e-Health and Resource Management Scheme for a Deep Learning-based Detection of Tumor in Wireless Capsule Endoscopy Videos

Tariq Rahim
ICT Convergence Research Center
Kumoh National Institute of Tech.
Gumi, South Korea
tariqrahim@ieee.org

Arslan Musaddiq
ICT Convergence Research Center
Kumoh National Institute of Tech.
Gumi, South Korea
arslan@kumoh.ac.kr

Dong-Seong Kim*
Dept. of IT Convergence Engr.
Kumoh National Institute of Tech.
Gumi, Korea
dskim@kumoh.ac.kr

***Abstract*—** Recently, a lot of concentration is on how early diagnosis for critical diseases can be accommodated with deep learning (DL). e-health is an emerging area in the junction of medical informatics, public health, and business, indicating health assistance and data delivered or improved by the Internet and associated technologies. Resource management as bandwidth allocation problem is a key problem while transmitting processed medical data where both data integrity and quality are of utmost importance. To address the early intelligent detection and diagnosis of the diseases, an end-to-end DL model i.e., You Only Look Once (YOLOv3-tiny) is selected for the detection of the tumor within with wireless capsule endoscopy videos. The DL mode is an improved version of the YOLOv3-tiny wherein each convolutional layer, different convolutional filters, is employed to extract both local and global features. The motivation is early detection of the critical disease followed by remote physician diagnosis where resource management as a bandwidth allocation is investigated using encoders like H.265/HEVC and VP9. The proposed scheme controls the frame rate, video resolution, and compression ratio as quantization based on the intelligent decision from the DL model. The performance of the improved YOLOv3-tiny model is benchmarked with YOLOv3-tiny and our previous work in terms of precision, sensitivity, F1-score, and F2-score. Furthermore, the resource management results are shown in terms of bandwidth and storage for both encoders.

Keywords—Deep learning; bandwidth; e-health; H.265/HEVC; wireless capsule endoscopy; VP9

I. INTRODUCTION

Healthcare is regarded as a prime preference worldwide as the patient's number having distinct diseases grows quickly causing an increment in spending on healthcare. Recently, the number of patients that require constant monitoring rose speedily, and considering the common paradigm of providing healthcare facilities will not be ample; as the number of doctors is not equivalent to the patient's numbers. One main solution is to exploit e-health which is an emerging area in the junction of medical informatics, public health, and business, indicating health assistance and data delivered or improved by the Internet and associated technologies [1]. Diagnosing different diseases inside the small intestine timely is a tumultuous and time-consuming process for doctors. This has driven the opening of technologies such as wireless capsule endoscopy (WCE) and colonoscopy [2], [3]. A critical step in every WCE screening process is to handle the number of images generated during one process i.e., more than 55,000 images, and then to classify and detect pre-malignant or malignant tumors. Computer-aided techniques such as computer vision especially provide services for automated examination of WCE videos leading to the reduction of time taken by doctors for processing and analyzing videos to find malignant tissues.

Deep learning (DL) performs a significant part in several areas, including healthcare, image recognition, self-driving cars, and text recognition tasks [3]. A unique model of Convolutional Neural Network (CNN) as regression-based comprising of two stages i.e., spatial features analysis and temporal information tracking is employed for the detection of polyps [4]. CNNs demonstrate encouraging outcomes for segmentation and object detection such as unmanned aerial vehicles, solar panels, etc. [5], [6]. Both low-level image processing approaches, DL-based approaches are discussed for the detection of tumors in WCE images in a detailed fashion [3]. For the past two decades, the region-based CNN methods, such as R-CNN, Fast R-CNN, and Faster R-CNN exhibited encouraging outcomes for objects and polyp's detection [2], [7]. Regression-based efforts are conducted by using the You Only Look Once (YOLO) [8] and

single-shot multi-box detector (SSD) [9] for the detection of polyps.

