been shown that when data is represented well, it performs better than when it is represented poorly. Therefore, feature engineering has been a major trend in ML research for a considerable amount of time, guiding a great

deal of research. The goal of this strategy is to create features out of unprocessed data. It also usually takes a lot of people working together and is highly domain-specific. In the realm of computer vision, many different feature types have been introduced and studied. Some examples include the histogram of oriented gradients (HOG) [15], the scale-invariant feature transform (SIFT) [16], and the bag of words (BoW) [17]. When something new is developed and proven effective, it becomes the focus of research for decades to come.

When it comes to DL algorithms, feature extraction is largely accomplished mechanically. This motivates academics to find ways to extract discriminative characteristics with as little time, effort, and domain expertise as feasible [18]. Lower-level feature extraction is performed in the first layers of these methods' multi-layer data representation architecture, whereas higher-level feature extraction is performed in the last layers. Keep in mind that AI served as the original inspiration for this architecture's emulation of the human brain's primary sensory regions (AI). The human brain has the innate ability to automatically extract data representation from a variety of scenarios. In particular, the objects' classifications are the outcome, while the scene data received serves as input. The method used here is a close approximation of how the human brain operates. By extension, it highlights DL's primary advantage.

Due to its significant success, DL has quickly become one of the most talked-about topics in the field of ML. In this study, we provide a high-level introduction to DL, covering its fundamental ideas and architectures as well as its difficulties, practical applications, computational tools, and evolution matrix. One of the most common types of DL networks is the convolutional neural network (CNN)

[19, 20]. As a result of CNN's coverage, DL is now widely used. CNN's primary benefit over its forerunners is that it can recognise important elements automatically, without human supervision. Thus, we have provided a comprehensive overview of CNN by detailing its essential features. We have also supplied comprehensive explanations of the most common CNN architectures, beginning with the AlexNet network and ending with the High-Resolution network (HR.Net).

Many articles that provide a comprehensive overview of DL have been written and presented in recent years. One viewpoint has been presented on a variety of topics, such as a review of CNN architectures [21], deep learning for plant disease classification [22], deep learning for object recognition [23], deep learning applications in medical picture analysis [24], and so on. While covering some of the most significant parts of the subject, these assessments do not provide readers with a thorough understanding of DL ideas, detailed research gaps, computational tools, or DL applications. The fundamentals of DL—its principles, problems, and applications—must be mastered before moving on to the more advanced topics. Learning about DL, its research gaps, and its applications takes a lot of work and a mountain of research papers. Instead of trying to learn everything about DL from multiple sources, we advise using a single document as a thorough overview as a starting point. Through this analysis, we hoped to shed light on some of the most important topics in DL, including its open challenges, applications, and computational tools. Our findings can potentially be used as a starting point for research into related areas of DL.
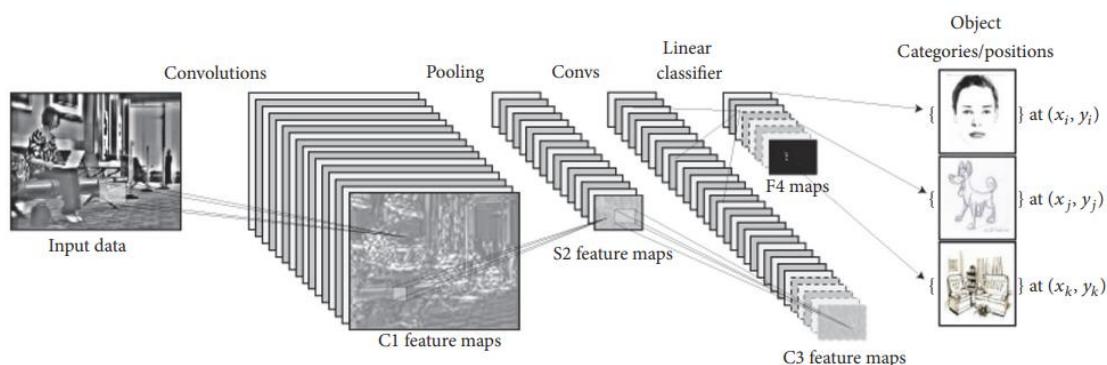
3. RESEARCH METHODOLOGY

Figure 2: An example CNN architecture for a computer vision task (object detection).

