

Influence of color correction on pathology detection in Capsule Endoscopy

Bidosse Emmanuel Agossou, Marius Pedersen, Kiran Raja, Anuja Vats, and
Pål Anders Floor

Department of Computer Science, Norwegian University of Science and Technology
(NTNU)

`{bidosse Emmanuel Agossou, marius.pedersen, kiran.raja, anuja.vats,
paal.anders.floor}@ntnu.no`

Abstract. Pathology detection in Wireless Capsule Endoscopy (WCE) using deep learning has been explored in the recent past. However, deep learning models can be influenced by the color quality of the dataset used to train them, impacting detection, segmentation and classification tasks. In this work, we evaluate the impact of color correction on pathology detection using two prominent object detection models: Retinanet and YOLOv5. We first generate two color corrected versions of a popular WCE dataset (i.e., SEE-AI dataset) using two different color correction functions. We then evaluate the performance of the Retinanet and YOLOv5 on the original and color corrected versions of the dataset. The results reveal that color correction makes the models generate larger bounding boxes and larger intersection areas with the ground truth annotations. Furthermore, color correction leads to an increased number of false positives for certain pathologies. However, these effects do not translate into a consistent improvement in performance metrics such as F1-scores, IoU, and AP50. The code is available at <https://github.com/agossouema2011/WCE2024>.

Keywords: Wireless Capsule Endoscopy, · Color correction · Retinanet · YOLOv5 · Detection

1 Introduction

Digestive system diseases are widespread around the world and cause considerable distress which can be fatal. In 2019, about 2276.27 million estimated prevalent cases and 2.56 million deaths were counted for digestive system diseases [1]. Endoscopy is the procedure commonly used to visualize the gastrointestinal tract and detect diseases [2]. However, traditional endoscopy is uncomfortable and painful for patients [3], discouraging it for wider preventive screening programs. A new type of gastrointestinal endoscopy, known as Wireless Capsule Endoscopy (WCE) [4], was introduced in 2000, which is an alternative and minimally invasive screening method. In WCE, the patient swallows a pillcam (pill-sized camera) which moves through the gastrointestinal tract and records videos

[5]. The whole process can take on average eight to twelve hours, generating around 60,000 frames [6]. Considering the long procedure duration, the demand placed on gastroenterologists for WCE screening is huge, leaving room for missing pathologies. Therefore, methods to improve pathology detection accuracy and reduce the time spent in screening WCE images are of great importance.

However, WCE has some drawbacks compared to traditional endoscopy like low frame rate and images of low resolution with poor quality [7]. In addition, colors are distorted in WCE, making it challenging to perform color reproduction and color consistency [8]. Correcting or enhancing the color of the WCE images before training a deep learning model could help improve pathology detection accuracy. Faithfulness to color of human tissues and blood vessels is key to accurately detecting abnormalities that often deviate from normal in their appearances [8]. Post-processing algorithms can distort colors that may lead to misinterpretation. Thus making color reproduction accuracy and consistency all the more significant for human inspection as well as automated diagnosis. Color accuracy refers to the ability of a system to produce exact color matches from input to output, while color consistency refers to the ability to produce image data with a similar response as that in a human interpreter [8].

This work evaluates the influence of color correction on pathology detection in WCE. Detection in our context means pathology localization inside the image using bounding boxes. To the best of our knowledge, little attention is given to this direction in the literature. The contributions of this work are the following:

- We present two new color corrected versions of the publicly available SEE-AI [9] dataset. The dataset will be made publicly available to other researchers for investigating approaches for classification and detection tasks.
- We present a detailed analysis evaluating how slight changes in colors could potentially impact the behaviour of detection models in WCE. The analysis points to strengths and weakness of color correction for WCE for future works to build upon.

2 Related works

Object detection has gained huge attention [10] and consists of identifying objects or instances of objects in an image or a video using bounding boxes. There are two main groups of object detection methods: *one-stage detectors* and *two-stage detectors* [11][12]. One-stage objects detectors train models using a large amount of ground truth data. i.e., accurate human-annotated data. They require only a single pass through the neural network and predict all the bounding boxes in one go. One-stage methods include You Only Look Once (YOLO) [13], and Single Shot Detector (SSD) [14]. The two-stage detectors on the other hand generate a set of regions of interest with a region proposal network, and then perform classification and bounding box regression. Popular two-stage models include Faster-RCNN [15] and Mask R-CNN [16].

Ding et al. [17] used a Convolutional Neural Network (CNN) based model to first classify WCE images into normal and abnormal frames, and further catego-

size the abnormal images into different pathology classes including inflammation, ulcer, polyps, lymphangiectasia, and bleeding. Aoki et al. [18] used a CNN-based model to detect blood content in WCE images. A CNN-based detection system for various abnormalities using WCE videos from multiple centers was developed by Aoki et al. [19]. The system combines four CNN-based models: three SSD models and one ResNet50, to detect among other, mucosal breaks, angioectasia, polyps, submucosal tumors, venous structures, blood contents. Yokote et al. [9] created an annotated dataset for WCE containing 12 types of small bowel pathologies. In addition, an object detection model based on YOLOv5 was used to detect the pathologies.