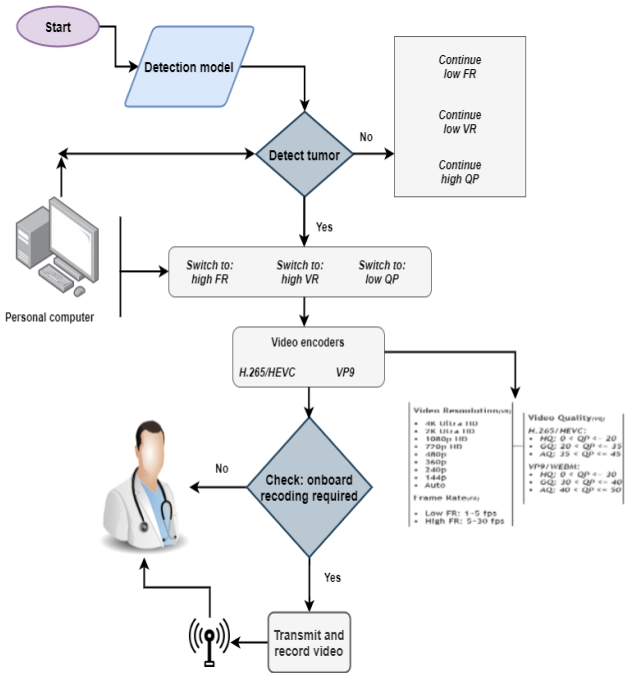

Figure 1. Schematic of the proposed e-Health resource management system.

The wide usage of digital video for communication and entertainment drives researchers to obtain a commending perception of videos [10]. With the progression in technology for capturing, storing, sending, and displaying videos, high-quality video services have lately become widespread. A direct relationship is observed as the frame rate increase with the issue that high-quality video contents are normally in large file size. This led to the introduction of the different encoders such as H.264, High-Efficiency Video Coding (HEVC) /H.265, and VP9, etc. providing better compression rates with the almost same perceptual quality of experience [11]. The encoding of the videos at different frame rates and video resolutions becomes important when there is either limited bandwidth od data is of crucial importance as in case of medical imaging data.

Considering these aspects, in this paper a DL based model i.e., YOLOv3-tiny is improved for the detection of tumor within the WCE images. The model is improved by adding different size of convolutional filters within the same convolutional layer that can extract both local and global features. A proposal scenario is given for e-health application where resource management has been addressed while transmitting these processed medical data. Encoding of the medical data based on the DL model decision is carried out using encoders like HEVC and VP9, where a condition-based decision is performed accordingly for switching the frame rates (FR), video resolution (VR), and quantization parameter (QP) selection.

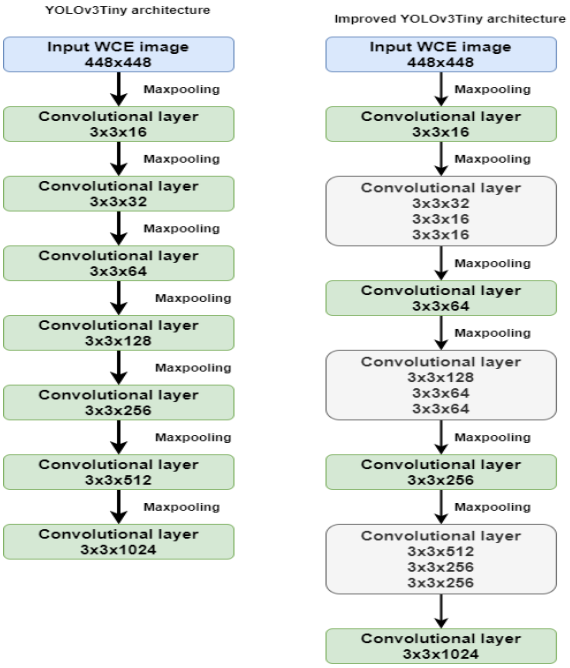

Figure 2. DL model where both default and improved YOLOv3-tiny models are shown.

