Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/47/20241113

An investigation of machine learning-based video compression techniques

Ye Zhu

School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China

zhuye1@sjtu.edu.cn

Abstract. As video technology continues to seamlessly weave itself into the fabric of daily life, there is a growing need for enhanced storage and efficient video transmission. This surge in demand has led to heightened expectations and standards for video compression technology. Machine learning as an up-and-coming technology can play its advantages in the field of video compression. This article reviews the current state of research on combining video compression techniques with machine learning. The article provides an overview of various research avenues for enhancement, spanning from conventional video compression algorithms to the fusion of traditional compression frameworks with machine learning methodologies, and even the development of novel end-to-end compression algorithms. In additional, the article explores the possible various application scenarios of machine learning-based video compression algorithms based on the characteristics of such non-standard and arithmetic demanding algorithms. At the end, the article speculates on the future of video compression algorithms based on the content of the various studies reviewed in the article.

Keywords: Video Compression, Machine Learning, End-to-end Framework.

1. Introduction

In recent times, the influence of video technology has significantly transformed all facets of human existence. Ranging from leisure to learning, and from interaction to professional endeavors, videos have assumed a crucial position in the day-to-day routines. The earliest video technology was film. Film was switched at certain speeds to capture moving images and show them as a projection. But video technology was still very primitive. Movies were still a very niche affair due to the rather high production and broadcasting costs. Subsequently, the advent of digital video brought television into millions of homes, allowing viewers to watch a variety of programs and news from the comfort of their homes. The introduction of videotape technology allowed people to record and playback their favorite content, further boosting the popularity of video entertainment. Furthermore, with the development of the industry, the price of video cameras keeps decreasing. Camcorders have gone from being professional movie making equipment to being an affordable electronic product for the general public. With the popularity of smart phones, people can easily record and make a video with their cell phones. People began to use video to record their lives in place of the traditional photo-taking.

Video technology is a large and complex discipline. Thereinto, video compression is an important technique that is possible to cope with the increasing amount of video and reduce the pressure on video

DOI: 10.54254/2755-2721/47/20241113

storage. With video compression technology, it is possible to transmit clearer video in limited network bandwidth, to meet the needs of more people online playback, live broadcast.

Nowadays, hardware is evolving very rapidly and the amount of arithmetic provided is constantly increasing. The compression algorithms formulated in the past times are limited by the computing power and storage and transmission limitations of the hardware in the past times. In the new times compression algorithms should fully utilize more computing power to achieve more complex but more efficient compression. For example, AOMedia Video 1 (AV1) is a next-generation video coding format initially designed for video transmissions over Internet. It has higher compression rate compare to former coding standard like H.264 or VP9. It uses complex prediction modes and improved motion estimation. These techniques help reduce distortion during the compression process, resulting in better video quality. And LZ4 is a new compression method. It takes advantage of modern hardware, provides a very fast speed in both compression and decompression. Also, it provides streaming support and low overhead for both compression and decompression. These features make LZ4 suitable for real-time applications like video streaming.

In this article, the author will synthesize the latest scientific research progress on the current application of neural networks to video compression and look for possible breakthrough directions.

2. Review of video compression techniques

2.1. Overview of traditional video compression techniques

Traditional video compression algorithms use many techniques for video compression. Inter-frame prediction captures the differences between frames by motion estimation and uses I-frames (stores a complete image), P-frames (need to reference to previous I or P frame), and B-frames (need to referenct to both previous and future frames to decode) to reduce the redundancy of information between frames. Intra-frame prediction allows the algorithm to reduce intra-frame information redundancy by predicting by neighbouring pixel points [1-3]. Advanced methods in this domain encompass an array of techniques, such as transformation and quantization methods, control of quantization parameters, entropy coding, predictive residual coding, variable-size coding units, multi-frame referencing, and various other approaches.

There are two main ways to apply machine learning to video compression. The first one is to follow the compression process of MPEG [1], H.264 [2], H.265 [3], and replace one or two handcraft modules with machine learning method to get higher compression ratio or faster compression speed. The second approach is to entirely disregard the constraints of the current framework and construct a video compression framework from the ground up. Leveraging the capabilities of machine learning, one can create an end-to-end video compression framework.

2.2. Traditional compression scheme combined with machine learning tools

The first approach has the advantage that it can be more easily integrated into existing video compression algorithms. Only a few modules need replacement to achieve the performance improvement. There have been many research digging into this direction. This article will review studies that replacing intra-prediction with machine learning techniques, using neural network to improve inter-prediction performance, establishing a new method for entropy coding, learning-based filtering and machine learning based frame enhancement.

