



# Improved Object Detection Artificial Intelligence Using the Revised RetinaNet Model for the Automatic Detection of Ulcerations, Vascular Lesions, and Tumors in Wireless Capsule Endoscopy

Ayako Nakada <sup>1</sup>, Ryota Niikura <sup>1,2,\*</sup>, Keita Otani <sup>3</sup>, Yusuke Kurose <sup>3</sup> , Yoshito Hayashi <sup>4</sup>, Kazuya Kitamura <sup>5</sup>, Hiroyoshi Nakanishi <sup>6</sup>, Seiji Kawano <sup>7</sup>, Testuya Honda <sup>8</sup>, Kenkei Hasatani <sup>9</sup>, Tetsuya Sumiyoshi <sup>10</sup> , Tsutomu Nishida <sup>11</sup> , Atsuo Yamada <sup>1</sup>, Tomonori Aoki <sup>1</sup>, Tatsuya Harada <sup>3</sup>, Takashi Kawai <sup>2</sup> and Mitsuhiro Fujishiro <sup>1</sup>

- <sup>1</sup> Department of Gastroenterology, Graduate School of Medicine, The University of Tokyo, Tokyo 1138655, Japan  
<sup>2</sup> Department of Gastroenterological Endoscopy, Tokyo Medical University Hospital, Tokyo 1600023, Japan  
<sup>3</sup> Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo 1538904, Japan  
<sup>4</sup> Department of Gastroenterology and Hepatology, Osaka University Graduate School of Medicine, Osaka 5650871, Japan  
<sup>5</sup> Department of Gastroenterology, Kanazawa University Hospital, Kanazawa 9208641, Japan  
<sup>6</sup> Department of Gastroenterology, Ishikawa Prefectural Central Hospital, Ishikawa 9208530, Japan  
<sup>7</sup> Department of Gastroenterology and Hepatology, Okayama University Graduate School of Medicine, Dentistry, and Pharmaceutical Sciences, Okayama 7008558, Japan  
<sup>8</sup> Department of Gastroenterology, Nagasaki Harbor Medical Center, Nagasaki 8508555, Japan  
<sup>9</sup> Department of Gastroenterology, Fukui Prefectural Hospital, Fukui 9108526, Japan  
<sup>10</sup> The Center for Digestive Disease, Tonan Hospital, Sapporo 0600004, Japan  
<sup>11</sup> Department of Gastroenterology, Toyonaka Municipal Hospital, Toyonaka 5608565, Japan  
\* Correspondence: rniikura@triton.ocn.ne.jp; Tel.: +81-3-3342-6111



**Citation:** Nakada, A.; Niikura, R.; Otani, K.; Kurose, Y.; Hayashi, Y.; Kitamura, K.; Nakanishi, H.; Kawano, S.; Honda, T.; Hasatani, K.; et al. Improved Object Detection Artificial Intelligence Using the Revised RetinaNet Model for the Automatic Detection of Ulcerations, Vascular Lesions, and Tumors in Wireless Capsule Endoscopy. *Biomedicines* **2023**, *11*, 942. <https://doi.org/10.3390/biomedicines11030942>

Academic Editor: Shaker A. Mousa

Received: 3 January 2023

Revised: 10 February 2023

Accepted: 15 March 2023

Published: 17 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** The use of computer-aided detection models to diagnose lesions in images from wireless capsule endoscopy (WCE) is a topical endoscopic diagnostic solution. We revised our artificial intelligence (AI) model, RetinaNet, to better diagnose multiple types of lesions, including erosions and ulcers, vascular lesions, and tumors. RetinaNet was trained using the data of 1234 patients, consisting of images of 6476 erosions and ulcers, 1916 vascular lesions, 7127 tumors, and 14,014,149 normal tissues. The mean area under the receiver operating characteristic curve (AUC), sensitivity, and specificity for each lesion were evaluated using five-fold stratified cross-validation. Each cross-validation set consisted of between 6,647,148 and 7,267,813 images from 217 patients. The mean AUC values were 0.997 for erosions and ulcers, 0.998 for vascular lesions, and 0.998 for tumors. The mean sensitivities were 0.919, 0.878, and 0.876, respectively. The mean specificities were 0.936, 0.969, and 0.937, and the mean accuracies were 0.930, 0.962, and 0.924, respectively. We developed a new version of an AI-based diagnostic model for the multiclass identification of small bowel lesions in WCE images to help endoscopists appropriately diagnose small intestine diseases in daily clinical practice.