II. Proposed e-Health Resource Management Scheme

As shown in Fig. 1 the proposed e-health resource management system comprised of a DL model for the detection of the tumor within the WCE images followed by switching frame rate (FR), resolution, and quantization parameter (QP).

The DL model employs improved YOLOv3-tiny considering the color, textural, and spatial appearance of the tumor within the image and gastrointestinal tract of human body. The YOLOv3 employs darknet-53 which is a complex model. In the training stage, the input WCE images are divided into grid cells $A \times A$ having cell size of 8×8 producing 64 cells, and every grid cell will define the probability of occurrence of the tumor inside the image. Using RectLabel tool, the dataset is labeled that generate a bounding box "B" comprising of five predictions. Individual grid cell predicts "B" which depicts the box that includes an object and confidence score "C_s" for a specific class of the object inside. These predictions include the horizontal, vertical, height, width components as $x, y, h, and\ w$ respectively. Lastly, a "C_s" is determined for each specified grid cell that exhibits the confidence or how positive the DL model is that the rectangle includes an object [2]. The intersection over union (IoU) also described as the Jaccard index is an essential expression for computing the distance within the predicted and the ground truth "B" and employed as a performance metric to relate the similarity between two random shapes [12]. The IoU is given as follow:

$$IoU = \frac{|J \cap K|}{|J \cup K|} = \frac{I}{U} \qquad (1)$$

where, "J" and "K" describes the predicted and ground truth respectively.

Table. I Some of the simulation parameters employed in the training phase for improved YOLOv3-tiny model.

Parameters used	Configurations values
Image size	448 × 448
Optimizer	SGD
Learning rate (η)	0.0001
Batch size	32
Momentum	0.9
Iterations	10,000

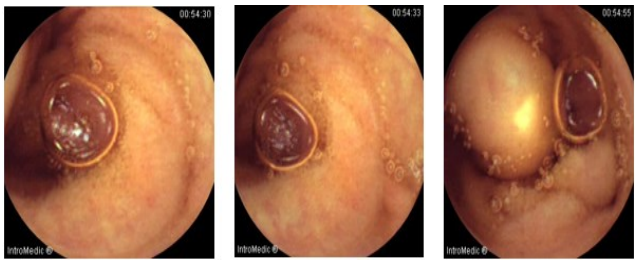

Figure 3. Some of the representative tumor frames from the Intromedic dataset [14].

Fig. 2 shows the improved YOLOv3-tiny model comprising of 7 convolutional layers with 6 pooling layers. The model is improved from default YOLOv3-tiny by adding different convolutional filter sizes such as 16, 32, 64, 128, 512, and 1024. This improvement was achieved by adding the different convolutional filter size within the same layer such as 128, 64, 64 and then followed by conventional approach of 512 convolutional filter size as shown in Fig. 2. This modification is performed for the extraction of the useful features within the WCE images that can be present at different spatial positions.

For faster convergence stochastic gradient descent (SGD) is employed as an optimizer that makes the training faster. For the activation function, rectified linear unit (ReLU) is used that lessen the vanishing of gradient descent and obtaining a better sparsity. For achieving a singular linear continuous vector as a flattening, two fully connected layers are used that are followed by *Softmax* for obtaining the required detection outcome. The detailed simulation parameters for the improved YOLOv3-tiny model are presented in Table. I.