There are three types of neural layers that make up a convolutional neural network (CNN): (i) convolutional layers, (ii) pooling layers, and (iii) fully connected layers. There are different kinds of layers, and they all serve different purposes. Figure 2 depicts a common CNN architecture used for this type of object detection in pictures task. At each layer of a convolutional neural network (CNN), the input volume is turned into an output volume of neuron activity, with the final fully connected layers producing a mapping from the input data to a 1D feature vector. Face identification, object detection, powering vision in robotics, and self-driving automobiles are just a few examples of the many areas where CNNs have proven to be highly effective in computer vision applications.

(i) Convolutional Layers. Convolutional networks (CNNs) generate multiple feature maps by convolving the entire input image with the intermediate feature maps, each of which is generated by a different kernel in the convolutional layer. Many papers (e.g., [23, 24]) have suggested using convolution operations instead of fully connected layers to speed up the learning process.

(ii) Pooling Layers. The purpose of the pooling layer is to make the input volume for the subsequent convolutional layer smaller in both width and height. The depth of the volume is unaffected by the pooling layer. This layer performs an operation also known as subsampling or downsampling due to the inherent loss of data in downsampling. In the long run, however, this type of loss is beneficial to the network because it helps to prevent overfitting and lightens the workload of the subsequent layer in the network. Typically, people use either average pooling or maximum pooling.

(iii) Fully Connected Layers. After a series of convolutional and pooling layers, fully connected layers are used to carry out the higher-level reasoning necessary for a neural network to function. Neurons in a completely connected layer are directly coupled to all of the activation in the layer below them, as the name implies. As a result, their activation can be calculated by multiplying a matrix by a bias offset. After a while, the 2D feature maps are transformed by the fully connected layers into a 1D feature vector.