**Keywords:** artificial intelligence; small intestine erosions and ulcers; small intestine tumors; small intestine vascular lesions; wireless capsule endoscopy

## 1. Introduction

Wireless capsule endoscopy (WCE) is a revolutionary examination method that can evaluate the entire 6 m long small intestine [1]. The American, European, and Japanese societies for gastrointestinal endoscopy recommend WCE as the primary examination for patients with obscure gastrointestinal bleeding, small intestine tumors, and inflammatory bowel disease. Although a single WCE examination can acquire 10,000–80,000 images,

only a few abnormal images are required to diagnose small intestine lesions. The most important issue related to WCE images is low inter- and intra-observer agreement [2]. A recent meta-analysis reported 0.6–0.79 inter-observer agreement in 56% of the WCE examinations for small intestine lesions. The diagnostic yield of WCE depends on the examination time [3] and the endoscopist's skill and experience [4]. Longer examination times and lack of experience may lead to lower diagnostic yields. Furthermore, there are currently no standardized diagnostic protocols or reporting systems. Thus, new diagnostic solutions are required to improve the accuracy of WCE diagnoses.

Artificial intelligence (AI) models can be used to improve the diagnostic accuracy of diseases of the small intestine [5,6]. We previously reported on the high diagnostic accuracy of an AI model that we developed, RetinaNet, for identifying small intestine erosions and ulcers, angioectasias, and tumors [5]. Although the model has high diagnostic yield, it may at times indicate false positive/negative results from the images. Generally, improvements in the diagnostic accuracy of AI models require an increase in the data size. Therefore, we developed a new RetinaNet model using the largest dataset in the world, consisting of >10,000,000 WCE images obtained from nine hospitals.

## 2. Materials and Methods

### 2.1. Study Sample and Preparation of the Image Set

We performed a retrospective study using a WCE database. First, we collected WCE images acquired between April 2009 and July 2019 from the University of Tokyo Hospital from patients with obscure gastrointestinal bleeding, possible small intestine tumors, or abdominal symptoms. We previously used these data to develop the RetinaNet model [5]. We expanded the angioectasia WCE database by adding images acquired between 2009 and 2019 at Ishikawa Prefectural Central Hospital, Fukui Prefectural Hospital, Tonan Hospital, the University of Okayama Hospital, the University of Kanazawa Hospital, Nagasaki Medical Center, the University of Osaka Hospital, and Toyonaka Municipal Hospital for obscure gastrointestinal bleeding, examination for small intestine tumors, or abdominal symptoms (Table 1). All WCE procedures used the PillCam SB2 or SB3 capsule endoscope (Medtronic, Minneapolis, MN, USA) and were carried out after patients had fasted for 12 h. Oral simethicone (40 mg) was administered before the WCE examinations [7].