Traditional intra-prediction methods create linear predictions using predefined directions. A Progressive Spatial Recurrent Neural Network (PS-RNN) is designed to improve traditional intra-prediction. It utilizes three spatial recurrent units to iteratively produce predictions, transmitting information from previous content to the blocks awaiting encoding. This newly proposed intra prediction scheme achieves an average bit-rate reduction of 2.5% under variable-block-size settings while maintaining the same reconstruction quality as HEVC. Another way is to compress the intra-frame information using neural network. Deepcoder [4], a Convolutional Neural Network (CNN) based video

DOI: 10.54254/2755-2721/47/20241113

compression network is designed to compress intra-frame information. It employs distinct CNN networks for processing predictive and residual signals, followed use Huffman coding to code into bits.

For inter-prediction, a neural network-based enhancement to inter prediction (NNIP) [5]is proposed to play the role of inter prediction during compression. NNIP comprises three key components: a residue estimation neural network, a combination neural network, and a deep refinement neural network. The residue estimation network estimates the residue between neighboring blocks. Subsequently, the combination network extracts and combines the feature maps of the estimated residue generated from previous process with those of the predicted block. Finally, the refinement network produces an improved residue to enhance the accuracy of the predicted block.

In terms of entropy coding, both H.264 and H.265 use Context-Adaptive Binary Arithmetic Coding (CABAC) to perform entropy coding. But CABAC relies on manually choosed binarization process and handcrafted context. Neural network method is applied to optimize CABAC. CNN can be used in probability estimation in replace of handcrafted context model. It can save 9.9% bits than CABAC does [6].

Due to the rate-distortion trade-off in video compression, artifacts may be present in the decompressed video. Therefore, a filtering method is employed within the video compression framework to mitigate these artifacts. This study combines neural network with Kalman filter network. The approach takes use of the recursive feature of the Kalman model and highly non-linear mapping ability of deep neural network (DNN) [7]. In further study, multiple DNNs are employed to estimate the respective states within the Kalman filter, which are then integrated within the framework of the filtering network [8].

A similar way to improve quality of decoded video is use neighboring high-quality frame to enhance low quality frame [9]. This approach designed a support vector machine based identifier to determine which are high quality frames in a decompressed video. Then a multi-frame CNN is used to improve the general quality of decompressed video, which low quality frame and its nearest two high quality frame are as the input, and output an enhanced low quality frame.

2.3. Designing a new compression scheme with machine learning method

The second approach leverages both the strengths of conventional coding schemes and the potent non-linear mapping capabilities of neural networks.

The first approach is to rebuild the whole compression scheme with machine learning techniques.

The Deep Video Compression (DVC) framework [10], use the full ability of neural networks. This study use machine learning to get motion information from frames. Then two separate auto-encoder neural network is used in compress both motion information and residual information to smaller size. All the neural networks are jointly optimized with one loss function, so they collaborate with other neural networks by balancing the compression rate and the compression distortion.

In the first end-to-end deep video compression frame work [11], pixel-wise motion information is first generated using a neural network. And then using an auto-encoder style network to compress information to bits. Other parts of the compression process are well designed to work with neural network part to ensure optimal efficiency. Furthermore, this video compression framework offers remarkable flexibility and can readily accommodate extensions through the utilization of lightweight or advanced networks to enhance speed or efficiency.

A new approach is to operate the compression not in traditional way but in feature-space. The Feature-space Video Coding framework (FVC) [12] is created to mitigate imprecise motion estimation or enhance the effectiveness of motion compensation while using learning-based approaches to video compression. This framework conducts primary operations within the feature space. The framework incorporates a novel deformable compensation module. to be more effective in motion compensation. First it calculates motion information. Then the motion information is sent to auto-encoder style network and compressed by it. After that, a deformable convolution operation is employed to produce the predicted information for compensation in decompression process.

DOI: 10.54254/2755-2721/47/20241113

3. Applications of machine learning-based video compression

Video compression improved by machine learning techniques can have a great effect in various aspects of people's lives and industries. But the most important point is that the machine learning based video compression algorithm is a "non-standard compression algorithm". The user of the algorithm needs to have control over the entire process from compression to decoding in order to apply a non-standard compression algorithm within the process. Here are some applications of machine learning-based video compression in real life and industry:

- Online Video Services: Video platforms like Netflix, YouTube or Bilibili, rely on video compression to deliver lots of video content to users over the internet. They can re-encode uploaded video with machine learning-based compression algorithm and decode them with the algorithm in user terminal. Machine learning-based compression algorithms help in reducing the bandwidth required for streaming while ensuring a good viewing experience.
- Video Conferencing: Applications like Zoom, Microsoft Teams, and Skype use video compression to transmit video and audio data efficiently during video conferences. This ensures smooth communication even on low bandwidth connections.
- Security and Surveillance: In the security and surveillance industry, video compression is used to store and transmit video footage from security cameras. This reduces storage requirements and allows for remote monitoring.
- Medical Imaging: In healthcare, video compression is used in medical imaging applications like telemedicine, where high-resolution images and videos need to be transmitted efficiently while preserving diagnostic quality.