Cherepkova and Hardeberg [20] investigated the impact of image color enhancement on the performance of a lesion classification algorithm for skin cancer detection. Six different image enhancement algorithms were selected to investigate which algorithms have more influence on lesion detection improvement, focusing especially on color correction. A CNN based classifier and support vector machine were used on a set of dermoscopy image dataset. Color enhancement generally yielded slightly better results for color versions of the images in their analysis [20]. In order to solve color distortion problem in WCE, Watine et al. [8] considered the color correction of endoscopy videos through post-processing using a dedicated color checker to select the colors that typically appear in the human colon. The developed color checker based on the colon gamut was named ColonColorChecker (CCC). Two color correction matrices were derived, one based on a standard Gretag Macbeth ColorChecker (CC) and the other bases on the CCC. These were applied to enhance color in WCE images. From their objective and subjective evaluations with two experienced gastroenterologists, the color corrections provide a clear improvement of color in clinical videos. However, the color corrections decreased the contrast in the color corrected videos reducing the diagnostic value of the approach [8].

3 Methodology

Motivated by the limited attention to localizing and detecting the pathology when color correction is applied using CC and CCC, we present a series of analysis in this work. We make use of the publicly available WCE SEE-AI dataset [9] and create two color corrected versions using CC and CCC [8]. As a result, we get two color corrected datasets. We then benchmark Retinanet and YOLOv5 models [21] on all the three datasets: the original SEE-AI, and the two color corrected, to understand the impact on pathology detection. We briefly present the WCE SEE-AI dataset [9] and relevant details of color correction. Further, we present our experimental set up and discuss relevant metrics used for analysis.

3.1 Dataset

SEE-AI Dataset: The SEE-AI dataset [9] is an annotated dataset for WCE with bounding box coordinates that contains 12 classes of small bowel pathologies. The images were collected with WCE PillCam SB-3 [22]. The dataset

contains 18,481 images extracted from 523 small bowel WCE procedures performed at Kyushu University Hospital (Japan). 12,320 images were annotated with 23,033 disease lesions, and combined with 6,161 normal images. The SEE-AI dataset exhibits imbalance in pathology categories with fewer images of some pathologies (diverticulum, stenosis, SMT) as opposed to others (lymph-follicle, erosion, polyp-like) as shown in Fig. 1.

Color Corrected Datasets: Making use of two color correction matrices proposed by [8], i.e., CC and the CCC corrections, we create two additional datasets for our analysis. The CC was generated from the standard 24 patch color checker, while the CCC was generated from a selection of 30 colors (24 colors that appear within the gastrointestinal system and 6 gray scale).

- SEE-AI ColorChecker Corrected Dataset (CCD), which is the SEE-AI Dataset from [9] corrected with the CC . We refer to this dataset as CCD.
- SEE-AI Colon Color Corrected Dataset (CCCD), which is the SEE-AI Dataset from [9] corrected with the CCC . We refer to this dataset as CCCD.

Fig. 2 presents two samples of the original SEE-AI images (in the middle), and the corresponding CC images (on the left side) and CCC images (on the right side). As seen in Fig. 2, color correction leads to a significant enhancement in the colors, particularly the reddish hues. However, compared to the original images, a loss of contrast can be observed.

3.2 Experimental Protocols

Challenges faced by object detection methods with images include *class imbalance* [23, 21] and *multiple aspect ratios* [24]. Class imbalance is when the dataset has only few labeled images for some classes, or when the spatial area of object of interest is smaller compared to the background. Multiple aspect ratios (aspect ratio imbalance) are when objects vary in aspect ratio with a variety of sizes: small, medium, and large objects with varying orientations[25]. These challenges make it difficult for object detection methods to achieve good detection results. Considering the nature of the dataset used for this analysis, we make use of Retinanet model as it is reported to have better performance under class imbalance problems due to incorporated focal loss [21]. To verify the consistency of our results, we also benchmark YOLOv5 with focal loss on our datasets. We further study each of the two models in three different settings that focus on pathology detection:

1. Retinanet and YOLOv5 on SEE-AI Dataset (R-OrigD)
2. Retinanet and YOLOv5 on the CCD (R-CCD)
3. Retinanet and YOLOv5 on the CCCD (R-CCCD)

3.3 Training and performance measures

For each case, we train each model using 80% of data and test with the remaining 20%. In addition, we ensured that both the training and testing sets have images