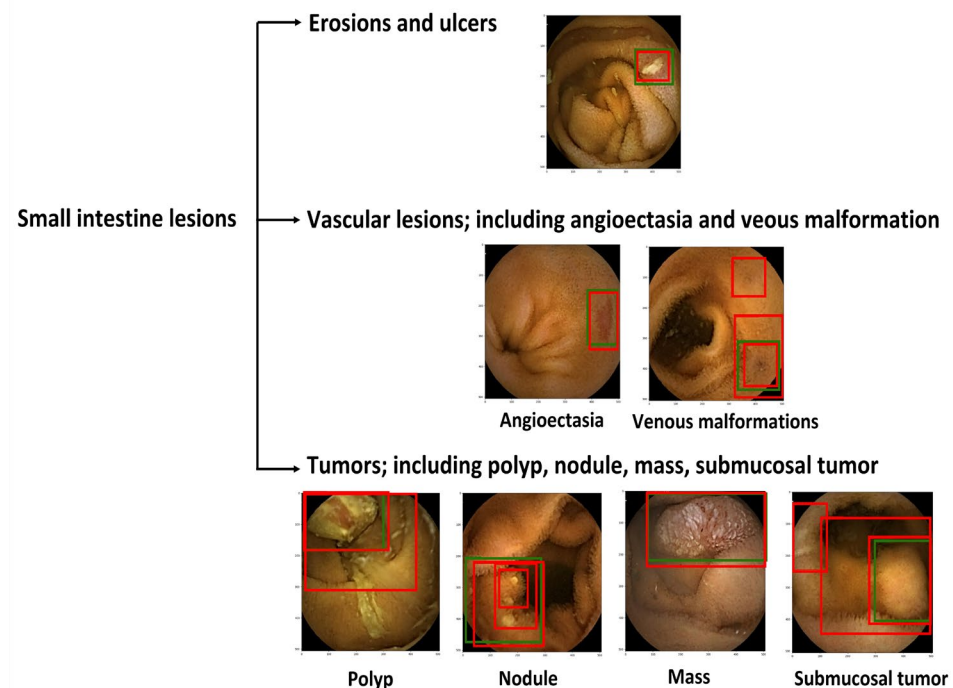
**Table 1.** Number of patients from each of the nine institutions.

Hospital	Erosions and Ulcers	Vascular Lesions	Tumors	Normal
University of Tokyo Hospital	161	19	73	314
Ishikawa Prefectural Central Hospital	51	21	20	142
Fukui Prefectural Hospital	6	0	3	0
Tonan Hospital	2	2	0	2
Osaka University Hospital	22	3	16	0
University of Kanazawa Hospital	127	32	26	6
Nagasaki Minato Hospital	11	6	4	13
University of Osaka Hospital	96	28	26	0
Toyonaka Hospital	1	0	1	0
Total	477	111	169	477

From the database, we extracted a case group of 651 patients with erosions and ulcers, angioectasias, or tumors. We also randomly extracted a control group of 482 patients and normal images from these patients.

The WCE images were used to develop a dataset, consisting of 6476 images of erosions and ulcers, 1916 of angioectasias, 7127 of tumors, and 14,014,149 normal images. This study was approved by all of the participating hospitals (No. 12016-1). A vascular lesion was defined as angioectasias and venous malformations; a tumor was defined as a polyp, nodule, mass, and/or submucosal tumor (Figure 1). Four expert WCE endoscopists (AN,

RN, TA, and AY) manually annotated all lesions with bounding boxes (gold-standard boxes). All annotations were performed independently, and any disagreement was resolved by consensus.



**Figure 1.** Classification of small intestine lesions. Green boxes, gold-standard bounding boxes; red boxes, AI-detected bounding boxes.