In all these applications, machine learning-based video compression methods are continually evolving to improve efficiency, reduce bandwidth requirements, and enhance video quality. These developments greatly influence the quality of the digital interactions and enhance the efficiency of diverse industries.

4. Discussions

The future direction remains uncertain, as it is unclear whether the traditional compression scheme with machine learning tools will be replaced by a completely new end-to-end machine learning-based video compression approach. To this question, for now, the answer is "no", because new scheme did not outperform traditional scheme in general [13]. However, as additional research progresses, end-to-end compression may eventually surpass traditional compression schemes, potentially introducing a superior compression algorithm to the world.

5. Conclusion

This article has provided an overview of the traditional video compression techniques and explored the avenues where machine learning has been integrated to enhance these methods. Traditional video compression techniques, when combined with machine learning tools, have shown substantial promise in enhancing compression efficiency and video quality. Neural networks have driven notable advancements in different facets of video compression, including intra-frame prediction, inter-frame prediction, entropy coding, and artifact reduction. Conversely, the rise of end-to-end video compression frameworks, entirely propelled by machine learning, has demonstrated the remarkable potential of non-linear representations in revolutionizing the field. These pioneering methods harness neural networks for tasks such as motion estimation, residual compression, and other critical compression elements. They provide a degree of flexibility and hold the potential to deliver an enhanced balance between compression rates and distortion. The applications of machine learning-based video compression span across streaming services, video conferencing, security and surveillance, and medical imaging, revolutionizing the way people consume and interact with visual content in the daily lives and industries. In this era of burgeoning video content and ever-increasing demands for efficient video transmission, machine learning-based video compression represents a dynamic field that continues to shape the way people experience and interact with the visual world. The ongoing synergy between traditional methods

DOI: 10.54254/2755-2721/47/20241113

and machine learning-driven innovations ensures a promising future for video compression technology, where the pursuit of higher quality and greater efficiency remains at the forefront of research and development efforts.

References

[1] Le G D 1991 MPEG: A video compression standard for multimedia applications Communications of the ACM 34(4) 46-58
[2] Wiegand T Sullivan G J Bjontegaard G and Luthra A 2003 Overview of the H. 264/AVC video coding standard IEEE Transactions on circuits and systems for video technology 13(7) 560-576
[3] Sullivan G J Ohm J R Han W J and Wiegand T 2012 Overview of the high efficiency video coding (HEVC) standard IEEE Transactions on circuits and systems for video technology 22(12) 1649-1668
[4] Chen T Liu H Shen Q Yue T Cao X and Ma Z 2017 Deepcoder: A deep neural network based video compression In 2017 IEEE Visual Communications and Image Processing (VCIP) (pp. 1-4) IEEE
[5] Wang Y Fan X Xiong R Zhao D and Gao W 2021 Neural network-based enhancement to inter prediction for video coding IEEE Transactions on Circuits and Systems for Video Technology 32(2) 826-838
[6] Song R Liu D L H and Wu F 2017 Neural network-based arithmetic coding of intra prediction modes in HEVC In 2017 IEEE Visual Communications and Image Processing (VCIP) pp. 1-4 IEEE
[7] Lu G Ouyang W Xu D Zhang X Gao Z and Sun M T 2018 Deep kalman filtering network for video compression artifact reduction In Proceedings of the European Conference on Computer Vision (ECCV) pp 568-584
[8] Lu G Zhang X Ouyang W Xu D Chen L and Gao Z 2019 Deep non-local kalman network for video compression artifact reduction IEEE Transactions on Image Processing 29 1725-1737
[9] Yang R Xu M Wang Z and Li T 2018 Multi-frame quality enhancement for compressed video In Proceedings of the IEEE conference on computer vision and pattern recognition pp 6664-6673
[10] Lu G Ouyang W Xu D Zhang X Cai C and Gao Z 2019 Dvc: An end-to-end deep video compression framework In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition pp 11006-11015
[11] Lu G Zhang X Ouyang W Chen L Gao Z and Xu D 2020 An end-to-end learning framework for video compression IEEE transactions on pattern analysis and machine intelligence 43(10) 3292-3308
[12] Hu Z Lu G and Xu D 2021 FVC: A new framework towards deep video compression in feature space In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition pp 1502-1511
[13] Liu D Li Y Lin J Li H and Wu F 2020 Deep learning-based video coding: A review and a case study ACM Computing Surveys (CSUR) 53(1) 1-35