

博士論文

Implementation of Real-Time
Computer-Aided Diagnosis
System with Quantitative
Staging and Navigation
on Customizable Embedded
Digital Signal Processor

定量的ステージングとナビゲーションを有するリアルタイム診断支援システムのカスタマイザブルデジタルシグナルプロセッサへの実装

小田川 真之

広島大学大学院先端物質科学研究科

2021年9月

目 次

(Table of Contents)

1. 主論文 (Main Thesis)

Implementation of Real-Time Computer-Aided Diagnosis System with
Quantitative Staging and Navigation on Customizable Embedded Digital
Signal Processor

(定量的ステージングとナビゲーションを有するリアルタイム診断
支援システムのカスタマイザブルデジタルシグナルプロセッサへの
実装)

Masayuki Odagawa

2. 公表論文 (Articles)

- (1) A Hardware Implementation on Customizable Embedded DSP Core for
Colorectal Tumor Classification with Endoscopic Video toward Real-Time
Computer-Aided Diagnosis System.

Masayuki Odagawa, Takumi Okamoto, Tetsuhi Koide, Toru Tamaki, Bisser
Raytchev, Kazufumi Kaneda, Shigeto Yoshida, Hiroshi Mieno, Shinji Tanaka,
Takayuki Sugawara, Hiroshi Toishi, Masayuki Tsuji and Nobuo Tamba
IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and
Computer Sciences, Vol.E104.A, No.4, pp.691-701 (2021).

DOI: 10.1587/transfun.2020EAP1069

- (2) Feasibility Study for Computer-Aided Diagnosis System with Navigation
Function of Clear Region for Real-Time Endoscopic Video Image on
Customizable Embedded DSP Cores.

Masayuki Odagawa, Tetsushi Koide, Toru Tamaki, Shigeto Yoshida, Hiroshi
Mieno, Shinji Tanaka
IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and
Computer Sciences, Vol.E105-A, No.1, In press.

DOI: 10.1587/transfun.2021EAL2044

- (3) Classification with CNN features and SVM on Embedded DSP Core for
Colorectal Magnified NBI Endoscopic Video Image.

Masayuki Odagawa, Takumi Okamoto, Tetsushi Koide, Toru Tamaki, Shigeto

Yoshida, Hiroshi Mieno, Shinji Tanaka

IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and
Computer Sciences, Vol.E105-A, No.1, In press.

DOI: 10.1587/transfun.2021EAP1036

3. 参考論文 (Thesis Supplements)

- (1) A Hardware Implementation of Colorectal Tumor Classification for Endoscopic Video on Customizable DSP toward Real-time Computer-Aided Diagnosis System.

Masayuki Odagawa, Takumi Okamoto, Tetsuhi Koide, Toru Tamaki, Bisser Raytchev, Kazufumi Kaneda, Shigeto Yoshida, Hiroshi Mieno, Shinji Tanaka, Takayuki Sugawara, Hiroshi Toishi, Masayuki Tsuji and Nobuo Tamba
Proceedings of 2019 IEEE International Symposium on Circuits and Systems, Sapporo, Japan, pp.1-5 (2019).

DOI: 10.1109/ISCAS.2019.8702379

- (2) Classification Method with CNN features and SVM for Computer-Aided Diagnosis System in Colorectal Magnified NBI Endoscopy.

Masayuki Odagawa, Takumi Okamoto, Tetsuhi Koide, Toru Tamaki, Bisser Raytchev, Kazufumi Kaneda, Shigeto Yoshida, Hiroshi Mieno and Shinji Tanaka
Proceedings of IEEE Regional 10 Conference, Online, Japan, pp.1-6 (2020)

DOI : 10.1109/TENCON50793.2020.9293709

主 論 文
(Main Thesis)

Abstract

In this thesis, I presents a quantitative staging classification method for a real-time Computer-Aided Diagnosis (CAD) system in a colorectal magnified Narrow Band Imaging (NBI) endoscopy. For the classification of a histologic findings, a real-time video (30fps) CAD system on site is required. Since colorectal tumor classification is based on pit pattern of colorectal lesion surface and vessel, it is difficult to identify cancer staging for non-expert. Quantitative staging classification which provides quantitative staging and objective index for real-time video to the doctor is required, since the conventional CAD system is real-time video polyp detection and still image classification. In endoscopic video, unclear regions exist in lesion because of continuous moving and indefinite shape, and, affects quantitative staging and objective index. By moving staging region to clear region, quantitative staging is improved. Therefore, navigation function to clear region is indispensable for clinical doctor. However, polyp detection and classification in previous CAD systems are main subject and quantitative staging with navigation function has not been studied. In addition, a real-time video (30 fps) CAD system has not been reported. And, a real-time CAD system HW is desired to use on site, and, desired to store in the same rack of endoscopic system or smart-glass. Thus, we realize a 1) CAD system for real-time video on customizable DSP with 30 fps, 30 ms latency and 1 W, 2) real-time quantitative staging CAD for video with over 90% accuracy and 3) real-time navigation to improve quantitative staging quality with 30 fps and easy to operate for clinical doctor on site.

We implement a real-time CAD System with quantitative staging and navigation for real-time video on customizable DSP. Processing cycles and memory size is reduced for real-time processing on customizable DSP 8-bit quantized AlexNet and SVM implementation and avoiding system bus conflict by using hidden layer feature. We realize a CAD system for real-time video with 44.6 fps and 22 ms latency, 66.6 mW power. We achieve quantitative staging with 90% accuracy by pre-trained CNN instead of handcrafted feature extraction and multi-sizing and balancing in training data set. And, we implement real-time navigation for effective quantitative staging. We propose navigation by unclear region detection for one staging region with 30fps @ 525 MHz, and, navigation by multiple staging regions

with 39fps @525 MHz. From our research, quantitative and objective staging index are provided to the doctor more accurately in magnified NBI endoscopic observation, which is independent from the experience of doctors, and diagnostic support method is established such as a "second opinion" at magnified NBI endoscopic observation on site.

Contents

Chapter 1. Introduction	1
1.1. Background.....	1
1.2. Colorectal Cancer Classification	5
1.3. Requirements for Computer-Aided Diagnosis System	7
1.4. Research Objectives.....	10
1.5. Conclusion.....	14
References	16
Chapter 2. Computer-Aided Diagnosis System for Colorectal Endoscopic Image	23
2.1. Positioning of the Computer-Aided Diagnosis System in this thesis	23
2.2. Computer-Aided Diagnosis System based on Bag-of-Features.....	24
2.3. Computer-Aided Diagnosis System with Convolutional Neural	27
2.3.1. Convolutional Neural Network and Feature Extractor	27
2.3.2. Support Vector Machine as Classifier	31
2.4. Conclusion.....	34
References	36
Chapter 3. Improvement for Endoscopic Video Image by Endoscopic Image Data Set 40	40
3.1. Problems in endoscopic image data affect classification accuracy.....	40
3.2. Endoscopic image dataset from clinical doctors.....	42
3.3. Data Balancing	44
3.4. Evaluation Result	46
3.5. Evaluation of Endoscopic Video Image.....	49
3.6. Conclusion.....	55
References	56
Chapter 4. Proposed CAD System Implementation to Customizable DSP Core.....	58
4.1. Multiply and Accumulate Calculation in CNN	58

4.2. Requirements for Hardware Platform of the CAD System Implementation	59
4.3. Overview of Customizable DSP Core	62
4.4. Hardware Design and Processing Flow.....	65
4.5. Bottleneck Analysis.....	69
4.6. Optimization.....	71
4.6.1. Optimization of the pre-processing of the input image.....	72
4.6.2. Optimization of waiting time for loading coefficients of full connect layer	76
4.7. Evaluation of the Developed Prototype System.....	80
4.8. Conclusions	86
References	87
Chapter 5. Practical Improvement to the CAD System with Navigation Function	91
5.1. Issues in Video Image	91
5.2. Overview of Navigation function of clear region.....	93
5.3. Possibility of the Proposed CAD System Implementation on VP6 and similar DSP cores in the low power ASIC.....	97
5.4. Overview of Navigation Function by Multiple Staging Region	101
5.5. Conclusion.....	102
References	104
Chapter 6. Conclusion.....	108
6.1. Conclusion.....	108
6.2. Future Works	111
References.....	113
Publications.....	114
Presentations on Conferences	115
<i>International conference</i>	115
<i>Domestic conference</i>	117

Acknowledgement	119
------------------------------	-----

List of Figures

Figure 1. Trends in Incidence and Morality for Cancers in Japan, from 1975 to 2018 [1]	1
Figure 2. Trends in Incidence and Morality for Cancers in United States, from 1975 to 2018 [3]	2
Figure 3. 5-year Relative Survival Rate from 2006-2008 in Japan [1]	2
Figure 4. Colon and Rectum which Colorectal Cancer Occurs. (American Cancer Society) [7]	3
Figure 5. Wavelengths filtering for NBI endoscopic system (Olympus) [8].	4
Figure 6. Overview of Narrow Band Imaging Technique (Olympus) [8].	4
Figure 7. Endoscopic Image by (a) normal light and (b) NBI (Olympus) [8].	4
Figure 8. Hiroshima Classification [13-14].	6
Figure 9. The Japan NBI Expert Team (JNET) classification [21-22] for NBI Magnification Findings.	7
Figure 10. Examples of Type 1, Type 2A and Type 3 in JNET Classification.	7
Figure 11. Overview of Computer Aided Diagnosis System with NBI Magnified Endoscopic Camera.	9
Figure 12 Purpose of Computer-aided diagnosis system.	11
Figure 13 Difference of Detection, Classification and Quantitative Staging.	11
Figure 14 Modality of Endoscopic System by Light and Magnification.	12
Figure 15 Quantitative Staging is Different by Position	12
Figure 16 Positioning of CAD System (Algorithm)	13
Figure 17 Comparison of (a) performance vs diagnosis quality and (b) performance vs power.	13
Figure 18. Positioning of the Computer-Aided Diagnosis System.	24
Figure 19. Overview of Computer-Aided Diagnosis system for endoscopic images based BoF	25
Figure 20. SVM output for the images of non-neoplastic and neoplastic lesions [30].	27
Figure 21. Overview of BoF based and CNN-SVM CAD systems.	28
Figure 22. Architecture of AlexNet [60].	29
Figure 23. Extracted Features for Type 1 image input: (c) Visual Word from D-SIFT, (d) Feature Vector from AlexNet.	29
Figure 24. Extracted Features for Type 2A image input: (c) Visual Word from D-SIFT, (d) Feature Vector from AlexNet.	30
Figure 25. Extracted Features for Type 3 image input: (c) Visual Word from D-SIFT, (d) Feature Vector from AlexNet.	30
Figure 26 Comparison of Pre-trained AlexNet vs D-SIFT (Type1 vs Type not 1)	31
Figure 27 Comparison of Pre-trained AlexNet vs D-SIFT (Type 2A vs Type 3)	31
Figure 28. Optimal Classification Hyperplane by SVM.	32
Figure 29. SVM classifier Type 1 and Type not 1	34
Figure 30. SVM classifier Type 1 and Type not 1	34

Figure 31. Differences of endoscopic video image (a) and still image (b).	40
Figure 32. Unclear frames in endoscopic video image: (a) blurring, (b) color shift, (c) reflection of light.	41
Figure 33. Differences from Endoscopic Systems. (a) System for training data set (system: Olympus CV-260, scope: Olympus CF-H260AZL/I [72]). (b) System for Test Video image (system: Olympus CV-290, scope: Olympus CF-HQ290ZL/I [73]).	42
Figure 34. Example of image patches collected by a clinical doctor (Hiroshima University Hospital)	42
Figure 35. distribution of image size of each lesion type.	43
Figure 36. Result of 10-fold CV for Type 1 vs Type not 1. (Non-balanced training data set)	48
Figure 37. Result of 10-fold CV for Type 2 vs Type 3. (Non-balanced training data set)	48
Figure 38. Result of 10-fold CV for Type 1 vs Type not 1. (Balanced training data set)	49
Figure 39. Result of 10-fold CV for Type 2 vs Type 3. (Balanced training data set)	49
Figure 40. Example of frame images in Type 2A endoscopic video sequence.	50
Figure 41. Comparison result for video image including the lesion Type 2A for 227 x 227 pixel region.	51
Figure 42. Example of frame images in Type 2A endoscopic video	52
Figure 43. Pixel differences during Freeze Time.	52
Figure 44. Comparison result for video image including the lesion Type 2A for 454 x 454 pixel region.	53
Figure 45. Example of frame images in Type 2A endoscopic video (454 x 454 pixel)	54
Figure 46. Classification results of Mean and Std. values of SVM output for 227 x 227 and 454 x 454 pixel regions.	54
Figure 47. Throughput and power efficiency for CNN on several device platforms [96].	60
Figure 48. Example of Very Long Instruction Word (VLIW) [102].	61
Figure 49. Overview of Single Instruction, Multiple Data (SIMD) [103].	61
Figure 50. Scatter-Gather operation and Image Data.	62
Figure 51. Overview of Vision P6 Core Architecture [107].	63
Figure 52. SVM Classification Accuracy (True-Positive) with feature extraction for 64-bit and 8-bit AlexNets which are pre-trained by ImageNet database.	65
Figure 53. The Developed Prototyping CAD System Overview.	66
Figure 54. Overview of Protium S1 Prototyping System Board [118].	66
Figure 55. Hardware and Software co-debugging feature in Protium S1 [118].	67

Figure 56. Block Diagram of the Implemented System.	68
Figure 57. Flowchart of the Developed CAD Prototype System.	69
Figure 58. Cycle Analysis View of Xtensa Xplorer [121].	70
Figure 59. Pipeline Analysis View of Xtensa Xplorer [121].	70
Figure 60. Profiling results of processing cycles per one still image.	71
Figure 61. Comparison results of SVM classification processing cycles before and after optimization [57].	72
Figure 62. Optimization Flow Using Auto Vectorization of Xtensa Xplorer.	73
Figure 63. Code of Convolutional Layer as Nested Loops	73
Figure 64. Vectorization Assistant View of Xtensa Xplorer [121]	75
Figure 65. Example of re-coding for Xtensa auto-vectorization.	75
Figure 66. Result of Optimization.	76
Figure 67. Feature dimensions of pre-trained AlexNet, and, architecture of AlexNet.	77
Figure 68. True Positive and Precision Rate for each hidden layer of AlexNet	79
Figure 69. F-measure for each hidden layer of AlexNet.	80
Figure 70. Performance, Area, Power Estimation View in Xtensa Xplorer [121].	80
Figure 71. The Developed CAD System.	82
Figure 72. Staging Region Setting.	82
Figure 73. The classification result of the endoscopic video image on the real-time prototyping system (Type 2A).	83
Figure 74. Real-time classification results from the boundary to the center of the lesion of Type 2A.	84
Figure 75. Real-time classification results from the boundary (blur region) to the center of the lesion of Type 3.	84
Figure 76. The classification result of image with 454 x 454 pixel window.	85
Figure 77. Real-time classification results of image with 454 x 454 pixel window.	85
Figure 78. Example annotated training detection boxes illustrating the 7 different artifact classes in the EAD2019 challenge dataset [128].	92
Figure 79. Overview of proposed CAD system with navigation function of clear region.	95
Figure 80. Flowchart of the proposed CAD system with navigation function of clear region.	96
Figure 81. Block diagram of Mediatek's MT8385 with two Tensilica VP6 cores (red rectangle)	98
Figure 82. Evaluation board of Mediatek's MT8385.	98
Figure 83. Timeline of the proposed CAD system with navigation function of clear region.	99
Figure 84 Navigation function by multiple staging region.	102
Figure 85 Staging region color by probability of each pathological type.	102
Figure 86 Timeline of navigation function by multiple staging region.	102

Figure 87 Comparison of (a) performance vs diagnosis quality and (b) performance vs power. 110

List of Tables

Table 1 Comparison of Implementation Methods: GPU, FPGA and DSP.	10
Table 2 Previous Study using Machine Learning	13
Table 3. Correspondence of JNET classification to previous classifications. [41-42]	26
Table 4. Training data set trimmed by a clinical doctor.	44
Table 5. Training data set after balancing.	46
Table 6. Convolutional layer configuration in AlexNet [60].	59
Table 7. The comparison of CNN models based on computational parameters and MACs [92]	59
Table 8. Comparison of Vision P5 and P6 for 3-D Convolution kernel [88]	63
Table 9. Utilization Comparison to GPUs [111]	64
Table 10. AlexNet Performance Comparison of VP6 and XM6 [96,115].	64
Table 11. Memory size and Accuracy of the 8-bit Optimized AlexNet with ImageNet.	65
Table 12. Output Message from Auto Vectorization.	74
Table 13. Performance for AlexNet and power comparison to other platforms [122-123]	81
Table 14 Execution time of CNN on single VP6 core.	96
Table 15 List of ASICs including VP6 and similar DSPs for CNN execution.	97
Table 16 Execution time for CNNs on ASICs providing DSPs [147]. (MT8385 is almost same as MT6771).	100

Chapter 1. Introduction

1.1. Background

In recent years, the incidence and the mortality of colorectal cancer is increasing in Japan as shown in Fig.1 [1]. Colorectal cancer is the leading causes of cancer-related patients and the 2nd leading causes of cancer death in Japan. The increasing trend of cancer patients and deaths is also seen in Europe [2]. In United States, the incidence and the mortality of colorectal cancer is decreasing as shown in Fig.2 [3]. However, colorectal cancer is the 3rd leading cause of cancer-related deaths in men and in women, and the 2nd most common cause of cancer deaths when men and women are combined in U.S. The death rate (the number of deaths per 100,000 people per year) from colorectal cancer has been dropping in U.S. There are some reasons for this. One reason is that colorectal polyps are now being found more often by screening and removed before they can develop into cancers by using endoscopy, or cancers are being found earlier when they are easier to treat. In addition, treatment for colorectal cancer has improved over the last few decades. Figure 3 shows 5-year relative survival rate for each cancer from 2006 to 2008. The 5-year relative survival rate of colon and rectum cancer is higher than other cancers. Therefore, if the colorectal cancer is detected and treated at the early stage, it can be recovered almost completely [5-6].

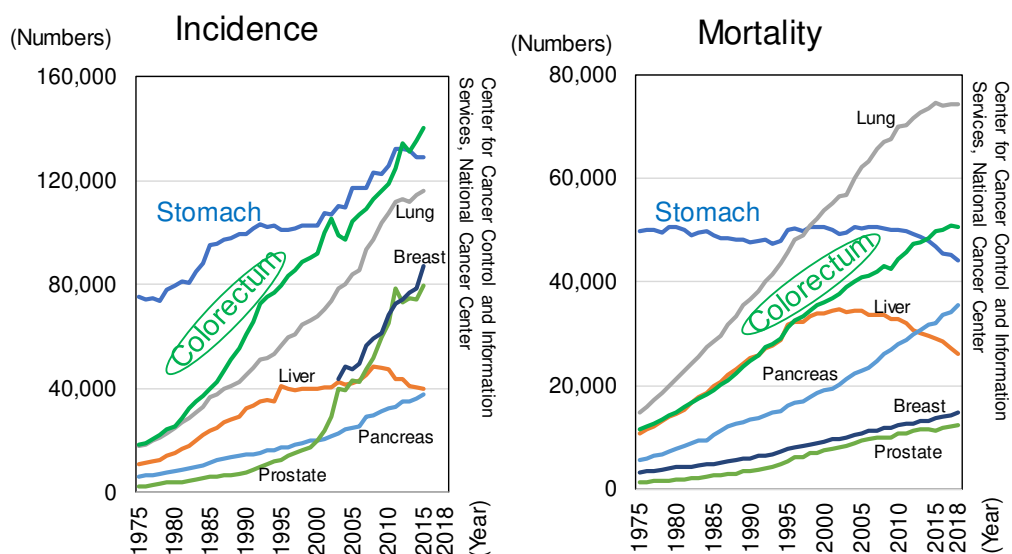
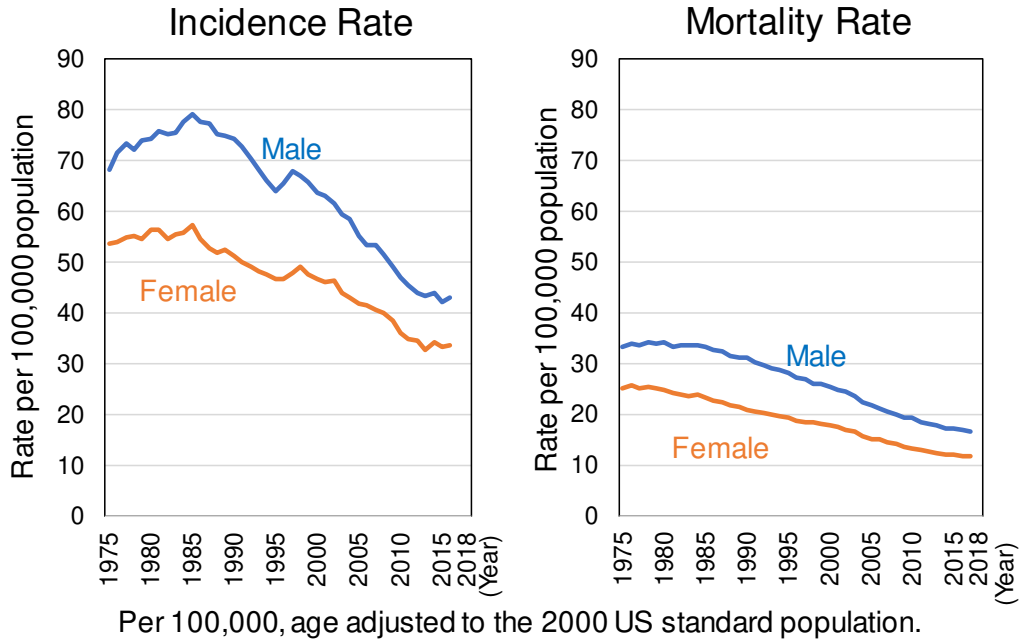


Figure 1. Trends in Incidence and Morality for Cancers in Japan, from 1975 to 2018 [1]



Data Source:

National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention, 2019

Figure 2. Trends in Incidence and Morality for Cancers in United States, from 1975 to 2018 [3]

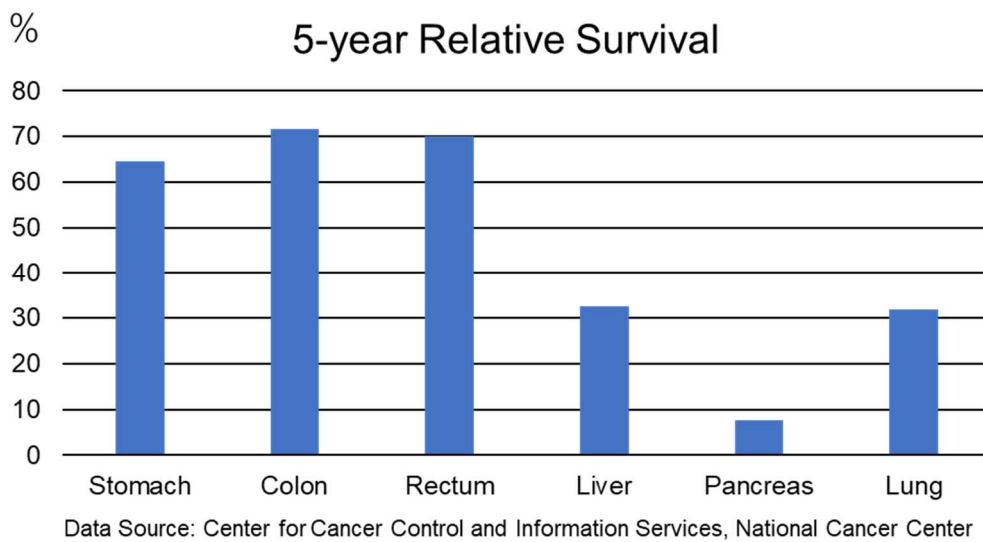


Figure 3. 5-year Relative Survival Rate from 2006-2008 in Japan [1]

Colorectal cancer is a cancer in the colon or rectum. The colon and rectum make up the large intestine (or large bowel), which is part of the digestive system. Most colorectal cancers start as a growth on the inner lining of the

colon or rectum. These growths are called polyps. If cancer forms in a polyp, it can grow into the wall of the colon or rectum over time. The wall of the colon and rectum is made up of many layers. Colorectal cancer starts in the innermost layer (the mucosa) in Fig.4 [7].

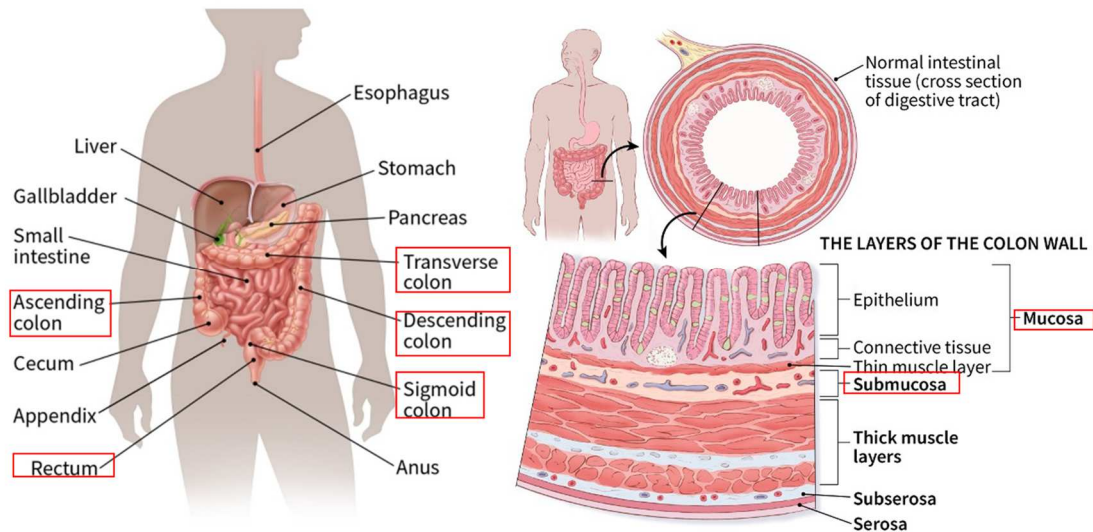


Figure 4. Colon and Rectum which Colorectal Cancer Occurs. (American Cancer Society) [7]

The observation of endoscopic images with Narrow Band Imaging (NBI) magnification and the pathology are major method of detection for colorectal cancer at early stage. NBI is an optical technique for endoscopic diagnosis, where light of specific blue and green wavelengths is used to visualize the detail of certain aspects of the surface of the mucosa. Conventionally, examining for lesions with an endoscope is performed using white light made up of the three primary colors: red, green, and blue (RGB). In contrast, NBI excludes red and uses blue and green only. A special filter is electronically activated by a switch in the endoscope to use of ambient light of wavelengths of 415 nm (blue) and 540 nm (green) as shown in Fig.5. When the blue and the green lights hit the tissue, the light is absorbed by the hemoglobin contained in the blood vessels as shown in Fig.6. The mucosal areas reflect the light back to the endoscope. As a result, the contrast of the blood vessels in relation to the mucosa is significantly enhanced in Fig.7 (b). The blue light only penetrates the superficial layers of the mucosa and is absorbed by the vessels on the surface. The other hands, the green light is absorbed by blood vessels within deeper mucosal layers (the submucosa). This is particularly

helpful for displaying highly vascularized suspect lesions in deeper tissue layers [8].

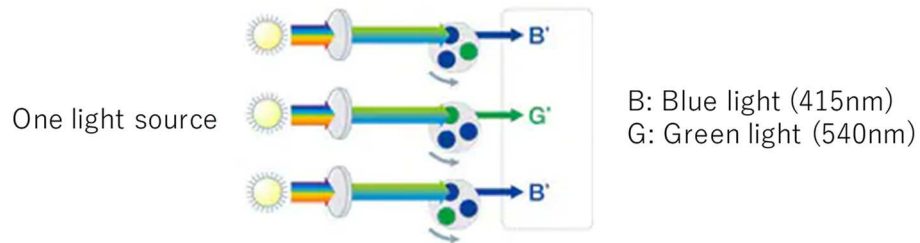


Figure 5. Wavelengths filtering for NBI endoscopic system (Olympus) [8].

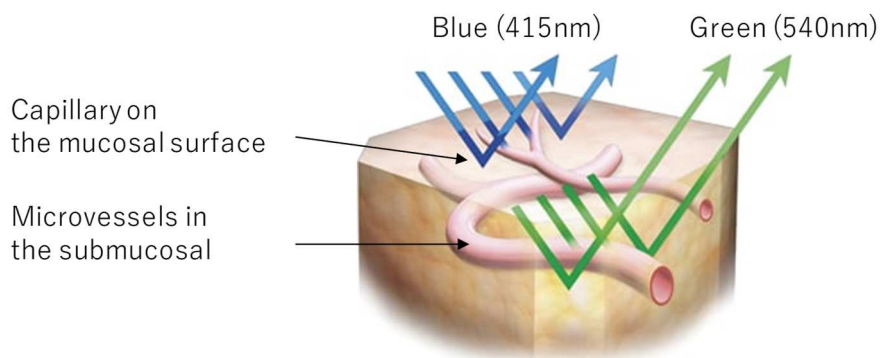


Figure 6. Overview of Narrow Band Imaging Technique (Olympus) [8].

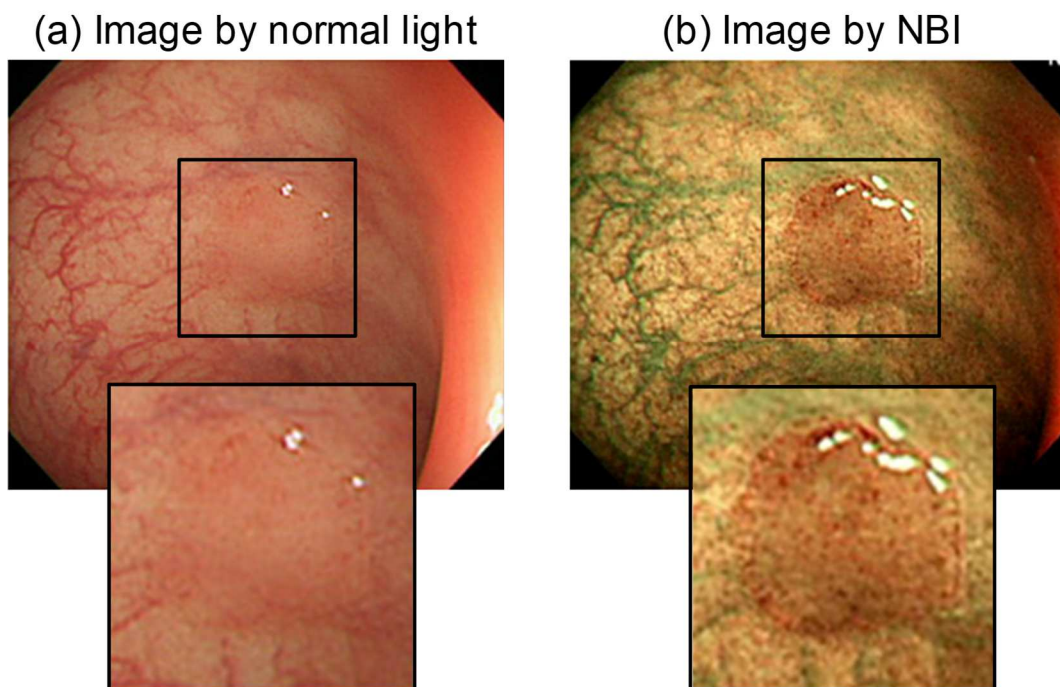


Figure 7. Endoscopic Image by (a) normal light and (b) NBI (Olympus) [8].

NBI endoscope has several advantages over chromoendoscopy. Endoscopic images with a uniform mucosal pattern across images without dye-spraying can be observed. Furthermore, NBI enables visualization of vascular features on the mucosal surfaces of lesions. This would be difficult to observe by chromoendoscopy [9-10]. In the diagnosis by the NBI magnification endoscope, an expertise clinical doctor diagnoses a tumor and the degree of cancer progression from the vessel pattern structure of the inner wall of the colon etc. The clinical doctor for diagnosis is required highly expertise and experience, and the number of expertise clinical doctors is limited. For colon neoplastic lesions, a cost-effective examination method and treatment method that takes treatment risk and post-treatment surveillance interval into consideration is desired. Therefore, a computer-aided diagnosis (CAD) system is required for improving the accuracy of diagnosis by objective judgment using computer image analysis.

1.2. Colorectal Cancer Classification

Since the observation by NBI endoscopic is general, some classification methods for colorectal cancer progress based on endoscopic images are proposed. Sano *et al.* reported first classification using NBI magnifying endoscopy in 2006 [11].

The Sano classification proposed in 2006 was the first published NBI magnifying endoscopic classification. It is based on the vascular findings alone. Its usefulness in qualitative and quantitative diagnosis have been reported by validation studies [11-14].

The Hiroshima classification published in 2008 is based on the surface pattern in addition to the vascular findings on NBI magnifying endoscopy to determine histological type of polyps [15-16]. Figure 8 shows Hiroshima classification. This classification consists of five categories: Type A, Type B, Type C1, Type C2 and TypeC3. Type A is classified as non-cancer, others are classified as cancer.

The Showa classification proposed in 2009 is based on the pit-like structure and microvascular architecture and the morphological expression of these characteristics on NBI magnifying endoscopy [17].

The Jikei proposed in 2009 is also based on mainly vascular findings on NBI magnifying endoscopy [18].

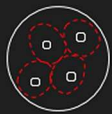
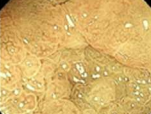







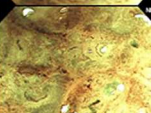
Non Cancer	A type			Microvessel intensity are vague or invisible. None or isolated lacy vessels may be present coursing across the lesion. Brown or black dots, star or round shaped surrounded by white.	
	B type			Regular surface pattern is observed by the increased microvessel intensity around the pits and image enhance. Or regular meshed microvessel network pattern is observed.	
Cancer	C type	1			Irregular surface pattern is observed by the increased microvessel intensity around the pits and image enhance. Thickness and distribution of vessels are homogenous.
		2			More irregular surface pattern is observed by the increased microvessel intensity around the pits and image enhance. Thickness and distribution of vessels are heterogenous.
		3			Surface pattern is completely unclear. Thickness and distribution of vessels are heterogenous. Avascular area (AVA) and scattered microvessel fragments are observed.

Figure 8. Hiroshima Classification [13-14].

The NICE (NBI International Colorectal Endoscopic) classification proposed in 2009 is based on the color, vessels, and surface pattern on endoscopy. Validation studies of NICE classification have proved useful in qualitative and quantitative diagnosis [19-20].

At present, the Japan NBI Expert Team (JNET) classification that unifies these previous classifications has been proposed through consensus [21-22]. Figure 9 shows the JNET classification. This classification consists of four categories: Type 1, Type 2A, Type 2B, and Type 3. Normal, hyperplastic lesions and sessile serrated polyp are classified into Type 1 as non-cancer. The others are classified into Type 2A, Type 2B or Type 3 as cancer. Type 2A, Type 2B, and Type 3 represent progress of colorectal cancer. Figure 10 shows example images of each pathological type, Type 1, Type 2A and Type 3. These are categorized by vessel pattern and surface pattern on NBI magnifying findings. Also, validation studies of the JNET classification have been reported and proved useful in diagnosis on NBI magnifying endoscopy [23-24]. In this thesis, the JNET classification is used.

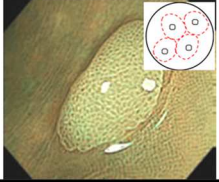
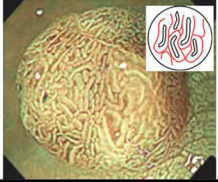
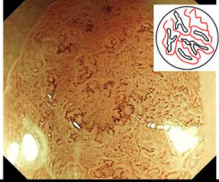
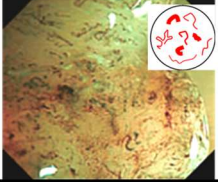
Quantitative Staging Classification				
Non Cancer		Cancer (Type not 1)		
	Type 1	Type 2A	Type 2B	Type 3
Vessel pattern	•invisible	•Regular caliber •Regular distribution (meshed/spiral)	•Variable caliber •Irregular distribution	•Loose vessel areas •Interruption of thick vessels
Surface pattern	•Regular dark or white spots •Similar to surrounding normal mucosa	•Regular (tubular/branched/papillary)	•Irregular or obscure	•Amorphous areas
Endoscopic image				

Figure 9. The Japan NBI Expert Team (JNET) classification [21-22] for NBI Magnification Findings.

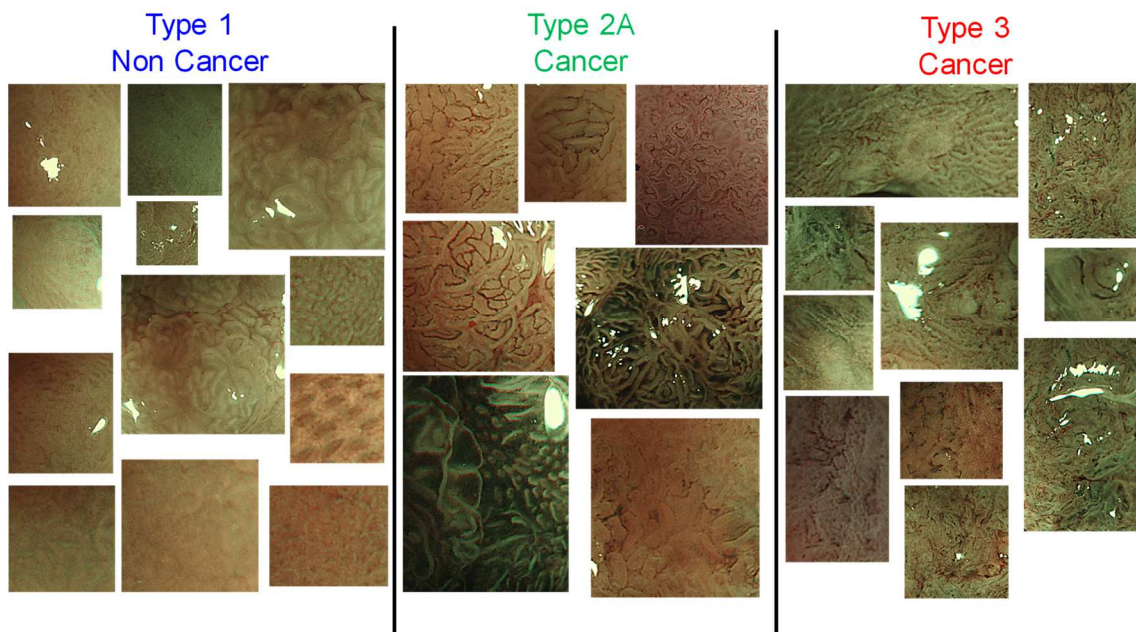


Figure 10. Examples of Type 1, Type 2A and Type 3 in JNET Classification.

1.3. Requirements for Computer-Aided Diagnosis System

System specifications required for the CAD system from clinical doctors are as follows. This requirement is the worst that doctors can tolerate, and for real-time video endoscopic observation on-site, higher requirements is needed.

(I) High speed and responsiveness

The required throughput is more than 1-5 fps and the required latency is less than 1 sec in worst case.

(II) High recognition rates

The required accuracy is more than 90% matches with results by the clinical doctors.