The detection model is followed by a resource management system where the two latest prominent encoders i.e., H.265/HEVC and VP9 are employed for encoding the videos at different QP levels. H.265/HEVC which is established by Joint Collaborative Team on Video Coding and founded by ITU-T Video Coding Experts Group and the ISO/IEC Moving Pictures Expert Group in 2010 provides a better compression efficiency than its prior encoders [13]. VP9 which is introduced recently as a WebM project and is an open-source encoder for compression provides an approximate similar efficacy when compared to H.265/HEVC [10]. The idea is to employ both these encoders after the DL detection model to encode the videos and transmit them to a physician at a high frame rate (FR), high video resolution (VR), and low QP if there is successful detection. In this way, the bandwidth savage can be achieved as only those frames that are crucial are transmitted at high FR, VR, and low QP while the rest of the frames are encoded and transmitted at low FR, VR, and high QP. The choice of using two encoders i.e., H.265/HEVS is to check the impact of compression, perceptual quality at different frame rates i.e., 15fps, 30fps, and 60fps, and bandwidth saving. The preliminary results of bandwidth saving are presented in this work. So, for the bandwidth saving the formulation can be done as follow:

$$Bandwidth\ saved\ (\%) = \left(1 - \frac{B_p}{B_h}\right) \times 100 \qquad (2)$$

where, B_p is the consumed bandwidth by single stream of video after presented model and B_h is the consumed bandwidth by single stream of video that is encoded at high FR, high VR, and low QP value.

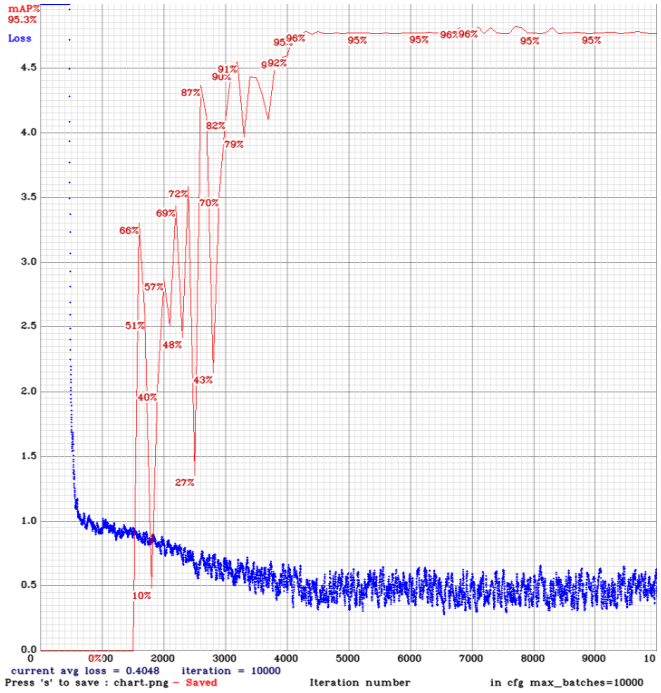

Figure 4. Real-time training stage of the proposed improved YOLOv3-tiny model for the detection of tumor.

III. Expeirmenal Resulst and Discussion

In this section, we detail the results generated from the proposed improved YOLOv3-tiny model DL model of the tumor detection and resource management scheme as bandwidth saving.

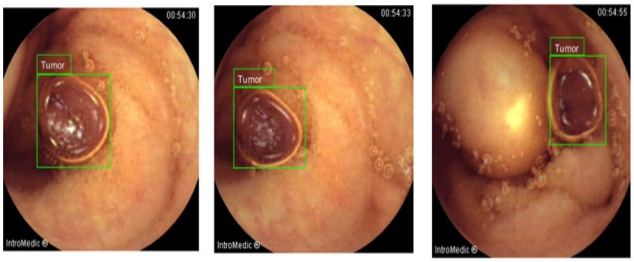

Figure 5 Tumor detection by the proposed improved YOLOv3-tiny model.

For the detection of the tumor within WCE images, the proposed YOLOv3-tiny model is trained and tested with the dataset acquired from the Intromedic cooperation as shown in Fig.3. For the specific implementation, 1000 images are selected that were divided into 70% and 30% for training and testing, respectively. The training was performed according to the Table. I parametric configuration. The real-time training stage is depicted in Fig. 4 where after every 1000 iterations, weights are updated and the average loss is minimized while the average precision is updated. As it can be observed that an average precision as mean average precision (mAP) of 95.30% is achieved after 10,000 iterations with an average loss of 0.4048. After the training phase completion, the model is tested and the detection results are shown in Fig. 5 where it can be seen as a successful detection.