3.1 Data Analysis Procedure

Graphs are a great way to visualise data for easy comprehension and analysis. It's useful for drawing conclusions, making decisions, creating forecasts, etc. In this paper, we use the programmes "VoSViewer," "Gephi," and "BibExel" to conduct a bibliometric analysis of the literature on deep neural network techniques for video compression. Common multidimensional data visualisation software falls under this category. VoSViewer is widely used in the field of bibliometrics as a visualisation tool. Keywords, citations, publication venues, authors, co-authorship, shared references, etc. can all be used to construct unique networks. All things have been simplified into circles. They are connected to one another by linkages. In this case, the relationship between two things is indicated by their relative distance from one another. The closer two things are, the less the gap between them. Graphical clustering tools like Gephi have recently gained in popularity. OpenGL is a popular 3D graphics engine, and this programme is cross-platform software that makes use of it. Data can be set up in a variety of ways depending on their size, characteristics, classification, etc. Information scientist Olle Persson created the programme BibExel. BibExel is open-source software that is free for educational use. It's another resource for bibliometric analysis that helps scientists in their work.
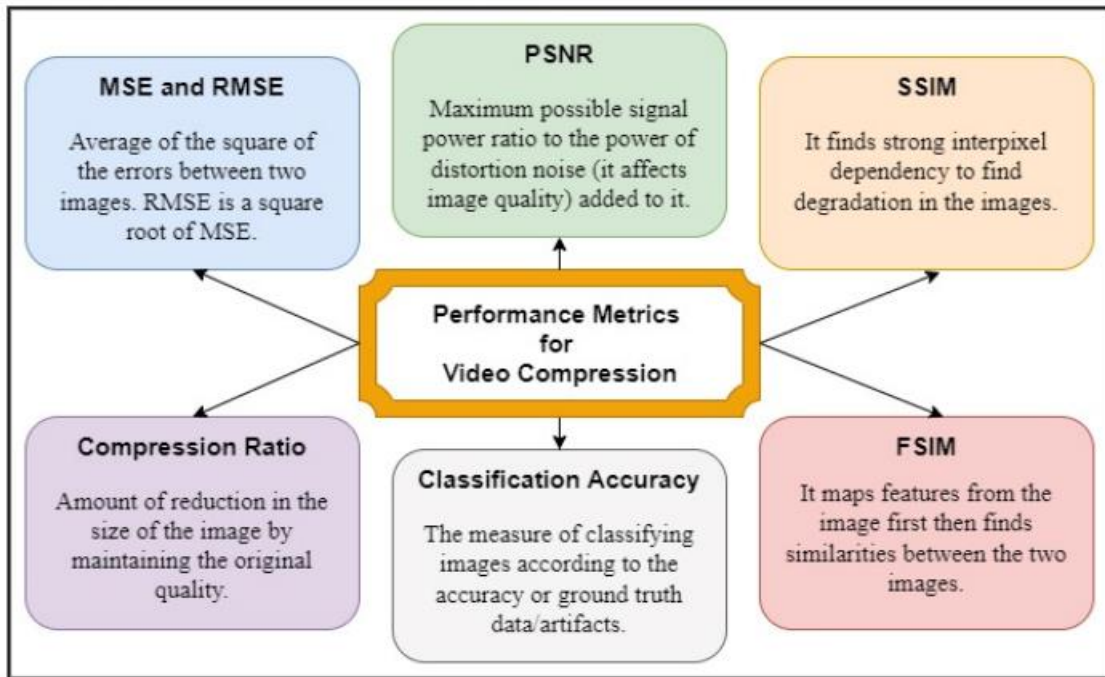
Figure 3. Performance metrics for video compression.

Videos undergo procedures before they may be viewed by the general audience. The video is first converted into a series of still images, as was explained earlier. The photos then need to be processed. Changes to the images may be apparent after processing. These changes add artificial, distracting features to the picture that take away from its overall quality. Images may become fuzzy, deform geometrically, and appear blocky due to compression standards. Therefore, evaluating the image or video's quality is essential to any data compression procedure. Most of the standard measures for assessing the quality of a video or image are depicted in Figure 3.
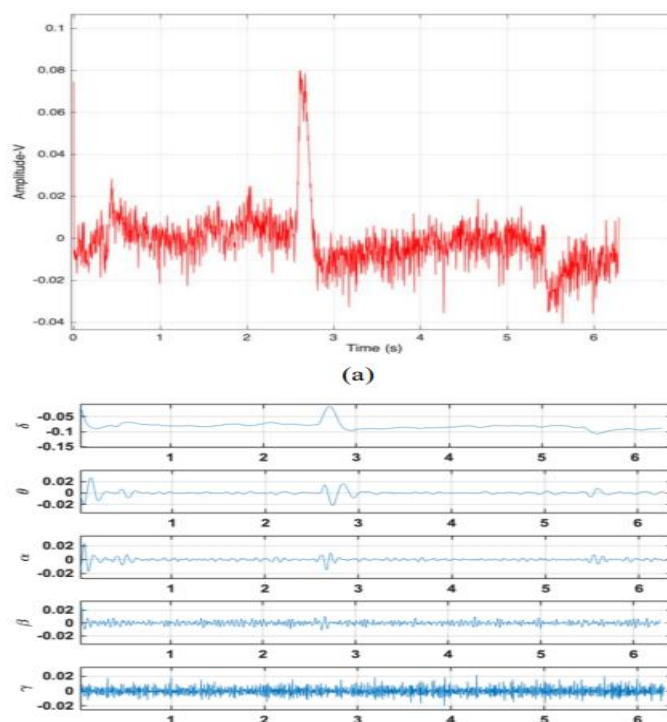
4. RESULTS AND DISCUSSION

Figure 4. A sample of the EEG signal: (a) the raw signal and (b) the decomposed and processed components.