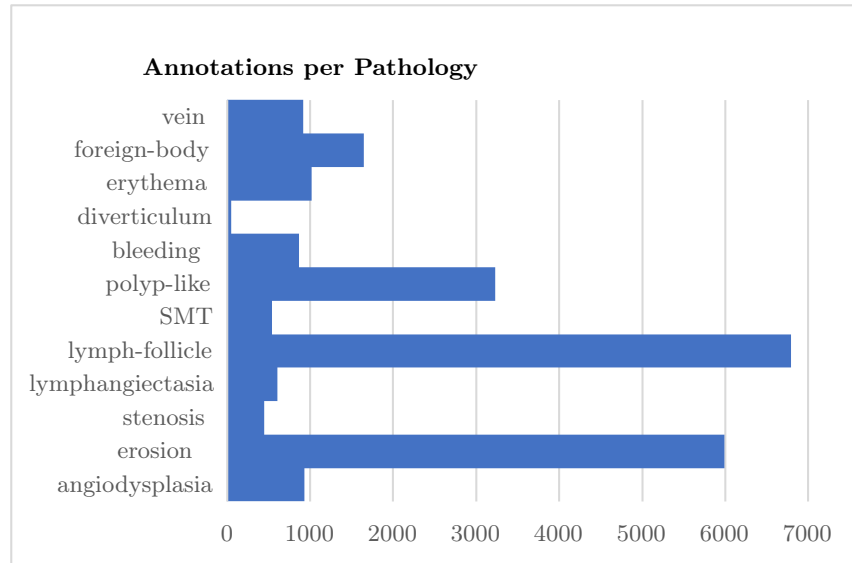


Fig. 1. SEE-AI Dataset Annotations per class. We can observe that classes like lymph-follicle, erosion and polyp-like have more annotations, where as SMT, stenosis, and diverticulum have very few annotations.

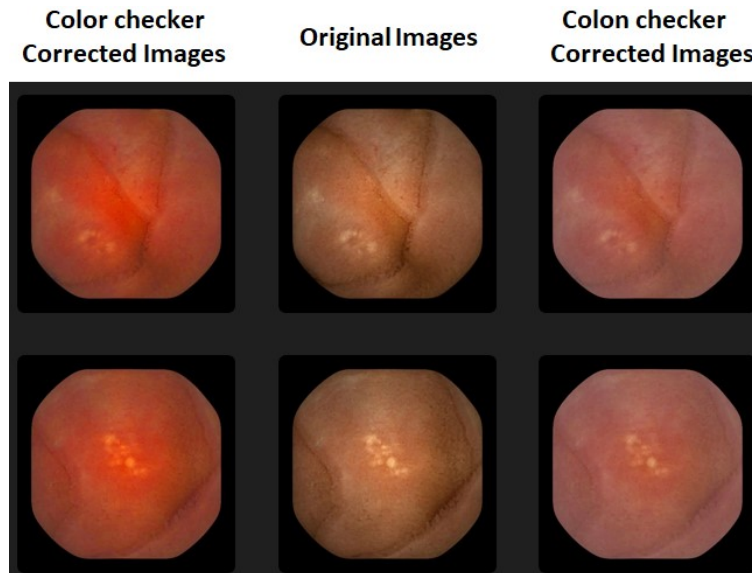


Fig. 2. Original images with their CC and CCC corrected versions. Compared to the original images, a loss of contrast can be observed with color corrected images.

from all 12 pathology classes. Retinanet and YOLOv5 were trained in each setting and the best weights were used for the testing.

The Retinanet method was evaluated using three metrics: F1-score, Intersection over Union (IoU) and Average Precision (AP) with an IoU threshold 0.5. YOLOv5 was evaluated using F1-score and AP. F1-score is a performance metric used for the classification task. It measures the model's accuracy by combining both the precision and recall scores. The precision in Equation (1) is a metric that measures the model's ability to classify a sample as positive. The recall in Equation (1) is the ability of a classification model to find all the relevant data points. TP is the total number of True Positives for a given class, FP the total number of False Positives, and FN is the total number of False Negatives. The range of F1-score in Equation (2) is between 0 and 1 (1 being the best performance).

$$Precision = \frac{TP}{TP + FP} \quad Recall = \frac{TP}{TP + FN} \quad (1)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (2)$$

$$AP_i = \sum_n (Recall_n - Recall_{n-1}) \times Precision_n, \quad (3)$$

where i represents the indice for a given image. and n represents the thresholds across which TP, FP, TN, FN are determined. The AP for a class $c := \{x_1, x_2, \dots, x_M\}$ is then given by:

$$AP_c = \frac{1}{M} \sum_{i=1}^M AP_i. \quad (4)$$

The metrics for each class are weighted based on their distribution within the dataset to compute the overall AP and F1 scores.

$$AP = \sum_{c=1}^N AP_c \times w_c; \quad F1 = \sum_{c=1}^N F1_c \times w_c \quad (5)$$

$$\text{where, } w_c = \frac{M_c}{\sum_{i=1}^N M_i} \quad (6)$$

where, N is the total number of classes (pathologies)

AP is a well-known performance metric which is frequently used to evaluate the performance of the detection model. AP is high when both precision and recall are high, and low when either of them is low across a range of confidence

threshold values. The range for AP is between 0 to 1 (1 being the best performance) [26]. In this work, we report AP with 50% threshold because 50% is a reasonable overlap between the predicted and actual bounding boxes and a good indication for a detection model performance [27]. F1 and AP in Tables 1-2 are computed using Equation (5). IoU is the ratio of the overlap area (intersection) to the combined area of prediction and ground truth called *union* [28]. Table 3 presents the performance metrics for IoU across the three experiments and the increase in percentages against the R-OrigD.

4 Results and Discussion

The results of the models performance metrics on the test set are summarized in Tables 1- 3 and Fig. 3- 9. Classes such as *diverticulum* and *foreign-body* have been exempted from our analysis due to very few and non-diverse samples of the pathologies, that could potentially bias the results.