Generally clinical doctors use a colorectal endoscopic camera as shown in Fig.11. The colorectal endoscopic camera is stored in a rack, and the CAD system will be installed in the same rack, or head-mount for smart-glass in future. Therefore, a low power consumption is required for the CAD system at the clinical site. Since a head-mount for smart-glass is wearable device, 1 Watt of power consumption is required. The inner wall of the colon is moving constantly, and the video image frame rates from the colorectal endoscope are 30 - 60 fps (HD-SDI (High Definition – Serial Digital Interface) Standard [25]). And, the endoscopic observation is performed on two screens, one is for the image of endoscope directly (Endoscopic Monitor in Fig.11), and another is for diagnosis support (CAD Monitor in Fig.11). It becomes a problem if the frames displayed on each screen are significantly different. Thus, the higher performance than (I) is required for more accurate and seamless diagnosis. For processing real time endoscopic video (30fps), 30fps of throughput, 33 ms of latency is needed. By providing the probability of each pathological types, it is possible to perform careful observation and take multiple still images with different probabilities. Since these data can be used at later conferences with diagnosis doctors, 90% of the classification accuracy (Requirement (II)) can be obtained close judgement by expertise clinical doctors' classification. Therefore, requirements for the real-time CAD system that we aim to realize are as follows:

(I) High speed and responsiveness for real-time endoscopic video

30fps of throughput, 33 ms of latency and 1 Watt of power consumption.

(II) High recognition rates for real-time endoscopic video

Over 90% accuracy for each frame in real-time endoscopic video

A CAD system using an endocytoscopy [26] with very high magnification (360x) has been reported [27]. The reported system can classify cancer or non-cancer based on vessel pattern. Our research group focus on NBI images magnified from medium-magnification (1.3-1.9x) to strong-magnification (3.6x) to classify the degree of cancer progression based on both vessel pattern and surface pit pattern. And, a CAD system using NBI endoscopic images has been proposed [28]. The system can classify colorectal endoscopic images into three pathological types in Hiroshima classification (Type A, Type B and Type C3). The system is consisted with Dense Scale-Invariant Feature Transform (D-SIFT) feature extraction, feature transformation based on Bag of Features (BoF) and Support Vector Machine (SVM) classification. The system can realize over 90 % classification rate for 120 x 120 pixels size of still endoscopic images by software processing [28-31]. Also, real time CAD systems which are implemented into FPGA have been proposed for processing of the above software systems [32-35].

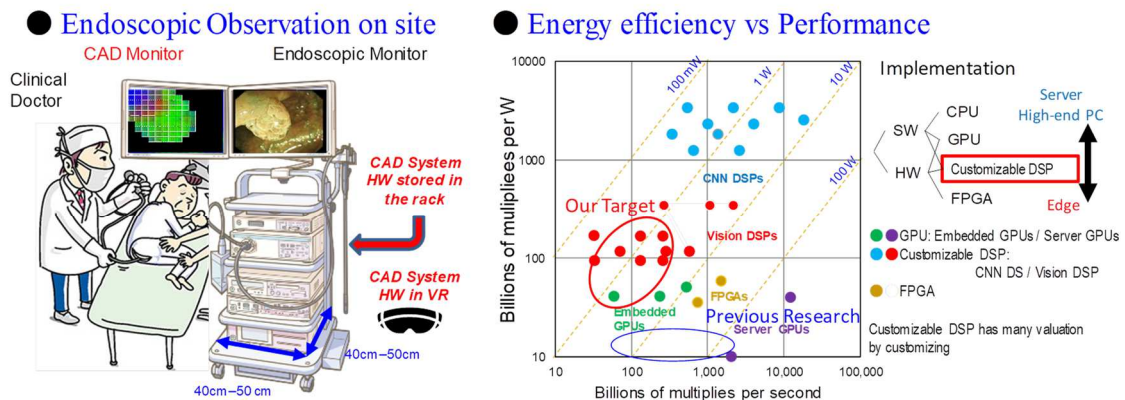


Figure 11. Overview of Computer Aided Diagnosis System with NBI Magnified Endoscopic Camera.

In general, there are customizable Digital Signal Processor (DSP), Graphics Processing Unit (GPU) and Field Programmable Gate Array (FPGA) as options for hardware (HW) implementation. Table 1 shows comparison of each implementation method. The performance is increased by applying several optimizations for each method, however the power consumption is not reduced so much. It also takes time for designing HWs in FPGA. Therefore, we selected a DSP for our implementation target.

In previous CAD systems, still NBI endoscopic images are used. However, the clinical doctor observes the inner wall of the colon by switching magnification and non-magnification on-the-fly for endoscopic video images. Therefore, a CAD system that supports colorectal tumor classification for endoscopic video image with stable objective information is much more required. On the other hand, several frames of the endoscopic video are not clear like a still image captured by freezing, because of color shift, blurring, reflection of light, and so on. A robust CAD system for endoscopic video images that is affected by these unclear images is required.

Table 1 Comparison of Implementation Methods: GPU, FPGA and DSP.

	Performance	Power Consumption	Development Period
GPU	○	△	○
FPGA	○	○	×
Customizable DSP	○	○	○

1.4. Research Objectives

CAD systems has been researched and reported CAD systems with detection and classification of polyps for the purpose of preventing oversight by screening as shown in Fig.12 [36-42]. Real-time video polyp detection and still image classification are main function in conventional CAD system. Our proposed CAD system provides quantitative staging and objective index for real-time video. Figure 13 shows difference of real-time polyp detection, classification and quantitative staging. Real-time polyp detection provides position of polyp by bounding box and label (Fig.13 (a)) [42]. Polyp classification classify polyp type, adenoma or hyper plastic with confidence level of AI (Fig.13 (b)) [42]. On the other hand, we aim to realize a CAD system with a quantitative staging classification that provides quantitative and objective index such as classification quantitative measure and staging quantitative measure to recognize the degree of progression of colorectal (Fig.13 (c)).

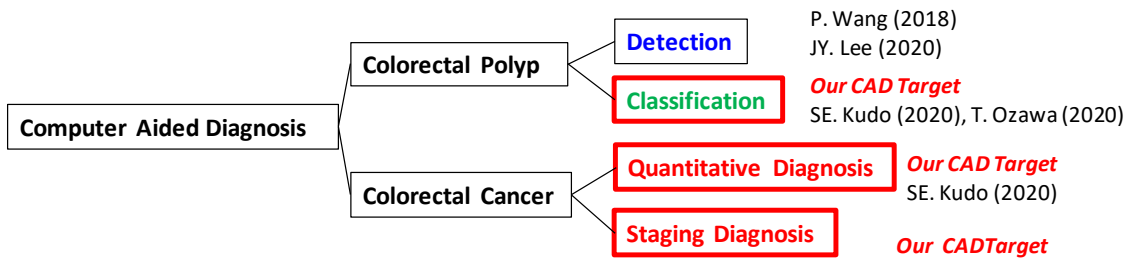


Figure 12 Purpose of Computer-aided diagnosis system.

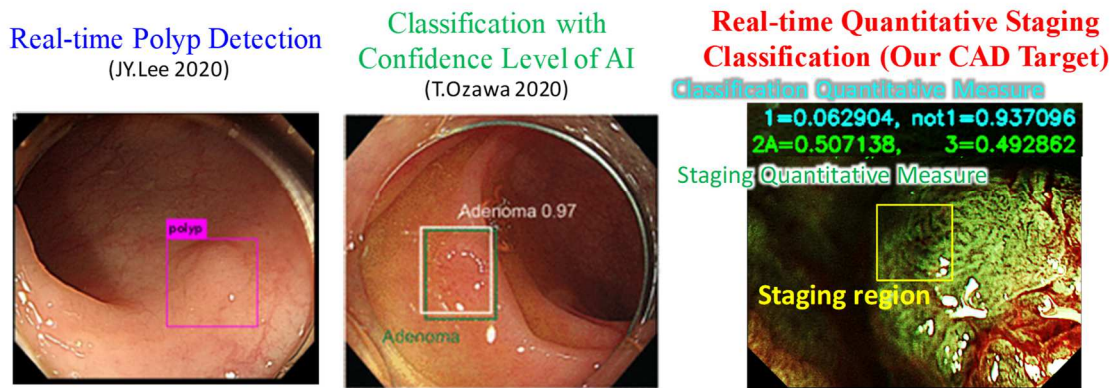


Figure 13 Difference of Detection, Classification and Quantitative Staging.

Figure 14 shows modality of endoscopic system by light and magnification. Also, research on CAD systems is being conducted for each endoscopic imaging system. The appearance of the inner wall of the colon differs depending on the magnification, and at high magnification as shown in Fig.14 (b), the pattern of the surface structure of the inner wall can be clearly observed. By highlighting with narrow band light (NBI system), it is possible to observe the structure of vessels in the inner wall of the colon. We aim to realize a real-time CAD system for magnified narrowband imaging system with 1.9x – 3.6x magnification ratio.

In magnified NBI video, unclear regions exist in lesion because of continuous moving and indefinite shape. When non-lesional areas or unclear regions such as blur, color shift, and reflection in a staging region, it is difficult to provide exact quantitative staging. Figure 15 shows quantitative staging index to the doctor is improved by moving staging region. Thus, navigation function to clear region is indispensable and effective for providing exact quantitative staging index to the doctor.

Magnification ratio in magnifying endoscopes (Type 2A) **JNET classification / Our Target**

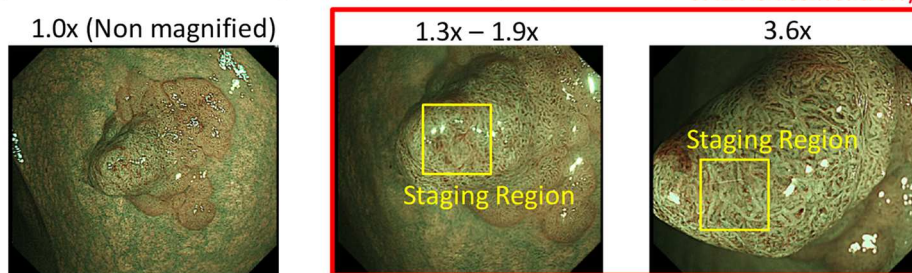


Figure 14 Modality of Endoscopic System by Light and Magnification.

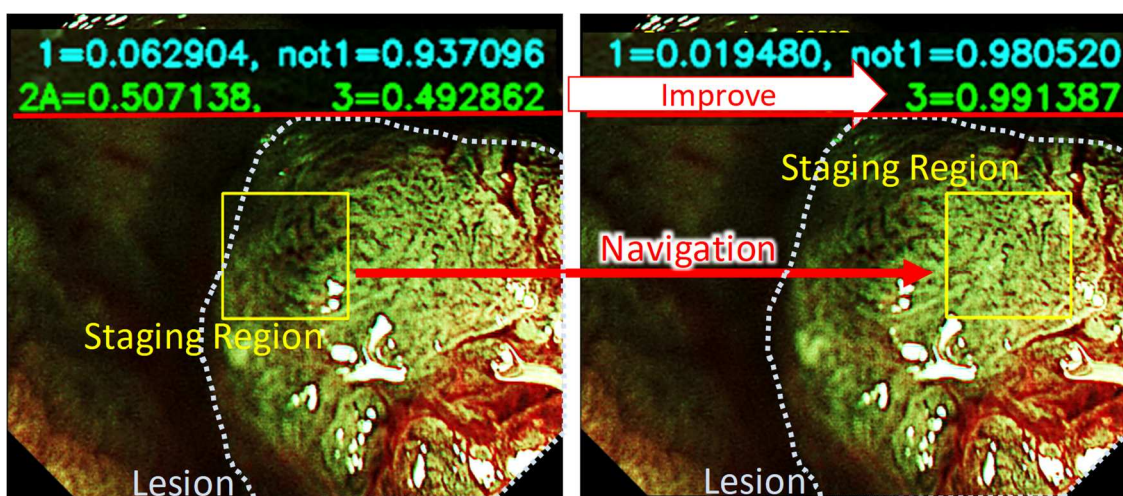


Figure 15 Quantitative Staging is Different by Position

Machine learning has been applied to build CAD systems, with shallow learning and deep learning approaches as show in Fig.16 and Table 2. Detection and classification in previous CAD systems are main subject, and, quantitative staging with navigation function has not been studied. Also, real-time video (30 fps) CAD system has not been reported. Figure 17 shows comparison of performance, power and diagnosis quality of previous studies.

While polyp detection and classification in CAD system is main subject, we aim to establish a diagnostic support method that provides quantitative and objective index of cancer stage to doctors based on the standard JNET (Japan NBI Expert Team) classification, not only classify cancer or not-cancer for real time endoscopic video (30fps).

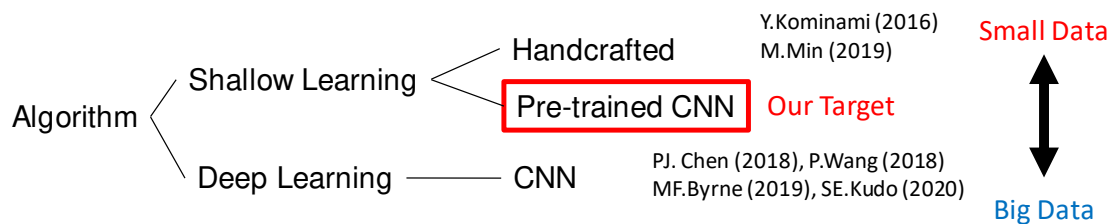


Figure 16 Positioning of CAD System (Algorithm)

Table 2 Previous Study using Machine Learning

	Machine Learning	Diagnosis Type	Endoscope	Training Data	Processing Time (fps)	Accuracy (%)	Sensitivity (%)	Platform
Y.Kominami (2016)	D-SIFT-SVM	Staging	Magnified NBI	2247 images	20	94.9	95.9	CPU
P.J.Chen (2018)	CNN	Classification	Magnified NBI	2157 images	2	90.1	96.3	
P.Wang (2018)	CNN	Detection	White Light	5545 images	25	-	94.4	Server
M.F.Byrne (2019)	CNN	Classification	NBI	223 videos	20	94.0	98	GPU
M.Min (2019)	GMM	Classification	BLI	208 images	-	78.4	83.3	
SE.Kudo (2020)	CNN	Classification	Endocyto-NBI	25945 images	2.5	96.0	96.9	CPU
Our Target	Pre-trained CNN + SVM	Quantitative Staging + Navigation	Magnified NBI	10209 images	30	> 90 Video	> 90	Customizable DSP

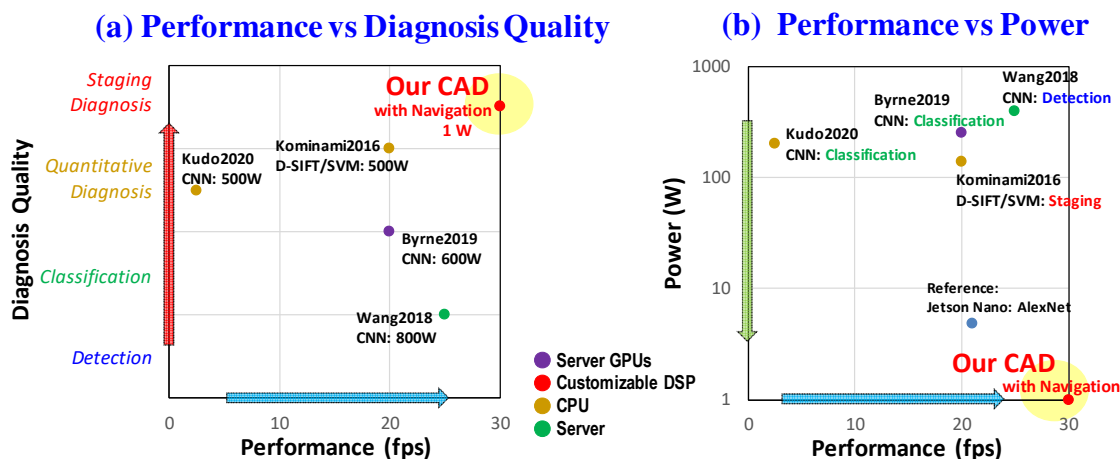


Figure 17 Comparison of (a) performance vs diagnosis quality and (b) performance vs power.

Our objects are realizing 1) CAD system for real-time video (30 fps) on customizable DSP (30 fps, 30 ms latency and 1 W), 2) Real-time CAD with Quantitative staging in video (Over 90% accuracy), 3) Real-time Navigation to improve quantitative staging quality (30 fps and easy to operate for clinical doctor on site).

We aim to Provide to the doctor more accurately in clinical observation independent from the experience variation of doctors, and, establish real-time confidence diagnosis method such as a "second opinion" on site.

1.5. Conclusion

Since colorectal cancer is major leading causes of cancer-related patients and deaths in Japan and overseas, development of computer aided diagnosis system is required with high accurate diagnosis at clinical site. For realizing a high accurate computer-aided diagnosis system, high performance image processing and low power hardware is needed. We develop a system that classifies colorectal NBI magnification endoscopic images into three histologic findings types (Type 1, Type 2A, and Type 3) out of the four types of JNET classification shown in Fig. 9. Our research group reported the CAD system which classifies endoscopic images into three types (Type A, Type B, and Type C3) based on Hiroshima classification [29]. These pathological types are corresponding to Type 1, Type 2A and Type 3 in JNET classification respectively. In this thesis, I use same data in previous research as a part of training data and the developed CAD system classifies into three pathological types (Type 1, Type 2A, and Type 3).

In Chapter 2, I describe how does a previous CAD system with BoF and SVM work. Also, the application of Convolutional Neural Network (CNN) as feature extractor to a CAD system and SVM classification is described. Then, I described evaluation results of two SVMs for quantitative staging classification.

In Chapter 3, I describe problems which affect quantitative staging classification accuracy in endoscopic image data set for SVM training. Then, I describe how to deal the endoscopic image data set to improve quantitative staging classification accuracy. I describe application result to endoscopic NBI video images.

In Chapter 4, I describe the proposed CAD system with quantitative staging implementation on a customizable DSP core for real-time video, and, the optimization for performance and memory size. And, I describe a prototype of CAD system with CNN feature extraction and SVM classification using FPGA based prototyping system. Validation results and comparison results compared to other implementation methods are described.

In Chapter 5, Then, I describe two real-time navigation functions for improving quantitative staging and objective index to the doctor. I describe a navigation by unclear region detection using a neural network, and, navigation by multiple staging regions. I discuss the implementation on a

customizable DSP core for real-time navigation.

Finally, conclusions and summary are described in Chapter 6. In addition, I describe future works to improve the developed CAD system.

References

- [1] Cancer Registry and Statistics. Cancer Information Service, National Cancer Center, Japan, "<https://ganjoho.jp/public/index.html>".
- [2] ECIS -European Cancer Information System, "<https://ecis.jrc.ec.europa.eu>, accessed on day/month/year".
- [3] American Cancer Society. Cancer Statistics Center. <http://cancerstatisticscenter.cancer.org>. Accessed Month Day, Year.
- [4] R.L. Siegel, K.D. Miller, and A. Jemal, "Cancer statistics, 2020.," *A Cancer Journal for Clinicians*, vol.70, pp.7–30, 2020.
- [5] H. Brenner, L. Jansen, A. Ulrich, J. Chang-Claude, and M. Hoffmeister, "Survival of patients with symptom- and screening-detected colorectal cancer," *Oncotarget*, vol.7, No.28, pp.44695–44704, 2016.
- [6] J.S. Moore, and T.H. Aulet. "Colorectal Cancer Screening," *Surgical Clinics of North America*, vol.97, pp.487–502, 2017.
- [7] American Cancer Society medical and editorial content team, "About Colorectal Cancer.," American Cancer Society, pp.1-15, 2020.
- [8] Olympus Medical Systems Corp., <https://www.olympus-global.com/technology/technology/luceraelite/>
- [9] K. Gono, K. Yamazaki, N. Doguchi, T. Nonami, T. Obi, M. Yamaguchi, N. Ohyama, H. Machida, Y. Sano, S. Yoshida, Y. Hamano, and T. Endo, "Endoscopic Observation of Tissue by Narrowband Illumination.," *Optical Review*, vol.10, pp.211-215, 2003
- [10] K. Gono, T. Obi, M. Yamaguchi, N. Ohyama, H. Machida, Y. Sano, S. Yoshida, Y. Hamano, and T. Endo, "Appearance of enhanced tissue features in narrow-band endoscopic imaging.," *Journal of Biomedical Optics*, vol.9, No.3, pp.568–577, 2004.

- [11] Y. Sano, T. Horimatsu, K.I. Fu, A. Katagiri, M. Muto, and H. Ishikawa, "Magnifying observation of microvascular architecture of colorectal lesions using a narrow-band imaging system.," *Digestive Endoscopy*, vol.18, pp. S44-S51, 2006.
- [12] Y. Sano, H. Ikematsu, K.I. Fu, F. Emura, A. Katagiri, T. Horimatsu, K. Kaneko, R. Soetikno, and S. Yoshida, "Meshed capillary vessels by use of narrow-band imaging for differential diagnosis of small colorectal polyps.," *Gastrointestinal Endoscopy*, vol.69, pp.278-283, 2009.
- [13] A. Katagiri, K.I. Fu, Y. Sano, H. Ikematsu, T. Horimatsu, K. Kaneko, M. Muto, and S. Yoshida, "Narrow band imaging with magnifying colonoscopy as diagnostic tool for predicting histology of early colorectal neoplasia.," *Alimentary Pharmacology & Therapeutics*, vol.27, pp.1269-1274, 2008.
- [14] H. Ikematsu, T. Matsuda, F. Emura, Y. Saito, T. Uraoka, K.I. Fu, K. Kaneko, A. Ochiai, T. Fujimori, and Y. Sano, "Efficacy of Capillary Pattern type IIIA/IIIB by Magnifying Narrow Band Imaging for Estimating Depth of Invasion of Early Colorectal Neoplasms," *BioMed Central (BMC) Gastroenterology*, pp.1-6, 2010.
- [15] S. Tanaka, M. Hirata, S. Oka, H. Kanao, I. Kaneko, S. Oba, and K. Chayama, "Clinical Significance of Narrow Band Imaging (NBI) in Diagnosis and Treatment of Colorectal Tumor.," *Gastroenterological Endoscopy*, vol. 50, pp.1289-1297, 2008.
- [16] H. Kanao, S. Tanaka, S. Oka, M. Hirata, S. Yoshida, and K. Chayama, "Narrow-Band Imaging Magnification Predicts The Histology and Invasion Depth of Colorectal Tumors.," *Gastrointestinal Endoscopy*, vol.69, no.3, pp.631-636, 2009.
- [17] Y. Wada, S. Kudo, H. Kashida, N. Ikehara, H. Inoue, F. Yamamura, K. Ohtsuka, and S. Hamatani, "Diagnosis of Colorectal lesions with the Magnifying Narrow-Band Imaging System.," *Gastrointestinal Endoscopy*, vol.70, pp.522-531, 2009.

- [18] T. Nikami, S. Saito, H. Tajiri, and M. Ikegami, "The Evaluation of Histological Atypia and Depth of Invasion of Colorectal Lesions using Magnified Endoscopy with Narrow-Band Imaging.," *Gastrointestinal Endoscopy*, vol.51, pp.10-19, 2009.
- [19] D.G. Hewett, T. Kaltenbach, Y. Sano, S. Tanaka, B.P. Saunders, T. Ponchon, R. Soetikno, and DK. Rex, "Validation of a simple classification system for endoscopic diagnosis of small colorectal polyps using narrow-band imaging.," *Gastroenterology*, vol.143, pp.599-607, 2012.
- [20] N. Hayashi, S. Tanaka, D.G. Hewett, T.R. Kaltenbach, Y. Sano, T. Ponchon, BP. Saunders, DK. Rex, and RM. Soetikno, "Endoscopic prediction of deep submucosal invasive carcinoma: validation of the narrow-band imaging international colorectal endoscopic (NICE) classification.," *Gastrointestinal Endoscopy*, vol.78, pp.625-632, 2013.
- [21] Y. Saito, T. Matsuda, T. Nakajima, T. Sakamoto, M. Yamada, S. Saito, H. Ikematsu, Y. Wada, S. Oka, H. Kawano, Y. Sano, S. Tanaka, T. Fujii, S. Kudo, T. Uraoka, T. Uraoka, N. Kobayashi, H. Nakamura, K. Hotta, T. Horimatsu, N. Sakamoto, K.I. Fu, O. Tsuruta, H. Kashida, Y. Takeuchi, H. Machida, T. Kusaka, N. Yoshida, I. Hirata, T. Terai, H. Yamano, K. Kaneko, Y. Yamaguchi, N. Tamai, N. Nakano, N. Hayashi, M. Iwatate, H. Ishikawa, and S. Yoshida, "NBI Magnifying Endoscopic Classification of Colorectal Tumors Proposed by The Japan NBI Expert Team (JNET).," *Gastroenterological Endoscopy*, vol. 58, pp.2314-2322, 2016.
- [22] Y. Sano, S. Tanaka, S. Kudo, S. Saito, T. Matsuda, Y. Wada, T. Fujii, H. Ikematsu, T. Uraoka, N. Kobayashi, H. Nakamura, K. Hotta, T. Horimatsu, N. Sakamoto, K.I. Fu, O. Tsuruta, H. Kawano, H. Kashida, Y. Takeuchi, H. Machida, T. Kusaka, N. Yoshida, I. Hirata, T. Terai, H. Yamano, K. Kaneko, T. Nakajima, T. Sakamoto, Y. Yamaguchi, N. Tamai, N. Nakano, N. Hayashi, S. Oka, M. Iwatate, H. Ishikawa, Y. Murakami, S. Yoshida, and Y. Saito, "Narrow-Band Imaging (NBI) Magnifying Endoscopic Classification of Colorectal Tumors Proposed by the Japan NBI Expert Team," *Digestive Endoscopy*, vol.28, pp.526-533, 2016.

- [23] M. Iwatate, Y. Sano, S. Tanaka, S. Kudo, S. Saito, T. Matsuda, Y. Wada, T. Fujii, H. Ikematsu, T. Uraoka, N. Kobayashi, H. Nakamura, K. Hotta, T. Horimatsu, N. Sakamoto, K.I. Fu, O. Tsuruta, H. Kawano, H. Kashida, Y. Takeuchi, H. Machida, T. Kusaka, N. Yoshida, I. Hirata, T. Terai, H. Yamano, T. Nakajima, T. Sakamoto, Y. Yamaguchi, N. Tamai, N. Nakano, N. Hayashi, S. Oka, H. Ishikawa, Y. Murakami, S. Yoshida, and Y. Saito, "Validation study for development of the Japan NBI Expert Team classification of colorectal lesions.," *Digestive Endoscopy*, vol.30, pp.642-651, 2018.
- [24] S. Kobayashi, M. Yamada, H. Takamaru, T. Sakamoto, T. Matsuda, S. Sekine, Y. Igarashi, and Y. Saito, "Diagnostic yield of the Japan NBI Expert Team (JNET) classification for endoscopic diagnosis of superficial colorectal neoplasms in a large-scale clinical practice database.," *United European Gastroenterol Journal*, vol7, pp.914-923, 2019.
- [25] The Society of Motion Picture and Television Engineers, "SMPTE STANDARD for Television - Bit-Serial Digital Interface for High-Definition Television Systems.," ANSI/SMPTE 292M, 1996.
- [26] H. Inoue, M. Kaga, H. Hinami, S. Sugaya, Y. Sato, K. Sasajima, S. Hamatani, A. Shiokawa, and S. Kudo, "Ultra-High Magnifying Endoscopy; From Endomicroscopy to Endocytoscopy.," *Gastroenterological Endoscopy*, vol.49, pp.2811-2818, 2007.
- [27] M. Misawa, S. Kudo, Y. Mori, H. Nakamura, S. Kataoka, Y. Maeda, T. Kudo, T. Hayashi, K. Wakamura, H. Miyachi, A. Katagiri, T. Baba, F. Ishida, H. Inoue, Y. Nimura, and K. Mori, "Characterization of Colorectal Lesions Using a Computer-Aided Diagnostic System for Narrow-Band Imaging Endocytoscopy," *Gastroenterology*, vol.150, no.7, pp.1531-1532, 2016.
- [28] J. Yoshimuta, T. Takeda, T. Tamaki, R. Bisser, K. Kaneda, Y. Takemura, S. Yoshida, and S. Tanaka, "Colorectal NBI Image Recognition using Dense SIFT.," *IEICE Technical Report*, vol.110, pp.129-134, 2010.

- [29] T. Tamaki, J. Yoshimuta, M. Kawakami, B. Raytchev, K. Kaneda, S. Yoshida, Y. Takemura, K. Onji, R. Miyaki, and S. Tanaka, "Computer-Aided Colorectal Tumor Classification in NBI Endoscopy using Local Features.," *Medical Image Analysis*, vol.17, no.1, pp. 78-100, 2013.
- [30] Y. Kominami, S. Yoshida, S. Tanaka, Y. Sanomura, T. Hirakawa, B. Raytchev, T. Tamaki, T. Koide, K. Kaneda, and K. Chayama, "Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy.," *Gastrointestinal Endoscopy*, vol.83, no.3, pp.643-649, 2016.
- [31] S. Yoshida, Y. Kominami, S. Tanaka, T. Tamaki, T. Koide, and K. Chayama, "Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy.," *Intestine*, vol.21, no.2, pp.182-184, 2017.
- [32] T. Mishima, S. Shigemi, A. Hoang, T. Koide, T. Tamaki, B. Raytchev, K. Kaneda, Y. Kominami, R. Miyaki, T. Matsuo, S. Yoshida, and S. Tanaka, "FPGA Implementation of Feature Extraction for Colorectal Endoscopic Images with NBI Magnification.," *Proceedings of 2014 IEEE International Symposium on Circuits and Systems, Melbourne, Australia*, pp.2515-2518, 2014.
- [33] T. Okamoto, T. Koide, A. Hoang, T. Shimizu, K. Sugi, Toru Tamaki, T. Hirakawa, B. Raytchev, K. Kaneda, S. Yoshida, H. Mieno, and S. Tanaka, "An FPGA Implementation of SVM for Type Identification with Colorectal Endoscopic Images.," *Proceedings of the 20th Workshop on Synthesis and System Integration of Mixed Information technologies, Kyoto, Japan*, pp.81-86, 2016.
- [34] T. Koide, T. Okamoto, T. Shimizu, K. Sugi, A. Hoang, T. Tamaki, B. Raytchev, K. Kaneda, S. Yoshida, H. Mieno, and S. Tanaka, "Compact and High-Speed Hardware Feature Extraction Accelerator for Dense Scale Invariant Feature Transform.," *Proceedings of the 31st International Technical Conference on Circuits/Systems, Computers and*

Communications, Naha, Japan, pp.387-390, 2016.

- [35] T. Koide, T. Okamoto, K. Sugi, T. Shimizu, A. Hoang, T. Tamaki, B. Raytchev, K. Kaneda, S. Yoshida, H. Mieno, and S. Tanaka, "A Hardware Accelerator for Bag-of Features based Visual Word Transformation in Computer Aided Diagnosis for Colorectal Endoscopic Images.," Proceedings of the 31st International Technical Conference on Circuits/Systems, Computers and Communications, Naha, Japan, pp.233-236, 2016.
- [36] PJ Chen, MC Lin, MJ Lai, JC Lin, HH Lu, and VS Tseng, "Accurate Classification of Diminutive Colorectal Polyps Using Computer-Aided Analysis.," *Gastroenterology*, vol.154, pp. 568-575, 2018.
- [37] P Wang, X Xiao, JR Brown, TM Berzin, M Tu, F Xiong, X Hu, P Liu, Y Song, D Zhang, X Yang, L Li, J He, X Yi, J Liu, and X Liu, "Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy.," *Nature Biomedical Engineering*, vol2, pp. 741-748, 2018.
- [38] MF Byrne, N Chapados, F Soudan, C Oertel, M Linares Pérez, R Kelly, N Iqbal, F Chandelier, and DK Rex, "Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model.," *Gut*, vol.68, pp. 94-100, 2019.
- [39] M Min, S Su, W He, Y Bi, Z Ma, and Y Liu, "Computer-aided diagnosis of colorectal polyps using linked color imaging colonoscopy to predict histology.," *Scientific Reports*, vol.9, pp.2881.
- [40] SE Kudo, M Misawa, Y Mori, K Hotta, K Ohtsuka, H Ikematsu, Y Saito, K Takeda, H Nakamura, K Ichimasa, T Ishigaki, N Toyoshima, T Kudo, T Hayashi, K Wakamura, T Baba, F Ishida, H Inoue, H Itoh, M Oda, and K Mori, "Artificial Intelligence-assisted System Improves Endoscopic Identification of Colorectal Neoplasms.," *Clinical Gastroenterology and Hepatology*, vol.18, pp. 1874-1881, 2020.

- [41] JY. Lee, J. Jeong, EM. Song, C. Ha, HJ. Lee, JE. Koo, DH. Yang, N. Kim, and JS. Byeon, “Real-time detection of colon polyps during colonoscopy using deep learning: systematic validation with four independent datasets.” *Scientific Reports* vol.10, pp.8379, 2020.
- [42] T Ozawa, S Ishihara, M Fujishiro, Y Kumagai, S Shichijo, and T Tada. “Automated endoscopic detection and classification of colorectal polyps using convolutional neural networks.” *Therapeutic Advances in Gastroenterology*, vol.13, January 2020. doi:10.1177/1756284820910659

Chapter 2. Computer-Aided Diagnosis System for Colorectal Endoscopic Image

2.1. Positioning of the Computer-Aided Diagnosis System in this thesis

In the field of gastrointestinal endoscopy, the CAD system is the most investigated area, although it is still in the preclinical phase. Because colonoscopy is carried out by humans, it is inherently an imperfect procedure. The CAD system is expected to improve its quality regarding automated polyp detection and characterization (i.e. predicting the polyp's pathology). It could help prevent endoscopists from missing polyps as well as provide a precise optical diagnosis for those detected. Ultimately, these functions that CAD provides could produce a higher adenoma detection rate and reduce the cost of polypectomy for hyperplastic polyps.

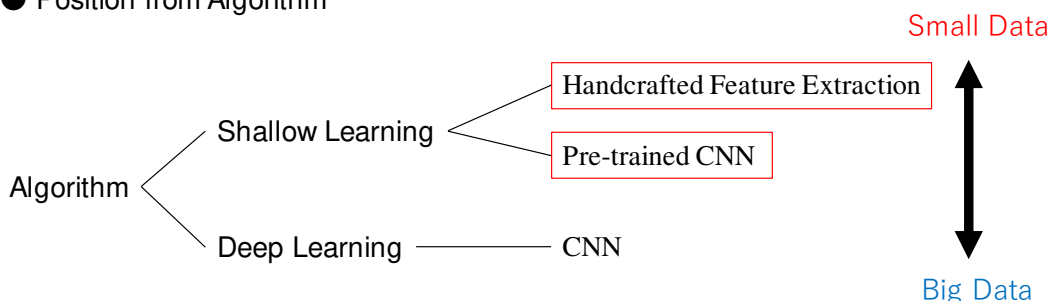
A computer-vision algorithm developed by computer scientists is adopted to a CAD system for the classification of colorectal polyps. There are two major approaches for realizing a CAD system, shallow learning and deep learning which are technologies of machine learning. Shallow learning, such as support vector machines (SVM), require extraction of features from training samples. For example, when developing a CAD system for colorectal polyps using shallow learning, an algorithm that extracts specific features of polyps from endoscopic images, such as color or shape, must be formulated. For feature extraction, there are two methods, hand-crafted feature extraction such as Bag-of-Features and pre-trained CNN. Shallow learning can be trained by a limited number of training image. In contrast, deep learning requires tremendous number of training image.

For a CAD system implementation, there are two approaches, software implementation and hardware implementation. For realizing a real-time CAD system which requires high performance, software implementation needs a server machine or a high-performance desktop PC. Thus, hardware implementation is preferable. A GPU and a customizable DSP can be executed software algorithm and also can be customized in HW. The other hand, implementation by FPGA takes long development period.

We aim to realize a high-accuracy, high-performance, low-power CAD system stored in the same rack of the endoscopic system such as an edge device that can be used in the small clinical site with limited training data. Therefore, our CAD system is based on shallow learning as shown red

rectangle in Fig.18. We selected a customizable DSP which power consumption is less compared to a GPU.

● Position from Algorithm



● Position from Implementation

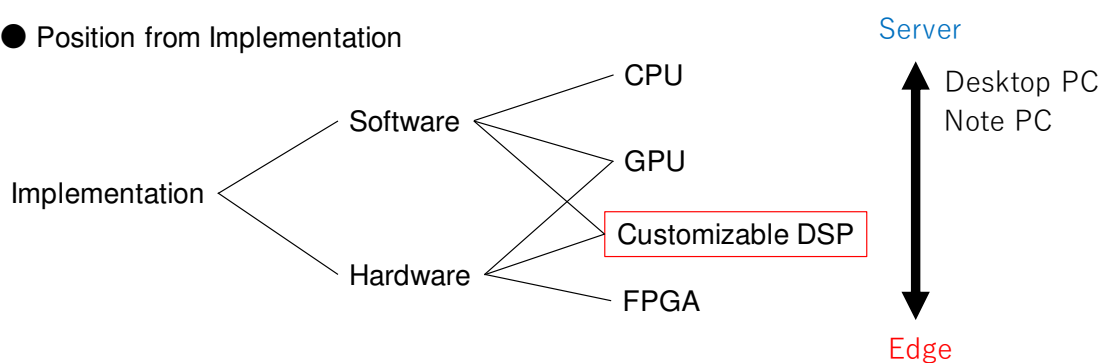


Figure 18. Positioning of the Computer-Aided Diagnosis System.

2.2. Computer-Aided Diagnosis System based on Bag-of-Features

Figure 19 shows outline of the CAD system based on Bag-of-Features (BoF) representation of local features in the endoscopic image. The BoF methods have been applied to image classification, object detection etc. The BoF applies a document search to an image. This method considers a feature vector achieved by extracting the feature of an image and classifies it by the appearance frequency [43-46].

The system has two stages, learning (Fig.19 (a)) and testing (Fig.19 (b)). The overview of processing flow of the system is as follows.

In the learning phase (Fig.19 (a)), a set of images with labels corresponding to JNET classification (Type 1, Type 2A, and Type3) is used. The local feature quantities of the endoscopic images for each type are extracted by Dense Scale-Invariant Feature Transform (D-SIFT) algorithm [47-48] (Fig.19 (a1)). The pit patterns of endoscopic images are complex compare to object recognition such as face recognition or pedestrian recognition etc. The D-SIFT

is a technique that divides and compute the features of an image in 512 dimensions from the gradient of luminosity value. An image with little color change such as a large intestine endoscopic image can be identified by taking a lot of features that the D-SIFT extracted. D-SIFT library in VLFeat (Vision Lab Features library) is provided as opensource [48]. This is a cross-platform open source of vision algorithms. Previous CAD systems compared in this thesis used this library [40-43].

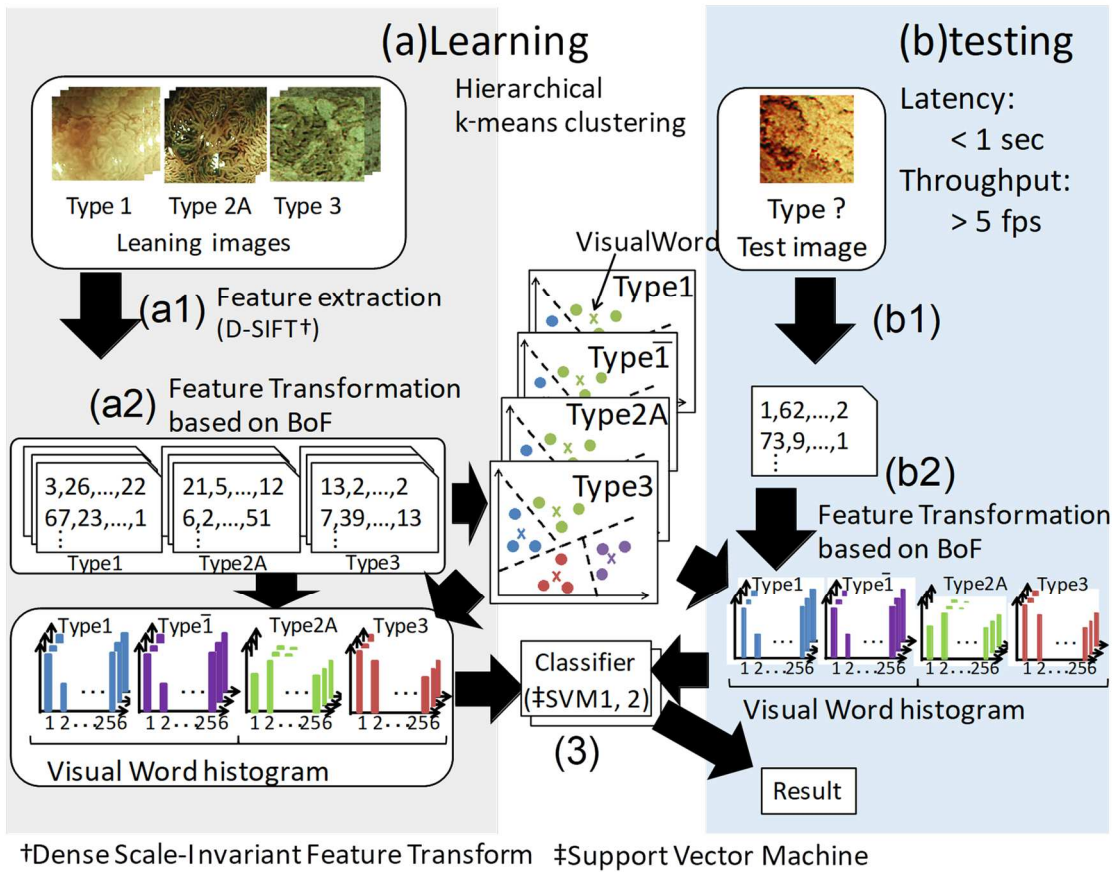


Figure 19. Overview of Computer-Aided Diagnosis system for endoscopic images based BoF

Then, the features attained at the learning phase are grouped into 256 ($=2^8$) clusters by each type. The center of each cluster is saved as a Visual Word (VW) for each type used for feature representation using k-means clustering (Fig.19 (a2)). VW histograms are input to Support Vector Machines (SVM1, SVM2) for the classification learning.

