Deep learning based combating strategy for COVID-19 induced increased video consumption

Sangeeta[#1], Preeti Gulia[#2]

[#]*Department of Computer Science & Applications, Maharshi Dayanand University, Rohtak, INDIA*

Abstract - *COVID-19 epidemic has brought tremendous changes globally. The adoption of prevention strategies like lockdown and remote working has suddenly changed all aspects of human life. A surge shift in online mode has been observed and it tremendously increased the internet traffic with a suitable rise in video content. Research and surveys state that these changes will become new normal and will last for a long time. As the bandwidths are limited and cannot be expanded instantly, there arose a need for alternate techniques to be explored to deal with growing video content efficiently. The lightweight and powerful deep learning based video compression and analytics techniques may help in efficiently processing video content. Deep learning based techniques are already giving potent results both in the video compression and video analytics domain independently. In this paper, the accelerated impact of COVID-19 on video compression methods has been demonstrated and proposed joint video compression-cum-analytics scheme which may significantly provide fast and efficient video analytics from the compressed video optimizing whole network.*

Keywords - *COVID-19, video, bandwidth, deep learning, compression.*

I. INTRODUCTION

COVID-19 has shaken the whole world. Though India is also suffering from this crisis but at present, its effect is relatively muted as compared to other developed countries. As no vaccine is available, it can only be combated by maintaining social distance. In the same tune, different countries including India have to take tough decisions of lockdown to save the lives of the people. As no one has witnessed such scenario earlier, sudden impose of lockdown hampered their lifestyle significantly. Fulfilling daily household requirements and keeping working from home emerged as a major challenge. Being unprepared for such an unexpected situation, a lot of sectors suffered heavily amid lockdown while others showed a sharp rise. The doable tasks are being carried out online.

The major sectors emerged as winners are e-learning, online gaming, at-home entertainment, health and wellness, tele-consultation, tele-medication., remote meetings and remote working As Consumers are not able to go to cinema halls and concerts for some time, they are spending more time watching videos, movies and TV shows at home. Some studies see 20–50% rise in OTTs in India during the lockdown. Online streaming services like Netflix, Amazon Prime etc. also experienced a big boost. The children are spending their playtime on online gaming. A surge of around 40% has been observed in online gaming. As gyms and clinics are closed, people prefer online fitness classes, tele-consultation and telemedicine. School and higher education got the worst hit during this period. The universities and schools are moving towards adoption of Learning Management Systems and virtual classrooms so that the students can learn being at home and their progress can be assessed online. We see a spark shift in digital education with use of learning apps like BYJU, online lectures, video tutorials etc. School education, which was primarily conventional classroom type in India before this crisis, is also looking for online solutions for learning and assessment. Amidst lockdown, online learning of extra-curricular courses like Guitar, Painting, etc. are seeing a huge growth. As all the official and administrative work conferences are moving online, video conferencing tools like Zoom saw a massive increase in adoption in businesses across the world. As retail is also one of the worst hit sectors, fashion retail is also shifting to live streaming commerce models which enable retailers to showcase and present their products on a larger commerce destination. It can be observed from the above trends that video content gets more and more adoption.

Digitalization in each sector has its associated infrastructure requirements, efficiency trade-offs, social impacts. Although India is moving towards digitalization but current digital remote requirements are at par. The bandwidth cannot be expanded instantly. These soaring trends are producing huge video content over the limited bandwidth. This crisis opens up large opportunities of research and innovation before the researchers for further exploration and advancements to meet such surged requirements.

Video Compression can play a significant role in dealing with the rising video content over limited bandwidth. The various platforms govern the mass consumption of videos by analytics based filtering and recommendations by various platforms like social media, education, news, marketing and advertising etc. Strong and Proficient compressions techniques may prove as an aid to to reduce internet traffic, save storage space, and increase throughput and provide fast and efficient analytics support. The emergence of new better compression algorithms could help service providers save the resources such as servers, network, and storage and provide clients an opportunity to have better video experience in lower bandwidths such as in mobile devices. It drives applications like cloud gaming, real-time high-quality video streaming 3D and 360-videos. Deep learning is already de-facto for analytics computation. Some of the recent researches in this domain are presenting comparable and better results in comparison to the existing traditional techniques. Further exploration in this avenue may result in a significant and powerful technique which would be suitably used in such circumstances of sudden beating stress of bandwidth.

II. COVID-19 PANDEMIC

Covid-19 is the most disastrous event of this year. This pandemic has impelled the whole world to a freeze. Such accidental and unpredicted events shook the whole economy and lifestyle of the people unbalancing the existing structure. It has created a global health crisis due to unavailability of its vaccine. It is highly infectious and spread by social contacts. Maintaining social distance and quarantining turned out to be the best preventive step. The countries have to adopt lockdown strategies to combat it.