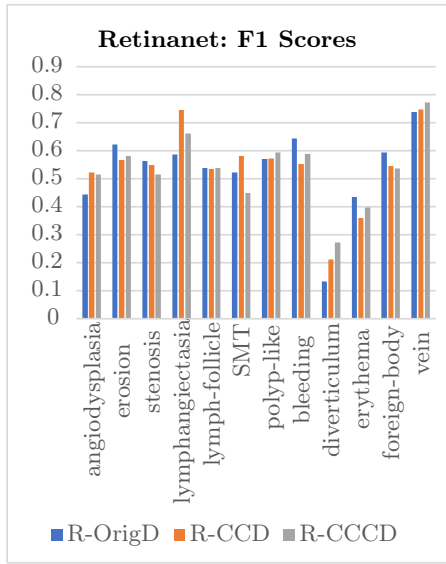


Fig. 3. Retinanet F1-scores across different color schemes. The F1 scores increase with both color corrections (CC and CCC) for angiodysplasia, lymphangiectasia and vein. Only CC increases F1-score for SMT while CCC increases F1-score polyp-like.

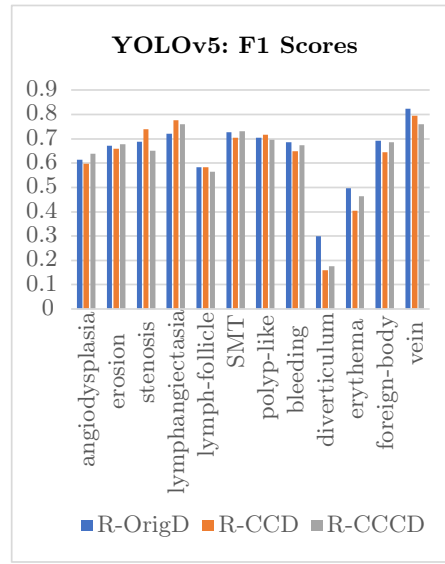


Fig. 4. YOLOv5 F1-scores across different color schemes. The F1 scores increase with both color corrections only for lymphangiectasia. Only CC increases F1-score for stenosis while CCC increases F1-score for angiodysplasia.

From the analysis of the performance metrics F1-score (Table 1) and AP (Table 2) for both Retinanet and YOLOv5, it is observed that overall, the color

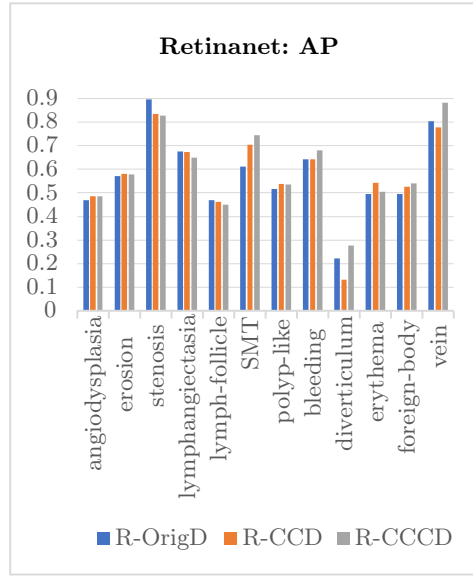


Fig. 5. Retinanet AP across different color schemes. The AP increase with both color corrections (CC and CCC) for angiodysplasia, SMT and polyp-like. Only CC increases AP for Erythema while CCC increases AP for bleeding and vein.

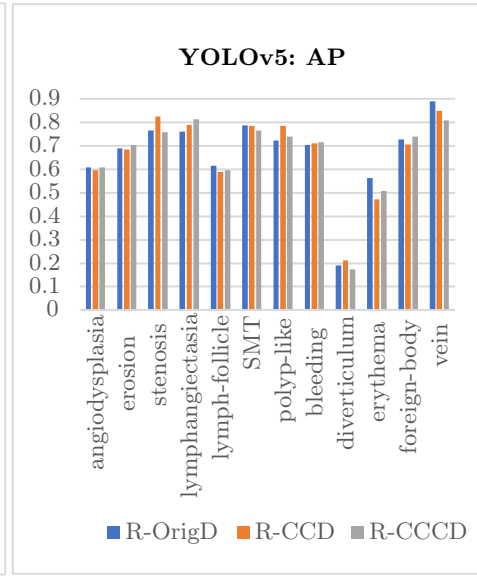


Fig. 6. YOLOv5: AP across different color schemes. The AP increase with both color corrections (CC and CCC) for lymphangiectasia, polyp-like and bleeding. Only CC increases AP for stenosis, while CCC increases AP for erosion and angiodysplasia.

Table 1. Models performances in F1-scores across the three different color schemes. Bold values indicate wherever color-correction results in better scores.

Pathologies	Weights	Retinanet			YOLOv5		
		R-OrigD	R-CCD	R-CCCD	R-OrigD	R-CCD	R-CCCD
angiodysplasia	0.036	0.444	0.523	0.515	0.629	0.601	0.634
erosion	0.252	0.623	0.567	0.581	0.676	0.667	0.677
stenosis	0.021	0.564	0.549	0.515	0.691	0.747	0.657
lymphangiectasia	0.025	0.587	0.746	0.662	0.718	0.776	0.759
lymph-follicle	0.299	0.538	0.535	0.539	0.584	0.583	0.563
SMT	0.024	0.522	0.582	0.449	0.736	0.705	0.733
polyp-like	0.144	0.570	0.572	0.593	0.697	0.718	0.696
bleeding	0.049	0.643	0.552	0.589	0.693	0.650	0.677
diverticulum	0.003	0.133	0.211	0.273	0.301	0.159	0.175
erythema	0.048	0.435	0.360	0.397	0.492	0.403	0.476
foreign-body	0.068	0.593	0.545	0.537	0.697	0.643	0.680
vein	0.031	0.739	0.747	0.773	0.816	0.797	0.761
F1-score		0.571	0.553	0.559	0.649	0.638	0.638

Table 2. Models performances in AP across the three different experiments schemes. Bold values indicate wherever color-correction results in better scores.