Twenty test respondents provided electroencephalographic (EEG) data signals. Eight men, nine women (all between the ages of 18 and 30), and three kids made up the sample size (with ages between 8 and 12 years). The test subjects were instructed to maintain a state of relative stillness and little eye blinking during the duration of the measurements. Each patient sat relaxed on an ergonomic chair; none of them had electrodes attached to their scalps. As soon as the subjects were comfortable in their surroundings, the electrodes were attached. A portion of the raw EEG signal collected for this study is displayed in Figure Figure 4 (a), along with its decomposed components in Figure 4 (b).
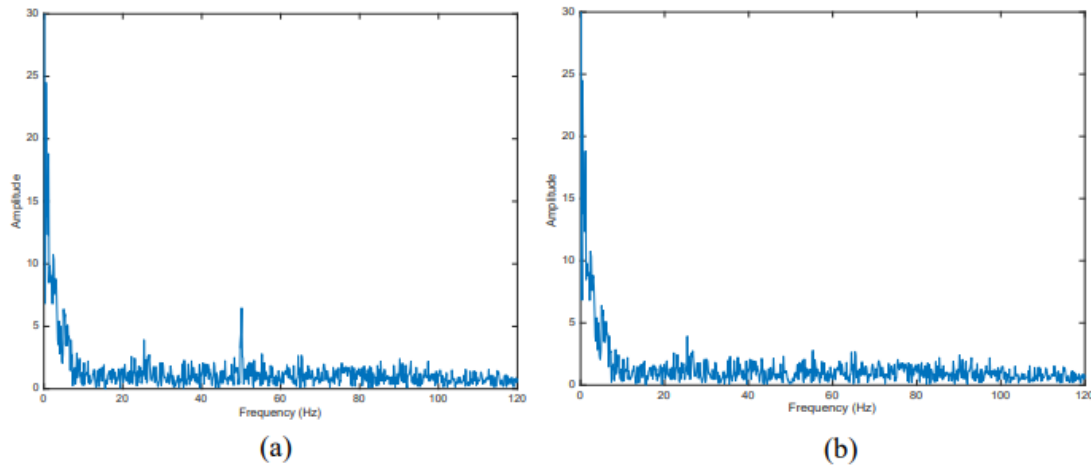

Figure 5. An example of a frequency-domain representation of the EEG signal: (a) raw signal and (b) filtered signal.

The signal before and after filtering is depicted in Figure 5. Also, the high-frequency noise was reduced with a low-pass filter whose corner frequency was set to 100 Hz. Due to the meticulous technique taken to eliminate sources of noise, other sources of noise that would ordinarily occur in any EEG readings, such as human artefact (low frequency) and electromagnetic noise superimposed on EEG, EMG, and EOG signals, did not exist. The waves that made up the EEG signals were decomposed.
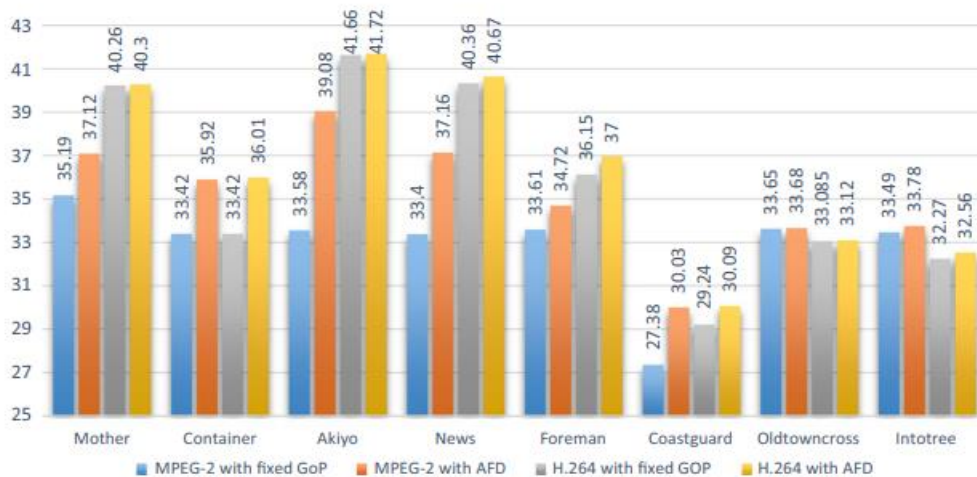