As the whole socio-economic structure got imbalanced, people have to look for alternative ways to meet all their requirements being at home only. The change in the lifestyle caused by this crisis indirectly affected the various sectors. Some sectors turned out to be in huge demand while others faced a setback. According to BCG Covid-19 consumer sentiment research conducted between March 23-26 in India, sectors like at-home entertainment, e-learning, telemedicine, online gaming, health and wellness etc. emerged as winners and others like travel and transport, automobiles, out of home entertainments turned out to be as losers [1].

Table 1: Some winner and loser sectors

Winner Sectors	**Loser Sectors**
At home entertainment	Travel
E-learning	Transport
Telemedicine	Automobiles
Online gaming	Out of home entertainment
Virtual Meetings	Discretionary spends
Work from home	Non-mobile electronics
Insurance	Home improvement
Health and wellness	Hotels and Restaurants

III. IMPACT OF COVID-19 ON VIDEO CONTENT

The increasing trend of doing more and more chores online puts the available bandwidth under stress. In the same tune, the internet traffic is soaring with the leading demand of online content consisting of text, audio, and video and others. Some of the major online platforms contribute to surging video content. As people stuck in their houses, they are spending more time in various video content based platforms.

Table 2: Sectors contributing superfluous video content

Social Media	Facebook, Instagram, WhatsApp, Tiktok, musically etc
Music, Movies and Television	Netflix, Amazon Videos, hotstar, torrent engines etc.
Education	Coursera,Khan Academy, Unacademy, Edx, Youtube channels etc.
News	CNN, NDTV, Hoststar, Google News, Youtube channels etc.
Sports and games	Steam, PS4, Xbox, ESPN, hotstar sports, PubG, online ludo etc.
Blogging, sharing and streaming	Youtube, Vimeo, Twitch etc.
Work from Home	Video Conferencing based meetings-ZOOM, Google Meet, Webinars etc.
Learning skills online	Cooking, Guitar etc.

The employees in both IT and non-IT sectors are working from home to a large extent. CCTV or drone based videos are also playing an important role in surveillance to look after the social distancing in specific and general public areas. Schools and universities are also adopting learning management systems and motivating their students for e-learning. All these personal and administrative activities, being conducted online, contribute significantly in the accretion of huge video traffic over the internet.

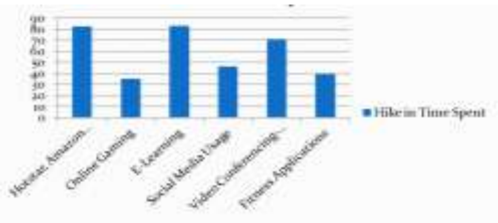

Fig 1: Hike in time spent on various online platforms

The above graph shows the percentage hike in the time spent in some online chores as observed by some surveys [2, 3]. A spark rise has been observed by several survey agencies ultimately leading to the voluminous video content. This crisis indirectly leads to a surge shift in online mode ultimately resulting in mass video content over limited bandwidth.

Taking the example of a virtual conference tool, Microsoft Team, its volume of online meetings increased by 1000% during COVID pandemic with more than 500% rise just in the month of March [30]. Out of all these meetings, 60% meetings were video based. But there are few countries where video usage was lower. People in India use video in 22 percent of meetings, Singapore 26 percent, South Africa 36 percent, France 37 percent, and Japan 39 percent. This was attributed in part to less access to devices and stable internet in some regions such as India and South Africa.

Table 3: Exponentially increased usage of Microsoft Team [30]

Date	Minutes spent in Microsoft team meetings per day
12 March 2020	560 Million
16 March 2020	900 Million
31 March 2020	2.7 Billion

IV. ROLE OF VIDEO COMPRESSION AND DEEP LEARNING

The sudden rise in mass consumption of videos forces various streaming platforms like YouTube, Netflix limits quality for beating bandwidth stress as the bandwidth cannot be suddenly scaled. Moreover, there is no segregation technique to differentiate between the video content related to work and entertainment. This is hampering the seamless e-learning, video-conferencing; work from home and many important tasks. In these circumstances, video compression can play a convincing role to deal with bandwidth stress up to some extent.

Video compression is required for efficient fast sharing of videos. The sharing platforms perform various analytics over the video to provide most relevant responses to search, block inappropriate content and disseminate it to relevant communities. Its heavy consumption puts challenges pertaining to video storage, transfer and analysis. A breakthrough in video compression and analytics techniques could address these challenges. The advancements in video compression techniques may help service providers save on resources such as servers and give clients an opportunity to have better video experience in lower bandwidths such as in mobile devices with more relevant videos.