## 2.2. RetinaNet Algorithm

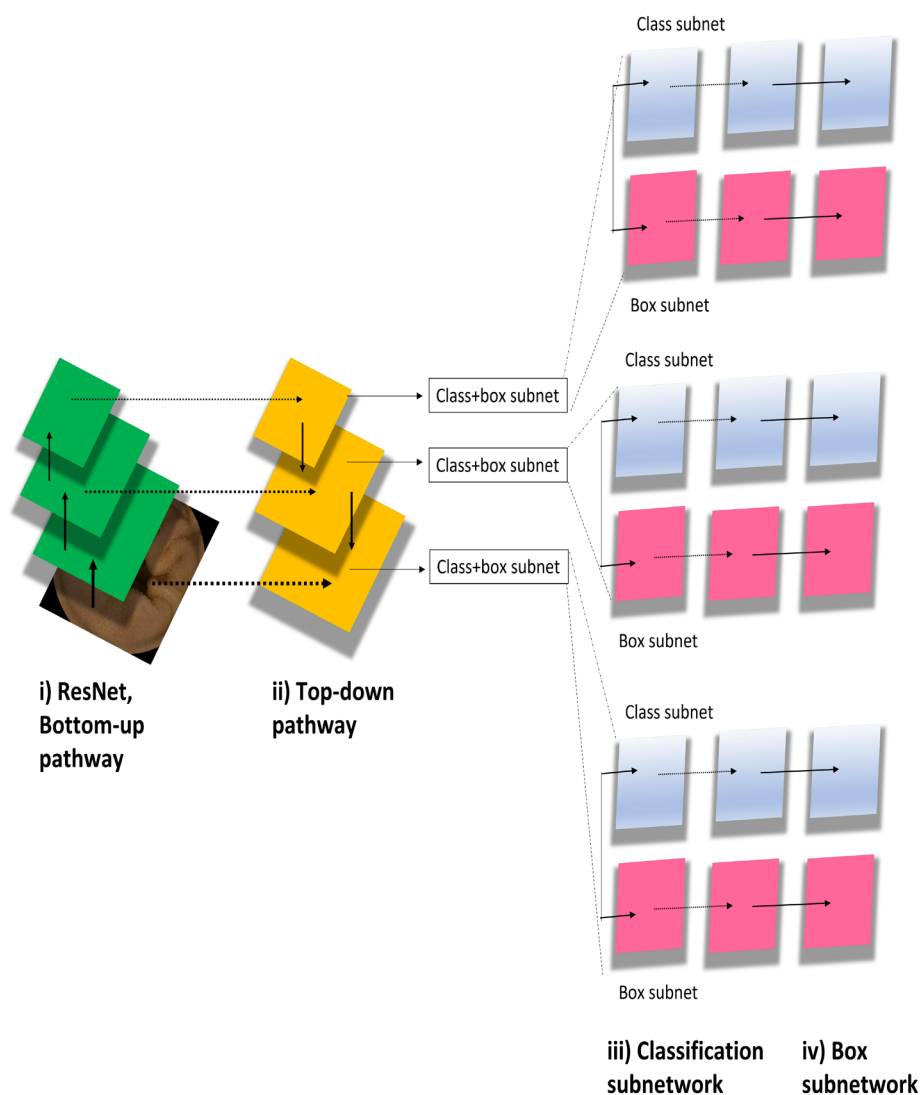
We used the deep neural network architectures of RetinaNet [8] to develop a new AI-based diagnostic model. The major RetinaNet network included ResNet, bottom-up pathway, top-down pathway, classification subnetwork, and box subnetwork (Figure 2). The RetinaNet network architecture uses a Feature Pyramid Network backbone on top of a feedforward ResNet architecture to construct a rich, multiscale convolutional feature pyramid. RetinaNet attaches two subnetworks: one for classifying anchor boxes and another for regressing from anchor boxes to ground-truth object boxes. We trained the RetinaNet model to detect areas within the bounding boxes as lesions and those outside of the boxes as background. The input image size was  $512 \times 512$ . Learning was carried out by penalizing incorrect outputs and iteratively minimizing this penalty. Notably, lesion detection differs from general object detection in that the boundaries of the detection targets are ambiguous. The penalty was relaxed to allow some positional shifting of the output boxes.

Previously, we had developed the RetinaNet model using the data of 398 erosion and ulceration images, 538 angioectasias images, 4590 tumor images, and 34,437 normal images from a single hospital [5]. In the current study, we further trained the model using 6476 erosion and ulcer images, 1916 angioectasias images, 7127 tumor images, and 14,014,149 normal images from nine hospitals.

## 2.3. Outcome Measures and Statistics

The primary outcome was a per-lesion image diagnosis of small intestine lesions including erosions and ulcers, vascular lesions, and tumors. The model accuracy was defined based on the overlap between the AI-drawn bounding boxes and the gold-standard boxes. We used five-fold stratified cross-validation to balance the lesion ratios to test the model (Figure 3). When generating the internal and external validation sets, random sampling was performed to avoid bias that could lead to false readings regarding the

model's performance. The trained RetinaNet model drew red bounding boxes (AI boxes) around lesions detected in the validation set, and output probability scores ranged from 0 to 1 for each erosion, ulceration, vascular lesion, and tumor; the higher the score, the greater the confidence that the region included a lesion of the specified type. The following definitions were used to assess model accuracy. First, any overlap between the AI box and the gold-standard box was considered positive. Second, if several AI boxes were created in a single image and even one of them detected a lesion, image classification was considered accurate.