In the classifier module (Fig.19 (3)), support vector (SV) for support vector

machine (SVM) is attained at the learning phase using the type information of leaning endoscopic images which are labeled by the professional diagnosis doctors.

Next, in the testing phase (Fig.19 (b)), the D-SIFT feature extraction is performed for a whole input endoscopic image. And, a visual-words histogram is created by voting for the nearest VW. Finally, the CAD system classifies the testing endoscopic image within an endoscopic video frame by pre-learned SVMs.

Also, our research group applied the system in real clinical test to show a utility of the real-time image recognition system. A set of 2,247 trimmed endoscopic images collected by a clinical doctor are used in the learning phase. The learning images are categorized into 504 Type A images (non-cancer) and 1,743 Type B and Type C3 images (cancer) by Hiroshima classification. So, the CAD system classifies into Type A or Type B/Type C3 in this experiment. The classification of type A, Type B and Type C3 defined by NBI magnification findings are corresponding to the JNET classification, Type 1, Type 2A, and Type 3, respectively as shown in Table 3 [49-50]. The feature quantities extracted from D-SIFT are 128-dimensional vectors. The output of the system is the SVM output, and a cutoff value of 0.5 is used to categorize the input images into two types. Figure 20 shows the result of evaluation. We verified that the SVM output of neoplastic lesions could identify the non-neoplastic lesions (non-cancer) or neoplastic lesions (cancer) [30].

Table 3. Correspondence of JNET classification to previous classifications. [41-42]

Sano Classification [11-14]	Type I	Type II	Type III A		Type III B
Jikei Classification [17]	Type 1	/		/	
Showa Classification [18]	Normal - Faint	/		/	
Hiroshima Classification [15-16]	Type A	Type B	Type C1	Type C2	Type C3
JNET Classification [19-20]	Type 1	Type 2A	Type 2B		Type 3

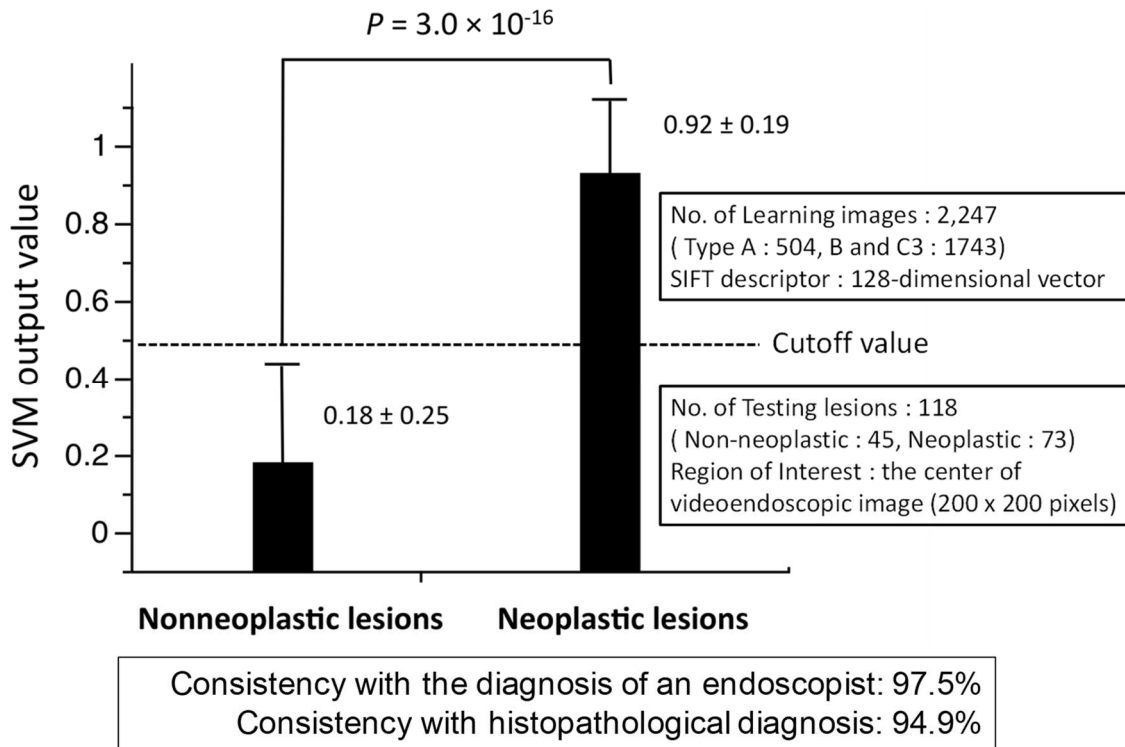


Figure 20. SVM output for the images of non-neoplastic and neoplastic lesions [30].

2.3. Computer-Aided Diagnosis System with Convolutional Neural

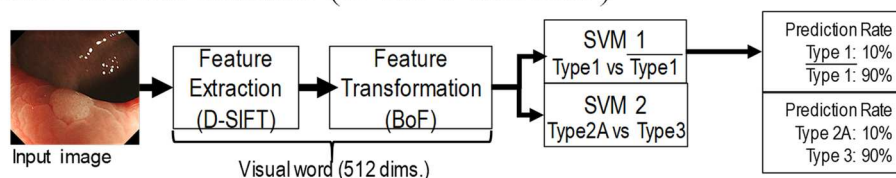
2.3.1. Convolutional Neural Network and Feature Extractor

Convolutional Neural Network (CNN) was proposed by LeCun, et al [51]. CNN have number of convolutional layers to extract features from images by repeating the resolution reduction of a feature map using the value of the local region. CNNs are known to high accurate results on image classification, recognition and segmentation in different fields of image application. In general image classification fields, it becomes capable to use a large number of image data needed to train a CNN, and but also several effective training methods have been proposed. However, in the medical application, there are not large number of images, so it is difficult to prepare training images for CNN approaches. The transferability of different layers in CNN, such as AlexNet and reuse them for a new task, offered a promising opportunity to overcome the lack of training samples issue and showed great performance results [52]. CAD systems with CNN as feature extractor and SVM as classifier have been proposed [53-55]. Also, our team proposed CAD systems for colorectal endoscopic image with CNN as feature extractor and SVM as

classifier [56-59].

Figure 21 shows overview of a CAD system with BoF using handcraft feature amount, and a CAD system with CNN feature extraction and SVM classification which is used in this thesis, respectively. Instead of D-SIFT and feature transformation, CNN is used. Our research group have evaluated pre-trained CNNs, AlexNet [60], CaffeNet [61] and GoogleNet [62] as feature extraction [63]. Our research group confirmed that output vector of 1000 dimensions from pre-trained CNNs trained by ImageNet data set provided in Large Scale Visual Recognition Challenge 2012 (ILSVRC 2012) dataset [64] is useful as a feature for other classification task without fine-tuning, and CAD systems with these pre-trained CNNs achieved over 90% accuracy. Since AlexNet has the simplest structure and considering the performance when it is installed in an embedded system, the pre-trained AlexNet, which is pre-trained by ILSVRC 2012 dataset, is used as a CNN feature extractor in this thesis. The ILSVRC is a benchmark in object classification and recognition on hundreds of object categories and millions of images. In the ImageNet dataset, there is no endoscopic image data [65]. There are two SVM classifiers which is connected sequentially for Type 1 vs Type not 1 (=Type 2A / Type 3) classification, and Type 2A vs Type 3 classification.

● Hand-crafted features (D-SIFT and BoF)



● Computer-Aided Diagnosis system using CNN feature (CNN-SVM)

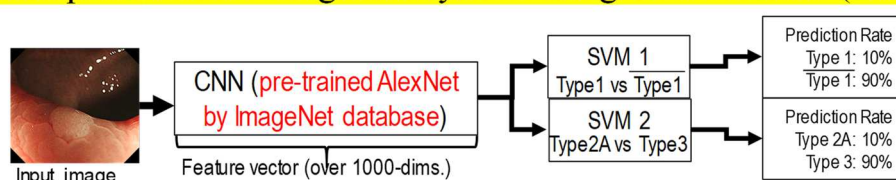


Figure 21. Overview of BoF based and CNN-SVM CAD systems.

Figure 22 shows the architecture of AlexNet. AlexNet has eight layers, five of them are convolution layers (*conv1* to *conv5*) with normalization (*norm1* and *norm2*) and pooling (*pool1*, *pool2*, and *pool5*), and, two of them are fully connected layers (*fc6* and *fc7*) followed by a softmax layer (*fc8*).

Figure 23 to Figure 25 show the top five categories and their probability values output when a Type 1, Type 2A and Type 3 endoscopic images are inputted to D-SIFT and AlexNet feature extractors, respectively. The AlexNet as the feature extractor isn't trained by endoscopic images, therefore, it outputs high probability values closely to learned categories such as bandages (Fig.23 (a)), jellyfish (Fig.24 (a)) and trilobites (Fig.25 (a)). Output values from D-SIFT have 512 dimensions (Fig.23 (b) to Fig.25 (b)), and, output values from AlexNet have 1000 dimensions (Fig.23 (c) to Fig.25 (c)).

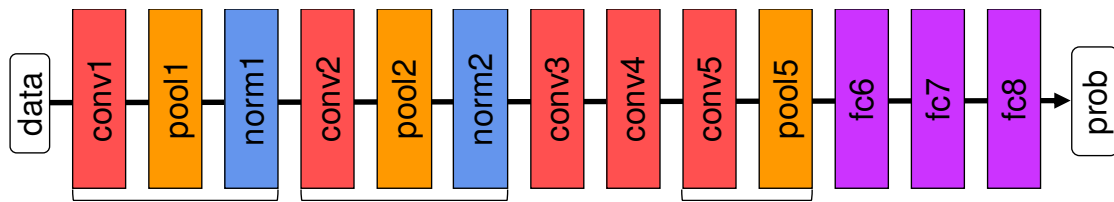


Figure 22. Architecture of AlexNet [60].

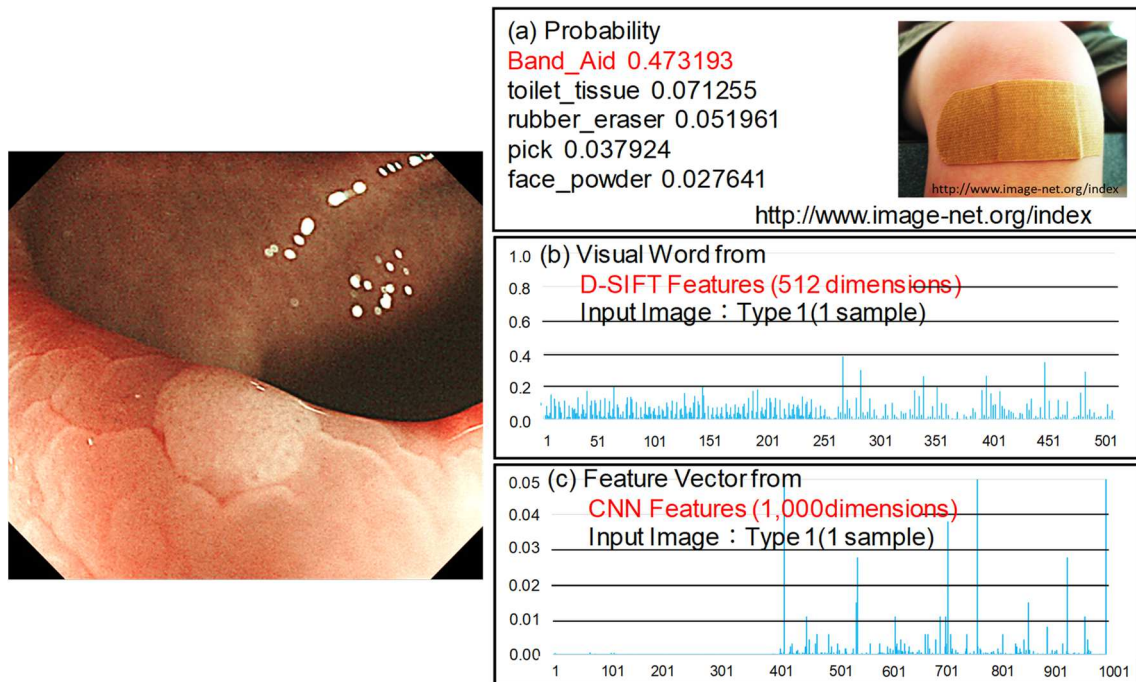


Figure 23. Extracted Features for Type 1 image input: (c) Visual Word from D-SIFT, (d) Feature Vector from AlexNet.

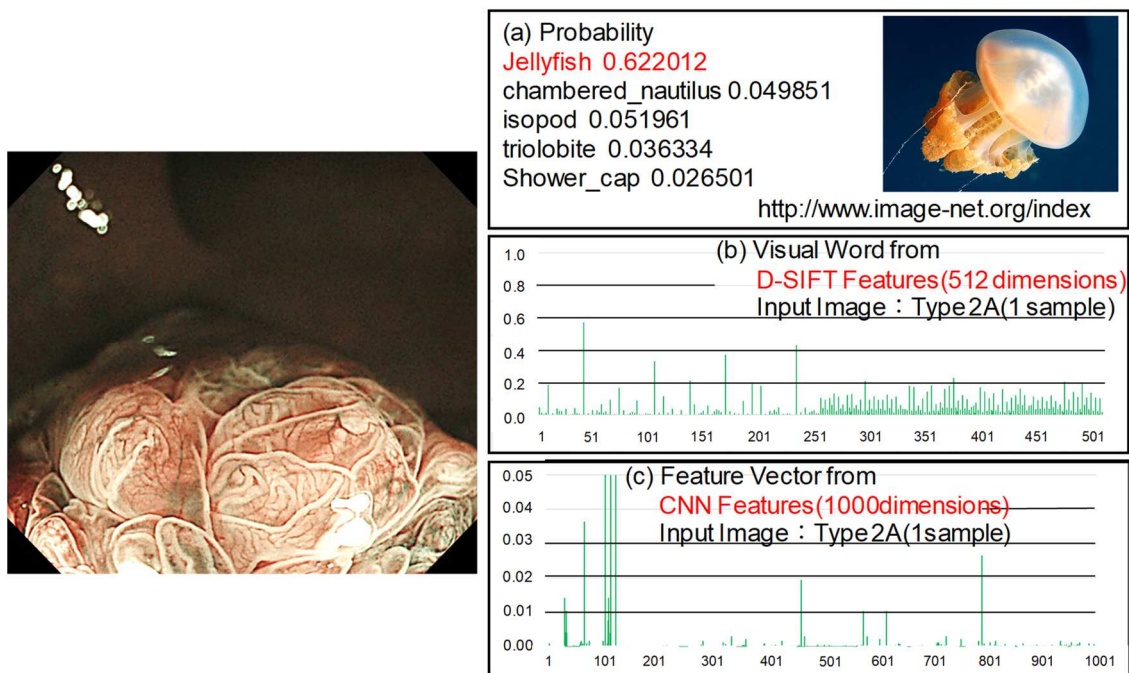


Figure 24. Extracted Features for Type 2A image input: (c) Visual Word from D-SIFT, (d) Feature Vector from AlexNet.

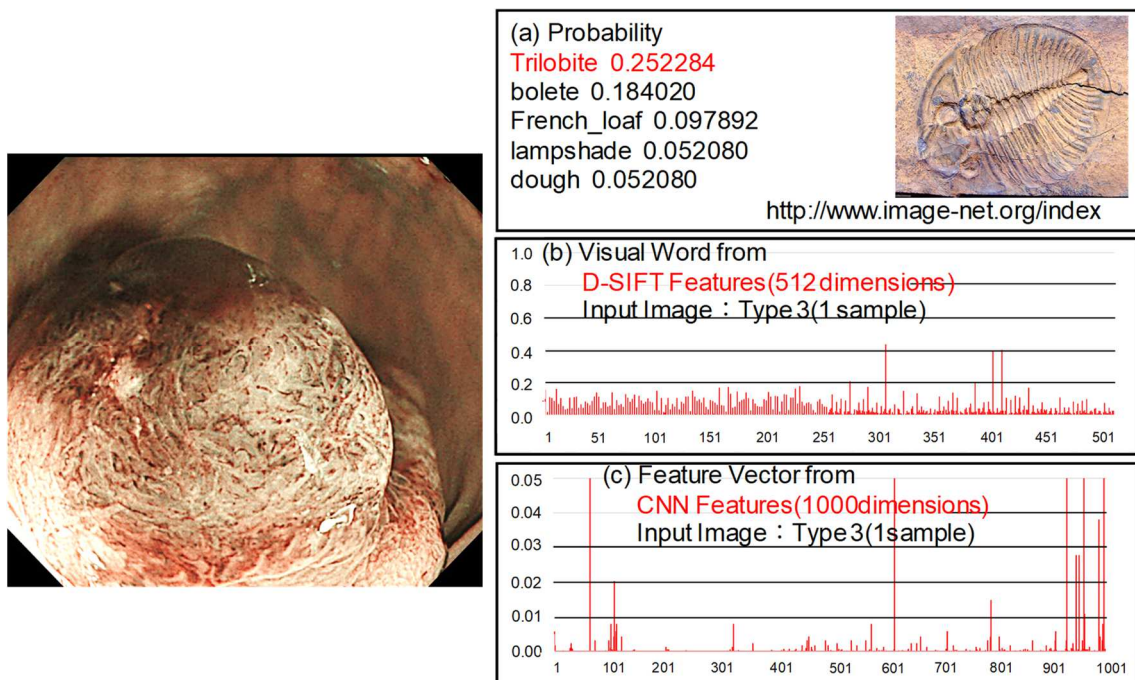


Figure 25. Extracted Features for Type 3 image input: (c) Visual Word from D-SIFT, (d) Feature Vector from AlexNet.

Figure 26 and Figure 27 shows the accuracy of the SVM trained by D-SIFT and the SVM trained by the output from *prob* of the pre-trained AlexNet.

Figure 26 shows the result of Type 1 vs Type not 1 classification. The output from *prob* of the pre-trained AlexNet is comparable to the conventional D-SIFT. In Fig.27, both the True Positive and Precision Rate of the SVM trained by the output from *prob* of the pre-trained AlexNet exceeded 85%, and these measures are improved from D-SIFT. It is considered to be used to quantitative staging classification.

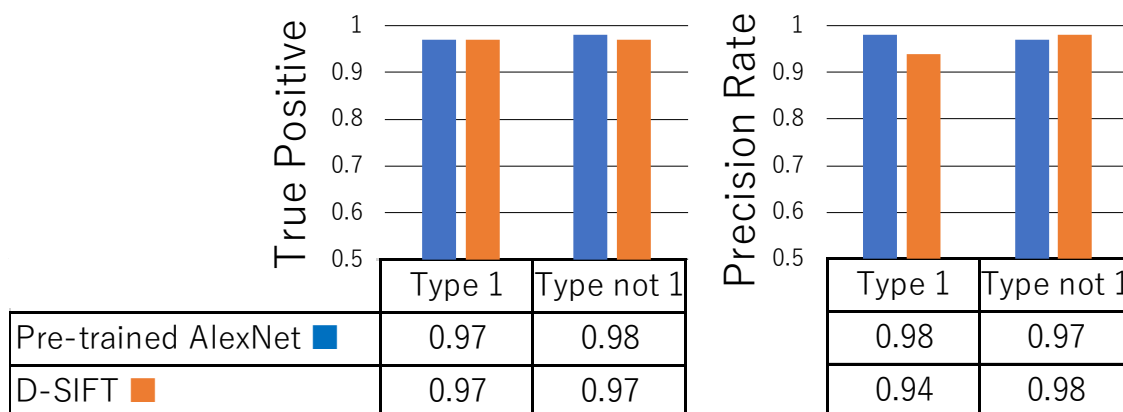


Figure 26 Comparison of Pre-trained AlexNet vs D-SIFT (Type1 vs Type not 1)

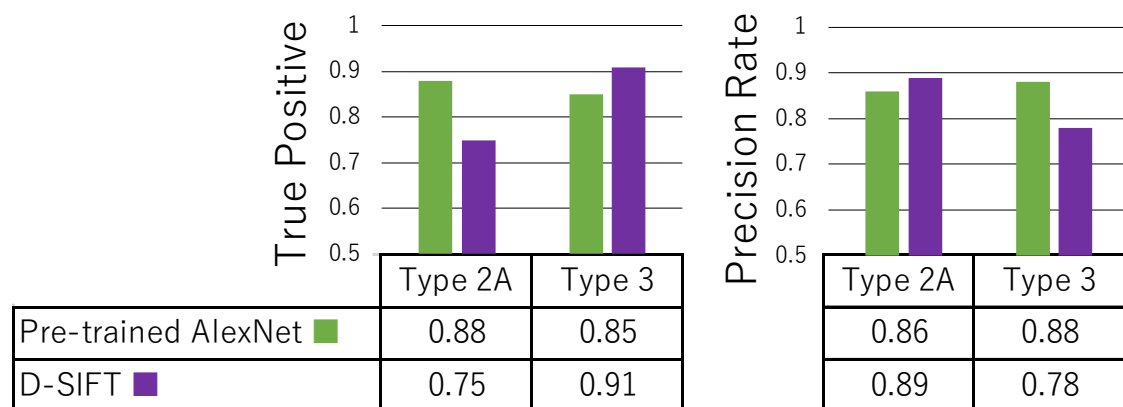


Figure 27 Comparison of Pre-trained AlexNet vs D-SIFT (Type 2A vs Type 3)

2.3.2. Support Vector Machine as Classifier

Support Vector Machine (SVM) has been introduced in the late 1990s [66]. SVM is a supervised machine learning method capable of classifying data into two categories shown in Fig.28. To classify data into the two categories, there are many possible hyperplanes that could be defined. Hyperplanes are decision boundaries that classify the data points into two categories. The closest data points of both classes to the hyperplane are called support vectors

(SV), as they define the margin and correspondingly the location and orientation of the hyperplane. The purpose of SVM is to find a hyperplane that has the maximum margin that is the maximum distance between data points of both classes. Therefore, future data points can be classified with more confidence. One of the SVM features is that it can be applied to problems that cannot be linearly separated by mapping the input space to a high-dimensional feature space (Kernel Trick) [67].

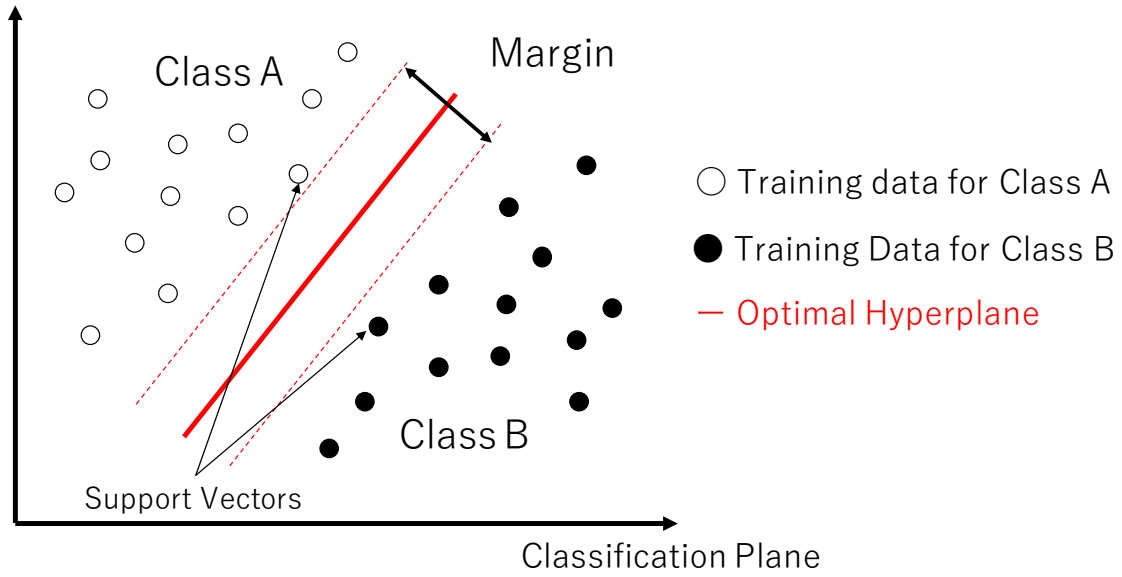


Figure 28. Optimal Classification Hyperplane by SVM.

The SVM classification function is shown in the equation (1). A classification hyperplane for linear separation is created by the supervised learning with SVM based on these VW histograms. The VW histograms of the output from pre-trained AlexNet show different feature vectors for each JNET Type as shown in Fig.23 (c) to Fig.25 (c). This shows possibility of classification by SVM. The equation is an example for determining whether the N-dimensional VWs, which is the input data to the SVM classifier, is Type 1 or Type not 1.

$$f_{1:\bar{1}}(\vec{x}) = \sum_{i=1}^{N_1+N_{\bar{1}}} coef_i \times (\overline{sv}_i \cdot \vec{x}) + \rho_{1:\bar{1}} \quad (1)$$

In equation (1), \overline{sv}_i is a support vectors (SV) determined during learning and constitutes a hyperplane which classifies data into two categories, Type

1 and Type not 1. In the CNN feature extraction, \overline{sv}_i and \vec{x} are N-dimensional Visual Word histogram data. In addition, $coef_i$ indicates the coefficient of each sv, and $\rho_{1:\bar{1}}$ indicates the coefficient of the discriminator, which are also determined during learning. N_1 and $N_{\bar{1}}$ indicate the number of SVs of Type 1 and Type not 1, respectively. In order to realize 3-class (3-type) classification, two classification stages by SVM are implemented. The first stage (Fig.21 SVM 1) classifies type 1 or Type not 1 (= Type 2A or Type 3), the second stage (Fig. 21 SVM 2) classifies type 2A or Type 3 from the images classified as Type not 1 in the first stage.

As libraries for SVM, the LIBSVM library [68] that supports multiple kernel functions and multi-class classification, and the LIBLINEAR library [69] that supports only linear kernel classification and speeds up processing have been released. Comparison studies of the LIBSVM library and the LIBLINEAR library have been reported [70-71]. The LIBLINEAR library is faster for large data classification.

Also, our research group reported comparison study for a CAD system. In this experiment, 908 trimmed endoscopic images collected by a technical doctor are used for SVM training. For LIBSVM, the number of support vector (SV) is 362 and each SV has 1000 dimensions. When storing one-dimensional data with double type 64 bit, the required amount of memory is about $362 \times 1000 \times 64 = 2.7$ MB. On the other hand, for LIBLINEAR, it is necessary to hold 1000 weighting coefficients for each dimension assuming that it is identified by logistic regression, and the required amount of memory data is $1000 \times 64 = 7.8$ KB. The amount of data for LIBLINEAR is small, almost 1/ 300 of that for LIBSVM. Figure 29 shows the comparison of classification accuracy for Type 1 vs Type not 1 when the same image for SVM training is inputted. Figure 30 shows, the comparison of classification accuracy for Type 1 vs Type not 1 when the image rotated 90 degree to right is inputted. From these results, it was confirmed LIBLINEAR is more accurate and achieve over 90% classification accuracy except for the Precision Rate of Type 1 when the image rotated 90 degree is inputted.

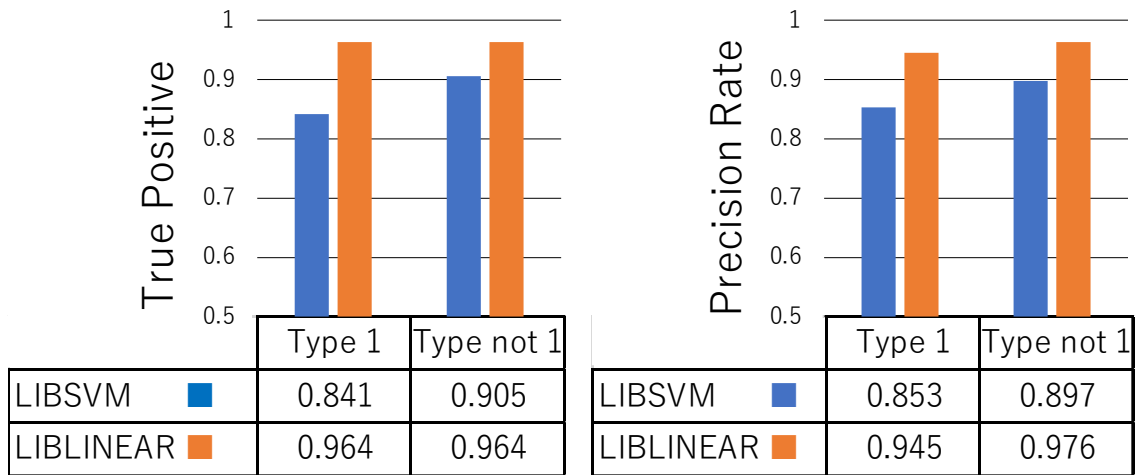


Figure 29. SVM classifier Type 1 and Type not 1
Test data: Same as trained data

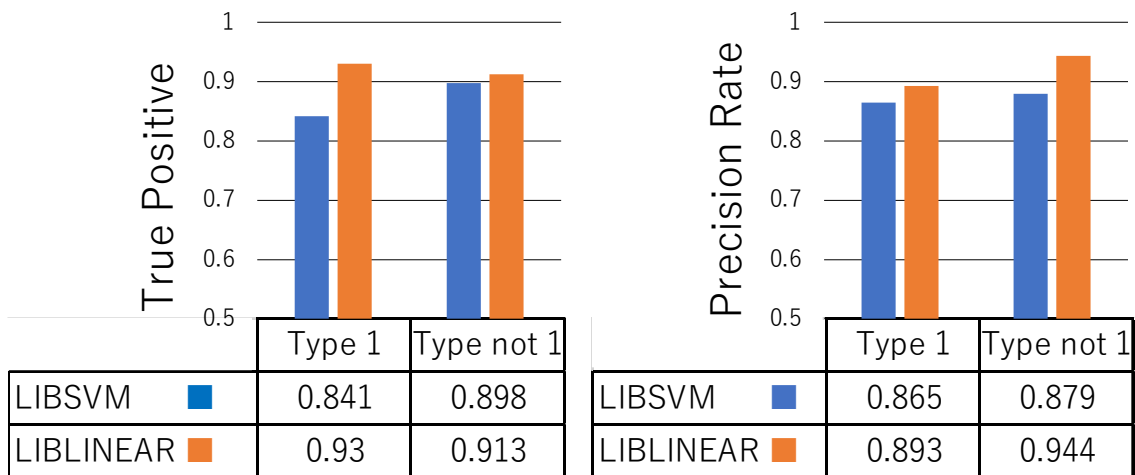


Figure 30. SVM classifier Type 1 and Type not 1
Test data: 90 degree rotated trained data to right

2.4. Conclusion

In this chapter, I explained a CAD system based on D-SIFT and BoF, and, a CAD system using CNN feature extraction and SVM classification which was developed by our research group. The output of each layer of CNN as a multi-dimensional vector expressing the feature quantity of the input endoscopic image is almost equivalence to the feature quantity based on D-SIFT. We use pre-trained AlexNet as a feature extractor instead of D-SIFT and BoF, and, realize a CAD system using CNN features and SVM classification. I confirmed over 90% accuracy for quantitative staging classification with pre-trained CNN instead of D-SIFT by limited training data. For realizing on an

embedded system, our group evaluated two SVM libraries, the LIBSVM and the LIBLINEAR. We confirmed that the LIBLINEAR is faster than LIBSVM with fewer amount of memory size and it is possible to realize three pathological types classification using two SVM classifiers connected sequentially.

References

- [43] G. Csurka, CR. Dance, L. Fan, J. Willamowski, and C. Bray, “Visual categorization with bags of keypoints.,” Proceedings of ECCV Workshop on Statistical Learning in Computer Vision 2004, Prague, Czech, pp.59-74, 2004.
- [44] J. Sivic, and A. Zisserman, “Video Google: a text retrieval approach to object matching in videos.,” Proceedings of the International Conference on Computer Vision, Nice, France, vol.2, pp.1470– 1477, 2003.
- [45] S. O'Hara, and B. Draper, "Introduction to the Bag of Features Paradigm for Image Classification and Retrieval.," CoRR abs/1101.3354, pp.1-25, 2011.
- [46] S. Lazebnik, C. Schmid, and J. Ponce, “Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories.,” Proceedings of IEEE Conference on Computer Vision & Pattern Recognition (CVPR'06), New York, USA, pp.2169-2178, 2006.
- [47] I. Kokkinos, M. Bronstein, and A. Yuille, “Dense scale invariant descriptors for images and surfaces.,” [Research Report] RR-7914, INRIA, pp.1-29, 2012.
- [48] A. Vedaldi, and B. Fulkerson, “VLFeat: An Open and Portable Library of Computer Vision Algorithms.,” <http://www.vlfeat.org/>, 2008.
- [49] H. Kashida, “Efficacy of Magnifying Endoscopy with Narrow Band Imaging in The Diagnosis of Colorectal Neoplasia.,” Gastroenterological Endoscopy, vol.54, issue.2, pp.235-243, 2012.
- [50] H. Tajiri, Y. Saito, and K. Sumiyama, “見えないものが観えてくる! 画像強調内視鏡の診断ロジック”, 羊土社, 2016.
- [51] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, “Backpropagation Applied to Handwritten Zip Code Recognition,” Neural Computation, vol.1, no.4, pp.541–551, Dec.

1989.

- [52] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation.," Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR'15), Boston, MA, USA, pp.3431-3440, 2015.

- [53] D.A. Ragab, M. Sharkas, S. Marshall, and J. Ren, "Breast cancer detection using deep convolutional neural networks and support vector machines.," PeerJ. pp.1-19, 2019.

- [54] K.M. Hosny, M.A. Kassem, and M.M. Foad, "Classification of skin lesions using transfer learning and augmentation with Alex-net.," PLOS ONE 14, pp.1-17, 2019.

- [55] R.A. Jasmine, and P.A.J. Rani, "Brain tumor classification using SVM based AlexNet.," Journal of University of Shanghai for Science and Technology, vol.22, pp.122-128, 2020.

- [56] T. Okamoto, T. Koide, T. Tamaki, B. Raytchev, K. Kaneda, S. Yoshida, H. Mieno, and S. Tanaka, "Classification Method for Real-Time NBI Colorectal Endoscopic Images with CNN features and SVM.," Proceedings of Design Automation Symposium 2017, Kaga, Japan, pp.27-32, 2017.

- [57] T. Okamoto, T. Kide, T. Tamaki, B. Raytchev, K. Kaneda, S. Yoshida, H. Mieno, S. Tanaka, H. Toishi, T. Sugawara, M. Tsuji, M. Odagawa, and N. Tamba, "Implementation of Computer-Aided Diagnosis System for Colorectal Endoscopic Images with CNN features and SVM on Customizable DSP Core.," Proceedings of Design Automation Symposium 2017, Kaga, Japan, pp.33-38, 2017.

- [58] M. Tsuji, H. Toishi, T. Sugawara, M. Odagawa, N. Tamba, T. Okamoto, T. Koide, T. Tamaki, B. Raytchev, K. Kaneda, S. Yoshida, H. Mieno, and S. Tanaka, "Real-time processing of computer-aided diagnosis system for colorectal tumor classification in NBI endoscopy using CNN features by implementing to Tensilica Vision P6 DSP.," Proceedings of the 2nd

International Symposium on Biomedical Engineering, Tokyo, Japan, pp.2425-2426, 2017.

- [59] M. Odagawa, M. Tsuji, H. Toishi, T. Sugawara, N. Tamba, T. Okamoto, T. Koide, T. Tamaki, B. Raytchev, K. Kaneda, S. Yoshida, H. Mieno, and S. Tanaka, "Tensilica DSP cores for neural networks and an application to computer-aided diagnosis system for colorectal tumor classification," Proceedings of the International Workshop on Nanodevice Technologies, Higashi-Hiroshima, Japan, pp.6-7, 2018.
- [60] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "ImageNet classification with deep convolutional neural networks," Advances in neural information processing systems, Vol.25, pp.1097-1105, 2012.
- [61] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional Architecture for Fast Feature Embedding.," Proceedings of the 22nd ACM international conference on Multimedia (MM '14), New York, NY, USA, pp.675–678, 2014.
- [62] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going Deeper With Convolutions.," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15), Boston, MA, USA, pp.1-9, 2015.
- [63] T. Tamaki, S. Sonoyama, T. Hirakawa, B. Raytchev, K. Kaneda, T. Koide, S. Yoshida, H. Mieno, and S. Tanaka, "Computer-Aided Colorectal Tumor Classification in NBI Endoscopy using CNN Features," Proceedings of the Korea-Japan joint workshop on Frontiers of Computer Vision (FCV2016), Takayama, Japan, pp.61-65, 2016.
- [64] J. Deng, A. Berg, S. Satheesh, H. Su, A. Khosla, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Competition 2012 (ILSVRC2012).," <https://image-net.org/challenges/LSVRC/2012/>
- [65] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang ,

- A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, and L. Fei-Fei, “ImageNet Large Scale Visual Recognition Challenge.,” *International Journal of Computer Vision*, vol.115, pp.211–252, 2015.
- [66] V. Vapnik, “Statistical learning theory,” John Wiley & Sons Inc., New York, 1998.
- [67] M. Hofmann, “Support Vector Machines — Kernels and the Kernel Trick.,” pp.1-6, 2006.
- [68] C.C. Chang, and C.J. Lin, “LIBSVM: A library for support vector machines.,” *ACM Transactions on Intelligent Systems and Technology*, vol.2, pp.27:1-27:27, 2011.
- [69] R.E. Fan, K.W. Chang, C.J. Hsieh, X.R. Wang, and C.J. Lin, “LIBLINEAR: A library for large linear classification”. *Journal of Machine Learning Research*, vol.9, pp.1871–1874, 2008.
- [70] H. Zou, and Z. Jin, “Comparative Study of Big Data Classification Algorithm Based on SVM”, *Proceedings of Cross Strait Quad-Regional Radio Science and Wireless Technology Conference (CSQRWC)*, Xuzhou, China, pp.1-3, 2018.
- [71] Y.W. Chang, and C.J. Hsieh, K.W. Chang, M. Ringgaard, C.J. Lin, “Training and Testing Low-degree Polynomial Data Mappings via Linear SVM.,” *Journal of Machine Learning Research*, vol. 11, pp.1471–1490, 2010.

Chapter 3. Improvement for Endoscopic Video Image by Endoscopic Image Data Set

3.1. Problems in endoscopic image data affect classification accuracy

In an endoscopic video image, the appearance is different from a still image since the distance between the lesion and the scope changes in each video frame. For an endoscopic video image, a robust system is required in appearance changes against a recognition of still image captured during video freezing. For example, in contrast to a still image, the following various events that affect the classification accuracy are observed in the actual endoscopic video image. Therefore, for the real-time computer-aided diagnosis of endoscope video images and, it is necessary to solve the following specific problems.