Pathologies	Weights	RetinaNet			YOLOv5		
		R-OrigD	R-CCD	R-CCCD	R-OrigD	R-CCD	R-CCCD
angiodyspasia	0.036	0.468	0.486	0.485	0.627	0.595	0.608
erosion	0.252	0.572	0.580	0.577	0.695	0.684	0.706
stenosis	0.021	0.895	0.834	0.828	0.766	0.827	0.765
lymphangiectasia	0.025	0.676	0.672	0.649	0.761	0.790	0.814
lymph-follicle	0.299	0.469	0.463	0.449	0.622	0.588	0.597
SMT	0.024	0.611	0.703	0.743	0.791	0.785	0.765
polyp-like	0.144	0.516	0.538	0.535	0.705	0.756	0.740
bleeding	0.049	0.643	0.643	0.679	0.706	0.709	0.714
diverticulum	0.003	0.222	0.133	0.278	0.190	0.213	0.173
erythema	0.048	0.496	0.543	0.504	0.536	0.464	0.497
foreign-body	0.068	0.496	0.525	0.541	0.727	0.706	0.739
vein	0.031	0.804	0.778	0.882	0.889	0.849	0.809
AP		0.540	0.548	0.548	0.677	0.666	0.674

corrections have not improved the models detection performance. Consequently, we conduct a detailed investigation into the effect of color correction for specific pathologies such as Erythema, Erosion, Polyp-Like and Bleeding, seeking statistical significance to more precisely attribute the observed effects. To achieve this, for each model, we analyzed the number of false positives detections, the detected bounding boxes areas (number of pixels), and the percentage of intersection area with the ground truth bounding box. Then, we conducted a combined sign tests for false positives and bounding boxes areas for the selected pathologies between the three different color schemes. The results of p-values of the combined sign test are presented in Tables 4-5. Green cells indicate where there is a significant difference between different color schemes (where p-values < 0.05). We also notice that with both models, there is significant difference between the three color schemes. All the tables and codes are available on our github link.

From the p-values, it is observed that there is a significant difference in the number of false positives and bounding boxes areas generated by the models between the three different color schemes. Therefore, color correction significantly impacts the bounding box size and the number of false positives. Further analysis of the FP results shows that, on average, R-CCD has significantly more FP compared to R-CCCD and R-OrigD, and that R-CCCD has significantly more than R-OrigD (Table 6). Analysis for each pathology reveals that the aggregate effect of increased false positives is dependent on the specific characteristics of the pathologies. For Erosion, Erythema, Bleeding and Polyp-like, color-correction schemes generally lead to an increase in the number of false positives generated by the model compared to without color correction (Table 6). For instance, for bleeding, without color correction, there are fewer false positives than after cor-

Table 3. Retinanet: IoU comparison across the three different color schemes. Green arrows indicate wherever there is an increase of the metric value whereas red arrows indicate wherever there is a decrease.

Pathologies	R-OrigD	R-CCD	R-CCCD	% I1	% I2
angiodyspasia	0.284	0.362 ↑	0.342 ↑	7.80	5.80
erosion	0.524	0.434 ↓	0.448 ↓	-9.00	-7.60
stenosis	0.367	0.338 ↓	0.325 ↓	-2.90	-4.20
lymphangiectasia	0.391	0.624 ↑	0.487 ↑	23.30	9.60
lymph-follicle	0.339	0.342 ↑	0.383 ↑	0.30	4.40
SMT	0.342	0.383 ↑	0.260 ↓	4.10	-8.20
polyp-like	0.452	0.462 ↑	0.486 ↑	1.00	3.40
bleeding	0.536	0.429 ↓	0.442 ↓	-10.70	-9.40
diverticulum	0.116	0.165 ↑	0.200 ↑	4.90	8.40
erythema	0.288	0.218 ↓	0.246 ↓	-7.00	-4.20
foreign-body	0.468	0.365 ↓	0.396 ↓	-10.30	-7.20
vein	0.585	0.607 ↑	0.580 ↓	2.20	-0.50

%I1 is the increase with R-CCD against R-OrigD

%I2 is the increase of R-CCCD against R-OrigD

recting the color. This may be due to the fact that color-corrected images tend to be more saturated, appearing visually more red.

Table 4. False Positives: p-values of the combined sign test. In green cells, p-values < 0.05, meaning there is a significant difference between different color schemes.