Fig. 6 Examination of the relative quality of AFD and fixed GoP video reconstruction techniques.

4.1 Investigation using AFD-based and fixed GoP-based conventional techniques

Standard methods used in the industry, such as H.264 and MPEG-2, were applied to further evaluate and test the results. The use of AFD in standard methodologies has been analysed and compared to the more traditional, fixed GOP "IBBPBBPBBPBBI" approach. The results of these analyses are depicted in Figure 6, which shows the quality of the reconstructed videos as determined by the PSNR (in dB) averaged over numerous studies using a wide variety of video sequences.

4.2 State-of-the-art of a system

It can be seen in Fig. 7 that the most recent video encoding standard, HEVC, uses neighbouring reconstructed pixels to predict the next coding block. The HEVC codec supports the DC mode and the planar mode, two of the 33 angular intra prediction choices. Inter-frame coding is one area where HEVC excels over its predecessor, H.264/AVC, with improvements such as more most probable modes (MPMs) for intra mode coding, advanced motion vector prediction (AMVP), and merge mode for motion vector predictor coding. As an added bonus, HEVC employs a greater number of interpolation filter taps to compensate for motion between samples in the absence of a corresponding sample.
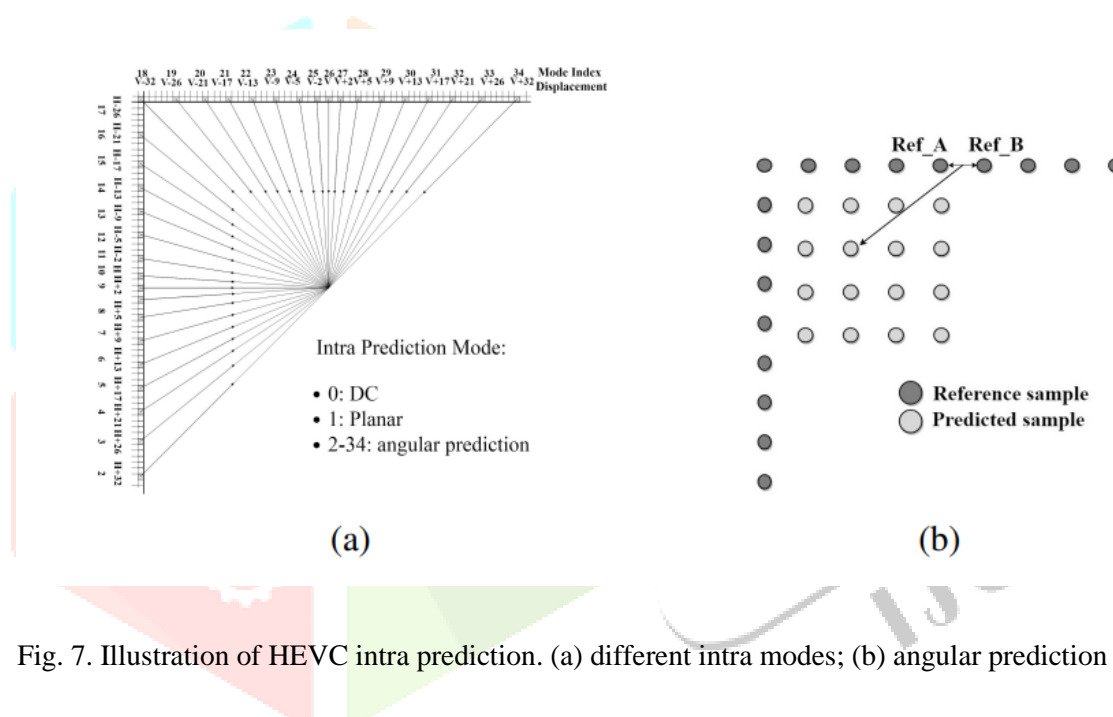