(1) Changing or defocusing of appearance of lesions owing to movement of the lesion in the depth direction: Figure 31 shows (a) an endoscopic video image and (b) a still image captured during freezing by a clinical doctor. Although the same lesion part is captured, microscopic defocus due to movement in the depth direction is shown in the video image. The appearance between (a) video image and (b) still image is different. In the endoscopic observation of the colon, the colon continually contracts and also the endoscope itself is not stopped completely. Because of this, the lesion moves in the depth direction. So, the endoscope is out of focus shown.

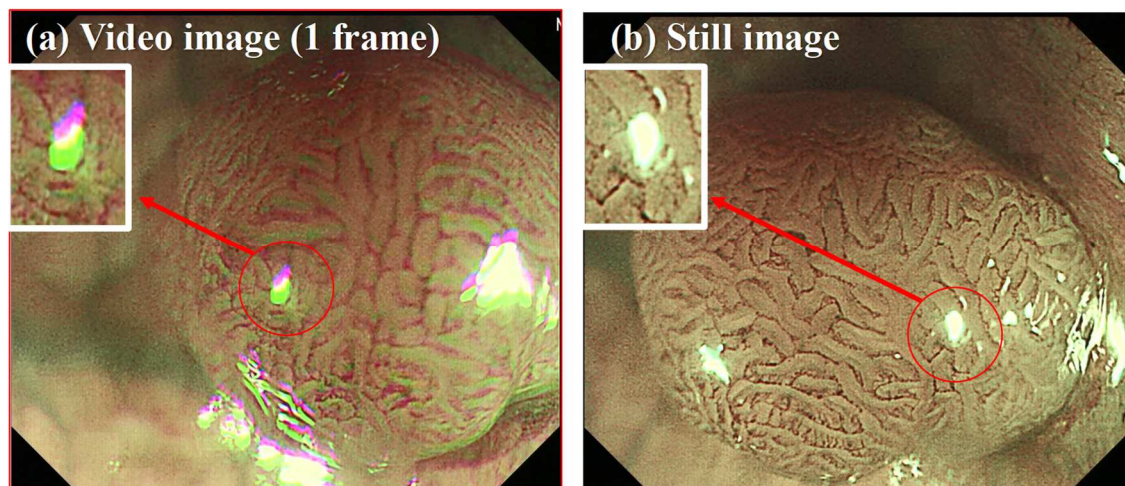


Figure 31. Differences of endoscopic video image (a) and still image (b).

(2) Unclear frames with blurring, color shift, reflection of light: Figure 32 shows examples of unclear frames with (a) blurring, (b) color shift, and (c) reflection of light. In the NBI magnified endoscopic system, light of blue (wavelength: 415 nm) and green (wavelength: 540 nm) is irradiated to the inner wall of the colon during observation by a clinical doctor. Light is irradiated from one light source. Thus, blue and green light is switched by an NBI filter [8]. A color shift occurs when this NBI filter is switched by a clinical doctor. Also, reflection of light occurs in any lesion part because light is irradiated always during observation. A clinical doctor does not consider color information of lesion for diagnosis. Therefore, the input image is gray scaled in the CAD system. So, edge patterns are emphasized which affect the classification result when gray scaled in frames with blur, color shift or reflection of light.

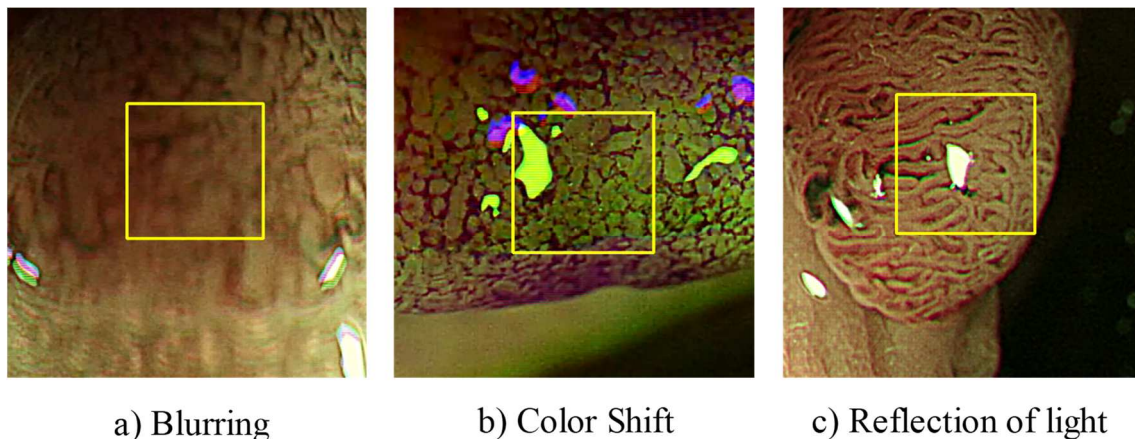


Figure 32. Unclear frames in endoscopic video image: (a) blurring, (b) color shift, (c) reflection of light.

(3) Different resolution and appearance between old and new endoscopic systems: According to the improvement of the endoscopic system, the resolution and appearance of old and new model endoscopic systems are different as shown in Fig.33, because a scope and an image processing technology are advanced. Thus, we have to confirm whether a classifier, which is learned by the endoscopic image data set from old model provided by a clinical doctor up to present, can be used to an endoscopic image from a different new model endoscopic system.

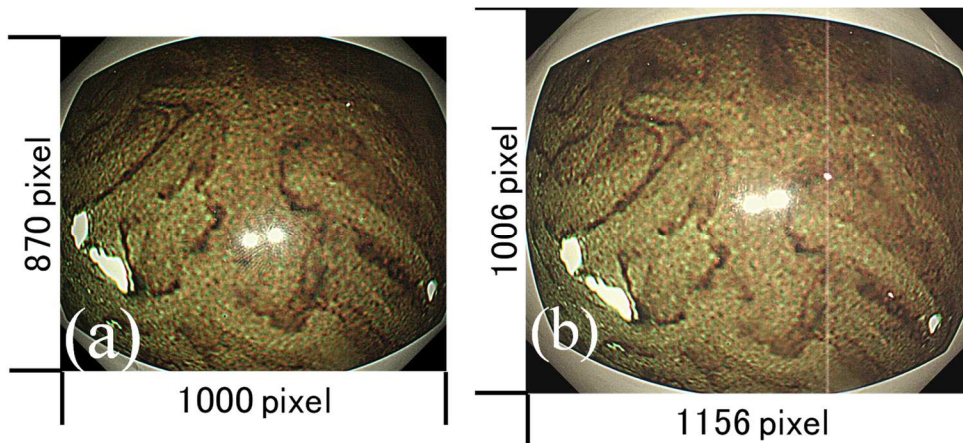


Figure 33. Differences from Endoscopic Systems. (a) System for training data set (system: Olympus CV-260, scope: Olympus CF-H260AZL/I [72]). (b) System for Test Video image (system: Olympus CV-290, scope: Olympus CF-HQ290ZL/I [73]).

3.2. Endoscopic image dataset from clinical doctors

For developing a robust CAD system with CNN feature extraction and SVM classification, a training data set considering appearances is prepared to solve problems in 3.1. The training data set in this thesis was collected at Hiroshima University Hospital with an old endoscopic system different from which used for testing video images. The data set is, which captured by the old endoscopic system, trimmed a clear part of lesions by the clinical doctor as shown in Fig.34.

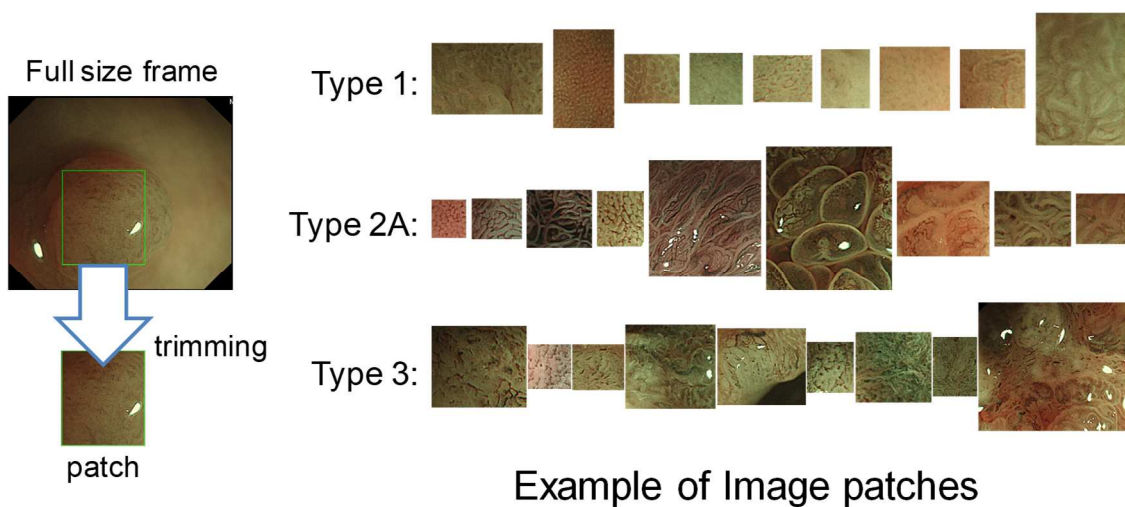


Figure 34. Example of image patches collected by a clinical doctor (Hiroshima University Hospital)

In NBI endoscopic observation, lesions moved in the depth direction are captured because of continually contracts of colon or fine movement of the endoscopic scope. In the learning by still images only from the old endoscopic system, it is probably that the appearance's change by the movement in the depth direction and the microscopic defocus affect the classification accuracy. Therefore, we decided to use data of multiple image sizes, which is considered the appearance's change in the depth direction of the image data set, as learning data for SVM. At first, we prepared 420 endoscopic still images obtained from the old endoscopic system for each pathological type which has trimmed by clearly capturing the lesion by a clinical doctor. As shown in Fig. 35, the height and width of the original still image are different. And, contour in the distribution graphs in Fig.35 indicate histogram of the height and width of original still images.

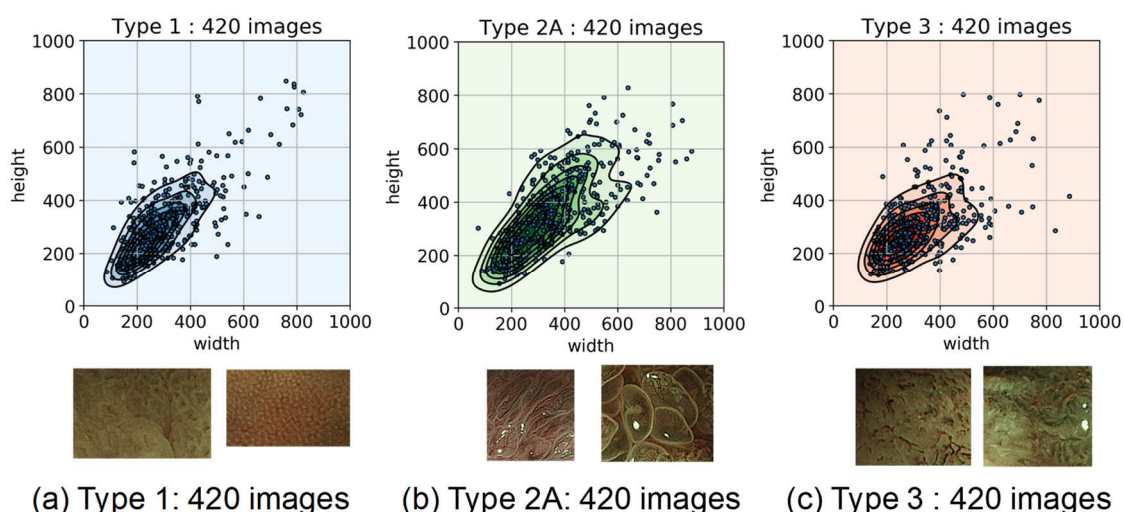


Figure 35. distribution of image size of each lesion type.

In the CAD system with CNN extraction and SVM classification, since the input image size of the pre-learned AlexNet used as the feature extractor is 227 x 227 pixels, it is necessary to resize the input image size to 227 x 227 pixels. The trimmed images are converted to grayscale images and resized to the input image size 227 x 227 pixels of the CNN feature extractor. At this time, image sizes of endoscopic image data used for SVM learning was trimmed with six sizes from 177 x 177 to 504 x 504 pixel for taking care of the appearance's change by the movement in the depth direction as shown in Table 4. By using multiple size of images as training data, for example, when

reducing from 504 x 504 pixel to 227 x 227 pixel, some information and pixels are lost, and when expanding from 177 x 177 pixel to 227 x 227 pixel, some pixels are interpolated and added. Therefore, we consider that it is assumed that the resized images represent changes of appearance due to the movement in the depth direction. Table 4 shows numbers of image data for each trimmed size. In Table 4, scan window size from the original still image is indicated by SW 177, SW 227, SW 277, SW 404, etc. Trimming step is indicated by Step30, Step40, etc.

Table 4. Training data set trimmed by a clinical doctor.

dataset	Type 1	Type 2A	Type 3
SW404_Step55_Grayscaled	713	1,199	447
SW454_Step40_Grayscaled	851	1,129	428
SW504_Step30_Grayscaled	958	932	433
SW177_Step200_Grayscaled	760	1,385	706
SW227_Step150_Grayscaled	705	1,392	663
SW277_Step100_Grayscaled	823	1,674	726

SW : Trimmed image size Step : Offset of trimmed image

3.3. Data Balancing

The size of the original endoscopic still image for each pathological type is different. So, the number of images for each pathological type differs. Especially, there are many Type 2 images, and the other hand there are few images of Type 3. For example, in SW404_Step55_Grayscaled, number of Type 2A and Type 3 images as cancer is 1,646, which is almost 2.3 times of number of Type 1 images, 713. Also, number of Type 2A images is 1,119, which is almost 2.7 times of number of Type 3 images, 447.

Generally, the imbalanced data set is known as one of major problems for real-world data [74-75]. If the data set is not evenly distributed in different classes or categories, then the data set is considered imbalanced. This is very common in real data sets. In such a data set, the class that has more data is defined as a majority class. On the other hands, the ones with much fewer data are called minority classes. Most classifiers are modeled by examining data statistics, and as a result, they are biased towards the majority class,

which can result in very poor classification accuracy in the minority class. In medical applications, one main concern of using data analysis for diagnosis is the imbalanced quality of majority and minority classes [74]. This problem is a particular interest in medical applications, where it is essential to correctly classify examples from an infrequent but important minority class. Analyzing data set with imbalanced class distribution is a basic challenge for learning classifiers [76-78].

To avoid problems of the imbalanced data set, some data balancing methods are proposed, and comparison studies are reported [79-80]. Data balancing methods are categorized into three groups.

(1) Under-sampling method: To balance the data set, the intrinsic samples in majority class are discarded. The simplest and yet most efficient method is Random Under-Sampling (RUS) which reduces randomly sample data in majority class [81].

(2) Over-sampling method: To balance the data set, new minority class data are created. Some approaches are proposed to create new minority class data such as SMOTE (Synthetic Minority Oversampling Technique) [82], ADSYN (Adaptive Synthetic sampling) [83] etc.

(3) Hybrid method: There are combination methods of under-sampling and over-sampling.

Therefore, by considering the balancing of the number of training images for each pathological type, we also created a data set to avoid training biased toward a specific type by random under-sampling method. "Balanced" in yellow rows of Table 5 is the prepared training data set based the smallest number of Type 3 (red character in Table 5) as a reference for eliminating the difference in the number of images for each pathological Type.

Then, the training images are gray scaled and resized by bilinear interpolation to the input image size of the CNN feature extractor to 227 x 227 pixel as the training data set for SVM classifiers.

Table 5. Training data set after balancing.

dataset		Type 1	Type 2A	Type 3
SW404_Step55_Grayscaled	Non-Balanced	713	1,199	447
	Balanced	447	447	447
SW454_Step40_Grayscaled	Non-Balanced	851	1,129	428
	Balanced	428	428	428
SW504_Step30_Grayscaled	Non-Balanced	958	932	433
	Balanced	433	433	433
SW177_Step200_Grayscaled	Non-Balanced	760	1,385	706
	Balanced	706	706	706
SW227_Step150_Grayscaled	Non-Balanced	705	1,392	663
	Balanced	663	663	663
SW277_Step100_Grayscaled	Non-Balanced	823	1,674	726
	Balanced	726	726	726

SW : Trimmed image size Step : Offset of trimmed image

3.4. Evaluation Result

The classification accuracy was evaluated by performing 10-fold Cross Validation (CV) ten times for comparing the classification accuracy of the proposed method based on the created Non-Balanced and Balanced training data set.

Cross-validation is a technique to evaluate predictive models by partitioning the original sample into a training set to train the model, and a test set to evaluate it. In 10-fold cross-validation, the original sample is randomly partitioned into 10 equal size subsamples. Of the 10 subsamples, a single subsample is retained as the validation data for testing the model, and the remaining 9 subsamples are used as training data. The cross-validation process is then repeated 10 times (the folds), with each of the 10 subsamples used exactly once as the validation data. The 10 results from the folds can then be averaged (or otherwise combined) to produce a single estimation. The advantage of this method is that all observations are used for both training and validation, and each observation is used for validation exactly once [84-86].

For evaluation metrics, we used *Recall*, *Precision*, and *F-measure* defined as equations (2), (3) and (4), respectively. The *Recall* shows correct recognition rate according to labels, and the *Precision* shows accuracy of recognition. In statistical analysis of binary classification, the *F-measure* is a measure of a test's accuracy. It is calculated from the *Recall* and *Precision* of the test, where

the *Precision* is the number of correctly identified positive results (TP (*True Positive*)) divided by the number of all positive results ($TP + FP$ (*False Positive*)), including those not identified correctly, and the *Recall* is the number of correctly identified positive results (TP) divided by the number of all samples that should have been identified as positive ($TP + FN$ (*False Negative*)). The *F-measure* represents the harmonic mean of the *Recall* and the *Precision* [87-89].

$$Recall(i) = \frac{TP(i)}{TP(i) + FN(i)} \quad (2)$$

$$Precision(i) = \frac{TP(i)}{TP(i) + FP(i)} \quad (3)$$

$$F - measure(i) = \frac{2 \times Recall(i) \times Precision(i)}{Recall(i) + Precision(i)} \quad (4)$$

i	Type 1, Type 2A, Type 3
$TP(i)$	<i>True Positive</i> , Number of images correctly identified i
$FP(i)$	<i>False Positive</i> , Number of images incorrectly identified i
$FN(i)$	<i>False Negative</i> , Number of images incorrectly identified not i

Figure 36 to Figure 39 show the results of 10-fold CV. The *pool5*, *fc6*, *fc7*, *fc8*, *prob* in figures are the results of SVM classifier learned by using the output of corresponding layer of AlexNet as the feature extractor shown in Fig.22. The black line indicates the number of dimensions of the feature quantity of each layer. Moreover, the results of method using D-SIFT feature extraction [29] described in Chapter 2 are shown.

Figure 36 and Figure 38 show the results of classification of Type 1 vs Type not 1 (non-cancer / cancer) by non-balanced dataset and balanced dataset, relatively. Regarding the classification of non-cancer and cancer, the CAD system with CNN extraction and SVM classification obtains a practically sufficient classification accuracy over 95% for Type 1 and Type not 1 by balancing data set.

Figure 37 and Figure 39 show the results of classification of Type 2A vs Type 3 by non-balanced dataset and balanced dataset, relatively. In Figure 39, the results of Type 2A is lower than Type 3. However, the proposed method by

balancing data set can achieve over 90% *F-measure* value for both Type 2A and Type 3. From these results, it is suggested that the classification accuracy of the proposed method is improved by balancing the number of training endoscopic data images for each pathological type.

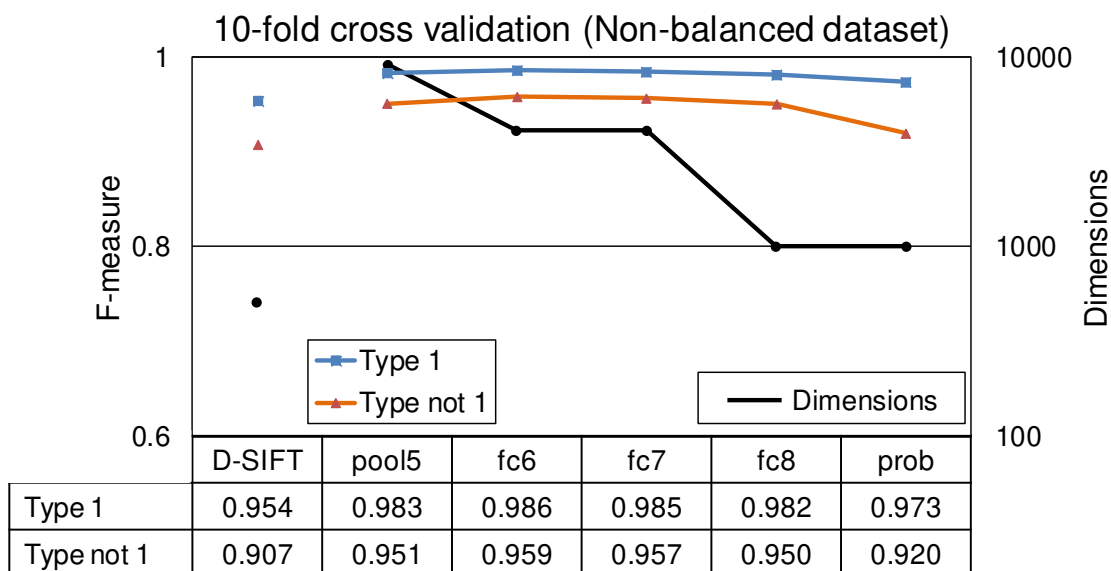


Figure 36. Result of 10-fold CV for Type 1 vs Type not 1. (Non-balanced training data set)

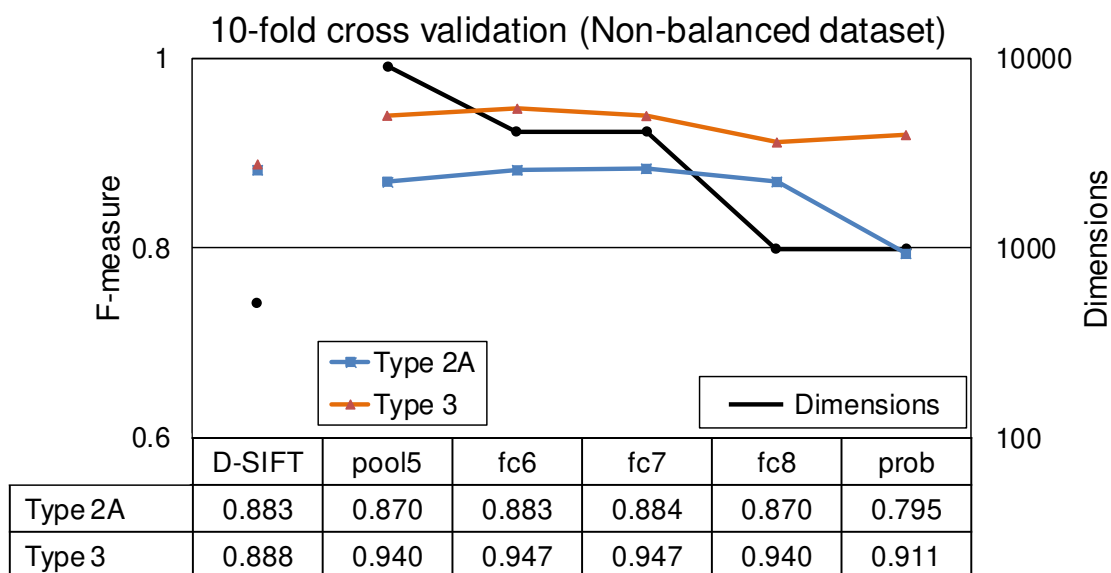


Figure 37. Result of 10-fold CV for Type 2 vs Type 3. (Non-balanced training data set)

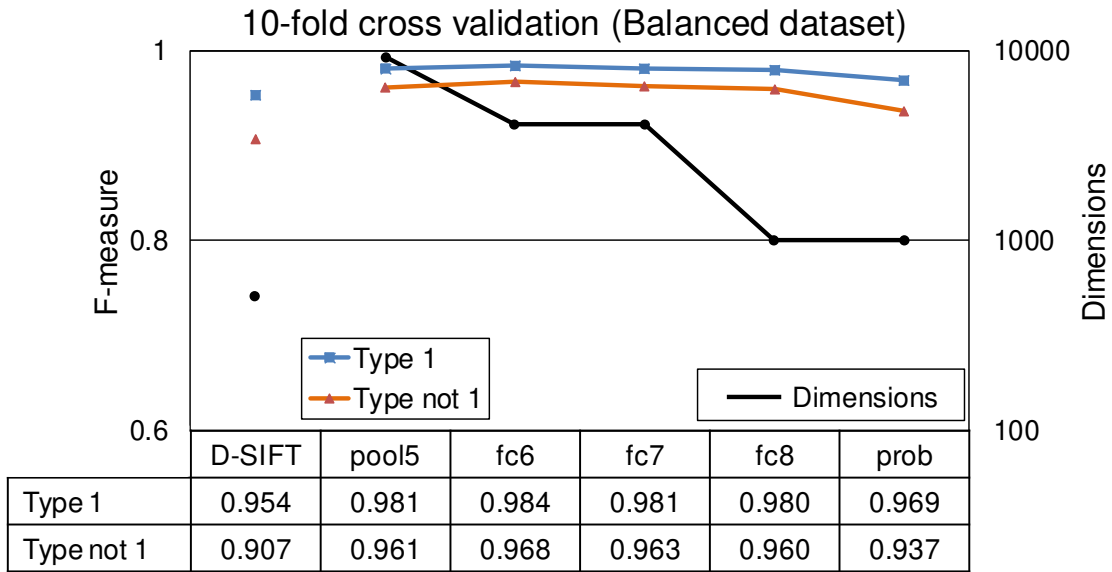


Figure 38. Result of 10-fold CV for Type 1 vs Type not 1.
(Balanced training data set)

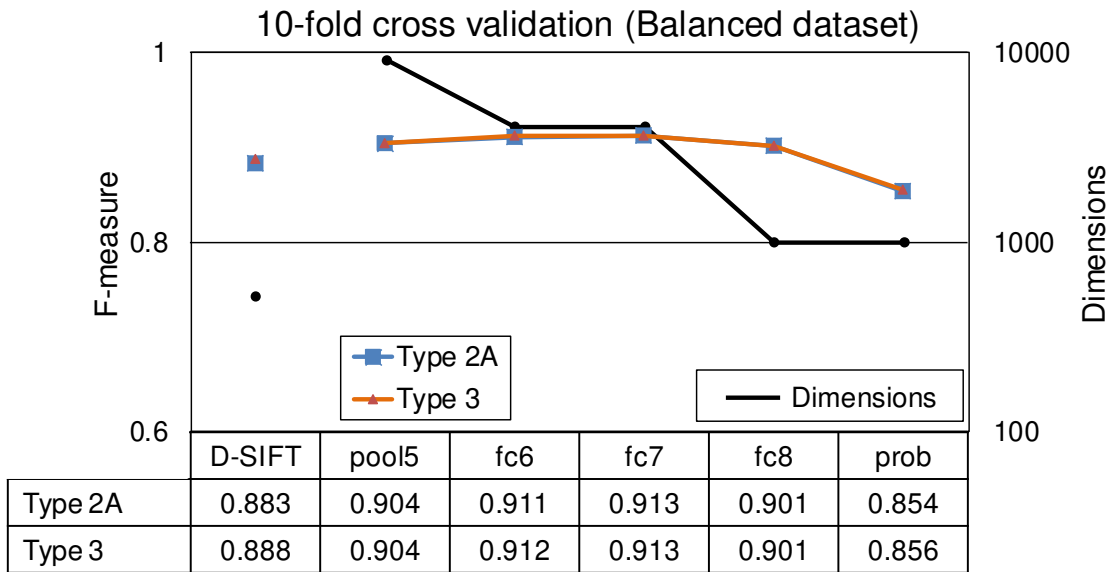


Figure 39. Result of 10-fold CV for Type 2 vs Type 3.
(Balanced training data set)

3.5. Evaluation of Endoscopic Video Image

We evaluated the classification result when inputting a video captured Type 2A lesions (300 frames) in the new endoscopic system into the developed CAD system with SVM classifiers trained by the endoscopic image dataset provided by the clinical doctor described in Chapter 3. Because it is a video

image, each frame also includes color shifts, blurs, light reflections, etc., and appearances also differ since the position of the endoscope and lesions change.

Figure 40 shows the 0, 150, and 300 frame images of the 300 frames video image used for validation. The green and red squares in the image indicate the 227 x 227 pixel (green) and 454 x 454 pixel (red) staging regions (ROI: Region of Interest) respectively that are input to the CNN function. In each frames, there are reflection of light and blurring.

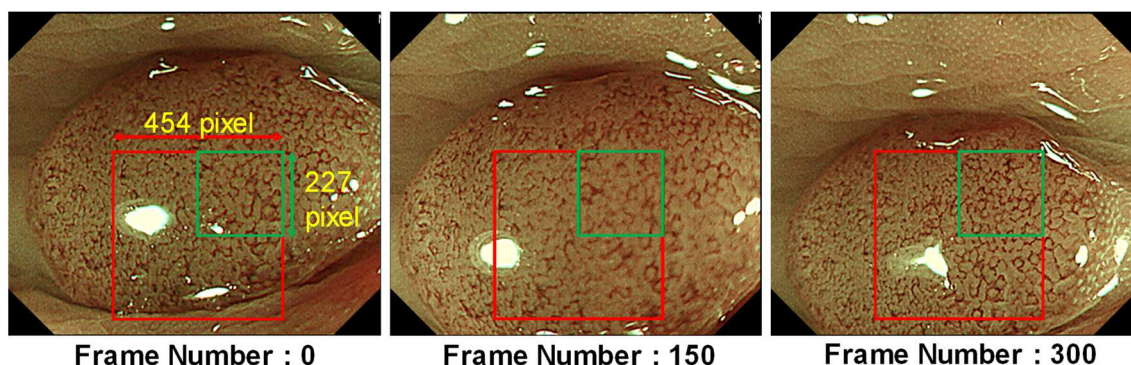


Figure 40. Example of frame images in Type 2A endoscopic video sequence.

Figure 41 shows the result of applying to the video image of 300 frames using the feature extracted from *pool5* layer when the green region of 227 x 227 pixels is input as the feature quantity of CNN. In graphs of Fig.41, the outputs from the four SVM classifiers which trained by only with 227 x 227 pixels (SW227 in Table 5) (*dataset_227_only(pool5)*) and trained by the data set balanced with the six kinds of sizes shown in Table 5 (Balanced) (*dataset_multisize(pool5)*) are plotted for each 10 frames of the Type 2A video image. The upper graph shows the output result from the SVM classifier of Type 1 vs Type not 1 (= Type 2A or Type 3), and in the Type 2A video image it is preferably 1.0 (= Type not 1), plotted on the upper side. The lower graph shows the output from the SVM classifier of Type 2A vs Type 3, and in the Type 2A video image it is better that it is plotted on the upper side close to 1.0 (= Type 2A). And, the light blue frame of the graph are frames in which a clinical doctor pushes the freeze button, and a clinical doctor takes a still image during these frames. The video sequence stops during the freezing, and it will be a part with no significant change in appearance although it is not the same frame image because of some noise or etc.

As the results in Fig.41, *dataset_multisize(pool5)* trained by multiple image

sizes shows better results than *dataset_227_only (pool5)* since the appearance of lesions changes from frame to frame. Figure 42 shows an entire image of frames #90 and #150 (upper side) and an original and gray scaled image of a green area of 227 x 227 pixels (lower side). Figure 46 shows the mean value and the standard deviation of the output of the SVM classifier for all 300 frames of the video image using the bar chart with error bars. From these results, it is confirmed that the Type 1 vs Type not 1 SVM classifier can classify Type not 1 with high accuracy regardless of image size of training data by using CNN feature extraction. Although the Type 2 vs Type 3 SVM classifier is generally difficult to classify, it is confirmed that the proposed method can classify robustly for all 300 frames since *dataset_multisize (pool5)* shows sufficient result compared to *dataset_227_only (pool5)*. The output results are changing even while the clinical doctor push the freeze button so as to stop video about 1-2 second and takes a still image (Fig. 41, Freeze frames)). It is considered that a video format encoding causes subtle pixel differences even though the frame appears to be stationary during the endoscope is frozen. Figure 43 shows pixel differences compared by XOR operation between frame #60 and frame #70. If frame #60 and frame #70 is same, image of differences #60 XOR #70 in Fig.43 shows all black.

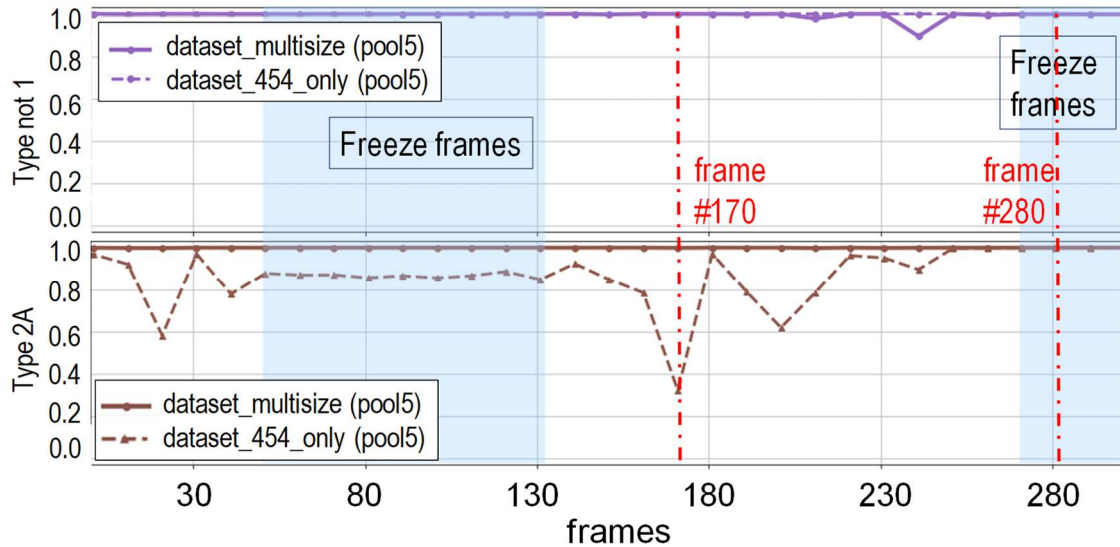


Figure 41. Comparison result for video image including the lesion Type 2A for 227 x 227 pixel region.

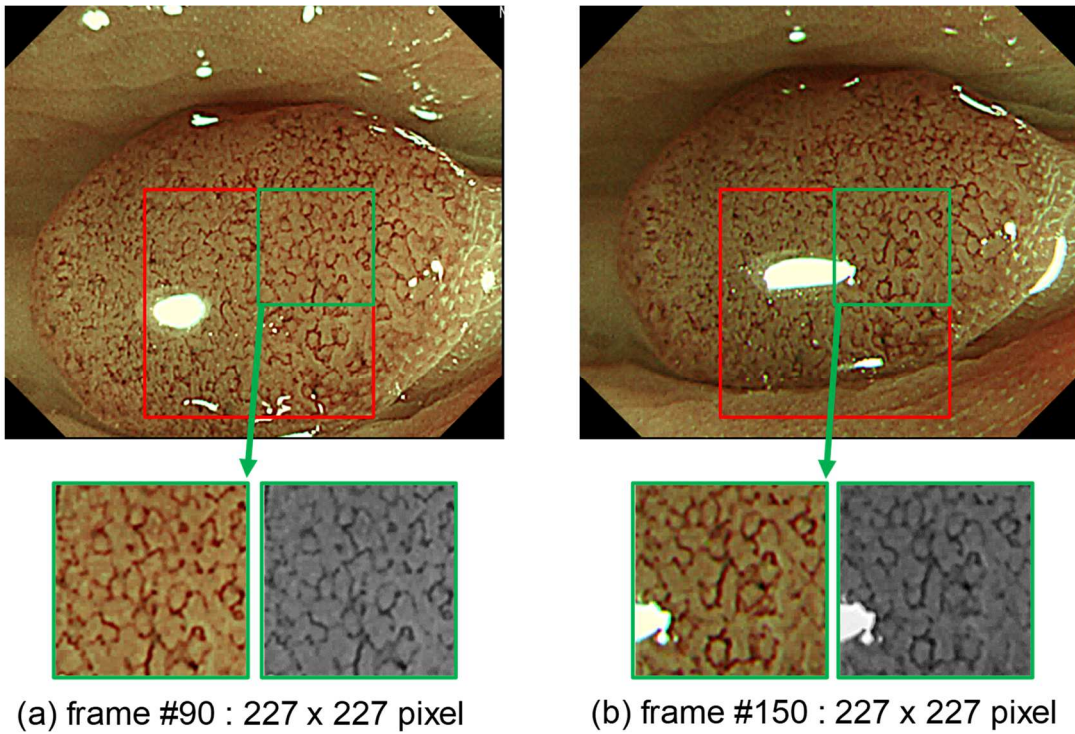


Figure 42. Example of frame images in Type 2A endoscopic video (227 x 227 pixel)

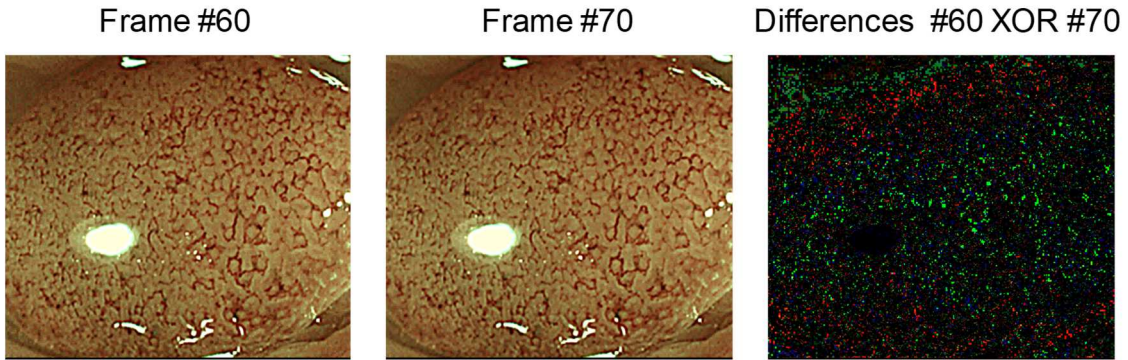


Figure 43. Pixel differences during Freeze Time.

In the same way, Figure 44 to Figure 46 show the results of the region (red) of 454 x 454 pixel of 300 frames as the feature quantity of CNN with applying *dataset_454_only (pool5)* and *dataset_multisize (pool5)*. In this evaluation, the region of 454 x 454 pixels is resized to 227 x 227 pixels by bilinear interpolation, since the pre-trained AlexNet is used as the CNN feature extractor. As shown by the red frame in Fig.45, since the area of 454 x 454 pixels is wider than 227 x 227 pixels, it includes color shift, blurring, reflection of light and so on. However, Figure 44 shows that the classification result of

dataset_454_only (pool5) in the freezing duration from 50 to 130 frames is improved and the result of Type 2A vs Type 3 SVM classification of *dataset_454_only (pool5)* in Fig.46 is improved. This result suggests that it is possible to classify robustly against noise with the feature extraction from the CNN by input image resizing.