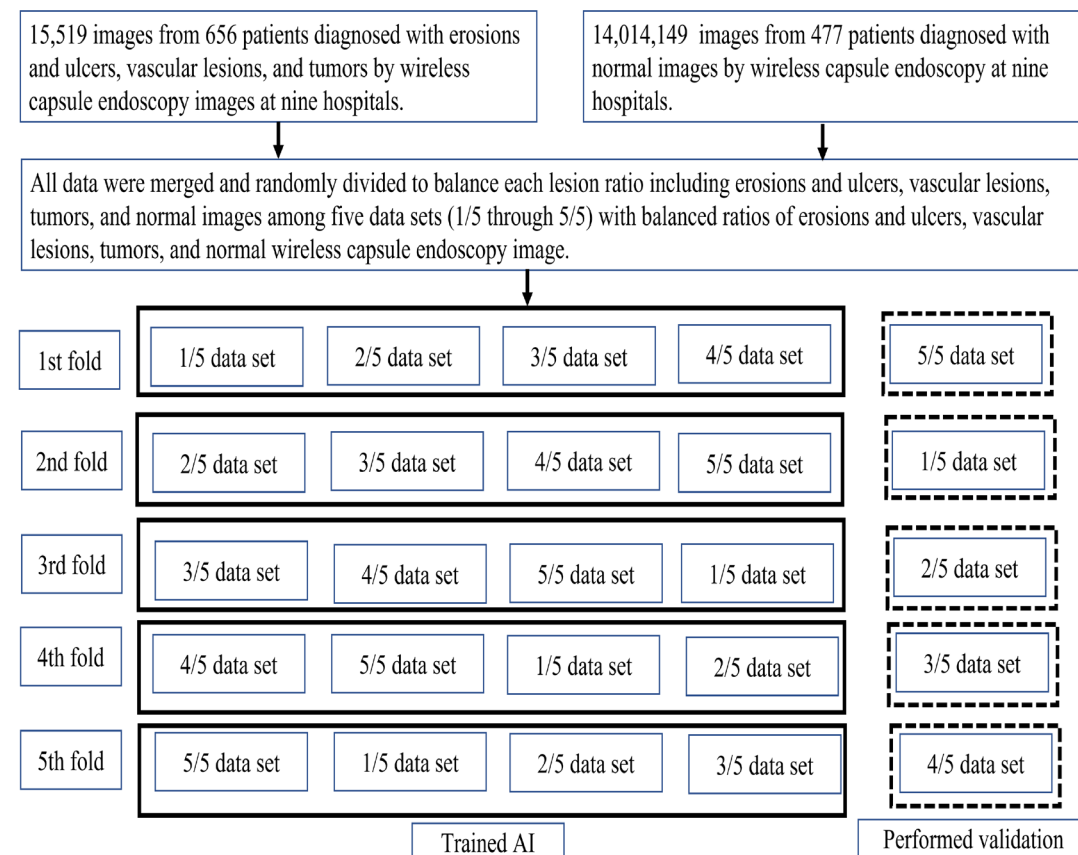


**Figure 2.** RetinaNet model algorithm.

We plotted receiver operating characteristic (ROC) curves and estimated areas under the ROC curve (AUC) and 95% confidence intervals (CIs), sensitivity, specificity, and the accuracy of the AI detection model for each lesion image for each probability score cutoff of the Youden index. The mean AUC, sensitivity, and specificity were estimated using the fold data.

The secondary outcomes were the per-intersection over union (IOU) and per-patient diagnosis of the three lesional types. During per-lesion IOU analyses, we defined the area of overlap divided by the area of union as the IOU. We calculated the IOUs for all lesions in each cross-validation set, and then estimated the mean IOU for each lesion. During per-patient analyses, we estimated the number of affected patients and the rates of AI-detected

lesions in each cross-validation set. All statistical analyses were performed with Python (ver. 3).