Fig. 7. Illustration of HEVC intra prediction. (a) different intra modes; (b) angular prediction instance.

In the context of the state-of-the-art video coding framework, loop filters are a relatively new video coding technique that have been proposed frequently since the year 2000. We have combined the sample adaptive offset and deblock filtering features of HEVC in this work (SAO). To improve coding efficiency, however, it is becoming more difficult to adapt algorithms for the traditional hybrid video coding framework based on local correlations in images and videos.

4.3 Future Directions

Many real-time applications, such as prediction or comprehending data's causal processes, need the identification of an intelligent model guiding the data. For a video call to go through, for instance, it's probably best to focus on the person(s) you're talking to rather than the background. A tennis match is another illustration. The uniqueness of the spectators is less essential than the condition of the court and players during a tennis match. From an information theoretical standpoint, a reliable predictor will naturally generate efficient compressors. When this is the case, the results from applying machine learning methods are as expected. Numerous machine learning algorithms may accomplish tasks such as regression, classification, clustering, decision trees, extrapolation, and many others. If you have data and need to complete a task that relies on that data, machine learning can teach an algorithm to extract the relevant information. Many types of machine learning methods, including supervised and unsupervised learning as well as reinforcement learning, can be applied to the development of such algorithms. With the current DNN methods, the rate-

distortion is greatly enhanced, however the model's robustness is greatly reduced. Additionally, it calls for additional memory, which restricts their usefulness. New solutions could be proposed as a result of research into this field of study.

CONCLUSION

Rapid advancements in computer vision are largely responsible for the recent popularity of machine learning. In this paper, we will discuss the three main types of machine learning for computer vision, and we will see how they have been used to achieve high performance rates in a variety of visual understanding tasks: convolutional neural networks (CNNs), the "Boltzmann family" (DBNs and DBMs), and supervised discovery algorithms (SdAs). A few examples are human pose estimation, image retrieval, semantic segmentation, action and activity recognition, face detection, and object detection. However, there are perks and drawbacks to every group. CNNs can automatically learn features depending on the provided dataset, giving them a distinct advantage over other models. For some computer vision tasks, CNNs' invariance to transformations is a huge benefit. But unlike DBNs/DBMs and SdAs, which can function in an unsupervised manner, they rely substantially on the presence of labelled data. While training CNNs and DBNs/DBMs require significant processing resources, SdAs are capable of being trained in real time under specific conditions.

REFERENCES

1. Rozenwald MB, Galitsyna AA, Sapunov GV, Khrameeva EE, Gelfand MS. A machine learning framework for the prediction of chromatin folding in Drosophila using epigenetic features. PeerJ Comput Sci. 2020;6:307.

2. Amrit C, Paauw T, Aly R, Lavric M. Identifying child abuse through text mining and machine learning. Expert Syst Appl. 2017;88:402–18.

3. Hossain E, Khan I, Un-Noor F, Sikander SS, Sunny MSH. Application of big data and machine learning in smart grid, and associated security concerns: a review. IEEE Access. 2019;7:13960–88.

4. Crawford M, Khoshgoftaar TM, Prusa JD, Richter AN, Al Najada H. Survey of review spam detection using machine learning techniques. J Big Data. 2015;2(1):23.