From these results, we confirmed that it is possible not only to classify Type not 1 by the Type 1 vs Type not 1 SVM classifier with high accuracy, by using a SVM classifier of *dataset_multisize (pool5)* that trained by images of multiple image sizes, to classify clearly by suppressing the influence of moving images with the Type 2A vs Type 3 SVM classifier. And, it is considered that the proposed method is robust against appearance change by microscopic movement in the depth direction and so on by preparing training data set of different multiple sizes and further resizing to 227 x 227 pixels by trimming from the same training still image. From the above, it was confirmed that the proposed CAD system with high robustness to the endoscopic video image can be constructed compared to the conventional D-SIFT method [25]. In addition, these results suggest that it is possible to apply to an endoscopic video image taken by a new endoscopic system even if the SVM classifier is trained by the data set using a still image acquired by the old endoscopic system.

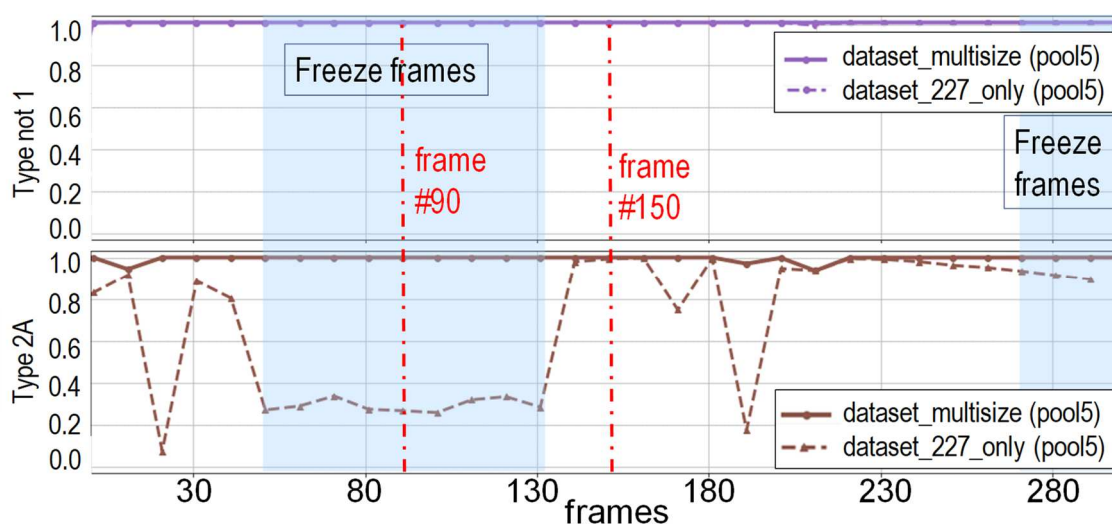


Figure 44. Comparison result for video image including the lesion Type 2A for 454 x 454 pixel region.

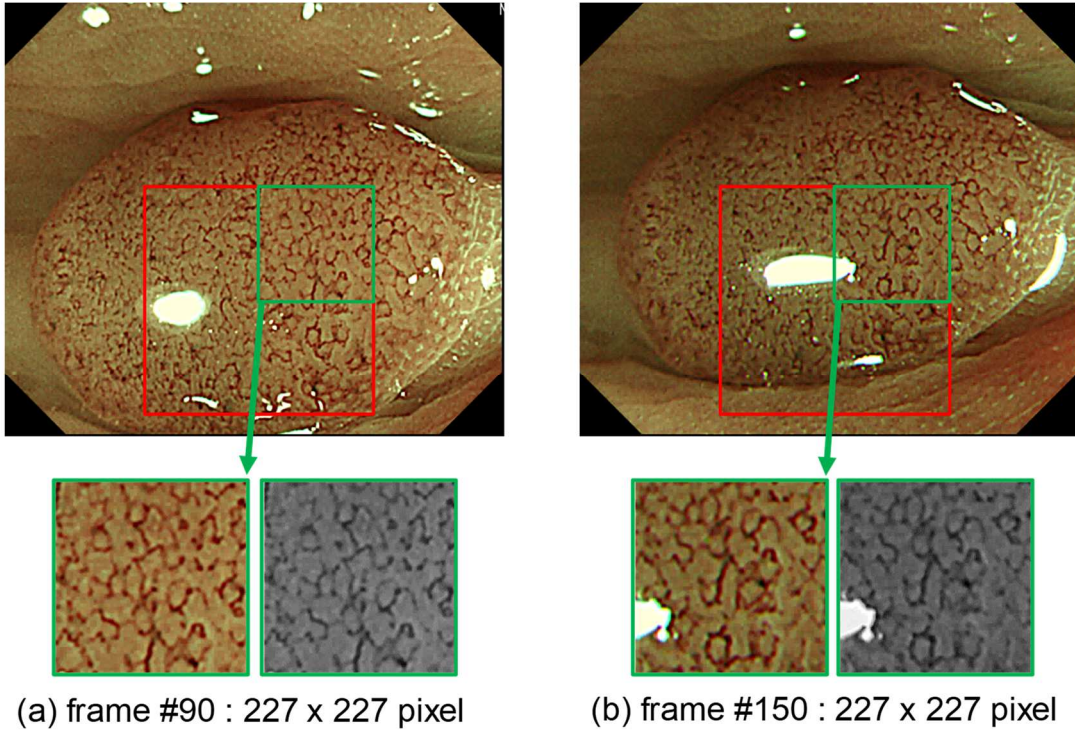


Figure 45. Example of frame images in Type 2A endoscopic video (454 x 454 pixel)

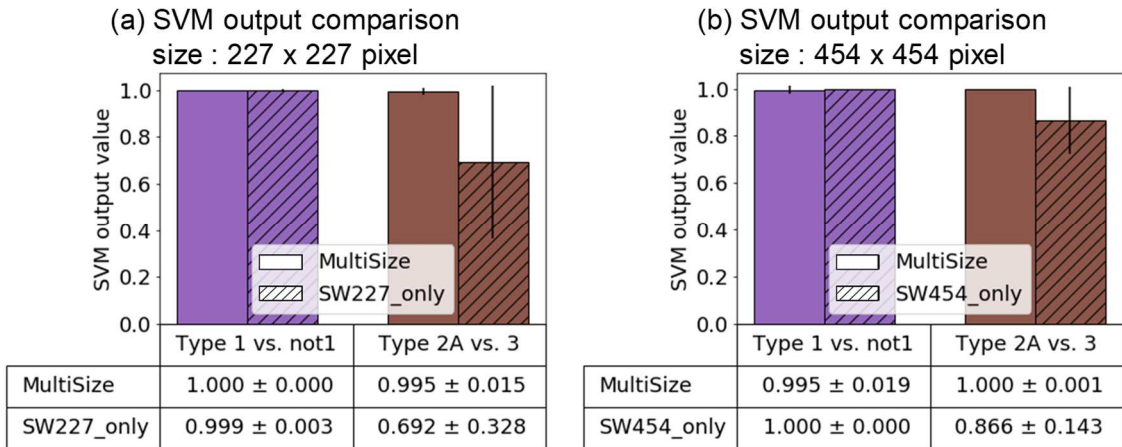


Figure 46. Classification results of Mean and Std. values of SVM output for 227 x 227 and 454 x 454 pixel regions.

By using multiple size of images as training data, for example, when reducing from 454 x 454x pixel to 227 x 227 pixel, some information and pixels are lost, and when expanding from 177 x 177 pixel to 227 x 227 pixel, some pixels are interpolated and added. As a result, it is assumed that the resized images are slightly unclear like blur. Since SVM classifiers are trained by

these images, it can be considered that the developed CAD system has become robust for endoscopic video sequence.

3.6. Conclusion

In this Chapter, I described the endoscopic data set and problems in endoscopic image data affect classification accuracy. Then, I described problems of imbalanced data set for SVM classification training. To avoid these problems, we prepared multiple sizes of endoscopic image, and, created balanced image data set by random under-sampling method. By preparing multiple sizes of endoscopic image and balancing the image data set, we confirmed classification accuracy for each pathological type achieved over 90%. This result suggested that the classification accuracy of the CAD system with CNN extraction and SVM classification is improved by balancing the number of training endoscopic data images for each pathological type. I confirmed a CAD system with quantitative staging classification for real-time video image achieved over 90% accuracy. I confirmed the average value and standard deviation of the output of SVM is improved by multiple size image data set for applying 300 frame video data.

References

- [72] Olympus Medical Systems Corp., “EVISLUCERA Colon videoscope Olympus PCF TYPE Q260AZI.”
- [73] Olympus Medical Systems Corp., “EVISLUCERA ELITE Colon videoscope Olympus CF-HQ290Z Series.”
- [74] M. Buda, A. Maki, and M.A. Mazurowski, “A systematic study of the class imbalance problem in convolutional neural networks.,” *Neural Networks*, vol.106, pp.249-259, 2018.
- [75] J. Diz, G. Marreiros, and A. Freitas, “Applying Data Mining Techniques to Improve Breast Cancer Diagnosis.,” *Journal of Medical Systems*, vol.40, pp.203, 2016.
- [76] S. Fotouhi, S. Asadi, and M.W. Kattan, “A comprehensive data level analysis for cancer diagnosis on imbalanced data.,” *Journal of Biomedical Informatics*, vol.90, pp.1-30, 2019.
- [77] N. Song and Q. Du, "Classification of Cervical Lesion Images Based on CNN and Transfer Learning.," *Proceedings of 9th International Conference on Electronics Information and Emergency Communication (ICEIEC2019)*, Beijing, China, pp.316-319, 2019.
- [78] X. Qin, F. M. Bui, and H. H. Nguyen, "Learning from an Imbalanced and Limited Dataset and an Application to Medical Imaging.," *Proceedings of 13th Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM 13)*, Victoria, Australia, pp.1-6, 2019.
- [79] G. Haixiang, Y. Li, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, “Learning from class-imbalanced data: Review of methods and applications.,” *Expert Systems with Applications*, vol.73, pp.220-239, 2016.
- [80] A. D. Pozzolo, O. Caelen, and G. Bontempi, "Comparison of balancing techniques for unbalanced datasets.," *Machine Learning Group*

Universit'e Libre de Bruxelles Belgium, pp.1, 2013.

- [81] M. A. Tahir, J. Kittler, K. Mikolajczyk, and F. Yan, "A Multiple Expert Approach to the Class Imbalance Problem Using Inverse Random under Sampling.," *Multiple Classifier Systems*, pp.82-91, 2009.
- [82] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique.," *Journal of artificial intelligence research*, vol.16, pp.321-357, 2002.
- [83] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," *Proceedings of International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, Hong Kong, China, pp.1322-1328, 2008.
- [84] A. M. David, "The Relationship between Variable Selection and Data Augmentation and a Method for Prediction.," *Technometrics*, vol.16, pp.125-127, 1974.
- [85] M. Stone, "Cross-Validatory Choice and Assessment of Statistical Predictions.," *Journal of the Royal Statistical Society: Series B (Methodological)*. vol.36, pp.111-147, 1974.
- [86] M. Stone, "An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion.," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol.39, pp.44-47, 1977.
- [87] N. Chinchor, "MUC-4 Evaluation Metrics.," *Proceedings of the Fourth Message Understanding Conference*, McLean, VA, USA, pp. 22-29, 1992.
- [88] C. J. van Rijsbergen, "Information Retrieval", Butterworth-Heinemann, London, 1979.
- [89] Y. Sasaki, "The truth of the F-measure.," *Teach Tutor Mater.*, pp.1-5, 2007.

Chapter 4. Proposed CAD System Implementation to Customizable DSP Core

4.1. Multiply and Accumulate Calculation in CNN

Convolutional Neural Networks (CNN) are composed of different layers such as convolutional layers (*conv*), normalization layers (*norm*), pooling layers (*pool*) and fully connected layers (*fc*). A convolution calculation is shown as equations (5) [90-91].

$$Y_{i,k,x,y} = \sum_{c=1}^C \sum_{v=1}^r \sum_{u=1}^r D_{i,c,x+u,y+v} G_{k,c,u,v} \quad (5)$$

A convolution calculation is multiplication and accumulation of corresponding elements of an input feature map and a kernel to generate a single output pixel. The kernel is then swept across the input feature map to generate a single channel of the output feature map. A convolutional layer takes an input feature map of N minibatch images, each with height, width and number of channels, H, W and C, respectively. A filter or kernel of size $r \times r$ pixels having C channels is also provided as input to the convolutional layer. K such kernel is applied to the input feature map to generate the output feature map. In this equation, a kernel map is denoted as $G_{k,c,u,v}$, a single tile of input feature map is denoted as $D_{i,c,x+u,y+v}$ and output pixel is denoted as $Y_{i,k,x,y}$. Where x and y are coordinates of the feature map tile, i is image number in the batch, u and v are iterators over kernel while c is the iterator over channels and k is kernel index.

Table 6 shows the configuration of convolutional layers in AlexNet. The total number of parameters can be calculated. For *conv1* layer, the input size is $227 \times 227 \times 3$, the kernel size is 11×11 , the stride is 4, and the output size is 55×55 . So, the *conv1* has 290400 ($55 \times 55 \times 96$) neurons and 364 ($11 \times 11 \times 3 + 1$) weights. Parameters are 105,705,600 (290400×364). Also, the number of multiplies and accumulate 100 million. In Table 7, the comparison of well-known models based on the number of weights and MACs are shown [92].

A convolutional layer computation consumes more than 90% of the total computation of CNNs. AlexNet has the least number of total MAC operations, however, it is over 700 million.

Table 6. Convolutional layer configuration in AlexNet [60].

	<i>conv1</i>	<i>conv2</i>	<i>conv3</i>	<i>conv4</i>	<i>conv5</i>
Input feature map size	227x227	27x27	13x13	13x13	13x13
Kernel Size	11x11	5x5	3x3	3x3	3x3
No. of kernels	96	256	384	384	256
No. of channels	3	48	256	192	192
Output feature map size	55x55	27x27	13x13	13x13	13x13
Stride	4	1	1	1	1

Table 7. The comparison of CNN models based on computational parameters and MACs [92]

	AlexNet[60]	VGG-16[93]	GoogLeNet[62]	ResNet 50[94]
Input Size	227 x 227	224 x 224	224 x 224	224 x 224
No. of conv layers	5	13	57	53
Depth of conv layers	5	13	21	49
Filter Size	3,5,11	3	1,3,5,7	1,3,7
No. of Channels	3-256	3-512	3-832	3-2048
No. of Filters	96-384	64-512	16-384	64-2048
Strides	1,4	1	1,2	1,2
No. of weights in conv layers	2.3M	14.7M	6.0M	23.5M
No. of MACs in conv layers	666M	15.3G	1.43G	3.86G
No. of fc layers	3	3	1	1
Filter Size	1,6	1,7	1	1
No. of Channels	256-4096	512-4096	1024	2048
No. of Filters	1000-4096	1000-4096	1000	1000
No. of weights in fc layers	58.6M	124M	1M	2M
No. of MACs in fc layers	58.6M	124M	1M	2M
Total Weights	61M	138M	7M	25.5M
Total MACs	724M	15.5G	1.43G	3.9G

4.2. Requirements for Hardware Platform of the CAD System Implementation

In general, there are some options for hardware implementation of CNN

feature extraction and SVM classification such as GPU, FPGA and DSP. Figure 47 shows a performance comparison based on multiply computation for CNN on these platforms. The X axis shows numbers of multiplies per second (execution performance) and the Y axis shows numbers of multiplies per watt (power efficiency). DSPs specified for CNN has high performance close to FPGAs and server GPUs. And, DSPs are most power efficient compare to FPGAs and server GPUs. As surveyed in References [95], there are many FPGA implementations of CNN architectures. The processing speed is increased by applying several optimization methods. However, the power consumption is not reduced so much. It also takes time for designing hardware in FPGA. Therefore, we selected a DSP for our prototype system implementation with low power dissipation.

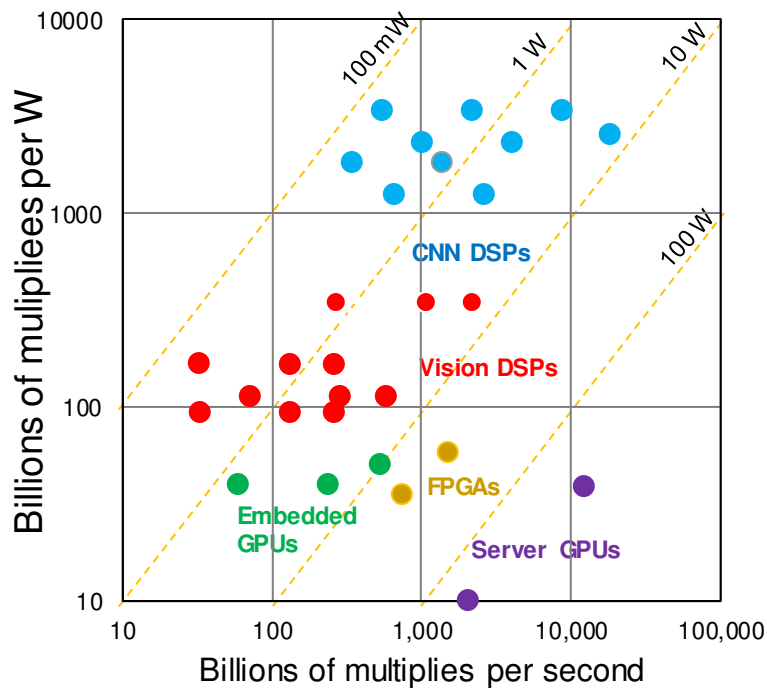


Figure 47. Throughput and power efficiency for CNN on several device platforms [96].

For efficient execution of CNN and SVM, simultaneous execution of multiple instructions by multiple numbers of MAC and VLIW/SIMD execution units are required. VLIW (Very Long Instruction Word) is one of instruction set

architectures designed to exploit instruction level parallelism as shown in Fig.48. Conventional processors mostly allow programs to specify instructions to execute in sequence only. On the other hand, a VLIW processor allows programs to explicitly specify instructions to execute in parallel [97-98]. SIMD (Single Instruction, Multiple Data) is a class of parallel computers in Flynn's taxonomy [99-100]. It describes computers with multiple processing elements that perform the same operation on multiple data simultaneously as shown in Fig.49. Such machines exploit data level parallelism [101].

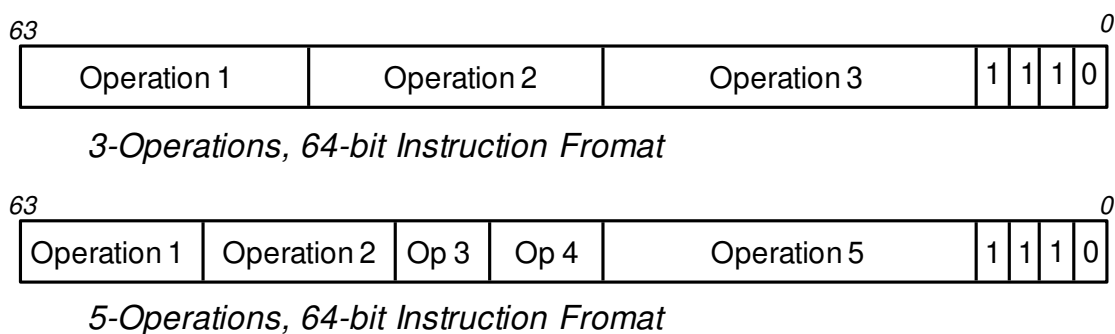


Figure 48. Example of Very Long Instruction Word (VLIW) [102].

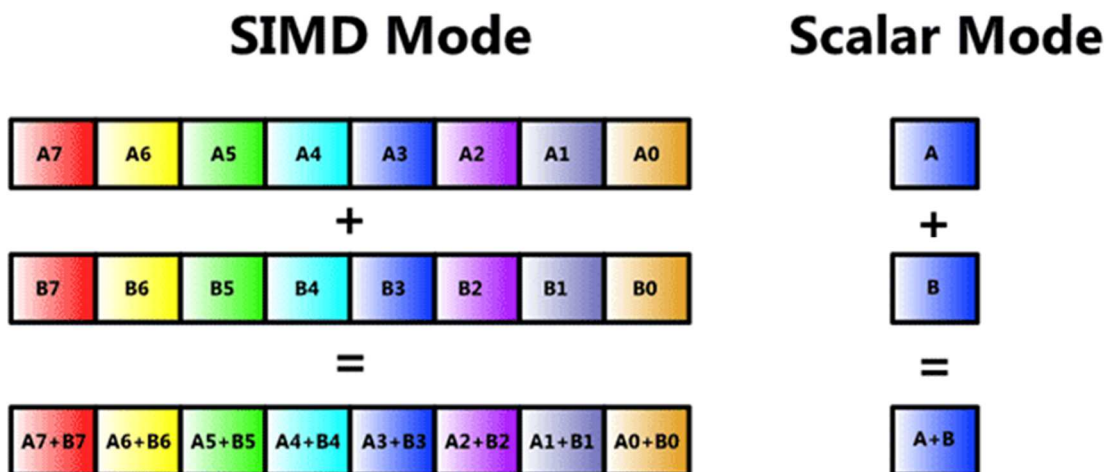


Figure 49. Overview of Single Instruction, Multiple Data (SIMD) [103].

Also, an efficient tiled image transfer by Scatter Gather Direct Memory Access [104] is required for video image processing in our target system. Direct Memory Access (DMA) is a technique for transferring blocks of data between system memory and peripherals without a processor having to be involved in each transfer. DMA not only offloads a system's processing

elements, but can transfer data at much higher rates than processor reads and writes. Scatter Gather DMA augments this technique by providing data transfers from one non-contiguous block of memory to another by means of a series of smaller contiguous-block transfers as shown in Fig.50. Scatter Gather DMA is also used in FPGAs for hardware wired neural network accelerator [105-106].

Image data is generally stored in memory with pixels, which is consisted of 8-bit red, green, blue data and 8-bit paddings, continuously as shown in Fig.50. Some operations use only one of RGB colors in a pixel. Therefore, accesses to memory are in-continuous.

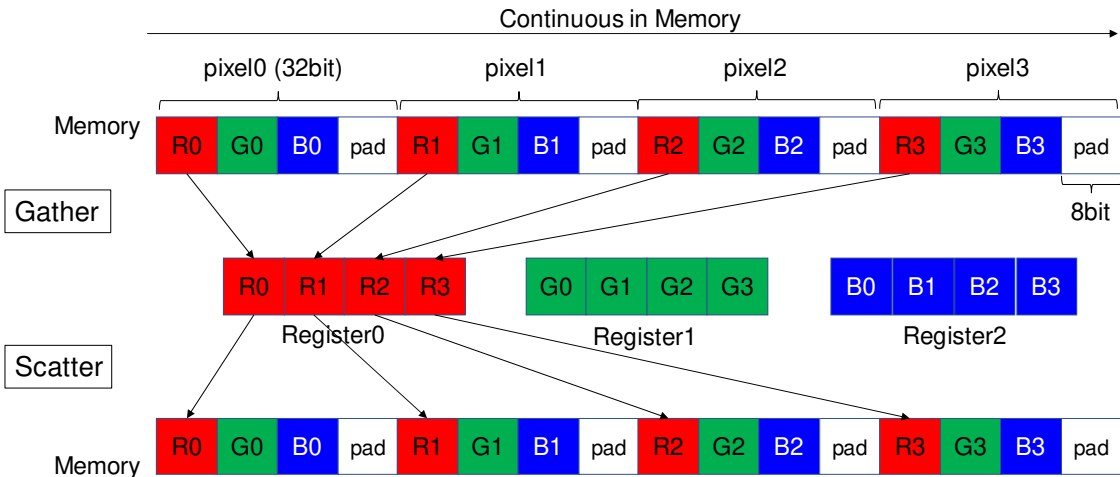


Figure 50. Scatter-Gather operation and Image Data.

4.3. Overview of Customizable DSP Core

In this thesis, we use the Cadence Tensilica® Vision P6 DSP core (VP6 core) [107], which is a customizable DSP core for embedded applications developed by Cadence Design Systems, as the target architecture to be implemented.

Figure 51 shows the architecture of VP6 core. The VP6 core has instruction sets specified for image processing and CNN processing, and has a 256 parallel multiply-accumulate (MAC) operation unit of 8-bit x 8-bit for high speed processing of convolution operation in CNN (Fig.51 (a)). The VP6 is a similar architecture of CNN DSPs in Fig.47. It adopts 5-slot VLIW for enabling execution of instructions in 5 parallels and 64-way 8-bit / 32-way 16-bit SIMD for enabling 64 parallel 8-bit data and 32 parallel 16-bit data operations. The integrated DMA (iDMA) allows the system to transfer high-resolution data directly into the local memory of the DSP, thus hiding the

data access latency associated with accessing data from an external DRAM (Fig.51 (b)). The SuperGather™ technology enhances memory interface on VP6 to efficiently read / write non-contiguous locations from local memory (Fig.51 (c)). There are built-in 512-bit dual load / store data memories for high performance video image processing.

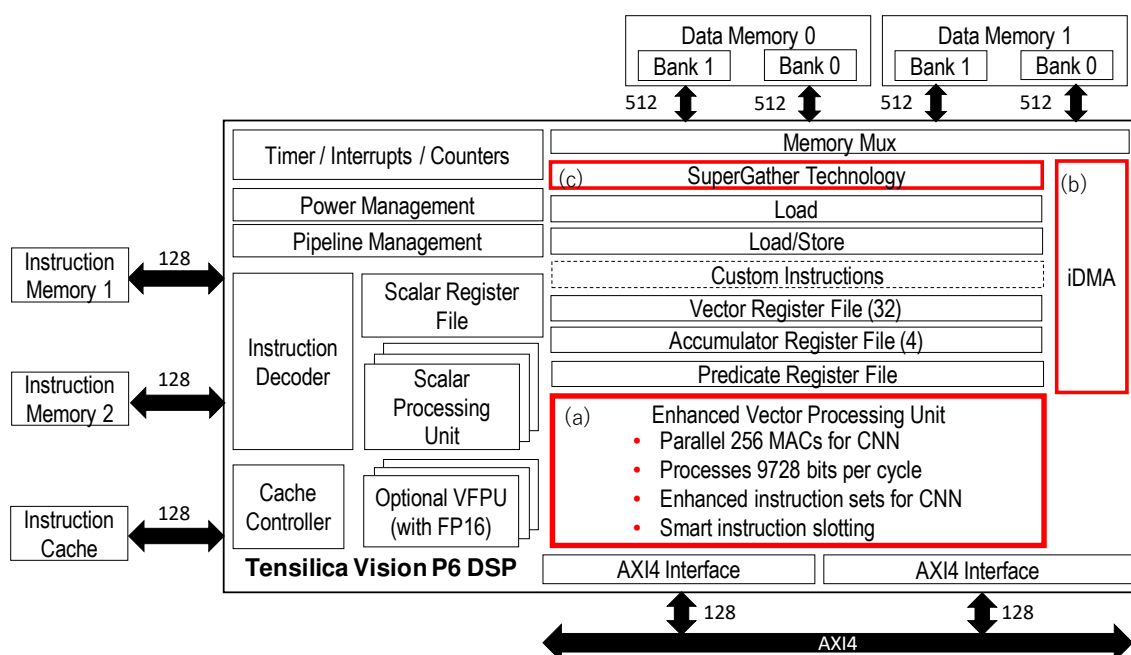


Figure 51. Overview of Vision P6 Core Architecture [107].

In the VP6 core, cache size, internal data memory size, internal instruction memory size and vector floating point unit are customizable. It also allows designers to extend and add instructions by Tensilica Instruction Extension (TIE) [108-110]. It achieves 3.3 times the peak performance compared to the previous version Vison P5 (VP5) core for 3-D convolution kernel because of four times number of MACs as shown in Table 8.

Table 8. Comparison of Vision P5 and P6 for 3-D Convolution kernel [88]

	Multiplier Utilization	Relative Performance
Vision P5	95%	1.0
Vision P6	80%	2.2

Input: 14x14x64 (8-bit); conv: 5x5x64, stride 1; output: 10x10x64 (8-bit)

VP6 also utilizes multiplier 57% for AlexNet [96]. Table 9 shows comparison

results of utilization of MAC resources for AlexNet to other GPUs reported in [111]. VP6 has higher utilization than other GPUs.

Table 9. Utilization Comparison to GPUs [111]

	Pascal Titan X[112]	GTX1080[113]	GTX1080Ti[113]	Vision P6[107]
Technology(nm)	16	16	16	16
Precision	32-bit float	32-bit float	32-bit float	8-bit fixed-point
Frequency(GHz)	1.41	1.6	1.5	1
Utilization	0.45	0.37	0.51	0.57

As a customizable DSP, CEVA XM6 DSP core (XM6) is also known [114]. The XM6 has 128 parallel multiply-accumulate (MAC) operation unit of 16-bit x 16-bit. For CNN, an external hardware accelerator NeuPro-S AI Processor with 512 parallel multiply-accumulate (MAC) operation unit of 16-bit x 16-bit is prepared as co-processor of the XM6. Table 10 shows the comparison result of AlexNet performance of the VP6 and XM6. The VP6 has fewer parallel multiply-accumulate (MAC) operation unit without necessary of CNN accelerate co-processor, however, throughput of AlexNet is almost same as the XM6 with NeuPro-S [115].

Table 10. AlexNet Performance Comparison of VP6 and XM6 [96,115].

	Vision P6[107]	Ceva XM6[114]
Frequency (GHz)	1.1	1.5
Multipliers	256	128+512
AlexNet Throughput / GHz (fps)	192	196

We use the pre-trained AlexNet for the Vison P6 DSP, which is reported by G. Efland et al, in [96] and optimized the bit width of the coefficient parameter to 8-bit for utilizing 256 parallel MAC of 8-bit x 8-bit. As a result, the amount of memory used in the embedded system is reduced and the processing speed is improved. It is known that high recognition accuracy is maintained even if the coefficient of AlexNet is reduced to 8-bit [116-117]. Table 11 shows comparison results of the memory usage and the top 1 and top 5 recognition accuracy of the original 64-bit AlexNet, and 8-bit optimized AlexNet, respectively. The recognition accuracy of 8-bit optimized AlexNet is 1.09% and 0.66% of the error compared to the top 1 accuracy and top 5 accuracy of

the original 64-bit AlexNet, respectively.

We compared classification accuracy (True-Positive) by using these pre-trained AlexNet with ImageNet database of the general object recognition contest as the feature extraction to SVM endoscopic image classification as shown in Fig.52. From these results, we consider it is possible to construct the CAD system by reducing the memory size of 75% and keeping the classification accuracy 90% using the SVM classifier relearning for the feature extractor with 8-bit optimized AlexNet.

Table 11. Memory size and Accuracy of the 8-bit Optimized AlexNet with ImageNet.

	Memory(Mbyte)	Top 1 Accuracy(%)	Top 5 Accuracy(%)
AlexNet(64-bit Original)	243.9	57.10	80.20
AlexNet(8-bit Optimized)	61.0	56.01	79.54

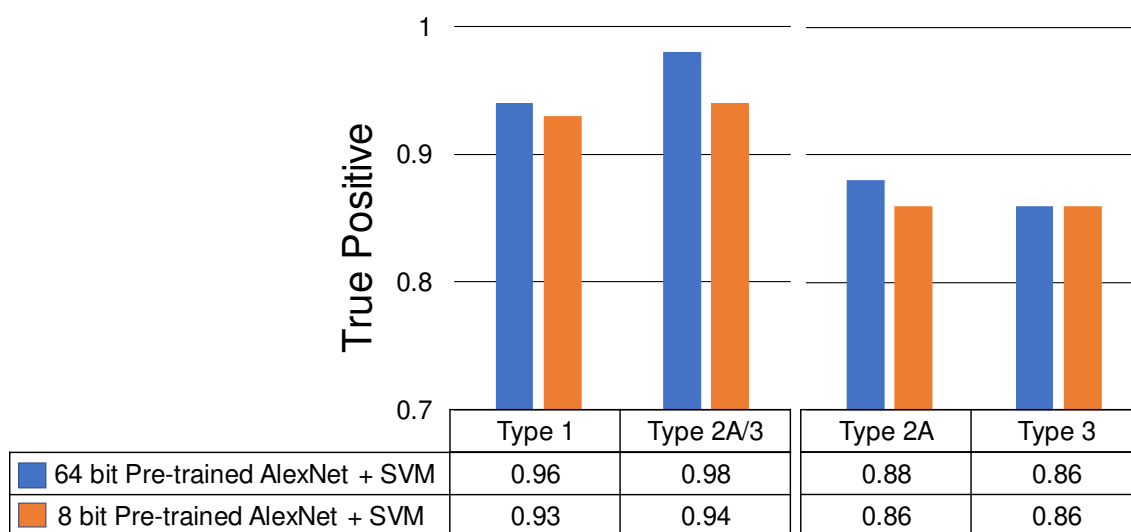


Figure 52. SVM Classification Accuracy (True-Positive) with feature extraction for 64-bit and 8-bit AlexNets which are pre-trained by ImageNet database.

4.4. Hardware Design and Processing Flow

For developing the proposed CAD system, we decided to use FPGA based prototyping system for enabling rapid implementation of HW and SW both, and, real-time validation of the CAD system. The prototyping system shown in Fig.53 is made up of two parts, a host computer which stores the endoscopic data and displays classification results and the Cadence rapid prototyping

platform Protium® S1 [118] in which the VP6 core and peripherals are installed. In the proposed CAD system, the main processing units such as the CNN (AlexNet) feature extraction and the SVM classification are executed on the VP6 core.

One Xilinx Virtex Ultrascale XU 440 FPGA [119] is installed in the Protium S1 as shown in Fig.54. Approximate total design capacity is up to 25M gates. The VP6 core and peripheral circuits are implemented on this FPGA. It is possible to observe waveform of arbitrary signal in the FPGA, and, the on-the-fly hardware debugging functionality is provided in the Protium S1 as shown in Fig.55. The software debugger is connected via JTAG to the VP6 core installed in the Protium S1. The hardware waveform monitor, and the software debugger are synchronized, so HW-SW co-debugging is possible.

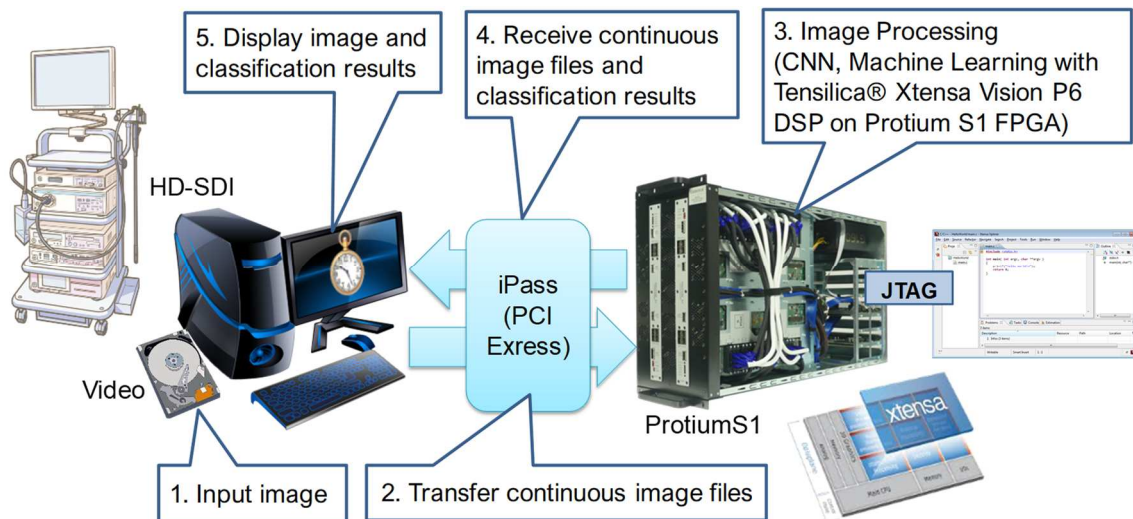


Figure 53. The Developed Prototyping CAD System Overview.

Xilinx UltraScale U440 [111]
Devices with heat sink.

For Host Machine,
X8 PCIe Ipass,USB,Ethernet

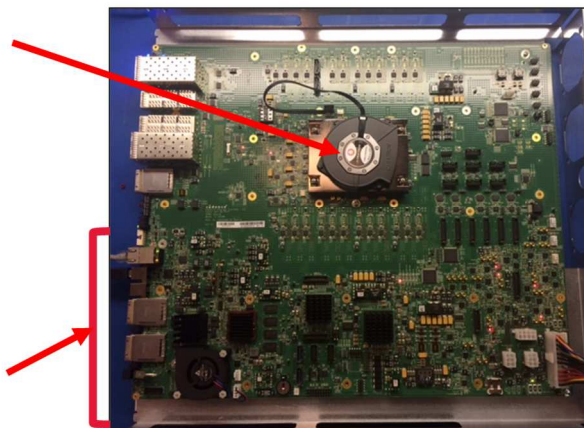


Figure 54. Overview of Protium S1 Prototyping System Board [118].



Figure 55. Hardware and Software co-debugging feature in Protium S1 [118].

Figure 56 shows a block diagram of the developed system including the VP6 core and peripherals. In this system, some memories such as two blocks of 1 Mbyte internal RAM (Fig.56 (g) (h)) for image buffering, 1 Mbyte internal ROM (Fig.56 (i)) for system ROM and 64 Mbyte work memory area for coefficients of CNN on external DRAM (Fig.56 (j)) are implemented. Furthermore, we customized VP6's configuration as follows: (1) 48 Kbyte instruction cache (Fig.56 (d)) and (2) 256 Kbyte x 2 blocks of built-in data memory (Fig.56 (b) (c)) are configured with VP6 core (Fig.56 (a)). (3) Optional VFPU (Vector Floating Point Unit) is not implemented because we use 8-bit fixed point data type in pre-trained AlexNet instead of 32-bit floating point data type in the proposed CAD system and there is the gate size limitation of FPGA. We measured the size of the required internal memory for image data and coefficient data of CNN and customized the internal memory size. Direct Memory Access (DMA) Controller [120] (Fig.56 (e)) is also implemented for high speed data transfer between the host PC and the Protium S1 for real-time processing. In this prototype system, the configuration of processing unit including the VP6 core can be debugged and optimized by cycle accurate co-simulation of software / hardware and emulation with actual machine operation.

Figure 57 shows the flowchart of the developed prototype system. When the image processing started, a frame data is read as an image data from the endoscopic video sequence (Fig.57 (1)). Next, any region of 227×227 pixels as the input data to the CNN is cropped from one frame of the endoscopic video as a classification region (Fig.57 (2)). If the classification region size is other than 227×227 pixels, and in that case, the classification region is resized to 227×227 pixels. The cropped image data of classification region is stored to

the image buffering memory (Internal RAM # 0) shared between the host PC and the VP6 (Fig.57 (3)). In the VP6, CNN feature extraction and SVM classification are executed to obtain the classification result for the classification region (Fig.57 (4)). The classification result is stored to the image buffering memory (Internal RAM # 1) shared between the host PC and the VP6 (Fig.57 (5)). Finally, the host PC reads the classification result and displays the classification region and the result together with the input frame image (Fig.57 (6)). The above process is repeated until the endoscopic video ends.

Data transfer between the host PC and the VP6 is performed via PCI (Peripheral Component Interconnect) Express connected by the iPass connector as shown in Fig.53 and Fig.56. In the developed system, a single unified bus by AXI (Advanced eXtensible Interface) (Fig.56 (f)) is used and all masters and slaves are connected to this single AXI bus. Data transfer in the developed system is performed by the this single AXI bus. The set of weight coefficients as learning data of the CNN and the model file data as learning data of the SVM are stored to the internal RAM, or to the work memory on the external DRAM as necessary. Therefore, we use VP6's features, Scatter Gather mechanism, and specific libraries for efficient data transfer.