The performance efficacy of the proposed improved YOLOv3-tiny model is shown in Table. II where performance metrics such as precision, sensitivity, F1-score, and F2-score are opted. These metrics are estimated after the calculation of true positive, true negative, false positive, and false negative of the model. Precision reflects that how much precisely the model is detecting a tumor within WCE image, while sensitivity shows the calculation of the proportion of the actual tumor detection correctly. To balance the precision and sensitivity as hormonic mean, F1-score and F2-score are computed. The detailed formulation can be found in [2], where we have shown how these metrics are calculated. As shown in Table. III, it can be observed that the performance efficacy of the proposed improved YOLOv3-tiny model outperforms the default YOLOv3-tiny model in terms of performance metrics reported within the Table. III.

Table. III Performance evaluation in terms of opted metrics.

Performance metrics	**YOLOv3-tiny (%)**	**[5] (%)**	**Our DL model (%)**
Precision	89.21	90.48	93.40
Sensitivity	87.47	88.02	92.08
F1-score	86.52	86.88	92.11
F2-score	88.37	89.15	91.85

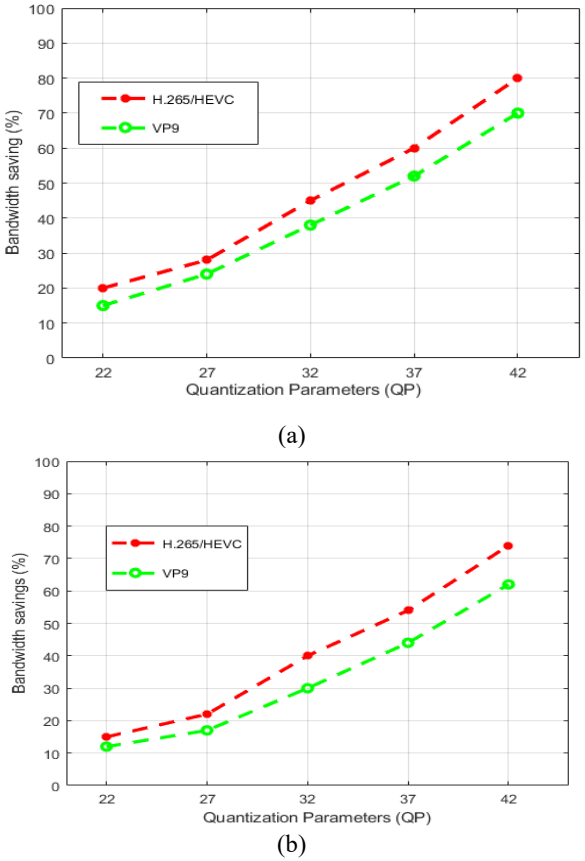

Figure 6. Bandwidth saving in percentage for both H.265/HEVC and VP9 encoders where (a) 30fps and (b) 60fps are shown.

Moreover, Fig. 6 depicts the attempt of a bandwidth saving scheme at two different rates i.e., 30fps and 60fps while employing H.265/HEVC and VP encoders. For encoding libraries of FFmpeg are used at different QP levels [10]. The bandwidth saving in percentage for 30fps is more when the proposed DL-based model operates and can be employed by choosing high QP levels that can additionally save storage issue of the file size. For 60fps that comes under the category of high frame rate and is favorable for high-quality videos also show a good performance in terms of bandwidth saving. The performance of H.265/HEVC is showing a better performance but as VP9 is a royalty-free open-source encoder, so its usage is preferable for the non-license scenario.