Traditional approach is sharing via compression of the video and analyzing by first decompressing into original raw format and then applying deep learning techniques. The introduction of high quality image and video formats has made the task of video compression very challenging. This led to the emergence of highly efficient and powerful video compression techniques. Existing video compression techniques are designed and optimized manually. Further improvement in the same direction is becoming very challenging.

Recent researches have shown that deep learning based video compression techniques are giving comparable and better results in comparison to the existing traditional techniques. Deep learning is a machine learning technique that teaches computers to do what comes naturally to humans i.e. learn by example. This makes it possible to achieve recognition accuracy at higher levels than ever before. Deep learning requires large amounts of labeled data. It performs end-to-end learning where a network is given raw data and a task to perform, such as classification, and it learns how to do this automatically. In neural network architectures, the number of hidden layers in these networks varies from three to thousands and the network is trained by using large sets of labeled data and neural network architectures that learn features directly from the data without the need for manual feature extraction. One of the most popular such deep neural networks is Convolutional Neural Network. Initially, a number of CNN based models for image compression have been designed and proposed and later the same extended for the videos. The positive results of these researches motivated further exploration in the application of deep learning concepts for its practical applicability. A number of advanced approaches have also been proposed to consume videos using compressed videos directly. The approaches are also shown to not only be faster than existing action recognition approaches but also give state-of-the-art results. Consuming compressed video already removes superfluous information. Deep learning based compression may emerge as a powerful and significant domain in future.

V. RELATED WORK

Video content is accumulating over the internet at a fast pace [4]. The emergence of ultra high quality definitions has made the task of compression more challenging and complex. The quality of compression directly corresponds to the video analytics. Efficient compression architecture may contribute a lot in the tasks related to object detection and action recognition [5]. In the traditional approaches, manually designed architectures are used to eradicate redundancies. The field of image and video compression emerged only a few years ago. Earlier, deep learning based approaches were experimented on images and some of the CNN based image compression schemes [6-15] resulted in commendable performance. The deep learning approaches have the capability to deal with huge multi-scaled non-linear data and can be trained, designed and optimized in an end to end manner. The initial models were designed only to reduce the mean square error [11, 12, 13]. But later models opted for rate distortion optimization techniques to improve the efficiency. Adaptive bit coding technique by importance map has also gained attention [16]. Moreover, Intra prediction based image compression approaches also resulted in efficient results [17].

Predictive coding is the core concept behind the traditional and widely used video compression architectures like H.264 and H.265 [18, 19]. In progression to video compression, several individual modules focused deep learning based enhancements and improvements to existing architectures like entropy and residual encoding have been presented [20-23]. As these models cannot be not end to end trained, though efficient but not giving optimal performance. This led to the invention of new end to end trained deep learning based video compression schemes.

Action recognition is an area of research where given a video; the goal is to recognize what action is being performed. Understanding videos is arguably the next frontier in the Deep learning and Computer vision field as Videos capture way more information than images can. The traditional way for object detection is expressed below:

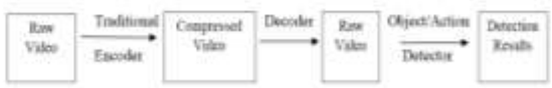

Fig 2: Traditional Video Analytics Scheme

Chao-Yuan Wu et al. gave the concept of Action recognition in Videos [24]. The approach is also shown to not only be faster than existing action recognition approaches but also gives state-of-the-art results. The key idea is the use of compressed video instead of uncompressing the video into RGB frames. Consuming compressed video already removes superfluous information. Motion vectors in video compression provide us the motion information that lone RGB images do not have. With compressed video, correlation in video frames is accounted, i.e. spatial view plus some small changes over time. In the modeling approach, firstly the video is encoded into it MPEG format. This video is then used to train 3 different models I-frame, motion vector and a residual model. All these 3 models are stacked to give final action prediction at the end. The researchers have also proposed various deep learning based object detection techniques which use the concept of CNN and R-CNNs [25, 26]. A new technique YOLO and its various up gradations were also proposed for fast and efficient object detection [27]. The researchers at Google also came up with a light weight model known as MobileNet [28]. Recently, Shiyao Wang et al. proposed a technique for Fast Object Detection in Compressed Video [29]. A fast object detection model was proposed incorporating a light motion-aided memory network called MMNet, which can be directly applied to compressed videos. Motion and appearance information stored and transmitted within a video stream are used instead of building another model to retrieve necessary motion cues. The video analytics task can be made more efficient and quality driven in tune with proficient video compression architecture.