5. Deldjoo Y, Elahi M, Cremonesi P, Garzotto F, Piazzolla P, Quadrana M. Content-based video recommendation system based on stylistic visual features. J Data Semant. 2016;5(2):99–113.

6. Al-Dulaimi K, Chandran V, Nguyen K, Banks J, Tomeo-Reyes I. Benchmarking hep-2 specimen cells classifca- tion using linear discriminant analysis on higher order spectra features of cell shape. Pattern Recogn Lett. 2019;125:534–41.

7. Liu W, Wang Z, Liu X, Zeng N, Liu Y, Alsaadi FE. A survey of deep neural network architectures and their applica- tions. Neurocomputing. 2017;234:11–26.

8. Pouyanfar S, Sadiq S, Yan Y, Tian H, Tao Y, Reyes MP, Shyu ML, Chen SC, Iyengar S. A survey on deep learning: algo- rithms, techniques, and applications. ACM Comput Surv (CSUR). 2018;51(5):1–36.

9. Alom MZ, Taha TM, Yakopcic C, Westberg S, Sidike P, Nasrin MS, Hasan M, Van Essen BC, Awwal AA, Asari VK. A state-of-the-art survey on deep learning theory and architectures. Electronics. 2019;8(3):292.

10. Potok TE, Schuman C, Young S, Patton R, Spedalieri F, Liu J, Yao KT, Rose G, Chakma G. A study of complex deep learning networks on high-performance, neuromorphic, and quantum computers. ACM J Emerg Technol Comput Syst (JETC). 2018;14(2):1–21.

11. Adeel A, Gogate M, Hussain A. Contextual deep learning-based audio-visual switching for speech enhancement in real-world environments. Inf Fusion. 2020;59:163–70.

12. Tian H, Chen SC, Shyu ML. Evolutionary programming based deep learning feature selection and network con- struction for visual data classifcation. Inf Syst Front. 2020;22(5):1053–66.

13. Young T, Hazarika D, Poria S, Cambria E. Recent trends in deep learning based natural language

processing. IEEE Comput Intell Mag. 2018;13(3):55–75.

14. Koppe G, Meyer-Lindenberg A, Durstewitz D. Deep learning for small and big data in psychiatry. Neuropsychop- harmacology. 2021;46(1):176–90.

15. Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: 2005 IEEE computer society confer- ence on computer vision and pattern recognition (CVPR'05), vol. 1. IEEE; 2005. p. 886–93.

16. Lowe DG. Object recognition from local scale-invariant features. In: Proceedings of the seventh IEEE international conference on computer vision, vol. 2. IEEE; 1999. p. 1150–7.

17. Wu L, Hoi SC, Yu N. Semantics-preserving bag-of-words models and applications. IEEE Trans Image Process. 2010;19(7):1908–20.

18. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44.

19. Yao G, Lei T, Zhong J. A review of convolutional-neural-network-based action recognition. Pattern Recogn Lett. 2019;118:14–22.

20. Dhillon A, Verma GK. Convolutional neural network: a review of models, methodologies and applications to object detection. Prog Artif Intell. 2020;9(2):85–112.

21. Khan A, Sohail A, Zahoora U, Qureshi AS. A survey of the recent architectures of deep convolutional neural net- works. Artif Intell Rev. 2020;53(8):5455–516.

22. Hasan RI, Yusuf SM, Alzubaidi L. Review of the state of the art of deep learning for plant diseases: a broad analysis and discussion. Plants. 2020;9(10):1302.

23. Xiao Y, Tian Z, Yu J, Zhang Y, Liu S, Du S, Lan X. A review of object detection based on deep learning. Multimed Tools Appl. 2020;79(33):23729–91.

24. Ker J, Wang L, Rao J, Lim T. Deep learning applications in medical image analysis. IEEE Access. 2017;6:9375–89.