IV. Conclusion and Future Work

For the detection of tumor within WCE images, an improved YOLOv3-tiny model is presented by adding different convolutional filter sizes that resulted in the extraction of local and global features. This extraction process enhanced the

detection capability of the model in comparison to the default YOLOv3-tiny model. The results are benchmarked with default YOLOv3-tiny model in detailed fashion. Furthermore, a resource management scheme is provided where upon the detection a decision is made for switching the video to high FR, high VR, and low QP resulting in bandwidth saving. Two encoders i.e., H.265/HEVC and VP9 are opted for encoding and resource management scheme. The initial results are only shown for 30fps and 60fps.

In future work, one can focus on a more efficient DL model for the detection of malignant tissues within the WCE images, and the storage capability of the encoders can be investigated. Also, different frame rates and different resolution of videos can be examined with a subjective analysis too.

ACKNOWLEDGMENT

This research was financially supported by the MSIT (Ministry of Science, ICT), Korea, under the Grand Information Technology Research Center support program (IITP-2020-2020-0-01612) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation), and Priority Research Centers Program (2018R1A6A1A03024003) through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology.

REFERENCES

[1] Chen, Toly. "Assessing factors critical to smart technology applications to mobile health care− the fgm-fahp approach." Health policy and technology 9, no. 2 (2020): 194-203.

[2] Rahim, Tariq, Syed Ali Hassan, and Soo Young Shin. "A deep convolutional neural network for the detection of polyps in colonoscopy images." Biomedical Signal Processing and Control 68 (2021): 102654.

[3] Rahim, Tariq, Muhammad Arslan Usman, and Soo Young Shin. "A survey on contemporary computer-aided tumor, polyp, and ulcer detection methods in wireless capsule endoscopy imaging." Computerized Medical Imaging and Graphics (2020): 101767.

[4] Zhang, Ruikai, Yali Zheng, Carmen CY Poon, Dinggang Shen, and James YW Lau. "Polyp detection during colonoscopy using a regression-based convolutional neural network with a tracker." Pattern recognition 83 (2018): 209-219.

[5] Hassan, Syed Ali, Tariq Rahim, and Soo Young Shin. "Real-time uav detection based on deep learning network." In 2019 International Conference on Information and Communication Technology Convergence (ICTC), pp. 630-632. IEEE, 2019.

[6] Han, Seung Heon, Tariq Rahim, and Soo Young Shin. "Detection of Faults in Solar Panels Using Deep Learning." In 2021 International Conference on Electronics, Information, and Communication (ICEIC), pp. 1-4. IEEE, 2021.

[7] Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. "Faster R-CNN: towards real-time object detection with region proposal networks." IEEE transactions on pattern analysis and machine intelligence 39, no. 6 (2016): 1137-1149.

[8] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767 (2018).

[9] Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. "Ssd: Single shot multibox detector." In European conference on computer vision, pp. 21-37. Springer, Cham, 2016.

[10] Rahim, Tariq, and Soo Young Shin. "Subjective Evaluation of Ultra-high Definition (UHD) Videos." KSII Transactions on Internet & Information Systems 14, no. 6 (2020).

[11] Rahim, Tariq, Muhammad Arslan Usman, and Soo Young Shin. "Comparing H. 265/HEVC and VP9: Impact of High Frame Rates on the Perceptual Quality of Compressed Videos." arXiv preprint arXiv:2006.02671 (2020).

[12] Kosub, Sven. "A note on the triangle inequality for the Jaccard distance." Pattern Recognition Letters 120 (2019): 36-38.

[13] Sullivan, Gary J., Jens-Rainer Ohm, Woo-Jin Han, and Thomas Wiegand. "Overview of the high efficiency video coding (HEVC) standard." IEEE Transactions on circuits and systems for video technology 22, no. 12 (2012): 1649-1668.

[14] Usman, Muhammad Arslan, Gandeva B. Satrya, Muhammad Rehan Usman, and Soo Young Shin. "Detection of small colon bleeding in wireless capsule endoscopy videos." Computerized Medical Imaging and Graphics 54 (2016): 16-26.