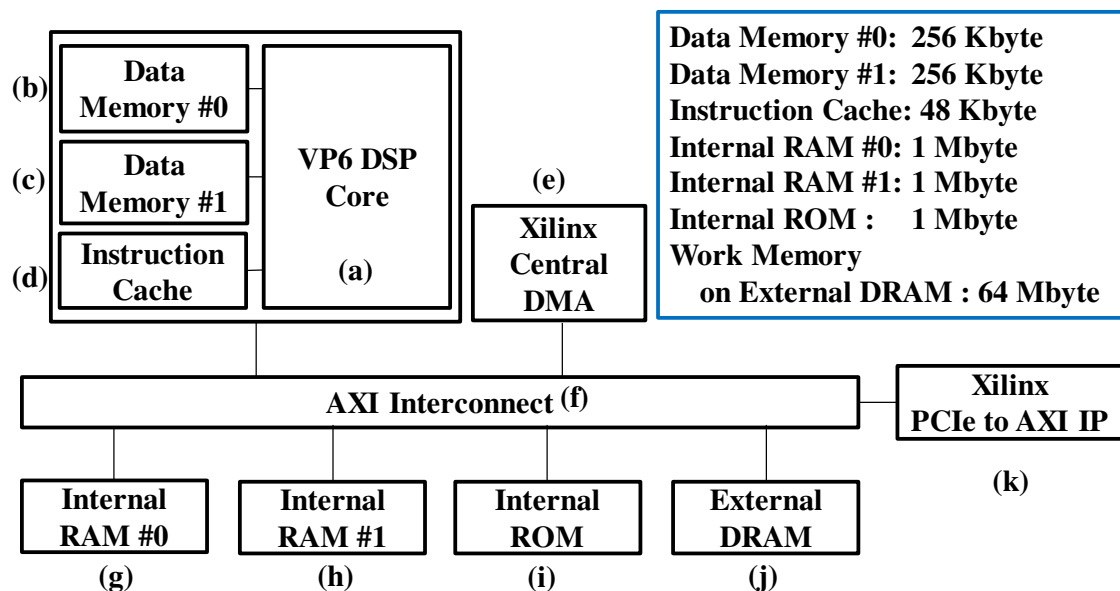


Figure 56. Block Diagram of the Implemented System.

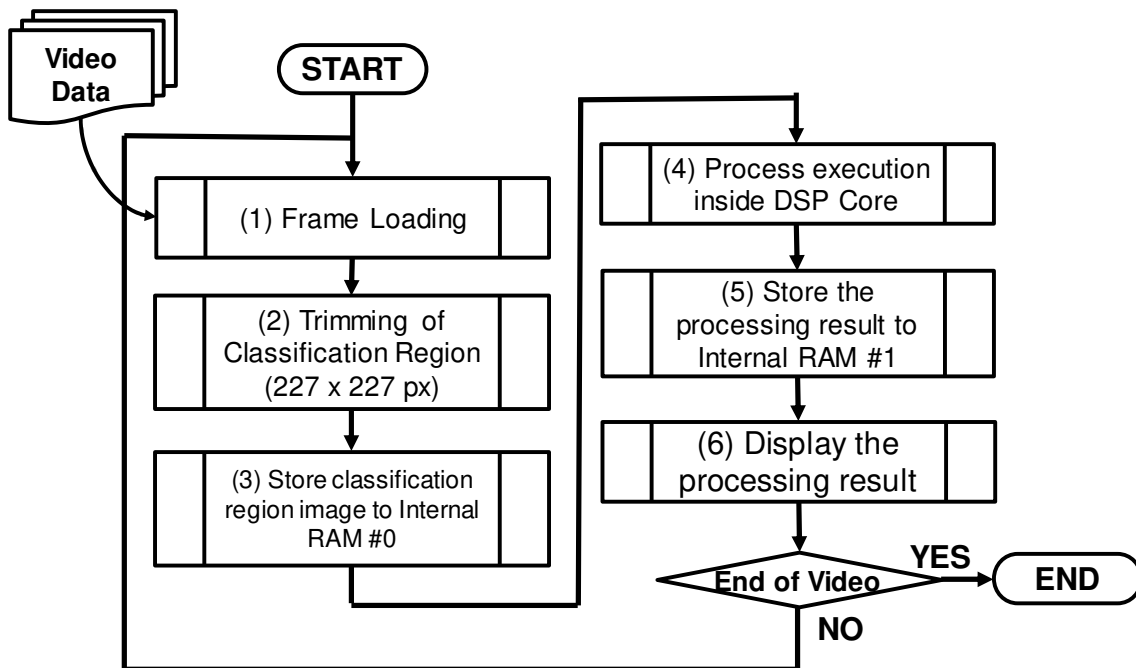


Figure 57. Flowchart of the Developed CAD Prototype System.

4.5. Bottleneck Analysis

For developing the real-time CAD system for endoscopic video images, we decided the optimization policy by profiling and analyzing bottleneck of still image based on processing cycles since processing of one frame in the video image is equivalent to the still image. For profiling of still image, we use the cycle accurate profiling functionality of Xtensa Xplorer [121], which is a software development toolkit for Tensilica Vision DSP family in the above simulation / emulation environment. Xtensa Xplorer generates profiling data such as instruction execution count, total cycles (Fig.58), subroutine calls, pipeline analysis (Fig.59), cache performance etc. It is possible to analyze the number of processing cycles for each executed function and estimate the total number of processing cycles with cycle accurate and pipeline modeled ISS (Instruction Set Simulator). Also, both local and system memory is modeled with programmable latencies specified for different transaction types, allowing an accurate system simulation for analyzing performance. Therefore, cache miss latencies, interlock or branch miss delay are estimated as shown in Fig.58.

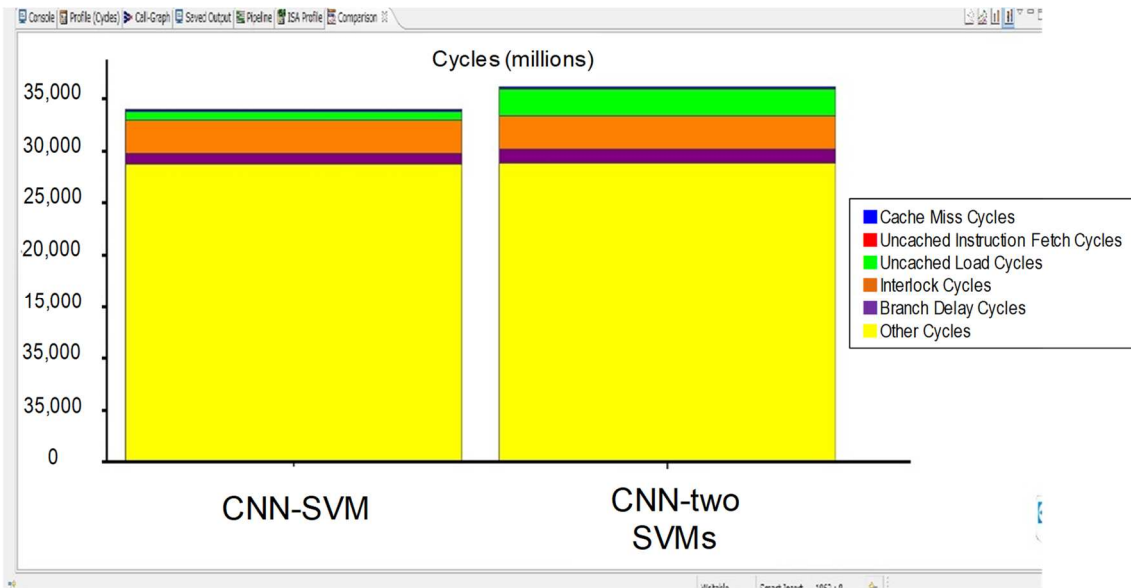


Figure 58. Cycle Analysis View of Xtensa Xplorer [121].



Figure 59. Pipeline Analysis View of Xtensa Xplorer [121].

Figure 60 shows the profiling result based on the number of processing cycles per one still image. The total number of processing cycles per image is almost 16.6 M cycles. When the clock frequency of the developed system is assumed 200 MHz, the latency is 83 ms. This latency corresponds to a delay of almost 2.5 frames when the frame rate of the endoscopic video image is 30 fps. The input image preprocessing shown as *image_preprocess* in Fig.60 is the maximum number of processing cycles (38%). In this image pre-processing, first an image data, which is transferred from the host PC, is separated into 3 color channel components (R, G, B), then these color channel components are stored into the predetermined place in the internal RAM #0, and finally the average color value (mean image) of each color channel of the input image is calculated.

And, the 2nd maximum of the total number of processing cycles is the waiting cycles to start *fc6* of Full Connect layer shown as *dma_wait_fc6* (12%) in Fig.60. As a result of further analysis, this wait time is not for the completion of the previous stage process. This wait time is the DMA waiting time for reading from the external DRAM of the weight coefficient of *fc6* as the learned AlexNet and storing it into the Data Memory. The size of the weight coefficient of *fc6* is 36 Mbyte. At the fully connected layer, the weight coefficients of all the nodes are required, since the calculation is performed with all the nodes of the preceding layer as inputs.

From this profiling results of the still image processing, if the input image preprocessing can be optimized and the intermediate data from the preceding layers before fully connected layer can be used as feature extractor, it is expected that the total number of processing cycles will be reduced by 60%.

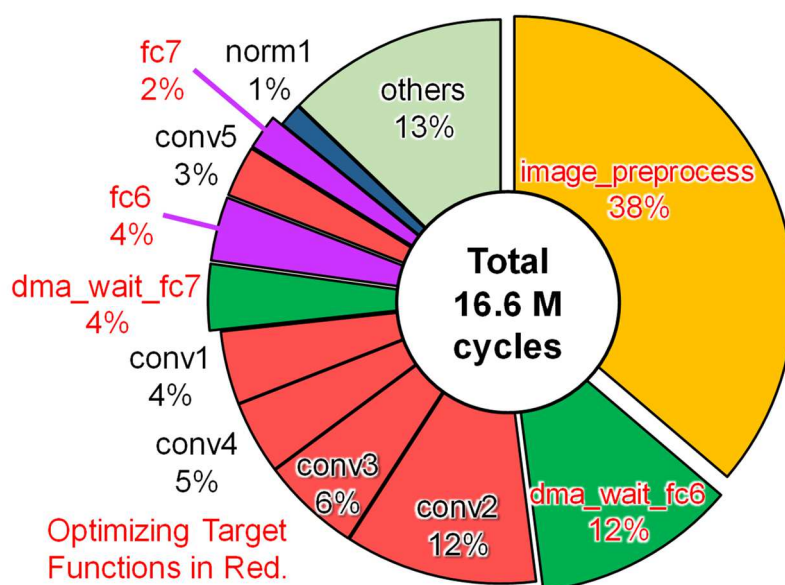


Figure 60. Profiling results of processing cycles per one still image.

4.6. Optimization

From the profiling results of the still image processing cycles in the previous section, we recognize that the pre-processing for the input image and the waiting time to read the weight coefficients of the fully connected layer are the main bottlenecks. Thus, we decided to optimize the video image processing by reducing the number of these two processing cycles.

4.6.1. Optimization of the pre-processing of the input image

For pre-processing of the input image, the specific instructions of VP6 DSP core are applied. The VP6 has 64-way/32-way SIMD instruction sets specified for vector operations, and specific libraries for efficient loading and storing to memories. These features are key for optimization.

The VP6's compiler has the auto-vectorization feature to vectorize operations and improve parallel execution efficiency in VLIW. Automatic vectorization is one of optimization technique. Computer languages and programs is designed to execute in sequence. So, many optimizing compilers perform automatic vectorization which converts computer programs written in C language from a scalar implementation, which processes a single pair of operands at a time, to a vector implementation, which processes one operation on multiple pairs of operands at once. where parts of sequential programs are transformed into parallel operations.

Our research group reported the results of optimization of the SVM classification by using this feature [57]. Figure 61 shows comparison results of the number of processing cycles of SVM classification when 908 endoscopic images are processed before and after optimization. We reduced the number of processing cycles to 1/30. Therefore, we implemented 2-SVM classifiers sequentially, Type 1 vs Type not 1 (= Type 2A/ Type 3) classification and Type 2A vs Type 3 classification, for 3 types classification. From this result, we considered that the number of processing cycles of pre-processing input images can be effectively reduced effectively by auto-vectorization feature.

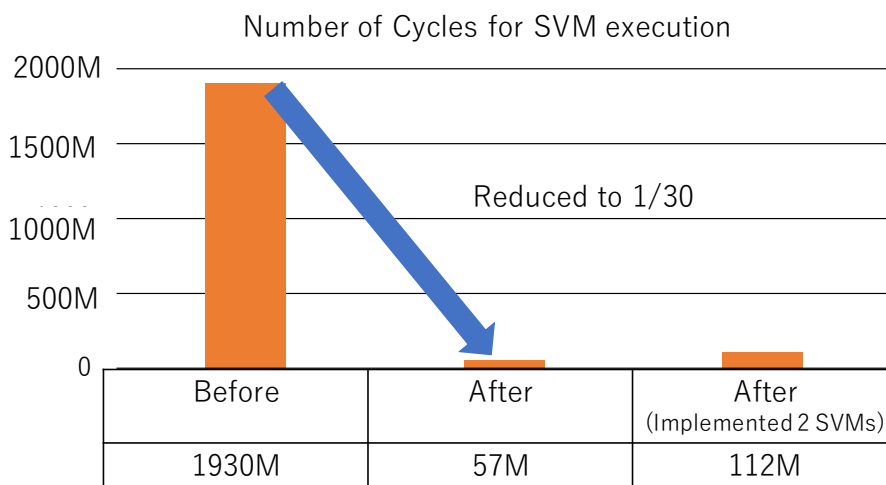


Figure 61. Comparison results of SVM classification processing cycles before and after optimization [57].

Figure 62 shows optimization flow using auto-vectorization feature of Xtensa Xplorer. Convolution operation is coded by the nested loops which have possibility to be vectorized as shown in Fig.63.

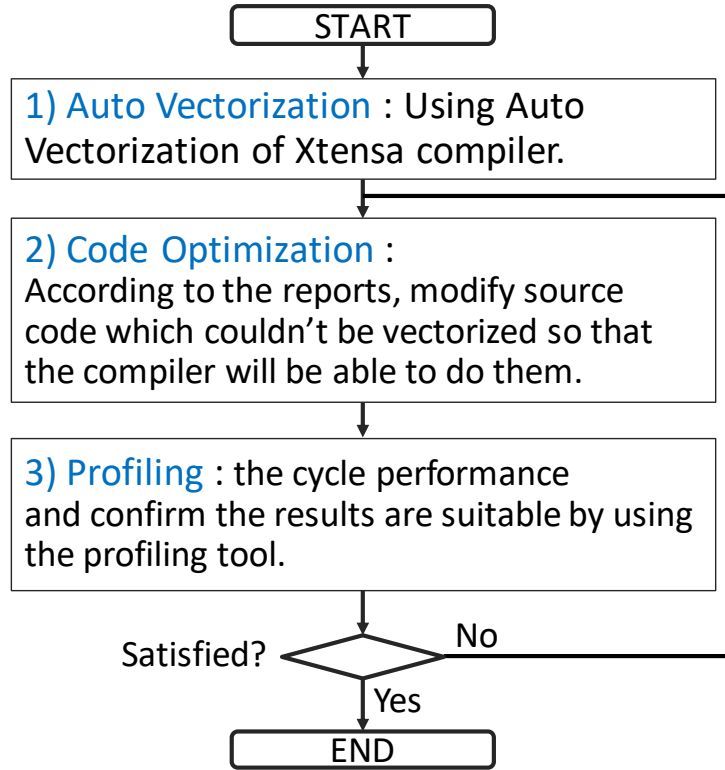


Figure 62. Optimization Flow Using Auto Vectorization of Xtensa Xplorer.

$$Y_{i,k,x,y} = \sum_{c=1}^C \sum_{v=1}^r \sum_{u=1}^r D_{i,c,x+u,y+v} G_{k,c,u,v}$$

Implementation as nested loops;

```

for (i = 0; i<N; i++)
  for (k=0; k<K; k++)
    for (y=0; y<H; y++)
      for (x=0; x<W; x++)
        Y(i,k,x,y) = 0
        for (c=0; c<C; c++)
          for (v=0; v<r; v++)
            for (u=0; u<r; u++)
              Y(i,k,x,y) += D(i,c,x+u,y+v)*G(k,c,u,v)
  
```

} Loops to be vectorized

← Multiply and Accumulate

Figure 63. Code of Convolutional Layer as Nested Loops

First, we use the auto-vectorization feature of VP6's compiler to vectorize operations and improve parallel execution efficiency in VLIW (Fig.62 1.)). Table 12 shows the output message from auto-vectorization feature which indicates “tried to vectorize the function predict, line number 2631 in Linear.cpp, in 16 parallel operations, and it failed.” The list of messages is initially sorted by the number of processor cycles used by a given loop, such that the most expensive loops appear first as shown in Fig.64.

Table 12. Output Message from Auto Vectorization.

File : Line	Function	Vectorizable or Not	Message
Linear.cpp : 2631	predict	Not	Retrying loop vectorization by 16

Then, we find codes that should be re-coded by the Vectorization Assistant feature of Xtensa Xplorer as shown in Fig.64 that helps to find codes that can be improved. Vectorization Assistant discovers and locates code that could not be vectorized along with an explanation that can help the programmer modify the code so that it can be vectorized (Fig.62 2.)). Xtensa Xplorer compiler is able to automatically detect opportunities for vectorization from standard C code. The dependence analysis of the compiler analyzes both inner and outer loops to see if they can safely be executed in parallel. For appropriate loops, the scalar operation in the loop are replaced with corresponding vector operations [121].

After finding codes that cannot be automatically vectorized by the auto-vectorization feature, we re-code the codes to improve the parallelization of operations and analyze total cycles weather the performance is satisfied for requirements (Fig.62 3.)). Figure 65 shows an example of our re-coding for auto-vectorization of Xtensa Xplorer. We unrolled unnecessary loop, clarified each variable data types as “float” and split one calculation into four calculations for auto-vectorization. After re-coding, auto-vectorization is successfully done.

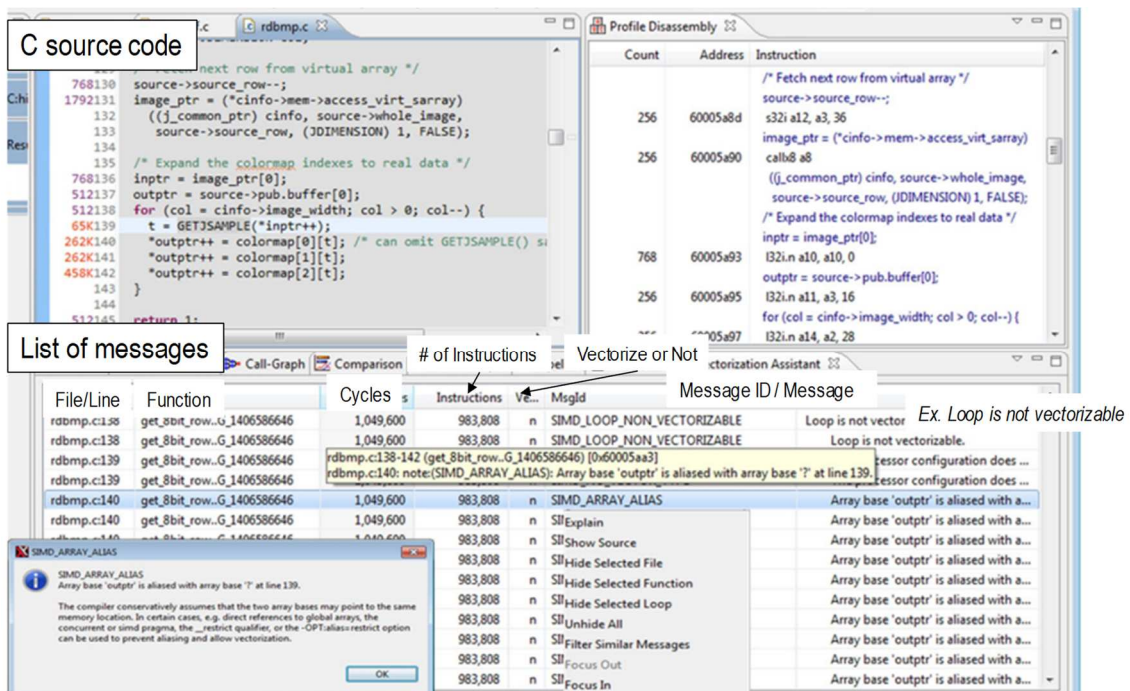


Figure 64. Vectorization Assistant View of Xtensa Xplorer [121]

a) Code that cannot be automatically vectorized

```
for(; (idx=lx->index) !=-1; lx++) {
  if(idx<=n)
    for(i=0;i<nr_w;i++) {
      dec_values[i] += w[idx*nr_w+i]*lx->value;
    }
}
```

b) After re-code

```
for(j=0; j<250; j++) // => 1000 / 4
{
  value_tmp0 += (float)w[j+ 0]*(float)PROB_SCALAR*(float)ResultfVEC[j+ 0];
  value_tmp1 += (float)w[j+250]*(float)PROB_SCALAR*(float)ResultfVEC[j+250];
  value_tmp2 += (float)w[j+500]*(float)PROB_SCALAR*(float)ResultfVEC[j+500];
  value_tmp3 += (float)w[j+750]*(float)PROB_SCALAR*(float)ResultfVEC[j+750];
}
dec_values[0] = value_tmp0 + value_tmp1 + value_tmp2 + value_tmp3;
```

Figure 65. Example of re-coding for Xtensa auto-vectorization.

The VP6 has specific instructions for convolution, pooling and sigmoid calculation [107]. We use these primitive instructions to effectively perform common functions that take into account the architecture and behavior of built-in hardware with intrinsic expression, which directly uses these primitive instructions like assembly language, and issue VLIW instructions

with high efficiency.

In addition, the VP6 has Scatter Gather DMA mechanism and the specified libraries and instructions for utilizing this DMA. The VP6 has 2 banks per local data RAM, and each bank has 8 sub-banks. These libraries allow to load and store into different sub banks in parallel. These libraries also allow multiple loads and stores to same sub-bank address combined for a single access operation. Furthermore, this DMA supports up to 64 outstanding requests for external memory access. The 1-Dimensional and 2-Dimensional data transfers between local and external memories are supported. Efficient memory accesses can be achieved by utilizing these libraries and instructions explicitly [107].

Figure 66 shows the result of optimization. In endoscopic image diagnosis, the clinical doctor does not take care the color information of lesion, therefore the image can be gray-scaled to reduce the amount of data. By this optimization, it is possible to reduce Convolution layer processing cycle by approximately 8 M cycles, almost 47 % (Fig.66).

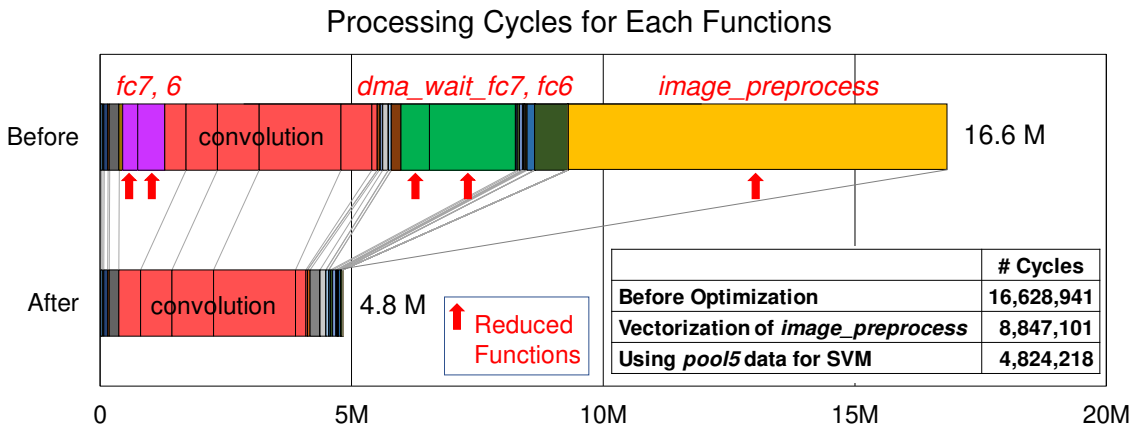


Figure 66. Result of Optimization.

4.6.2. Optimization of waiting time for loading coefficients of full connect layer

For reducing the waiting time of DMA transfer for reading the weight coefficients of the fully connected (*fc*) layer from the external memory to the internal data RAM, we will use the intermediate data as the extracted feature before the fully connected layer. And, the process of AlexNet itself will be quitted at that stage.

For implementing to an embedded system, the number of dimensions of the

intermediate data to be used is equivalent to the number of dimensions of the input data to the SVM classifier, which performs classification in the next stage. Therefore, we have to consider the trade-off between the classification accuracy and the number of dimensions of the intermediate data from hidden layer of AlexNet. Figure 67 shows the number of dimensions on intermediate data as feature quantity from each hidden layer of AlexNet. The prob layer output at the final stage is the probability of 1,000 categories as AlexNet's outputs value. We use *pool5*, *fc6*, *fc7*, *fc8* and *prob* as candidates for intermediate data as feature extraction. Thus, we evaluated the classification accuracy when these intermediate data are used as input data to the SVM classifier for comparison.

The classification accuracy is evaluated by *True Positive*, *Precision Rate*, and *F-measure*, equations (2) to (4) in Chapter 2. The *True Positive* shows correct recognition rate according to labels, and the *Precision Rate* shows accuracy of recognition. The *F-measure* represents the harmonic mean of the *True Positive* and the *Precision Rate*.

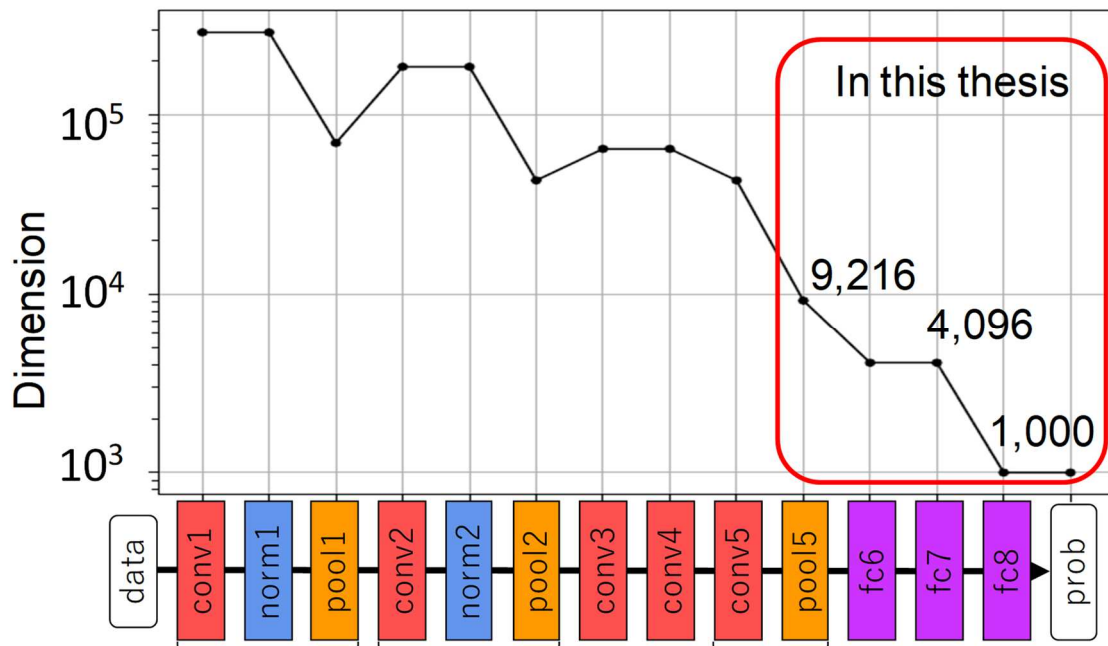


Figure 67. Feature dimensions of pre-trained AlexNet, and, architecture of AlexNet.

Figure 68 and Figure 69 show the evaluation results. The classification accuracy was evaluated by performing 10-fold Cross Validation (CV) ten times to compare the classification accuracy of the proposed method based on

the training data set created (Fig.34). The dataset for the evaluation is the still NBI image dataset containing 1,260 NBI patches obtained from endoscopic examinations at Hiroshima University. The test dataset is the same 1,260 NBI images, in addition these images are rotated 90 degrees, 180 degrees and 270 degrees to the right. For comparison, we use a previous study on endoscopic image recognition based on Bag-of-Features (BoF) [33]. The BoF approach uses Dense-Scale Invariant Feature Transform (D-SIFT) for feature extraction. The value of F-measure indicates over 97% of Type 1 and Type2A/Type3 classification. Compared to previous CAD system based on D-SIFT, classification accuracies using CNN feature extraction are higher for Type 2A and Type 3 classification. These results suggested it is possible to improve the system processing speed with keeping high classification accuracy. For example, the graph on the left side of Figure 68 shows the *True Positive* (TP) and *Precision Rate* (PR) for Type 1 and Type not 1, respectively. Since both TP and PR values are high, we can consider that the *False Positive* (FP) and *False Negative* (FN) of Type 1 and Type no 1 are small. *Precision Rate* (PR) is calculated by equations (6) besides equations (3) and equations (7) besides equations (2) in Chapter 3.

$$PrecisionRate(PR) = \frac{TruePositive(TP)}{TruePositive(TP) + FalsePositive(FP)} \quad (6)$$

$$TruePositive(TP) = 1 - FalseNegative(FN) \quad (7)$$

		Predicted Value by CNN	
		Positive	Negative
Correct Value	Positive	<i>True Positive</i> (TP)	<i>False Negative</i> (FN)
	Negative	<i>False Positive</i> (FP)	<i>True Negative</i> (TN)

Therefore, from this graph on the left side, for the CNN Features: *pool5*, *fc6*, *fc7*, and *fc8*, it is possible to classify with an accuracy of over 90%. In other words, this corresponds to the classification of whether the lesion is non-tumors or tumor.

Likewise, the right-hand graph of Figure 68 shows that the classification between Type 2A and Type 3 is more than 90% accurate with the CNN Features: *fc6*, *fc7* and *fc8*. In *pool5*, there is a slight difference in PR among

fc6, *fc7* and *fc8*. However, in the classification of lesions on endoscopic images, since cancer progresses to Type 2A, Type 2B, and Type 3, it is difficult to distinguish them clearly in the case of images on the border of the degree of progress. In this simulation, the output value of SVM classification is output in the range of 0 to 1. Because the threshold value for judgment is 0.5, it is the difficult when the difference between the outputs is small such as (0.6, 0.4) case. However, from the viewpoint of diagnosis support, for example, if the output value does not exceed 0.8, the CAD system provides the doctor with an objective index. Thus, the doctor can pay attention, stop the endoscope, and take still images for careful diagnosis.

The number of dimensions of the feature extracted from *pool5* increases by 9,216, which is about 9 times the number of dimensions from *prob*, the last layer in AlexNet. The number of processing cycles in SVM classification is approximately 0.06 M cycles when using the feature of *prob* as input, and approximately 0.56 M cycles when using the feature of *pool5* as input. The number of processing cycles is increased by 0.5M. However, the number of processing cycles reduced by removing the processing of *dma_wait_fc6*, *dma_wait_fc7*, *fc6*, and *fc7* is 4 M cycles. Using the 9,216-dimensional intermediate data as feature extracted from *pool5* is much more effective. The data transfer size by DMA is reduced to 56 Mbyte by using the output from *pool5* compared to using the output from *prob*. From these results, it is effective to use the output of *pool5* with dimension number 9,126 as the extracted feature of the input to SVM of the next stage.

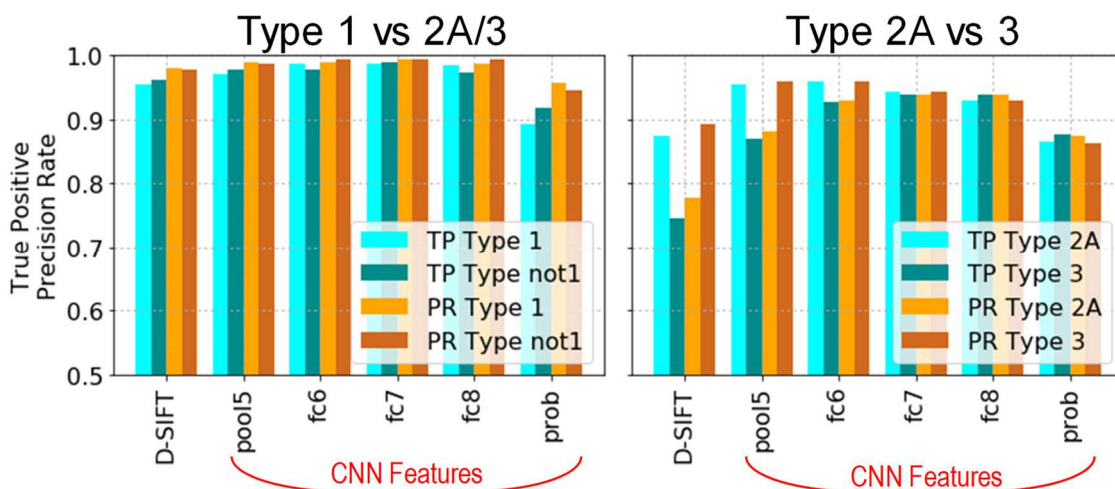
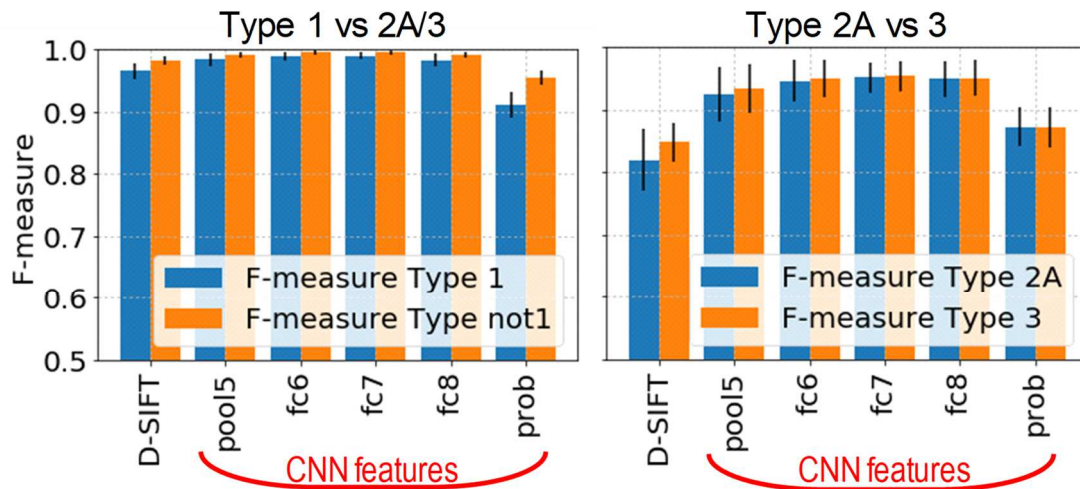


Figure 68. True Positive and Precision Rate for each hidden layer of AlexNet



Learning: 1260 NBI images (Type 1: 420, Type 2A,3: 840)
 Test: 90, 180 and 270 right degree rotated 1260 NBI images (Type 1: 420, Type 2A,3:840)

Figure 69. F-measure for each hidden layer of AlexNet.

4.7. Evaluation of the Developed Prototype System

Figure 66 shows the comparison of the numbers of processing cycles before and after optimization. This result shows that the number of total processing cycles can be reduced by approximately 70%. The reason for reduction of about 70% is that such as Load / Store of related data included in *Other* in Fig.60 is also reduced, while the reduction of 61% is expected. When the clock frequency is 200 MHz, the frame rate is 44.6 fps, which exceeds the input video frame rate of 30 fps, and also it is achieved the requirements from clinical doctors. Xtensa Xplorer has features of power, performance and chip area estimation for each typical processes like TSMC as shown in Fig.70 [121].

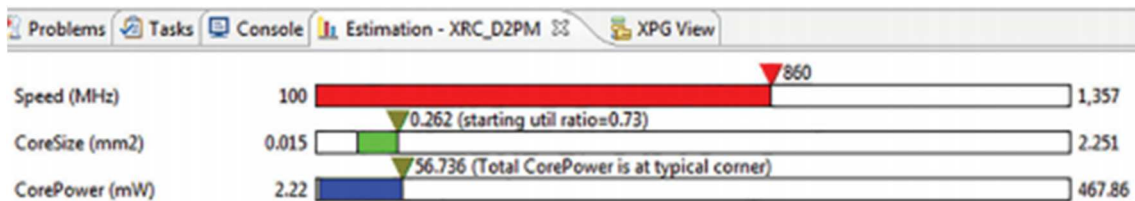


Figure 70. Performance, Area, Power Estimation View in Xtensa Xplorer [121].

We estimated power consumption is 66.6 mW under a 16 nm CMOS process. For example, Nvidia's Jetson AGX, an embedded GPU and LSI manufactured

in the 12 nm CMOS process of TSMC (Taiwan Semiconductor Manufacturing Company), executes AlexNet 299 fps at 14 W by 15 W mode [122]. It is equivalent to 2.08 W, when reducing the performance to the same frame rate of 44.6fps as VP6. We compared to other platforms and Table 13 shows the comparison results. As Table 13 shows, the power consumption of our system is lower compared to the latest embedded GPUs. In addition, the latency of our system is about 22 ms, from Step 2 to Step 4 in Fig.57, including video data transfer from the HDD in the host PC. The latency of the endoscopic camera transfer can be assumed to be almost equivalent, which corresponds to a delay of less than one frame when the frame rate of the input video image is 30 fps. There is no difference between the frame taken by the endoscope and the frame displaying the classification result. Thus, it is possible to provide the information about the classification result that is also easier for the doctor to understand.

Table 13. Performance for AlexNet and power comparison to other platforms [122-123]

	Vision P6 [107]	Jetson AGX [124]	Jetson Nano [125]	Raspberry Pi 3B [126]
Performance(fps)	44.6	299	21.7	0.35
Power(Watt)	0.0666	14	4.58	2.73
Power(Watt) per 1 fps	0.000149	0.0468	0.211	7.80
Clock Frequency (MHz)	200 (DSP)	670 (GPU)	640 (GPU)	1200 (ARM A53)

We constructed the proposed CAD system on the hardware prototyping platform, Protium S1, and validated using the endoscopic video images. Since Protium S1 has a function of observing waveform data from each signal for debugging, there is a limitation to improve the operating clock frequency. Therefore, the operating clock frequency is 22.97 MHz, which is equivalent to 4 fps as processing performance. Figure 71 shows the developed CAD system using the Protium S1. The developed CAD system is consisted of the host PC, the Protium S1 and the output monitor. The classification result is displayed on the output monitor in Fig.71. Figure 72 shows the setting feature of classification region by the clinical doctors in the CAD system. The clinical doctors can switch the size of the classification region, 227 x 227 pixels (green rectangle) or 454 x 454 pixels (red rectangle) shown as Fig.72 (a). The clinical

doctors can set the classification region position or move the clear part of endoscopic image into the classification region during observation (yellow rectangle) shown as Fig.72 (b).