**Figure 3.** Study flow diagram.

### 3. Results

#### 3.1. Per-Lesion Image Analyses

The number of patients from each of the nine institutions is shown in Table 1. Each cross-validation set consisted of between 6,647,148 and 7,267,813 images from 217 patients. The lesion and normal image ratios were well balanced among the cross-validation sets. Images of small intestine erosions and ulcers, vascular lesions, and tumors diagnosed by artificial intelligence (AI) are shown in Figure 4. The mean AUC values were 0.997 for erosions and ulcers, 0.998 for vascular lesions, and 0.998 for tumors (Figure 5). The mean sensitivity values were 0.919, 0.878, and 0.876; the mean specificities were 0.936, 0.969, and 0.937; and the mean accuracies were 0.930, 0.962, and 0.924, respectively (Table 2).

#### 3.2. Per-Lesion IOU Analyses

The mean IOU of RetinaNet was 0.839 (95% CI = 0.792, 0.886) for erosions and ulcerations, 0.833 (95% CI = 0.780, 0.886) for vascular lesions, and 0.798 (95% CI = 0.750, 0.846) for tumors. The IOU values for each type of lesion in each cross-validation set are shown in Table 3.

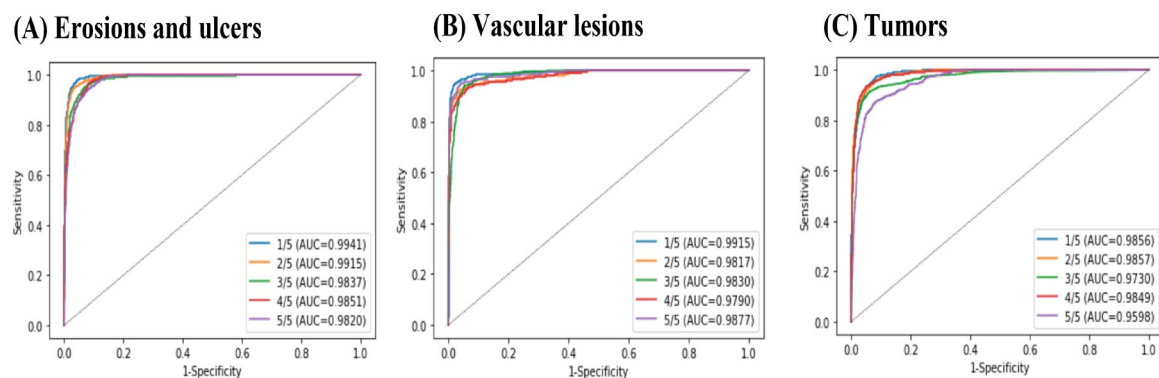
#### 3.3. Per-Patient Analyses

The per-patient diagnoses in each cross-validation fold are shown in Table 4. The AI model missed three, three, three, four, and one patient for erosions and ulcers in the first to fifth cross-validation folds, respectively; one patient for vascular lesions in the fourth fold; and one, one, and two patients for tumors in the third, fourth, and fifth folds, respectively.





**Figure 4.** Images of small intestine erosions and ulcers, vascular lesions, and tumors diagnosed by artificial intelligence (AI). (A) erosions and ulcers, (B) vascular lesions, and (C) tumors. Green boxes, gold-standard bounding boxes; red boxes, AI-detected bounding boxes.



**Figure 5.** Receiver operating characteristic (ROC) curves and area under the ROC curve (AUC) values for small intestine lesions: (A) erosions and ulcers, (B) vascular lesions, and (C) tumors.

**Table 2.** Per-image analysis of mean sensitivity, specificity, and accuracy.

	Sensitivity	Specificity	Accuracy
Erosions and ulcers	0.919 (0.896–0.942)	0.936 (0.914–0.957)	0.930 (0.912–0.953)
Vascular lesions	0.878 (0.823–0.933)	0.969 (0.958–0.979)	0.962 (0.951–0.973)
Tumors	0.876 (0.840–0.912)	0.937 (0.926–0.948)	0.924 (0.911–0.936)

Parentheses contain 95% confidence intervals.