VI. PROPOSED SCHEME

For video content based analytics, video is first decoded to a large raw format on the server and then fed to an analytics engine for metadata generation. These metadata are then stored and used for analytics purposes. This requires an analytics server to perform both decoding and analytics computation. Deep learning is already de-facto for analytics computation. There are numerous deep learning based object detection techniques to be applied on raw video. It is proposed that the whole process of decompression and object detection can be collectively made efficient by designing a combined architecture. The proposed architecture is as follows:

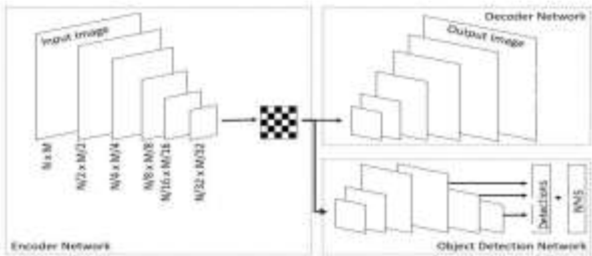

Figure 3: Video compression-cum-analytics scheme

This architecture will be designed on the code of separation of concerns and minimizes computing. Source will encode the frame with encoder network and broadcast. Analytics server can use encoded stream to compute object tags metadata to index videos and serve them to clients. Clients can use a decoder network to decode and view encoded videos.

Source side computing:

- Video Frame → Checkerbox → Entropy Encoding

Server side computing:

- Entropy Decoding → Checkerbox → Object Detection Network→ Objects detected

Client side computing:

- Entropy Decoding → Checkerbox → Decoder → Video frame

In this proposed work, multitask learning can be employed for both compression and analytics. This learning technique has proven to result in improved learning efficiency and prediction accuracy for the task-specific models, when compared to training the models separately. This architecture will help improve both compression quality and analytics accuracy.

The proposed scheme has benefits in all three video related tasks - video generation, analytics at server and consumption at client. With the development of highly efficient codecs, sources will be able to generate better video quality content with lower bandwidth requirements and similarly, clients will be able to consume those contents without putting too much stress on bandwidth. Analytics server will generate analytics metadata with information rich compressed format without the need of decoding those videos. it will result in lighter analytics models and faster response time.

Table 4: Advantages of proposed scheme

Video Generation	Video Analytics	Video Consumption
- High quality video generation - Low bandwidth requirement	- Low computation required - Highly scalable - Low latency analytics - Smaller analytics models	- High quality video - Lower bandwidth requirements - Possibility of real time analytics at client end

VII. CONCLUSION

The Covid-19 Pandemic has a long-lasting impact in the whole society. It led to a new normal lifestyle where social distancing and lockdown have forced people to work remotely. The sudden rise in online requirements leads to dissemination, storage and retrieval of huge video content resulting in a lot of stress over the limited bandwidth. Several platforms have to limit the quality of video content to beat the stress caused by accumulating voluminous video content. As video content becomes the primary need of this new normal, demand and research for highly efficient video platforms has increased in multiple sectors. The ultra efficient and lightweight video compression techniques may turn out to be significant tools in dealing with such circumstances. The proposed scheme relies on end-to-end deep learning based optimization and promises highly efficient architecture for video dissemination over the internet. As researches reveal that deep learning is giving promising results in analytics and video compression both, further exploration in the domain of its practicability and complexity will surely give potent results.

REFERENCES

[1] BCG Covid-19 consumer sentiment research India Survey Snapshot March 23-26 results

[2] https://www.nielsen.com/apac/en/insights/report/2020/the-impact-of-covid-19-on-media-consumption-across-north-asia/

[3] https://www.hindustantimes.com/tech/covid-19-lockdown-demand-for-netflix-whatsapp-zoom-and-other-ott-services-witness-major-growth-among-indians/story-KW0CBfBzCGnJJioTvowpnJ.html

[4] C. V. networking Index. Forecast and methodology, 2016-2021, white paper. San Jose, CA, USA, 1, 2016.

[5] C.-Y. Wu, M. Zaheer, H. Hu, R. Manmatha, A. J. Smola, and P. Kr¨ahenb¨uhl. "*Compressed video action recognition*". In CVPR, pages 6026–6035, 2018

[6] E. Agustsson, F. Mentzer, M. Tschannen, L. Cavigelli, R. Timofte, L. Benini, and L. V. Gool. "*Soft-to-hard vector quantization for end-to-end learning compressible representations.*" In NIPS, pages 1141–1151, 2017.