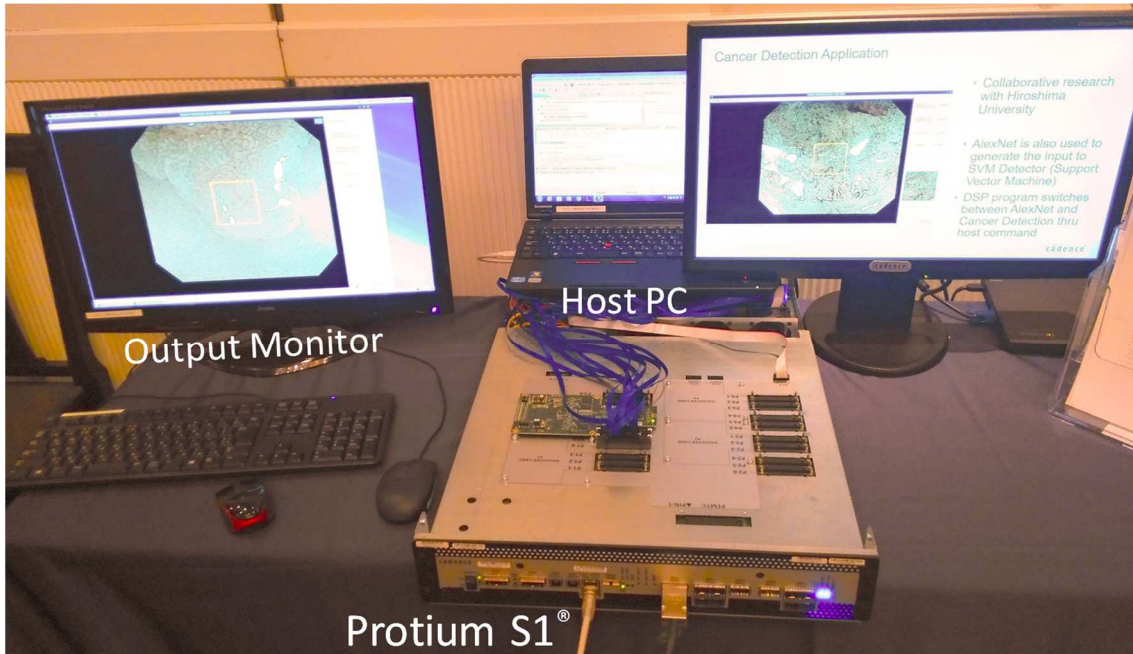
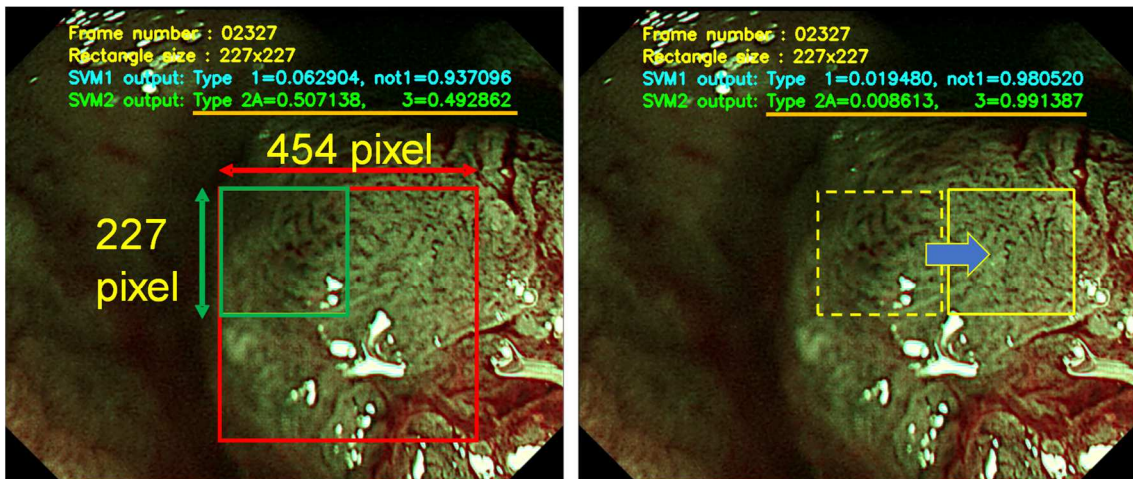


Figure 71. The Developed CAD System.



(a) Switch the Classification region sizes (227 x 227 or 454 x 454)

(b) Move the Classification region

Figure 72. Staging Region Setting.

Figure 73 shows the output display image of the developed system. The yellow rectangle area is the staging region (257 x 257 pixel size). The

classification results are displayed at the top of the endoscopic image. In the CAD system, it is much more important to show an objective numerical evaluation for a rectangular area, staging region that a clinical doctor wants to observe the corresponding lesion. Since it is possible to move the scope to the staging region where the clinical doctor wants to observe, in the first prototype system, the staging region set at any place such as the center region because the staging region can be moved at the doctor's decision by moving the scope.

Figure 74 and Figure 75 show classification results for Type 2A and Type 3, respectively. In the CAD system, the doctor stops the endoscope when it is difficult to see the staging region because of peristaltic motion of inner wall and confirm the staging region by a clear still image. Moreover, the doctor stops the endoscope and observe the lesion carefully by a clear still image when the recognition rate is going lower or the recognition rate is near 50%, in which it is difficult to the CAD system decide the Type. Then the doctor brings the staging region to the portion suspected of being a lesion. We confirmed that it can be recognized with higher accuracy by moving the yellow square of the staging region from the boundary of the lesion (Fig.74 (a), Fig.75 (a)) to the center of the lesion (Fig.74 (b), Fig.75 (b)) in real time in accordance with the lesion.

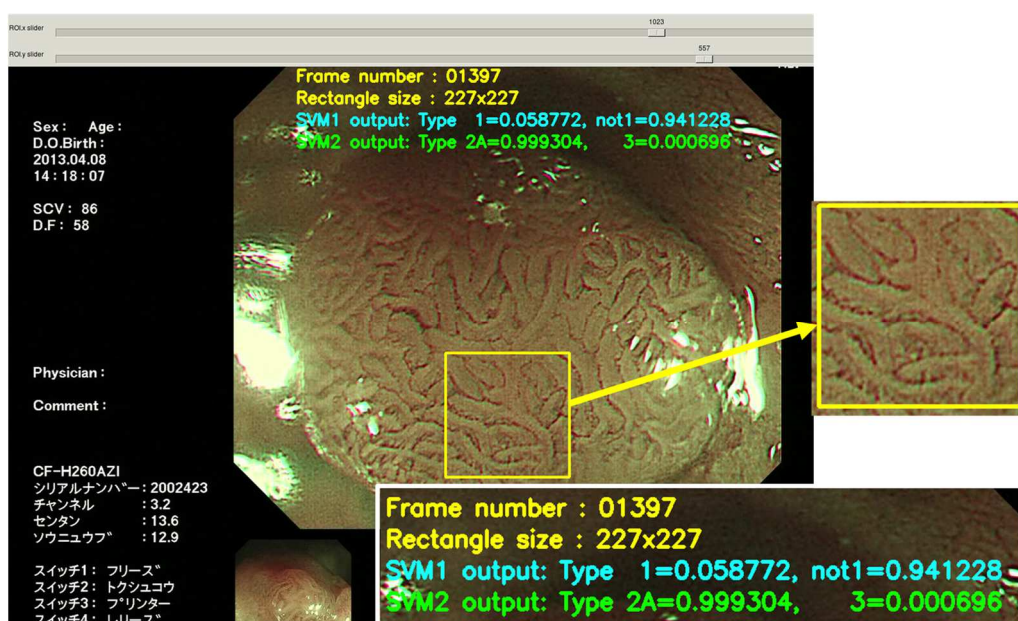
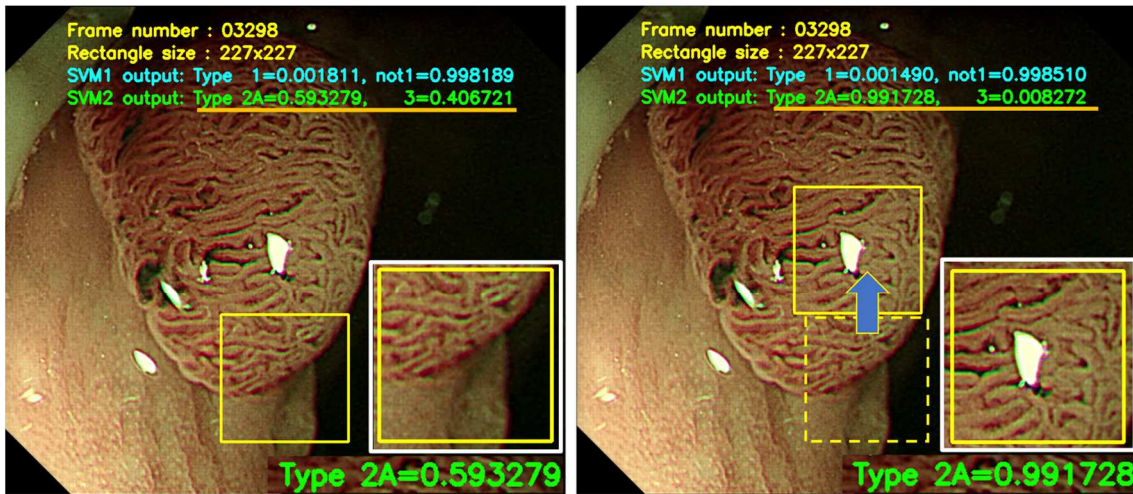
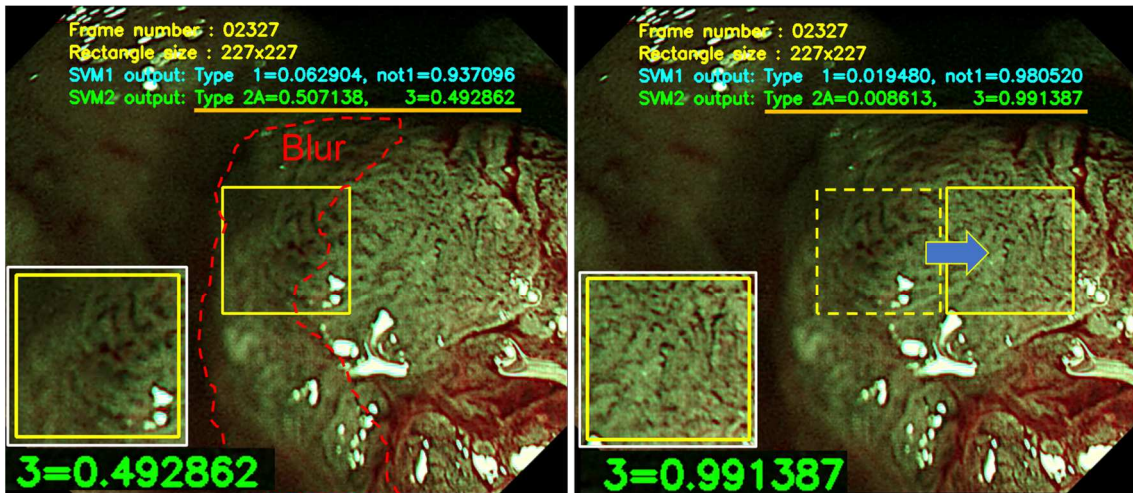


Figure 73. The classification result of the endoscopic video image on the real-time prototyping system (Type 2A).



(a) Classification results at the boundary of the lesion. (b) Classification results in the center of the lesion including reflection of light

Figure 74. Real-time classification results from the boundary to the center of the lesion of Type 2A.



(a) Classification results at the boundary of the lesion including blur region. (b) Classification results in the center of the lesion.

Figure 75. Real-time classification results from the boundary (blur region) to the center of the lesion of Type 3.

Figure 76 shows the output image in which the size of the yellow square is changed to 454 x 454 pixels and reduced with linear interpolation to 227 x 227 pixels as the input data. Even when the staging region (yellow rectangle in Fig.76) size is enlarged, it is confirmed that the prototype system can recognize lesion region correctly. Figure 77 shows the classification result of same frame image in Fig.74 and Fig.75 in which the size of the yellow square

is changed to 454 x 454 pixels. The clinical doctors can switch the size of the staging region, 227 x 227 pixels or 454 x 454 pixels.

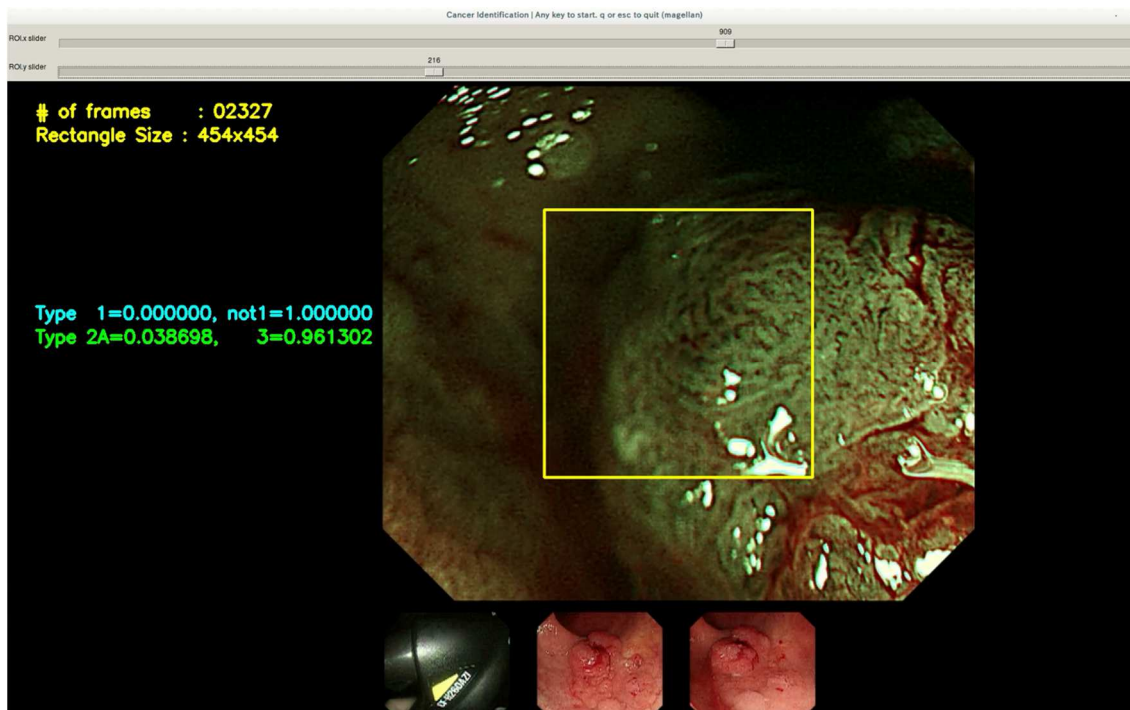
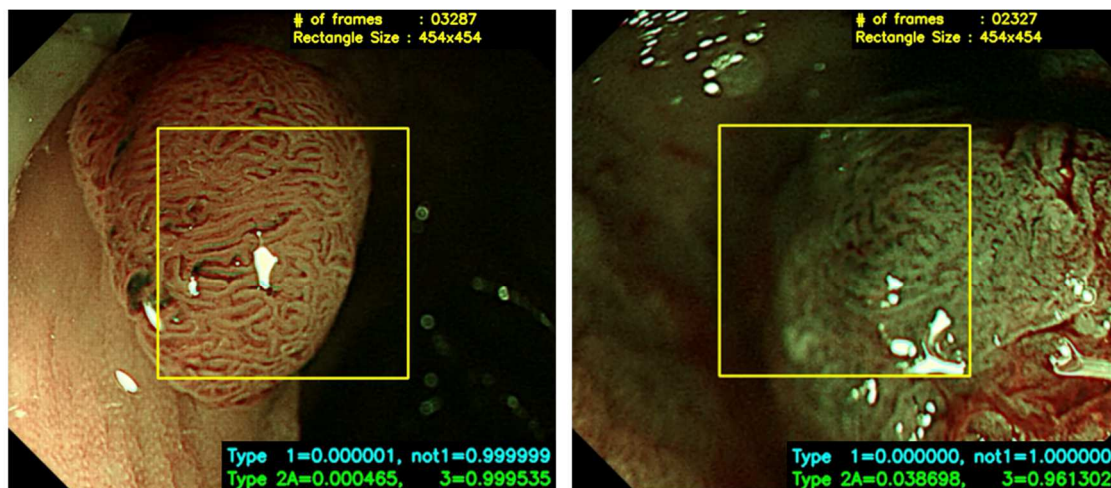


Figure 76. The classification result of image with 454 x 454 pixel window.



(a) Classification results of Fig. 57 with 454 x 454 pixel window

(b) Classification results of Fig. 58 with 454 x 454 pixel window

Figure 77. Real-time classification results of image with 454 x 454 pixel window.

4.8. Conclusions

In this Chapter, I explained hardware requirements for realizing a CAD system with CNN feature extraction and SVM classification on an embedded system. Then, we have carefully analyzed the profiling data of the still image processing of the CAD system for colorectal NBI endoscopic using CNN feature extraction and SVM classification. We have realized real time processing by improving the bottlenecks with over 70% reduction of the processing cycles per frame for video processing by optimizing the image pre-processing with VP6 DSP core and using the extracted feature of the hidden layer in AlexNet. Moreover, we have demonstrated that the developed system achieves real-time colorectal tumor classification for the endoscopic video image (44.6fps throughput / 22ms latency @200MHz, 66.6mW power consumption), and sufficient classification accuracy (> 90%).

References

- [90] A. Lavin, and S. Gray, "Fast Algorithms for Convolutional Neural Networks," Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR'16), Las Vegas, NV, USA, pp. 4013-4021, 2016.

- [91] A. Ahmad, and M. A. Pasha, "Towards Design Space Exploration and Optimization of Fast Algorithms for Convolutional Neural Networks (CNNs) on FPGAs.," Proceedings of 22nd Design, Automation & Test in Europe Conference & Exhibition (DATE19), Florence, Italy, pp.1106-1111, 2019.

- [92] V. Sze, Y. Chen, T. Yang, and J. S. Emer, "Efficient Processing of Deep Neural Networks: A Tutorial and Survey.," Proceedings of the IEEE, vol. 105, no. 12, pp. 2295-2329, 2017.

- [93] K. Simonyan, and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition.," Proceedings of 3rd International Conference on Learning Representations (ICLR) 2015, San Diego, CA, USA, pp.1-14, 2015.

- [94] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition.," Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR'16), Las Vegas, NV, USA, pp.770-778, 2016.

- [95] K. Abdelouahab, M. Pelcat, J. Serot, and F. Berry, "Accelerating CNN inference on FPGAs: A Survey", <https://hal.archives-ouvertes.fr/hal-01695375v2>, 2018.

- [96] G. Efland, S. Parikh, H. Sanghavi, and A. Farooqui, "High Performance DSP for Vision, Imaging and Neural Networks," IEEE Hot Chips 2016, Cupertino, CA, USA, 2016.

- [97] J. A. Fisher, "Very Long Instruction Word architectures and the ELI-512.," Proceedings of the 10th annual international symposium on Computer architecture (ISCA '83), New York, USA, pp.140–150, 1983.

- [98] J. A. Fisher, "The VLIW Machine: A Multiprocessor for Compiling Scientific Code.," *Computer*, vol.17, no.7, pp.45-53, 1984.
- [99] M. J. Flynn, "Some Computer Organizations and Their Effectiveness.," *IEEE Transactions on Computers*, vol.C-21, no.9, pp.948-960, 1972.
- [100] R. Duncan, "A survey of parallel computer architectures.," *Computer*, vol.23, no.2, pp.5-16, 1990.
- [101] J. L. Baer, "Multiprocessing Systems.," *IEEE Transactions on Computers*, vol.C-25, no.12, pp.1271-1277, 1976.
- [102] Cadence Design Systems, Inc., "Xtensa LX7 Processor.," https://ip.cadence.com/uploads/1099/TIP_PB_Xtensa_lx7_FINAL-pdf
- [103] Intel Corporation, "Introduction to Intel® Advanced Vector Extensions.," <https://software.intel.com/content/www/us/en/develop/articles/introduction-to-intel-advanced-vector-extensions.html>
- [104] B. Moyer, "How Does Scatter/Gather Work? Promises of Single-Cycle Access Are True, But....," *EE Journal*, <https://www.eejournal.com/article/20170209-scatter-gather/>, 2017.
- [105] L. Bai, Y. Zhao, and X. Huang, "A CNN Accelerator on FPGA Using Depthwise Separable Convolution.," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol.65, no. 10, pp.1415-1419, 2018.
- [106] Y. Ma, Y. Cao, S. Vrudhula and J. Seo, "An automatic RTL compiler for high-throughput FPGA implementation of diverse deep convolutional neural networks.," *Proceedings of 27th International Conference on Field Programmable Logic and Applications (FPL)*, pp.1-8, 2017.
- [107] Cadence Design Systems, Inc., "Vision DSPs for Imaging and Neural Networks." <https://ip.cadence.com/vision>
- [108] R. E. Gonzalez, "Xtensa: a configurable and extensible processor.," in

IEEE Micro, vol.20, no.2, pp.60-70, 2000.

- [109] G. Ezer, "Xtensa with user defined DSP coprocessor microarchitectures.," Proceedings of 2000 International Conference on Computer Design, Austin, TX, USA, pp.335-342, 2000.
- [110] C. Rowen and D. Maydan, "Automated processor generation for system-on-chip," Proceedings of the 27th European Solid-State Circuits Conference, Villach, Austria, pp.464-469, 2001.
- [111] C. Yang, S. Chen, J. Zhang, Z. Lv and Z. Wang, "A Novel DSP Architecture for Scientific Computing and Deep Learning.," IEEE Access, vol.7, pp.36413-36425, 2019.
- [112] Nvidia Corp., "NVIDIA TITAN X", <https://www.nvidia.com/en-us/geforce/10-series/titan-x-pascal/>
- [113] Nvidia Corp., "GEFORCE GTX 10 SERIES", <https://www.nvidia.com/en-us/geforce/10-series/>
- [114] CEVA, Inc, "CEVA-XM6", <https://www.ceva-dsp.com/product/ceva-xm6/>
- [115] J. Benndorf, "A New Computer Vision Processor Chip Design for Automotive ADAS CNN Applications in 22nm FDSOI", SOI Silicon Valley Symposium, Santa Clara, CA, USA, 2017.
- [116] T. Dettmers, "8-Bit Approximations for Parallelism in Deep Learning.," 4th International Conference on Learning Representations, San Juan, Puerto Rico, 2016, <http://arxiv.org/abs/1511.04561>
- [117] G. Phillip, "Ristretto: Hardware-Oriented Approximation of Convolutional Neural Networks," arXiv preprint arXiv:1605.060402, 2016.
- [118] Cadence Design Systems, "Protium S1 FPGA-Based Prototyping Platform," <https://www.cadence.com/news/protium>

- [119] Xilinx, Inc., “Virtex UltraScale.,”
<https://www.xilinx.com/products/silicon-devices/fpga/virtex-ultrascale.html>
- [120] Xilinx, Inc., “AXI Central DMA Controller.,”
https://www.xilinx.com/products/intellectual-property/axi_central_dma.html
- [121] Cadence Design Systems, “Tensilica Software Development Toolkit (SDK),” <https://ip.cadence.com/uploads/103/SWdev-pdf/>
- [122] Nvidia Corp., “Jetson AGX Xavier: Deep Learning Inference Benchmarks,” <https://developer.nvidia.com/embedded/jetson-agx-xavier-dl-inference-benchmarks>
- [123] R. Hadidi, J. Cao, Y. Xie, B. Asgari, T. Krishna, and H. Kim, “Characterizing the Deployment of Deep Neural Networks on Commercial Edge Devices,” Proceedings of the 2019 IEEE International Symposium on Workload Characterization, Orlando, FL, USA, pp.35-48, 2019.
- [124] Nvidia Corp., “Jetson AGX Xavier”, <https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-agx-xavier/>
- [125] Nvidia Corp., “NVIDIA Jetson Nano”, <https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-nano/>
- [126] Raspberry Pi Foundation, “Raspberry Pi 3 Model B”,
<https://www.raspberrypi.org/products/raspberry-pi-3-model-b/?resellerType=home>

Chapter 5. Practical Improvement to the CAD System with Navigation Function

5.1. Issues in Video Image

For the real-time processing of endoscope video images and the providing diagnostic support information, it is necessary to solve the problems described in Chapter 3.

(1) Microscopic defocus or changes of appearance of lesions due to movement of the lesion in the depth direction.

(2) Unclear frames with blur, color shift, reflection of light in the NBI endoscopy.

(3) Differences in resolution and appearance between old and new endoscopic systems.

Also, the improvements in the video processing and scoping systems of endoscopic systems are known to result in different resolutions and appearances (Fig.33). Therefore, it is desirable to have the generality that the image data of an old endoscope previously provided by a clinical doctor can be applied to another different system. In order to solve this problem, the transfer learning which is a method for the feature distribution of image data in a new endoscope system, which is closer to the feature distribution of the conventional endoscope system, is proposed the conventional system using the D-SIFT function. On the other hand, the trend survey in the case of the classification method using the CNN feature extraction has not been conducted.

Above problems in endoscopic video image are major challenges in endoscopic diagnosis. Detecting artifact in endoscopic video contest (Endoscopy Artifact Detection challenge: EAD2019) was held in 2019, Venice, Italy, co-located with the 16th International Symposium on Biomedical Imaging (ISBI 2019), and EAD2019 challenge dataset was released [127-128]. Figure 78 shows an example of EAD2019 challenge dataset. In this dataset, there are 7 classes of artifacts in endoscopic video image, specularity (reflection of light), blur, saturation, contrast, artifact (ex. message form

endoscopy system), medical instrument and bubbles for 5 organs, colon, stomach, oesophagus, bladder and liver. This training dataset obtained from 3 types of endoscopy, WLI (white light imaging), NBI (narrow band imaging) and AFI (auto-fluorescence imaging) consists in total 2,147 annotated frames over all these 7 classes. This dataset is assembled from 6 different data centers that includes John Radcliffe Hospital, Oxford, UK; ICL Cancer Institute, Nancy, France; Ambroise Paré Hospital of Boulogne-Billancourt, Paris, France; Istituto Oncologico Veneto, Padova, Italy; University Hospital Vaudois, Lausanne, Switzerland; Botkin Clinical City Hospital, Moscow. Some artifact detecting methods using CNN are proposed for this data set [129-132]. A comparison study for these methods has been reported in 2020 [133].



Figure 78. Example annotated training detection boxes illustrating the 7 different artifact classes in the EAD2019 challenge dataset [128].

To solve above problems, method which skips unclear frames till a clear

frame has been proposed [134]. By this method, there is a possibility that a significant lesion is missed if it is in an unclear frame. Also, methods for improving the image quality of endoscopic video images, such as deblurring of video images, have been proposed [135-136]. The deblurring of the video image is general in multimedia application, and interpolating technique between clear frames is well known for deblurring [137]. However, this method changes original endoscopic video image, so there is a risk that the critical and significant lesion information will be tampered.

Therefore, we aimed to develop a robust classifier by the proposed method using the CNN feature and SVM classification and consider solution of the above problem without changing of input endoscopy video image by preparing a training data set taking into account appearances in Chapter 3.

5.2. Overview of Navigation function of clear region

For real time CAD system, a robust system is required in appearance changes against a recognition of still image captured during video freezing. Shown as Fig.32 in Chapter 3, unclear frames with (a) blur, (b) color shift, (c) reflection of light that affect the classification accuracy are observed in the actual endoscopic video image compared to the still image. The clinical doctors make diagnosis based on surface pit pattern and vessel structure, but if the endoscopic image is unclear, the pit pattern and vessel structure cannot be seen, and this leads to errors of doctor's judgement. Also, there is a possibility that the classifier of the CAD system provides incorrect classification result. Thus, in our CAD system, the clinical doctor can switch the staging region size (227 x 227 or 454 x 454) and move the staging region to avoid unclear regions in the endoscopic video frames. Output results from our CAD system provides the probability of each pathological types. For example, after avoiding unclear regions, the probability of Type 3 in Fig.75 in Chapter 4 is improved from about 0.49 to 0.99.

Therefore, the real time navigation function in a CAD system, which is consist of unclear region detection and providing a direction of clear region, is useful for the clinical doctors' diagnosis. In a real-time CAD system, detecting unclear regions in endoscopic video image is one of major challenges.

The navigation function of clear region consists of two main steps, unclear region detection and classification by AlexNet and SVMs. Unclear region

detection in the endoscopic video frames is one of major challenges. Detecting artifact in endoscopic video contest (Endoscopy Artifact Detection challenge: EAD2019) was held in 2019, and EAD2019 challenge dataset was released [127-128]. Some artifact detecting methods using R-CNN [138] etc. are proposed for this data set. These methods are executed on high performance graphics board like Nvidia's 1080 Ti in a host machine [133]. It takes some seconds for artifact detection even executed on high performance graphic board because R-CNN is precise for detection but complex architecture.

Our research aims to realize a CAD system on the embedded system. Therefore, for unclear region detection in the endoscopic video frames, we consider that YOLO2 (You Only Look Once: Unified, Real-Time Object Detection) [138-140] can be used in the embedded systems because YOLO2 is known as simple architecture model for multi object detection compared to other object detection neural network models. In proposed CAD system, as shown in Fig.79, after (1) detecting unclear regions, if the staging region is overlapped unclear regions, the CAD system informs the clinical doctors by red rectangle and (2) providing the direction of clear region by red arrow, then the clinical doctors move the staging region and (3) colorectal tumor is classified. Input video image is 30 fps, and, Fig.79 shows from frame #2487 to frame #2517, almost 1 second. In Fig.79, Type 3 lesions are included in the almost whole image. However, the staging region is overlapped light reflection, and a red window indicates that the probability value is not accurate. Moreover, a red arrow indicates a direction of clear region.

Figure 80 shows the flowchart of the proposed CAD system with the navigation function of clear region. When the system starts the image processing, the frame data is read as an image data from the endoscopic video sequence (1). From next step, (a) Navigation function of clear region (blue rectangle in Fig.80) and (b) Classification (green rectangle in Fig.80) are executed in parallel on different VP6 cores. For unclear region detection, the input data is resized to 512 x 512 pixels, then STEP (a-1) unclear image region as blur, color shift or reflection of light is detected by YOLO2 as shown in Fig.79 (1). Next, STEP (a-2) the clear region which does not affect the classification accuracy is detected and direction is provided as shown in Fig.79 (2).

For classification process, as the input data to the CNN, an arbitrary clear region of 227 x 227 pixels is cut out from one frame of the endoscopic video as

a staging region. There are three cases for setting the staging region, (i) defined by unclear regions detection, (ii) set by the clinical doctor manually and (iii) the same staging region as previous frame when the doctor set no region and the unclear region detection cannot detect till the current frame (red rectangle in Fig.80).

Then, STEP (b-1) the image data of staging region is stored to the image buffering memory shared between the host and the VP6. In the VP6, STEP (b-2) CNN feature extraction and SVM classification are performed to obtain the classification result for the staging region. STEP (b-3) The classification result is stored to the image buffering memory shared between the host and the VP6. Finally, STEP (b-4) the host reads the classification result and displays the staging region and the result together with the input frame image. The above two processes, (a) and (b) are repeated until the endoscopic video ends.

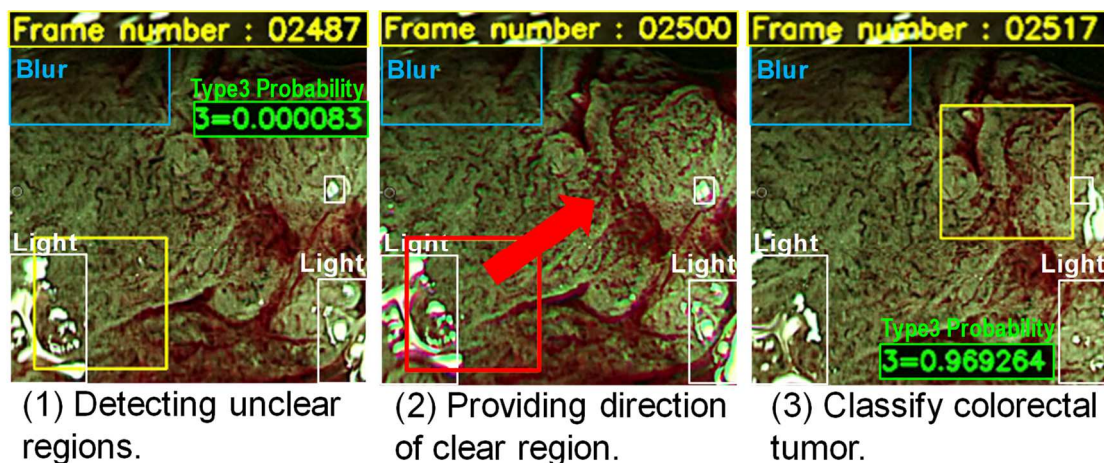


Figure 79. Overview of proposed CAD system with navigation function of clear region.

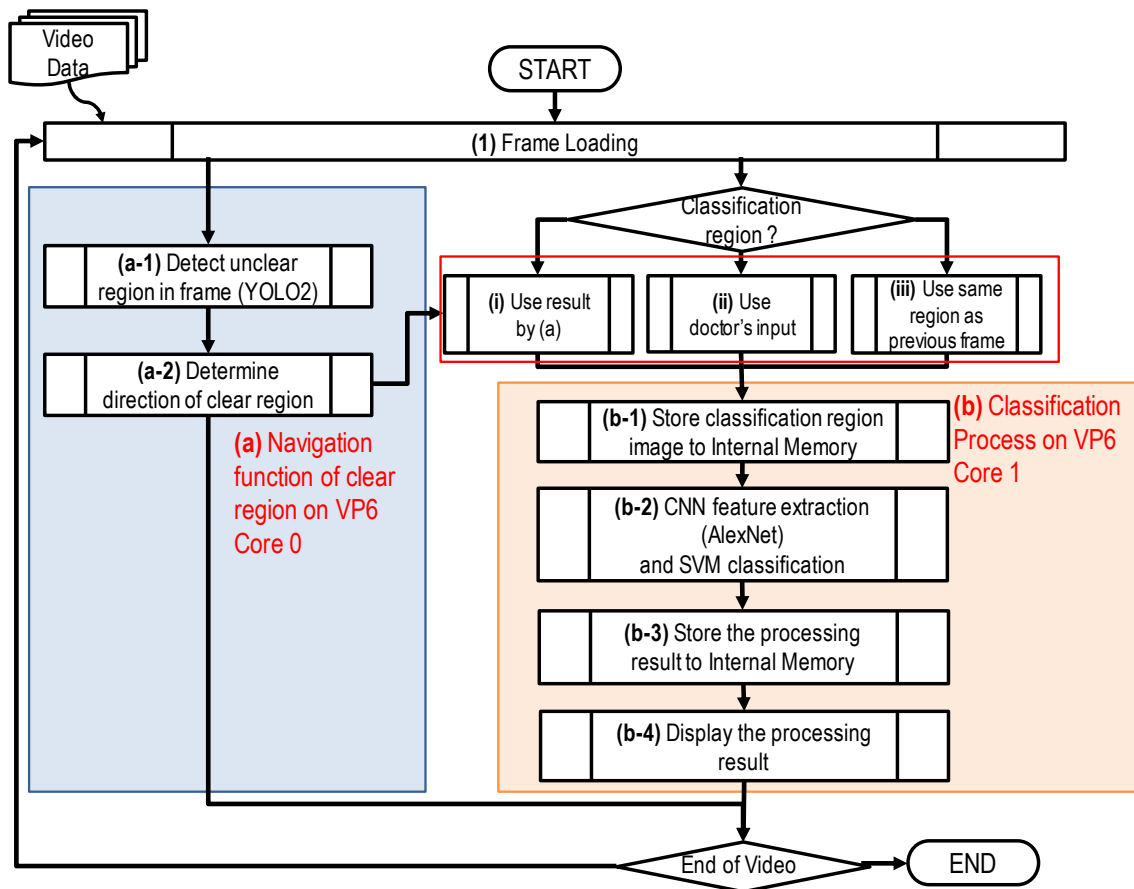


Figure 80. Flowchart of the proposed CAD system with navigation function of clear region.

We simulated YOLO2 and classification (AlexNet and SVMs) on the VP6 core by Xtensa Xplorer. AlexNet in this simulation is same pre-trained model used in our CAD system in Chapter 4. Table 14 shows the simulation results of execution time. From these results in Table 14, when the clock frequency is 200 MHz, the frame rate is less than 2.3 fps which is not achieved the clinical doctor's requirement (< 5 fps). When the clock frequency is 525 MHz and these two CNN processes performed sequentially, it takes 167.24 ms (8.54 ms + 158.7 ms), 5.97 fps (> 5 fps).

Table 14 Execution time of CNN on single VP6 core.

	YOLO2	AlexNet + SVMs	Total
Execution Time (200 MHz)	416.7 ms / 2.4 fps	22.4 ms / 44.6 fps	439.1 ms / 2.27 fps
Execution Time (525 MHz)	158.7 ms / 6.3 fps	8.54 ms / 117.4 fps	167.24 ms / 5.97 fps
Input Image Size	512 x 512 pixels	227 x 227 pixels	-

5.3. Possibility of the Proposed CAD System Implementation on VP6 and similar DSP cores in the low power ASIC

From the viewpoint of personal information protection, it is delicate issues to store the medical images outside a clinical site such as a cloud server. At the small clinical site, it is desirable to have a CAD system that can be added to the same chassis of the endoscope instead of PCs. Therefore, we are aiming to realize CAD system in embedded ASICs. From 2017, ASICs with VP6 cores or similar DSP cores for CNN execution are released for mobile and IoT applications. Table 15 shows lists of ASICs implemented VP6 cores or similar DSP cores for CNN execution. In these ASICs, ARM cores as main CPUs for application, a GPU for image processing and one or two DSP cores for CNN processing are implemented, and also interfaces for camera and display are implemented. Since it is premised on operating with even lower power consumption, it is considered to be suitable for use in CAD systems in the medical site.

Table 15 List of ASICs including VP6 and similar DSPs for CNN execution.

	Mediatek		Qualcomm		Mediatek
Chips	MT6771 [141]	MT6779 [142]	SDM710 [143]	SDM675 [144]	MT8385 [145]
Release	Q1 2018	Q1 2019	Q2 2018	Q1 2019	Q1 2020
Technology	TSMC 12nm	TSMC 12nm	Samsung 10nm	Samsung 11nm	TSMC 12nm
Main CPU (ARM)	A73/A53	A75/A55	A75/A55	A76/A55	A73/A53
# of Main CPU	4/4	4/4	2/6	2/6	4/4
Main CPU Clock (GHz)	2	2.2 / 2	2.2 / 1.7	2.0 / 1.7	2
DSP for AI (VP6 or similar core)	VP6	Vision Q6 + Accerlator	Hexagon 685	Hexagon 685	VP6
# of DSPs	2	2	1	1	2
DSPs Clock (MHz)	525	624	-	-	525
Application	Mobile				IoT

For example, Figure 81 shows the block diagram of Mediatek MT8385 [145] with two VP6 cores for CNNs (red rectangle in Fig.81). In MT8385, there are four ARM A73 and A53 as main CPUs for application, Mali G72 as GPU for graphic processing, display and camera interfaces and video encoder /decoder. Figure 82 shows the overview of the evaluation board of Mediatek MT8385

[146]. In our previous research, we have prototyped the CAD system on the VP6 core. Thus, at first, we examined possibility of implementation of the navigation function on MT8385 equivalent to MT6771.