**Table 3.** IOU values for each type of lesion in each cross-validation set.

	First Fold	Second Fold	Third Fold	Fourth Fold	Fifth Fold
Erosions and ulcers	0.8893	0.8972	0.7792	0.7959	0.8354
Vascular lesions	0.9155	0.8490	0.7913	0.7604	0.8511
Tumors	0.7991	0.8132	0.8455	0.8297	0.7061

**Table 4.** Number of diagnosed small intestine lesions according to the number of patients analyzed.

	First Fold	Second Fold	Third Fold	Fourth Fold	Fifth Fold
Erosions and ulcers	64/67	65/68	65/68	63/67	73/74
Vascular lesions	27/27	30/30	25/25	34/35	29/29
Tumors	36/36	29/29	36/37	27/28	26/28

#### 4. Discussion

We improved our AI model RetinaNet to detect all types of small-bowel lesions in WCE images. We further trained the model using a larger number of WCE images obtained from nine institutions. Currently, the model shows the highest performance for diagnostic yield for WCE examination among object-detection AI models [9].

##### 4.1. Improved Specificity and Accuracy of Tumor Detection

Our current RetinaNet model is better than the original in terms of tumor detection. For the previous RetinaNet model, the mean specificity was 0.918 (95% CI = 0.881–0.955) and the mean accuracy was 0.914 (95% CI = 0.879–0.950) [5]. For the current model, the mean specificity was 0.937 (95% CI = 0.926–0.948) and the mean accuracy was 0.924 (95% CI = 0.911–0.936). Small intestine tumors are rare. The increased number of WCE images from the nine hospitals improved the diagnostic yield.

##### 4.2. High Specificity of the RetinaNet Algorithm for All Types of Lesion

Our RetinaNet model has considerable strength with high specificity, given the extended AI training on each lesion type. It learns the features of normal images using weakly supervised learning; this allows for improved accuracy, as the total number of training images can be easily accommodated. In the current study, more than 10,000,000 normal images were used for training, as a previous meta-analysis reported an association between higher specificity and a larger number of training images in WCE AI models [9]; moreover, AI models trained using a total number of training images >20,000 show the lowest false-positive rate in small WCE examinations [9].

##### 4.3. Future Tasks

Further validation analyses are required to evaluate various AI models for WCE image assessment. Such analysis should ideally use open-source codes and directly compare AI models using the same dataset. Comparisons using publicly available common datasets or meta-analyses may also be effective.

We have also planned to use the current version of the RetinaNet model in a clinical setting. The model can diagnose WCE images, but not videos. Therefore, we have developed an original, comprehensive, user-interface network system, including video-to-image conversion, that shows abnormal images classified by the lesion type and identifies their location in the small intestine. Endoscopists can obtain RetinaNet-detected results using the web-based system anytime and anywhere. We plan to use this system in hospitals that may be willing to participate in the research.

##### 4.4. Limitations

First, this study used a retrospective design. Next, the diagnostic yield of our former RetinaNet model showed good performance, but had reached a plateau level regarding learning effects. Specifically, the improvement in diagnostic yield, in terms of sensitivity,

specificity, and accuracy for erosions, ulcers, and vascular lesions, was limited, although the number of training images increased by twenty-, four-, and two-fold, respectively. Furthermore, it missed several patients with erosions and ulcers, one patient with a vascular lesion, and one patient with a tumor. These missed features would be difficult to diagnose, even for expert endoscopists. The diagnostic yield of small intestine lesions using the current AI model does not reach that of expert endoscopists; however, in current clinical practice, we believe that it would be effective as a first screening tool for endoscopists reading WCE videos or as a means of cross-checking image findings after an endoscopist's reading of WCE videos.

## 5. Conclusions

We developed a new version of the RetinaNet model for the multiclass diagnosis of lesions in WCE images. The improved version of our model will be especially useful for endoscopists to appropriately diagnose small intestine disease in daily clinical practice.