[7] E. Agustsson, M. Tschannen, F. Mentzer, R. Timofte, and L. Van Gool. *"Generative adversarial networks for extreme learned image compression."* arXiv preprint arXiv:1804.02958, 2018.

[8] J. Ball´e, V. Laparra, and E. P. Simoncelli. *"End-to-end optimized image compression."* arXiv preprint arXiv:1611.01704, 2016.

[9] J. Ball´e, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston. *"Variational image compression with a scale hyperprior."* arXiv preprint arXiv:1802.01436, 2018.

[10] L. Theis, W. Shi, A. Cunningham, and F. Husz´ar. *"Lossy image compression with compressive autoencoders."* arXiv preprint arXiv:1703.00395, 2017.

[11] G. Toderici, S. M. O'Malley, S. J. Hwang, D. Vincent, D. Minnen, S. Baluja, M. Covell, and R. Sukthankar. *"Variable rate image compression with recurrent neural networks".* arXiv preprint arXiv:1511.06085, 2015.

[12] G. Toderici, D. Vincent, N. Johnston, S. J. Hwang, D. Minnen, J. Shor, and M. Covell. *"Full resolution image compression with recurrent neural networks."* In CVPR, pages 5435 – 5443, 2017.

[13] N. Johnston, D. Vincent, D. Minnen, M. Covell, S. Singh, T. Chinen, S. Jin Hwang, J. Shor, and G. Toderici. *"Improved lossy image compression with priming and spatially adaptive bit rates for recurrent networks."* In CVPR, June 2018.

[14] M. Li,W. Zuo, S. Gu, D. Zhao, and D. Zhang. *"Learning Convolutional networks for content-weighted image compression."* In CVPR, June 2018.

[15] O. Rippel and L. Bourdev. *"Real-time adaptive image compression."* In ICML, 2017.

[16] F. Mentzer, E. Agustsson, M. Tschannen, R. Timofte, and L. Van Gool. *"Conditional probability models for deep image compression."* In CVPR, number 2, page 3, 2018.

[17] M. H. Baig, V. Koltun, and L. Torresani. *"Learning to inpaint for image compression."* In NIPS, pages 1246–1255, 2017.

[18] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra. *"Overview of the h. 264/avc video coding standard."* TCSVT, 13(7):560–576, 2003.

[19] G. J. Sullivan, J.-R. Ohm, W.-J. Han, T. Wiegand, et al. *"Overview of the high efficiency video coding(hevc) standard."* TCSVT, 22(12):1649–1668, 2012.

[20] T. Chen, H. Liu, Q. Shen, T. Yue, X. Cao, and Z. Ma. *"Deepcoder: A deep neural network based video compression."* In VCIP, pages 1–4. IEEE, 2017.

[21] Z. Liu, X. Yu, Y. Gao, S. Chen, X. Ji, and D. Wang. *"Cu partition mode decision for hevc hardwired intra encoder using convolution neural network."* TIP, 25(11):5088–5103, 2016.

[22] G. Lu,W. Ouyang, D. Xu, X. Zhang, Z. Gao, and M.-T. Sun. *"Deep kalman filtering network for video compression artifact reduction."* In ECCV, September 2018.

[23] R. Song, D. Liu, H. Li, and F. Wu. *"Neural network-based arithmetic coding of intra prediction modes in hevc."* In VCIP, pages 1–4. IEEE, 2017.

[24] Chao-Yuan Wu, Manzil Zaheer, Hexiang Hu, R. Manmatha, Alexander J. Smola, Philipp Kr¨ahenb¨uhl. *"Compressed Video Action Recognition."* arXiv:1712.00636v2 [cs.CV] 29 Mar 2018.

[25] Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik. "*Rich feature hierarchies for accurate object detection and semantic segmentation Tech report (v5)."* arXiv:1311.2524v5 [cs.CV] 22 Oct 2014.

[26] Ross Girshick. *"Fast R-CNN."* arXiv:1504.08083v2 [cs.CV] 27 Sep 2015.

[27] Joseph Redmon, Ali Farhadi. *"YOLO9000: Better, Faster, Stronger."* arXiv:1612.08242v1 [cs.CV] 25 Dec 2016.

[28] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam. *"MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Application."* arXiv:1704.04861v1 [cs.CV] 17 Apr 2017.

[29] Shiyao Wang, Hongchao Lu, Zhidong Deng. *"Fast Object Detection in Compressed Video."* arXiv:1811.11057v2 [cs.CV] 31 Jan 2019.

[30] https://www.microsoft.com/en-us/microsoft-365/blog/2020/04/09/remote-work-trend-report-meetings/