	Retinanet			YOLOv5		
	R-OrigD	R-CCD	R-CCCD	R-OrigD	R-CCD	R-CCCD
R-OrigD	X	7.5267e-19	3.8419e-09	X	1.3781e-04	0.0031
R-CCD	7.5267e-19	X	0.0018	1.3781e-04	X	0.2125
R-CCCD	3.8419e-09	0.0018	X	0.0031	0.2125	X

Given that color correction leads models to vary the number of generated bounding boxes compared to the ground truth, with this effect being pathology-dependent and potentially impacting IoU and AP metrics, we analyzed the model’s ability to identify all abnormal regions in an image post-color correction. This was done to determine whether the model could still detect abnormal regions regardless of the variation in bounding box count from the original images. To capture the model’s response to abnormal regions independently of conventional metrics, we calculated the percentage of area of intersection between the original and generated bounding boxes. We present the violin plots of the percentage of the intersection area of the detected bounding boxes with the ground truth bounding box for different color schemes for Polyp-Like (Figure 7) and Bleeding (Figure 8). From the violin plots, it can be observed the following: For Polyp-Like, with both Retinanet and YOLOv5, R-CCD creates larger intersection area with the ground truth than the R-CCCD, which creates larger intersec-

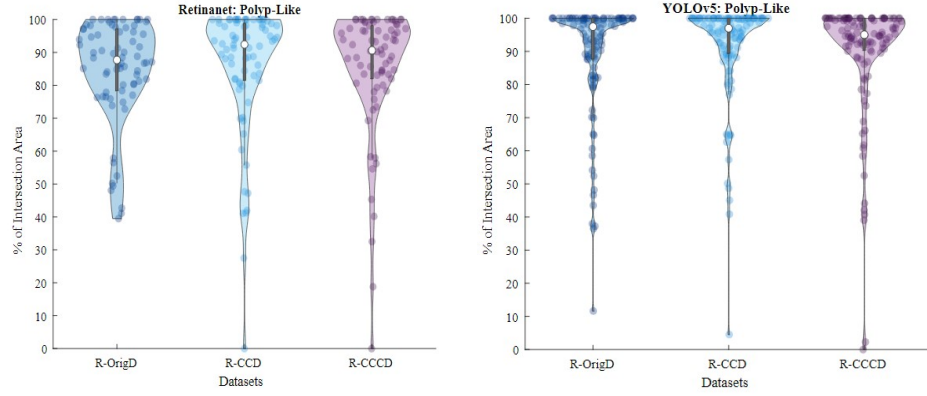


Fig. 7. Violin plots for Polyp-Like. With Retinanet and YOLOv5, R-CCD creates larger intersection area with the ground truth than the R-CCCD, which creates larger intersection area than the R-OrigD.

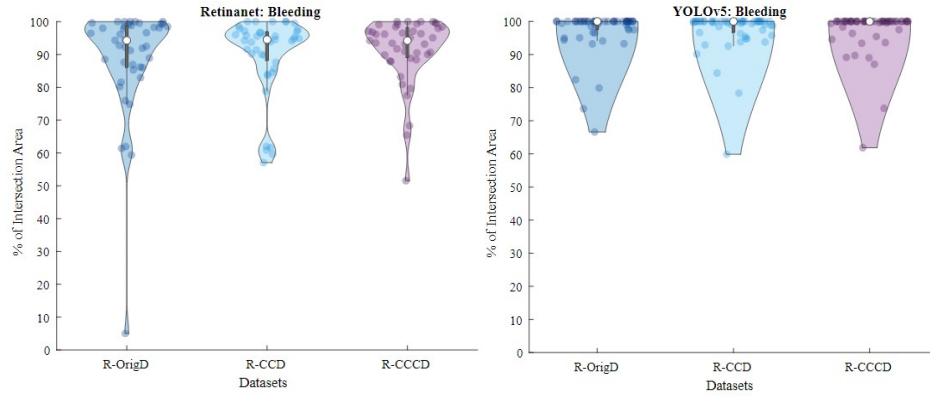


Fig. 8. Violin plots for Bleeding. With Retinanet and YOLOv5, R-CCCD creates larger intersection area than R-CCD and R-OrigD.