**Author Contributions:** Conceptualization, R.N.; methodology, R.N., A.Y., Y.K. and K.O.; software, K.O. and Y.K.; formal analysis, A.N., R.N. and Y.K.; investigation, Y.H., K.K., H.N., S.K., T.H., K.H., T.S., T.N., A.Y. and T.A.; resources, Y.H., K.K., H.N., S.K., T.H. (Tetsuya Honda), K.H., T.S., T.N. (Tetsuya Honda), A.Y. and T.A.; data curation, A.N. and R.N.; writing—original draft preparation, A.N. and R.N.; writing—review and editing, A.Y., T.A. and Y.K.; supervision, T.H. (Tatsuya Harada), T.K. and M.F.; project administration, R.N.; funding acquisition, R.N. and Y.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by AMED (grant no. 21cm0106485h0001) and the Japanese Foundation for Research and Promotion of Endoscopy (2022).

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committees of Tokyo University (12016-(1), 20 September 2019) and all participating hospitals.

**Informed Consent Statement:** Informed consent was obtained in the sense that patient opt-out was possible via the website.

**Data Availability Statement:** The data presented in this study are available on reasonable request to the corresponding author.

**Conflicts of Interest:** No author has any possible conflict of interest.

## References

- Iddan, G.; Meron, G.; Glukhovsky, A.; Swain, P. Wireless capsule endoscopy. *Nature* **2000**, *405*, 417. [[CrossRef](#)] [[PubMed](#)]
- Cortegoso Valdivia, P.; Deding, U.; Bjørsum-Meyer, T.; Baatrup, G.; Fernández-Urién, I.; Dray, X.; Boal-Carvalho, P.; Ellul, P.; Toth, E.; Rondonotti, E.; et al. Inter/Intra-Observer Agreement in Video-Capsule Endoscopy: Are We Getting It All Wrong? A Systematic Review and Meta-Analysis. *Diagnostics* **2022**, *12*, 2400. [[CrossRef](#)] [[PubMed](#)]
- Beg, S.; Card, T.; Sidhu, R.; Wronska, E.; Ragunath, K. UK capsule endoscopy users' group. *Dig. Liver Dis.* **2021**, *53*, 1028–1033. [[CrossRef](#)] [[PubMed](#)]
- Zheng, Y.; Hawkins, L.; Wolff, J.; Goloubeva, O.; Goldberg, E. Detection of Lesions During Capsule Endoscopy: Physician Performance Is Disappointing. *Am. J. Gastroenterol.* **2012**, *107*, 554–560. [[CrossRef](#)] [[PubMed](#)]
- Otani, K.; Nakada, A.; Kurose, Y.; Niikura, R.; Yamada, A.; Aoki, T.; Nakanishi, H.; Doyama, H.; Hasatani, K.; Sumiyoshi, T.; et al. Automatic detection of different types of small-bowel lesions on capsule endoscopy images using a newly developed deep convolutional neural network. *Endoscopy* **2020**, *52*, 786–791. [[CrossRef](#)] [[PubMed](#)]
- Aoki, T.; Yamada, A.; Kato, Y.; Saito, H.; Tsuboi, A.; Nakada, A.; Niikura, R.; Fujishiro, M.; Oka, S.; Ishihara, S.; et al. Automatic detection of various abnormalities in capsule endoscopy videos by a deep learning-based system: A multicenter study. *Gastrointest. Endosc.* **2021**, *93*, 165–173. [[CrossRef](#)] [[PubMed](#)]
- Niikura, R.; Yamada, A.; Maki, K.; Nakamura, M.; Watabe, H.; Fujishiro, M.; Oka, S.; Esaki, M.; Fujimori, S.; Nakajima, A.; et al. Associations between drugs and small-bowel mucosal bleeding: Multicenter capsule-endoscopy study. *Dig. Endosc.* **2018**, *30*, 79–89. [[CrossRef](#)] [[PubMed](#)]



8. Lin, T.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [[CrossRef](#)] [[PubMed](#)]
9. Iwata, E.; Niikura, R.; Aoki, T.; Nakada, A.; Kawahara, T.; Kurose, Y.; Harada, T.; Kawai, T. Automatic detection of small-bowel lesions from capsule endoscopy images using a deep conventional neural network: A systematic review and meta-analysis. *Prog. Dig. Endosc.* **2022**, *100*, 27–35. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.