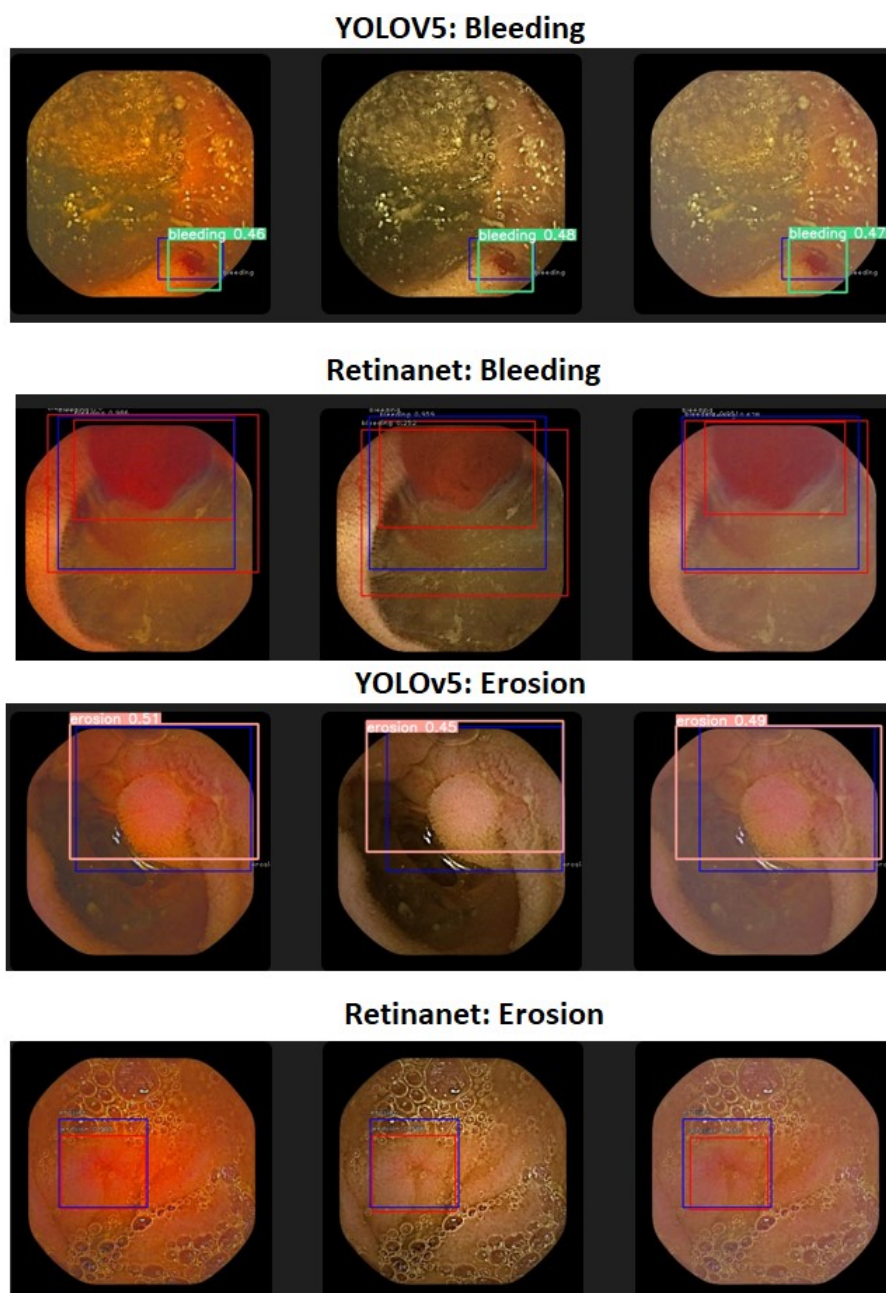


Fig. 9. Models detection results across the three different color schemes.

Table 5. Bounding boxes Areas: p-values of the combined sign test. In green cells, p-values < 0.05 meaning there is a significant difference between different color schemes.

	Retinanet			YOLOv5		
	R-OrigD	R-CCD	R-CCCD	R-OrigD	R-CCD	R-CCCD
R-OrigD	X	4.1079e-08	4.5910e-06	X	0.0094	0.0152
R-CCD	4.1079e-08	X	0.0142	0.0094	X	0.0056
R-CCCD	4.5910e-06	0.0142	X	0.0152	0.0056	X

Table 6. Total number of False Positives. For the selected pathologies, both color corrections (R-CCD, R-CCCD) generally increase the number of false positives generated by the model compared to without color correction (R- OrigD). In green cells are the increase in False Positives.

Pathology	Retinanet			YOLOv5		
	R-OrigD	R-CCD	R-CCCD	R-OrigD	R-CCD	R-CCCD
Erosion	130	250	225	220	239	247
Polyp-like	122	124	90	128	145	138
Erythema	148	234	212	87	93	105
Bleeding	34	65	63	18	17	22
Total False Positives	434	673	590	453	494	512

tion area than the R-OrigD. For Bleeding, with both Retinanet and YOLOv5, R-CCCD creates larger intersection area than R-CCD and R-OrigD. However, with Retinanet, R-CCD creates larger intersection area than R-OrigD where as with YOLOv5, R-CCD creates smaller intersection area than R-OrigD. Overall, the color corrections make the models create larger intersection area with the ground truth than the original dataset.

5 Conclusion

This work evaluates the impact of color on the performance of a deep learning model in wireless capsule endoscopy. We use the SEE-AI dataset, create two color corrected sets and benchmark Retinanet and YOLOv5 models on the three datasets. Given the metrics F1 scores, AP and IoU, the impact of color scheme variations on the performance of the object detection model is heterogeneous across different pathologies and color schemes, resulting in no consistent enhancement or deterioration of performance metrics. Therefore, we further investigated the influence of the color correction on the number of False Positives and the detected bounding boxes. As results, the color corrections generate more false positives and larger intersection areas with the ground truth. However, we observed that the color corrections degrade the contrast of the images in WCE, which leaves rooms for future research to investigate more on the impact of contrast enhancement of the color corrected images on the performance of detection models in WCE.

References

1. R. Wang, Z. Li, S. Liu, and D. Zhang, "Global, regional, and national burden of 10 digestive diseases in 204 countries and territories from 1990 to 2019," *Frontiers in Public Health*, vol. 11, 2023.
2. O. Bchir, M. M. Ben Ismail, and N. AlZahrani, "Multiple bleeding detection in wireless capsule endoscopy," *Signal, Image and Video Processing*, vol. 13, pp. 121–126, 2019.
3. L. J. Neilson, J. Patterson, C. von Wagner, P. Hewitson, L. M. McGregor, L. Sharp, and C. J. Rees, "Patient experience of gastrointestinal endoscopy: Informing the development of the newcastle endoprem™," *Frontline Gastroenterology*, vol. 11, no. 3, pp. 209–217, 2020.
4. G. Iddan, G. Meron, A. Glukhovsky, and P. Swain, "Wireless capsule endoscopy," *Nature*, vol. 405, no. 6785, pp. 417–417, 2000.
5. P. Muruganantham and S. M. Balakrishnan, "A survey on deep learning models for wireless capsule endoscopy image analysis," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 83–92, 2021.
6. T. Rahim, M. A. Usman, and S. Y. Shin, "A survey on contemporary computer-aided tumor, polyp, and ulcer detection methods in wireless capsule endoscopy imaging," *Computerized Medical Imaging and Graphics*, vol. 85, p. 101767, 2020.
7. B. Sushma and P. Aparna, "Recent developments in wireless capsule endoscopy imaging: Compression and summarization techniques," *Computers in Biology and Medicine*, vol. 149, p. 106087, 2022.
8. L. Watine, P. A. Floor, M. Pedersen, P. Nussbaum, B. Ahmad, and Ø. Hovde, "Enhancement of colour reproduction for capsule endoscopy images," in *2023 11th European Workshop on Visual Information Processing (EUVIP)*. IEEE, 2023, pp. 1–6.
9. A. Yokote, J. Umeno, K. Kawasaki, S. Fujioka, Y. Fuyuno, Y. Matsuno, Y. Yoshida, N. Imazu, S. Miyazono, T. Moriyama *et al.*, "Small bowel capsule endoscopy examination and open access database with artificial intelligence: The see-artificial intelligence project," *DEN open*, vol. 4, no. 1, p. e258, 2024.
10. Y.-C. Liu, C.-Y. Ma, Z. He, C.-W. Kuo, K. Chen, P. Zhang, B. Wu, Z. Kira, and P. Vajda, "Unbiased teacher for semi-supervised object detection," *arXiv preprint arXiv:2102.09480*, 2021.
11. Q. Zhou, C. Yu, Z. Wang, Q. Qian, and H. Li, "Instant-teaching: An end-to-end semi-supervised object detection framework," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 4081–4090.
12. P. Mi, J. Lin, Y. Zhou, Y. Shen, G. Luo, X. Sun, L. Cao, R. Fu, Q. Xu, and R. Ji, "Active teacher for semi-supervised object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 14482–14491.
13. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
14. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 21–37.
15. S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

16. K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.
17. Z. Ding, H. Shi, H. Zhang, L. Meng, M. Fan, C. Han, K. Zhang, F. Ming, X. Xie, H. Liu *et al.*, "Gastroenterologist-level identification of small-bowel diseases and normal variants by capsule endoscopy using a deep-learning model," *Gastroenterology*, vol. 157, no. 4, pp. 1044–1054, 2019.
18. T. Aoki, A. Yamada, Y. Kato, H. Saito, A. Tsuboi, A. Nakada, R. Niikura, M. Fujishiro, S. Oka, S. Ishihara *et al.*, "Automatic detection of blood content in capsule endoscopy images based on a deep convolutional neural network," *Journal of gastroenterology and hepatology*, vol. 35, no. 7, pp. 1196–1200, 2020.
19. —, "Automatic detection of various abnormalities in capsule endoscopy videos by a deep learning-based system: a multicenter study," *Gastrointestinal Endoscopy*, vol. 93, no. 1, pp. 165–173, 2021.
20. O. Cherepkova and J. Y. Hardeberg, "Enhancing dermoscopy images to improve melanoma detection," in *2018 Colour and Visual Computing Symposium (CVCS)*. IEEE, 2018, pp. 1–6.
21. T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.
22. M. logo. Pillcam™ sb 3 capsule endoscopy system. [Online]. Available: <https://www.medtronic.com/covidien/en-us/products/capsule-endoscopy/pillcam-sb3-system.html>
23. R. Zhao, B. Qian, X. Zhang, Y. Li, R. Wei, Y. Liu, and Y. Pan, "Rethinking dice loss for medical image segmentation," in *2020 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2020, pp. 851–860.
24. W.-Y. Hsu and W.-Y. Lin, "Ratio-and-scale-aware yolo for pedestrian detection," *IEEE transactions on image processing*, vol. 30, pp. 934–947, 2020.
25. S. Mukherjee, "Top ten challenges in object detection every data scientist should know." [Online]. Available: <https://analyticsindiamag.com/top-ten-challenges-in-object-detection-every-data-scientist-should-know/>
26. A. Anwar. What is average precision in object detection localization algorithms and how to calculate it? [Online]. Available: <https://towardsdatascience.com/what-is-average-precision-in-object-detection-localization-algorithms-and-how-to-calculate-it-3f330efe697b>
27. J. Tian, Q. Jin, Y. Wang, J. Yang, S. Zhang, and D. Sun, "Performance analysis of deep learning-based object detection algorithms on coco benchmark: a comparative study," *Journal of Engineering and Applied Science*, vol. 71, no. 1, p. 76, 2024.
28. Kukil. Intersection over union (iou) in object detection segmentation. [Online]. Available: <https://learnopencv.com/intersection-over-union-iou-in-object-detection-and-segmentation/>