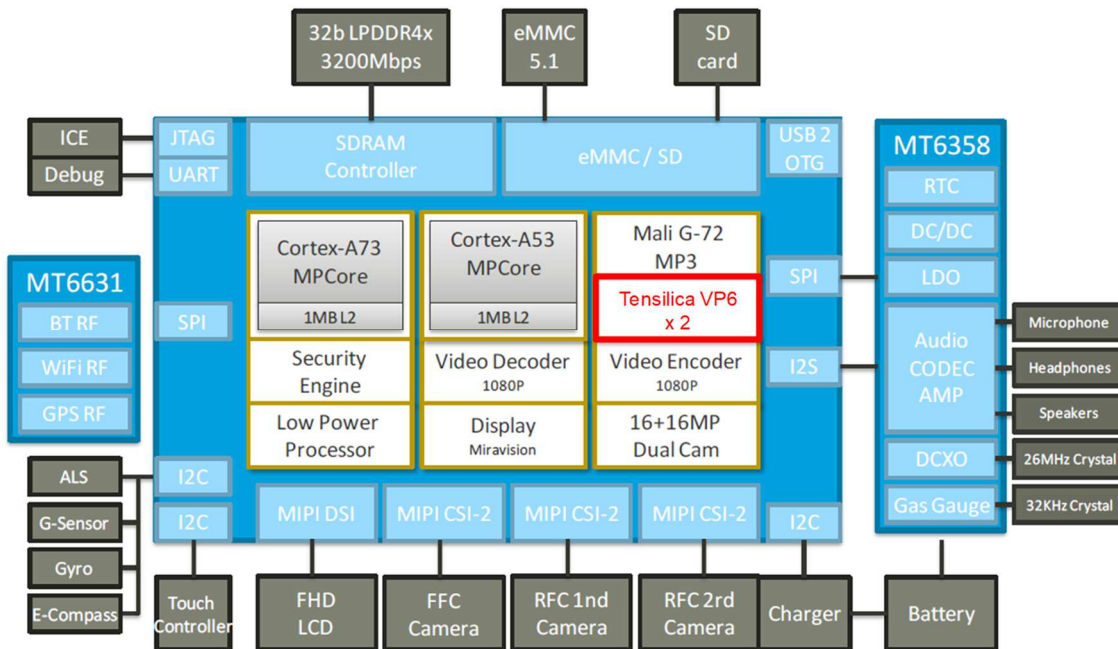


Figure 81. Block diagram of Mediatek's MT8385 with two Tensilica VP6 cores (red rectangle)

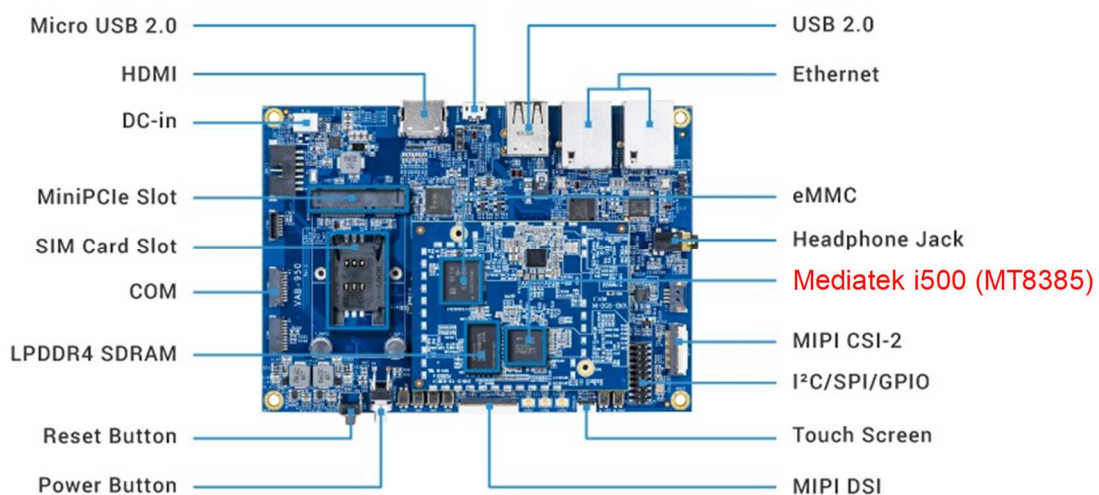


Figure 82. Evaluation board of Mediatek's MT8385.

From the simulation result in Table 14, if the navigation function and the colorectal tumor classification performed sequentially, the performed frame

rate is 5.97 fps at 525 MHz as same clock frequency of VP6 cores in MT8385. In the real CAD system, we have to consider pre- / post-processes, image size reduction for inputting YOLO2, image trimming for AlexNet, overlaying the staging region area or classification result etc. Thus, we consider that navigation function which consists of (1) detecting unclear regions and (2) providing the direction of clear region is executed in one of the two VP6 cores, and (3) colorectal tumor classification is executed in the remaining VP6 core on MT8385 simultaneously-

Figure 83 shows the timeline of the proposed CAD system performed by (A) sequential execution mode and (B) parallel execution mode. Execution time of two processes in Fig.80 are different. (a) Navigation function treats 6 fps and (b) Classification process treats over 30 fps. In the observation of the colon by the endoscope, the colon continually contracts and the endoscope itself is not completely stopped. Since the movement of the colon is minute and the movement of endoscopic scope is not sudden, it is considered that there is small affection to the classification accuracy even if the staging region detection performed once in 6 frames.

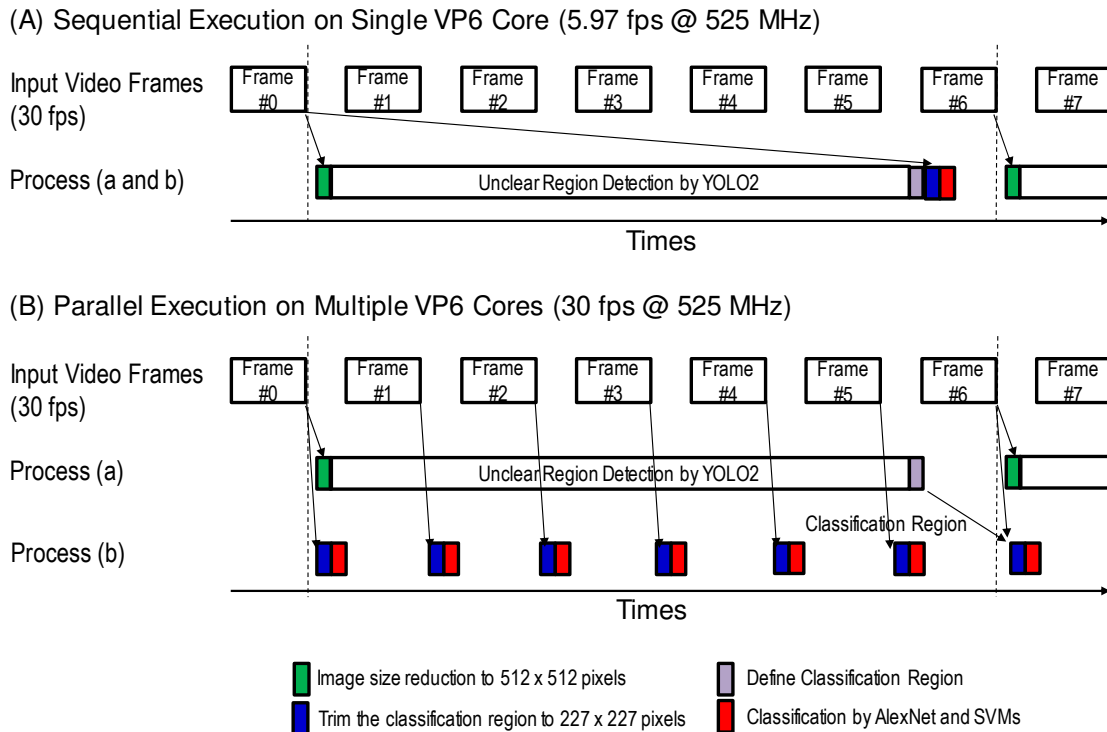


Figure 83. Timeline of the proposed CAD system with navigation function of clear region.

Thus, we confirmed both sequential execution and parallel execution of the unclear region detection and colorectal tumor classification has enough performance on MT8385 or MT6771. In the case of parallel processing, it is possible to continue to provide colorectal cancer classification information to the clinical doctors even during navigation function processing, frame number 0 to 5 in Fig.83. The clinical doctors can obtain classification information by moving the staging region by themselves or using the same area of the previous frame during navigation function processing.

Other ASICs (MT6779/SDM675/SDM710) listed in Table 15 have higher performance than MT8365/MT6771. AI benchmark tests application is released and execution performance results on these ASICs are reported [147]. The AI benchmark is designed to check the performance and the memory limitations associated with running AI and deep learning algorithms. It consists of several computer vision tasks performed by neural networks. Table 16 shows the CNN execution results on these ASICs. In AI benchmark tests, there are some CNN models for classification, recognition and so on. However, AlexNet and YOLO2 used in the proposed CAD system are not included in the AI benchmark tests. Thus, we compared four CNN models similar to AlexNet and YOLO2 in the proposed CAD system. MobileNet v2 [148] and Inception V3 [149] are tests for classification task, and, Inception-Resnet v1 [150] is a test for face recognition task. These three models represent a core set of architectures for classification problems that are suitable for mobile application in the AI benchmark tests. SRCNN [151] is a test for deblurring task which removes Gaussian blur from images.

Table 16 Execution time for CNNs on ASICs providing DSPs [147]. (MT8385 is almost same as MT6771).

	MobileNet v2	Inception V3	Inc-ResNet v1	SRCNN	Proposed CAD
	Classification	Classification	Face Recognition	Deblurring	-
MT6771	27 ms / 37 fps (x 1.0)	89 ms / 11.2 fps (x 1.0)	181 ms / 5.5 fps (x 1.0)	163 ms / 6.1 fps (x 1.0)	5.97 fps
MT6779	4 ms / 250 fps (x 6.6)	23 ms / 43.5 fps (x 3.9)	38 ms / 26.3 fps (x 4.8)	22 ms / 45.6 fps (x 7.5)	> 23.3 fps
SDM710	12 ms / 83.3 fps (x 2.3)	48 ms / 20.8 fps (x 1.9)	95 ms / 10.5 fps (x 1.9)	70 ms / 14.3 fps (x 2.3)	> 11.3 fps
SDM675	10 ms / 100 fps (x 2.7)	34 ms / 29.4 fps (x 2.6)	73 ms / 13.7 fps (x 2.5)	53 ms / 18.9 fps (x 3.1)	> 14.9 fps

From the AI benchmark tests report, it can be seen that various CNN models can be executed on low power embedded ASICs. There is no result of MT8385, however, MT8385 architecture is almost same as MT6771. We can consider the performance for CNNs of MT8385 is almost same as MT6771. For example, MT6779 has over 3.9 times performance of MT6771. Thus, the proposed CAD system can be expected to process at over 23.3 fps, and, to provide classification information while performing navigation function by MT6779 at sequential processing. We consider that the CAD system using these embedded ASICs can be added as an option in the chassis of the endoscope and can be used compactly and usable with low power consumption.

5.4. Overview of Navigation Function by Multiple Staging Region

Figure 84 shows overview of a navigation function by multiple region. As shown in Fig.84, multiple staging regions are set with overlapping in one frame. We proposed two settings: (1) one staging region is set on the center of endoscopic video frame and other four staging regions are set on upper, lower, left and right of the staging region on center, and (2) one staging region is set on the center of endoscopic video frame and other four staging regions are set on diagonals of the staging region on center. During observation, the quantitative staging classification is executed, and Type 1 probability is provided over 0.5 when lesion part is not included, or unclear region is included in the staging region. The color of rectangle which indicated staging region is changed by probability of each pathological type as shown in Fig.85. These colors of rectangles give the doctor information about which areas to look for colorectal observation. Figure 86 shows the timeline of the proposed CAD system with navigation function by multiple staging regions on two VP6 cores. Both Process for region 1,2 and Process for region 3,4,5 can be executed over 30 fps. The navigation function by multiple staging region is based on quantitative staging classification. Thus, the execution time of the navigation function is same as the execution time of quantitative staging classification. The navigation function by multiple staging region can provide navigation information to doctors every frame.

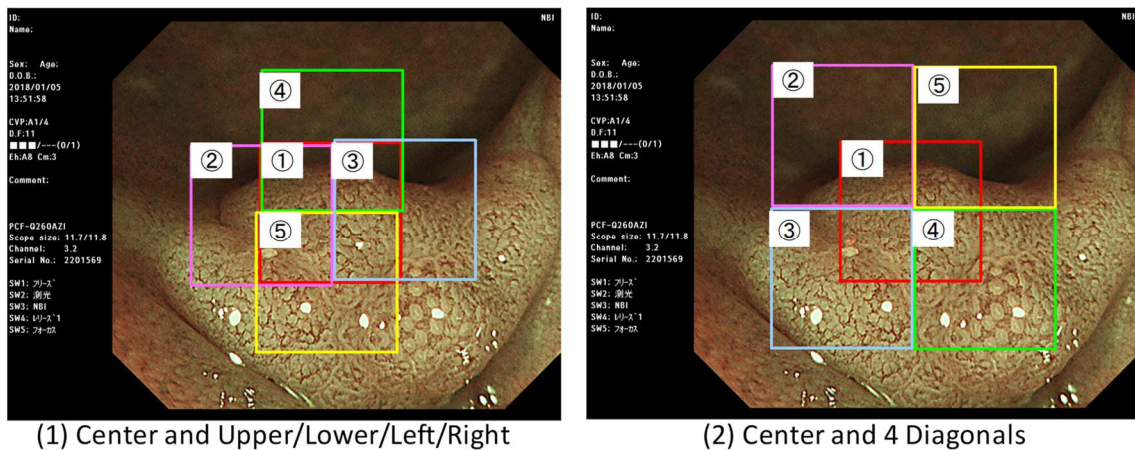


Figure 84 Navigation function by multiple staging region.

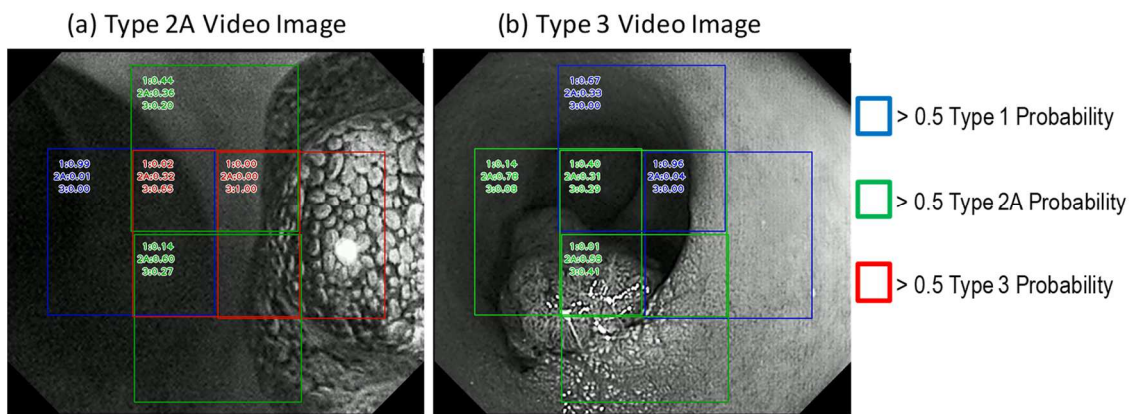


Figure 85 Staging region color by probability of each pathological type.

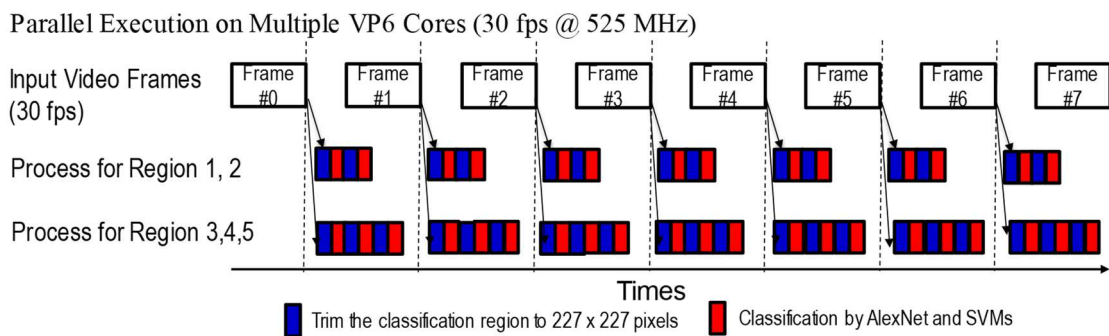


Figure 86 Timeline of navigation function by multiple staging region.

5.5. Conclusion

In this Chapter, I proposed two navigation function methods in CAD system as useful assistant feature to doctors for quantitative endoscopic observation. Two methods for navigation function is proposed, navigation function by

unclear region detection and navigation function by multiple staging region. Execution time of the navigation function by unclear region detection is estimated 5.97 fps @ 525 MHz on a customizable DSP. I compared parallel execution and sequential execution on multiple customizable DSP and confirmed parallel execution provides quantitative staging for video images of colon NBI magnification endoscopy. I confirmed the navigation function by multiple staging region is executed over 30 fps @ 525 MHz.

References

- [127] Endoscopic artefact detection challenge (EAD2019), <https://ead2019.grand-challenge.org/>
- [128] S. Ali, F. Zhou, C. Daul, B. Braden, A. Bailey, S. Realdon, J. East, G. Wagnieres, V. Loschenov, E. Grisan, W. Blondel, and J. Rittscher, "Endoscopy artifact detection (EAD 2019) challenge dataset," Computing Research Repository, CoRR, vol. abs/1905.03209, 2019.
- [129] S. Yang and G. Cheng, "Endoscopic artefact detection and segmentation with deep convolutional neural network.," Proceedings of the 2019 Challenge on Endoscopy Artefacts Detection (EAD2019), Venice, Italy, vol.2366 of CEUR Workshop Proceedings. CEUR-WS.org, 2019.
- [130] I. Oksuz, J. R. Clough, A. P. King, and J. A. Schnabel, "Artefact detection in video endoscopy using retinanet and focal loss function.," Proceedings of the 2019 Challenge on Endoscopy Artefacts Detection (EAD2019), Venice, Italy, vol.2366 of CEUR Workshop Proceedings. CEUR-WS.org, 2019.
- [131] M. A. Khan and J. Choo, "Multi-class artefact detection in video endoscopy via convolution neural networks.," Proceedings of the 2019 Challenge on Endoscopy Artefacts Detection (EAD2019), Venice, Italy, vol.2366 of CEUR Workshop Proceedings. CEUR-WS.org, 2019.
- [132] C. Zhang, N. Zhang, D. Wang, Y. Cao and B. Liu, "Artifact Detection in Endoscopic Video with Deep Convolutional Neural Networks.," Proceedings of Second International Conference on Transdisciplinary AI (TransAI), Irvine, CA, USA, pp.1-8, 2020.
- [133] S. Ali, F. Zhou, B. Braden, A. Bailey, S. Yang, G. Cheng, P. Zhang, X. Li, M. Kayser, R.D. Soberanis-Mukul, S. Albarqouni, X. Wang, C. Wang, S. Watanabe, I. Oksuz, Q. Ning, S. Yang, M.A. Khan, X.W. Gao, S. Realdon, M. Loshchenov, J.A. Schnabel, J.E. East, G. Wagnieres, V.B. Loschenov, E. Grisan, C. Daul, W. Blondel, and J. Rittscher, "An objective comparison of detection and segmentation algorithms for artefacts in

clinical endoscopy.," Scientific Reports, 10, pp.1-15, 2020.

- [134] X. Gao, B. Braden, S. Taylor and W. Pang, "Towards Real-Time Detection of Squamous Pre-Cancers from Oesophageal Endoscopic Videos.," Proceedings of 18th IEEE International Conference On Machine Learning And Applications (ICMLA), Boca Raton, FL, USA, pp.1606-1612, 2019.
- [135] L. Peng, S. Liu, D. Xie, S. Zhu and B. Zeng, "Endoscopic video deblurring via synthesis," Proceedings of 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, pp.1-4, 2017.
- [136] Sharib Ali, Felix Zhou, Adam Bailey, Barbara Braden, James East, Xin Lu, and Jens Rittscher, "A deep learning framework for quality assessment and restoration in video endoscopy.," CoRR, vol. abs/1904.07073, 2019.
- [137] F. Tan, S. Liu, L. Zeng and B. Zeng, "Notice of Removal: Kernel-free video deblurring via synthesis," Proceedings of 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, pp. 2683-2687, 2016.
- [138] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 580-587, 2014.
- [139] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'16), Las Vegas, NV, USA, pp.779-788, 2016.
- [140] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'17), Honolulu, HW, USA, pp.6517-6525, 2017.

- [141] MediaTek Inc., “MediaTek Helio P60.” <https://www.mediatek.jp/products/mediatek-helio-p60>
- [142] MediaTek Inc., “MediaTek Helio P90.” <https://www.mediatek.com/products/smartphones/mediatek-helio-p90>
- [143] Qualcomm Technologies, Inc., “Snapdragon 710 Mobile Platform.” <https://www.qualcomm.com/products/snapdragon-710-mobile-platform>
- [144] Qualcomm Technologies, Inc., “Snapdragon 675 Mobile Platform.” <https://www.qualcomm.com/products/snapdragon-675-mobile-platform>
- [145] MediaTek Inc., “i500 (MT8385).” <https://www.mediatek.com/products/aiot/i500>
- [146] VIA Technologies, Inc., “VIA SOM-9X50” <https://www.viatech.com/en/products/boards/modules/som-9x50/>
- [147] A. Ignatov, R. Timofte, A. Kulik, S. Yang, K. Wang, F. Baum, M. Wu, L. Xu, L. V. Gool, “AI Benchmark: All About Deep Learning on Smartphones in 2019,” arXiv:1910.06663v1 [cs.PF], 2019.
- [148] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” arXiv preprint arXiv:1704.04861, 2017.
- [149] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, “Rethinking the inception architecture for computer vision,” Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR’16), Las Vegas, NV, USA, pp. 2818–2826, 2016.
- [150] C. Szegedy, S. Ioffe, V. Vanhoucke, A.A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning,” AAAI. vol. 4, p. 12, 2017.

- [151] C. Dong, C. C. Loy, K. He and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295-307, 1 Feb. 2016

Chapter 6. Conclusion

6.1. Conclusion

In this thesis, I described the hardware implementation of a computer-aided diagnosis support system that supports endoscopic video using a customizable DSP. I clarified the problem of unclear frames in endoscopic video images and described the balancing of training dataset to build a robust CAD system with CNN feature extraction and SVM classification. I carefully analyzed bottlenecks when the CAD system is installed into the customizable DSP core and clarified optimization methods so that processing can be executed in real time in the hardware prototype system to be developed. I evaluated the developed CAD system with these improvement approaches by results of classification. And, I confirmed the developed CAD system with CNN feature extraction and SVM classification can be executed in real-time for endoscopic video images with high classification accuracy.

In Chapter 1, I explained the background of this research. First, I described trends in colorectal cancer incidence and mortality worldwide. After that, I explained the principle of NBI magnified endoscopy and described the classification of colorectal cancer by the NBI endoscopy observation.

In Chapter 2, I explained how the conventional BoF-based CAD system works, then explained the CAD system with CNN feature extraction and SVM classification. I described the difference between outputs from CNN layers equivalence to feature quantity and the features by D-SIFT, and further described SVM libraries to be implemented referred the report of our research group. I confirmed over 90% accuracy for quantitative staging classification with pre-trained CNN instead of D-SIFT by limited training data.

In Chapter 3, I clarified the problems that affect the classification accuracy in endoscopic video images. Also, the endoscopic image dataset used in this study was explained, and the problems caused by the imbalanced dataset were described. To solve the problem of imbalanced dataset, I prepared the balancing dataset and evaluated the result of classification accuracy from SVM trained by the balancing dataset. As a result, I confirmed that the classification accuracy of SVM is improved by using the balancing dataset. I described the application to endoscopic video images. I explained some issues such as blurring, color shift etc. in the endoscopic video image. I evaluated the result of classification accuracy when the endoscopic video image was

input to the developed CAD system. It was shown that the classification accuracy was improved by using multiple size image dataset for SVM training. I confirmed average value and standard deviation of the output of SVM for quantitative staging classification ($0.692 \pm 0.328 \rightarrow 0.995 \pm 0.015$ for Type2A/3 staging).

In Chapter4, I showed that the acceleration of convolutional layer consists of multiply and accumulate is the key to CNN processing. I clarified the requirement for hardware to be implemented a CAD system with CNN and SVM. I decided to implement a CAD system on a customizable DSP, and performed a detailed analysis of bottlenecks, and explained the optimization method. I implemented the optimized CAD system with CNN and SVM on the FPGA-based prototyping system and confirmed that it is possible to classify endoscopic images in real time. I confirmed real-time processing of CAD system with staging classification for edge device at small clinics. (44.6 fps / 22 ms latency @ 200MHz, 66.6 mW)

In Chapter 5, And, I proposed two navigation function methods by unclear region detection and by multiple staging region. I confirmed that it is possible to realize the navigation function by unclear region detection using YOLO2 and staging classification by AlexNet and SVMs executed at 30 fps on the customizable embedded DSP core. I confirmed that it is possible to realize the navigation function by multiple staging region at 39fps on the customizable DSP core.

I described the real-time CAD system with quantitative staging and navigation on customizable embedded DSP. I proposed a CAD system that provides quantitative and objective index of cancer stage to doctors based on the standard JNET classification, not only classify cancer or not-cancer and implemented on a customizable DSP. And, I confirmed the developed CAD system achieved real-time quantitative staging classification for the endoscopic video image (44.6fps throughput / 22ms latency @200MHz, 66.6mW power consumption), and sufficient classification accuracy (> 90%). And, I proposed two navigation functions which provide unclear region information to doctors in the CAD system, and I confirmed that it is possible to realize the CAD system with the navigation function on the customizable embedded DSP core.

Academic and Industrial impact

- 1) CAD system for real-time video on customizable DSP
 - Memory and cycles reduction by 8-bit quantized AlexNet and SVM implementation. (Academic)
 - Avoiding system bus conflict by using hidden layer feature. (Academic)
 - 44.6 fps and 22 ms latency, 66.6 mW power (Industrial)
- 2) Real-time CAD with over 90% accuracy
 - Pre-trained CNN instead of Handcrafted feature Extraction (Academic)
 - Multi-sizing and balancing in training data set (Academic)
 - Quantitative staging with 90% accuracy (Industrial)
- 3) Real-time Navigation for effective quantitative staging
 - unclear region detection for one staging region (Academic), 30fps @ 525 MHz (Industrial)
 - multiple staging regions (Academic), 39fps @525 MHz (Industrial)

Figure 87 shows comparison of performance, power and diagnosis quality to previous studies. We have achieved high performance, power consumption and diagnostic quality compared to previous studies.

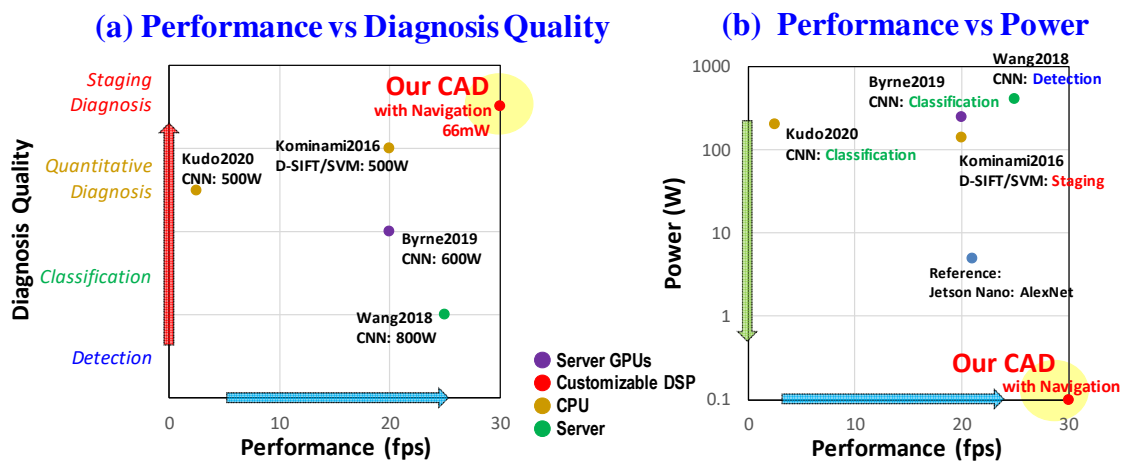


Figure 87 Comparison of (a) performance vs diagnosis quality and (b) performance vs power.

From the above research, quantitative and objective staging index are provided to the doctor more accurately in magnified NBI endoscopic observation, which is independent from the experience of doctors, and

diagnostic support method is established such as a "second opinion" at magnified NBI endoscopic observation on site.

6.2. Future Works

It has been shown that it is possible to construct a CAD system that can process endoscopic video images in real time, however some problems still remain.

In our research group, there are enough endoscopic images for each pathological type which has clearly captured and trimmed the lesion by the clinical doctor. Currently, there is a few images for unclear parts including blur, reflection of light or color shift. Therefore, as future research subjects, 1) we accumulate the training data sets for navigation function by CNN, classification, and verify the practical CAD system capability. Unclear region can be detected by YOLO2, however, there are lesion part or normal mucosa in the bounding box output by YOLO2. Thus, 2) we evaluate not only YOLO2 for unclear region detection quantitatively using statistical measure such as mAP (mean Average Precision) and implement with classification into the customizable DSP core. In the proposed CAD system, direction to clear region is navigated by a red arrow. 3) We consider how to provide the navigation and improve the navigation function which allows clinical doctors to freely switch the warning display according to their level of experience.

Expansion to endoscopic systems other than NBI (Narrow Band Imaging) is also a future research topic. The proposed CAD system can classify NBI images magnified from medium-magnification (1.3-1.9x) to strong-magnification (3.6x). It is considered to apply non-magnified endoscopic images and high magnification (360x) by endocytoscopy. In addition, it is considered to apply endoscopic system using engineering methods different from NBI such as WLI (White Light Imaging) and BLI (Blue Laser Imaging).

Optimization of the customizable DSP core by adding user-defined instruction set is also remained. This makes it possible to perform more complicated processing in neural networks proposed in recent years not only AlexNet or YOLO2. By using a neural network proposed in recent years, a CAD system with higher classification accuracy can be expected.

Furthermore, it has been reported an endoscopic system controlled by voice recognition [152] and voice recognition implemented on the customizable DSP for medical devices [153]. Therefore, assistance by voice recognition in the

operation of a doctor's endoscope is also conceivable.

References

- [152] S. Afonso, I. Laranjo, J. Braga, V. Alves, and J. Neves, "Endoscopic Procedures Control Using Speech Recognition.," *Advances in Information Science and Applications*, vol.2, pp.404-409, 2014.
- [153] A. Yamada, M. Tsuji, N. Tamba, M. Odagawa, and Y. Fujinaga, "Development of intra-operative sterile image reference system manipulated by voice recognition.," *Proceedings of the 79th Annual Meeting of the Japan Radiological Society*, Yokohama, Japan, 2020

Publications

- (1) A Hardware Implementation on Customizable Embedded DSP Core for Colorectal Tumor Classification with Endoscopic Video toward Real-Time Computer-Aided Diagnosis System.
Masayuki Odagawa, Takumi Okamoto, Tetsuhi Koide, Toru Tamaki, Bisser Raytchev, Kazufumi Kaneda, Shigeto Yoshida, Hiroshi Mieno, Shinji Tanaka, Takayuki Sugawara, Hiroshi Toishi, Masayuki Tsuji and Nobuo Tamba
IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences, Vol.E104-A77, No.4, pp. 691-701 (2021).
DOI: 10.1587/transfun.2020EAP1069

- (2) Feasibility Study for Computer-Aided Diagnosis System with Navigation Function of Clear Region for Real-Time Endoscopic Video Image on Customizable Embedded DSP Cores.
Masayuki Odagawa, Tetsushi Koide, Toru Tamaki, Shigeto Yoshida, Hiroshi Mieno, Shinji Tanaka
IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences, Vol.E105-A, No.1, pp. -, In Press
DOI: 10.1587/transfun.2021EAL2044

- (3) Classification with CNN features and SVM on Embedded DSP Core for Colorectal Magnified NBI Endoscopic Video Image.
Masayuki Odagawa, Takumi Okamoto, Tetsushi Koide, Toru Tamaki, Shigeto Yoshida, Hiroshi Mieno, Shinji Tanaka
IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences, Vol.E105-A, No.1, pp. -, In Press
DOI: 10.1587/transfun.2021EAP1036

Presentations on Conferences

International conference

First Author

1. A Hardware Implementation of Colorectal Tumor Classification for Endoscopic Video on Customizable DSP toward Real-time Computer-Aided Diagnosis System.
Masayuki Odagawa, Takumi Okamoto, Tetsuhi Koide, Toru Tamaki, Bisser Raytchev, Kazufumi Kaneda, Shigeto Yoshida, Hiroshi Mieno, Shinji Tanaka, Takayuki Sugawara, Hiroshi Toishi, Masayuki Tsuji, and Nobuo Tamba
Proceedings of 2019 IEEE International Symposium on Circuits and Systems, Sapporo, Japan, pp.1-5 (2019), Oral Session.
DOI: 10.1109/ISCAS.2019.8702379
2. Classification Method with CNN features and SVM for Computer-Aided Diagnosis System in Colorectal Magnified NBI Endoscopy.
Masayuki Odagawa, Takumi Okamoto, Tetsuhi Koide, Toru Tamaki, Bisser Raytchev, Kazufumi Kaneda, Shigeto Yoshida, Hiroshi Mieno, and Shinji Tanaka
Proceedings of IEEE Regional 10 Conference, Online, Japan, pp.1-6 (2020), Oral Session.
DOI : 10.1109/TENCON50793.2020.9293709
3. Tensilica DSP cores for neural networks and an application to computer-aided diagnosis system for colorectal tumor classification.
Masayuki Odagawa, Masayuki Tsuji, Hiroshi Toishi, Takayuki Sugawara, Nobuo Tamba, Takumi Okamoto, Tetsushi Koide, Toru Tamaki, Bisser Raytchev, Kazufumi Kaneda, Shigeto Yoshida, Hiroshi Mieno, and Shinji Tanaka
Proceedings of the International Workshop on Nanodevice Technologies 2018, Higshi-Hiroshima, Japan, pp.6-7 (2018), Oral Session.
4. Image Identification System on Rapid Prototyping Platform with Customizable DSP Core for Biomedical Application.
Masayuki Odagawa, Masayuki Tsuji, Hiroshi Toishi, Takayuki Sugawara, Nobuo Tamba, Takumi Okamoto, Tetsushi Koide, Toru Tamaki, Bisser Raytchev, Kazufumi Kaneda, Shigeto Yoshida, Hiroshi Mieno, and Shinji Tanaka
The 3rd International Symposium on Biomedical Engineering, Higshi-Hiroshima, Japan, November 2018, Poster Presentation.

Coauthor

1. Implementation of Computer-Aided Diagnosis System on Customizable DSP Core for Colorectal Endoscopic Images with CNN Features and SVM.
Takumi Okamoto, Tetsuhi Koide, Toru Tamaki, Bisser Raytchev, Kazufumi Kaneda, Shigeto Yoshida, Hiroshi Mieno, Shinji Tanaka, Takayuki Sugawara, Hiroshi Toishi, Masayuki Tsuji, Masayuki Odagawa, and Nobuo Tamba
Proc. of 2018 IEEE Regional 10 Conference, Jeju, pp.1663-1666 (2018), Oral Session
DOI: 10.1109/TENCON.2018.8650331
2. Feature Extraction of Colorectal Endoscopic Images for Computer-Aided Diagnosis with CNN.
Takumi Okamoto, Masayuki Odagawa, Tetsuhi Koide, Toru Tamaki, Bisser Raytchev, Kazufumi Kaneda, Shigeto Yoshida, Hiroshi Mieno and Shinji Tanaka
Proc. of 2019 2nd International Symposium on Devices, Circuits and Systems, Higashi-Hiroshima, pp.1-4 (2019), Oral Session
DOI: 10.1109/ISDCS.2019.8719104
3. Real-time processing of computer-aided diagnosis system for colorectal tumor classification in NBI endoscopy using CNN features by implementing to Tensilica Vision P6 DSP.
Masayuki Tsuji, Hiroshi Toishi, Takayuki Sugawara, Masayuki Odagawa, Nobuo Tamba, Takumi Okamoto, Tetsushi Koide, Toru Tamaki, Bisser Raytchev, Kazufumi Kaneda, Shigeto Yoshida, Hiroshi Mieno, and Shinji Tanaka
Proc. of the 2nd International Symposium on Biomedical Engineering, Tokyo, Japan, pp.2425-2426 (2017), Oral Session
4. Implementation of Computer-Aided Diagnosis System on Customizable DSP Core for Colorectal Endoscopic Images with CNN features and SVM.
Takumi Okamoto, Tetsuhi Koide, Toru Tamaki, Bisser Raytchev, Kazufumi Kaneda, Shigeto Yoshida, Hiroshi Mieno, Shinji Tanaka, Masayuki Odagawa, Takayuki Sugawara, Hiroshi Toishi, Masayuki Tsuji, and Nobuo Tamba
The 2nd International Symposium on Biomedical Engineering, Tokyo, Japan, November 2017, Poster Presentation.

5. An Improvement of Real-Time Computer-Aided Diagnosis System for Colorectal Endoscopic Video.
Hiroki Iwata, Masaya Ueda, Guan Juangang, Tetsushi Koide, Toru Tamaki, Bisser Raytchev, Kazufumi Kaneda, Shigeto Yoshida, Hiroshi Mieno, Shinji Tanaka, Masayuki Odagawa, Hiroshi Toishi, Takayuki Sugawara, Masayuki Tsuji, and Nobuo Tamba
The 4th International Symposium on Biomedical Engineering, Hamamatsu, Japan, November 2019, Poster Presentation

6. A Method to Improve Recognition Rate of Computer-Aided Diagnosis System for Colorectal Endoscopic Images.
Masaya Ueda, Hiroki Iwata, Masayuki Odagawa, Guan Juangang, Tetsushi Koide, Toru Tamaki, Bisser Raytchev, Kazufumi Kaneda, Shigeto Yoshida, Hiroshi Mieno, and Shinji Tanaka
The 4th International Symposium on Biomedical Engineering, Hamamatsu, Japan, November 2019, Poster Presentation

Domestic conference

Coauthor

1. CNN 特徴と SVM 分類を適用した大腸内視鏡画像がん診断支援システムのカスタマイザブル DSP コアへの実装(Implementation of Computer-Aided Diagnosis System for Colorectal Endoscopic Images with CNN features and SVM on Customizable DSP Core.)Takumi Okamoto, Tetsuhi Koide, Toru Tamaki, Bisser Raytchev, Kazufumi Kaneda, Shigeto Yoshida, Hiroshi Mieno, Shinji Tanaka, Takayuki Sugawara, Hiroshi Toishi, Masayuki Tsuji, Masayuki Odagawa and Nobuo Tamba
Proc. of Design Automation Symposium 2017, Kaga, pp.33-38 (2017), Oral Session

2. 大腸がん診断支援のためのカスタマイザブル DSP による内視鏡動画リアルタイム CNN 特徴抽出と SVM 分類(Implementation of Computer-Aided Diagnosis System for Colorectal Endoscopic Images with CNN features and SVM on Customizable DSP Core.)
Takumi Okamoto, Tetsuhi Koide, Toru Tamaki, Bisser Raytchev, Kazufumi Kaneda, Shigeto Yoshida, Hiroshi Mieno, Shinji Tanaka, Takayuki Sugawara, Hiroshi Toishi, Masayuki Tsuji, Masayuki Odagawa and Nobuo Tamba

Proc. of Design Automation Symposium 2018, Kaga, pp.39-44 (2018), Oral Session

3. 機械学習による内視鏡動画像リアルタイム診断支援システムのプロトタイピング(Prototyping of Real-time Computer-Aided Diagnosis System for Colorectal Endoscopic Movies and Images with Machine Learning.)

Takumi Okamoto, Masayuki Odagawa, Koujirou Takebayashi, Mikihisa Nagano, Tetsushi Koide, Toru Tamaki, Bisser Raytchev, Kazufumi Kaneda, Shigeto Yoshida, Hiroshi Mieno, Shinji Tanaka, Takayuki Sugawara, Hiroshi Toishi, Masayuki Tsuji and Nobuo Tamba

デザインガイア 2018, Hiroshima, Japan

IEICE Technical Report, vol.118, no.334, VLD2018-42, pp.13-18 (2018), Oral Session

Acknowledgement

I'm working for Cadence Design Systems (Japan) B.V. In October 2018, I joined Koide laboratory as a doctoral student. I have received many mentorships, advices and supports from many people in working this study until today. Therefore, I would like to take this opportunity to show my greatest appreciation to all of them.

First and foremost, I would like to show my deepest appreciation to my supervisor Associate Professor Tetsushi Koide for immense mentorships and supports. I appreciate all his contributions of time, advices, ideas and stimulating my experiences.

I would like to express again my deepest appreciation to my sub-supervisors, Professor Shin-Ichiro Kuroki, Professor Akinobu Teramoto, Professor Minoru Fujishima and Professor Suguru Kameda for reviewing my doctoral thesis and advising on my research.

In addition, I would like to express my deep gratitude to Professor Seiichiro Higashi, Professor Masakazu Iwasaka, Associate Professor Anri Nakajima, Associate Professor Shuhei Amakawa, Associate Professor Takeshi Yoshida, Associate Professor Masataka Miyake, Associate Professor Hiroaki Hanafusa, Associate Professor Mamoru Sasaki, at Department of Semiconductor Electronics and Integration Science, Graduate School of Advanced Sciences of Matter, Hiroshima University. I also would like to express my deep gratitude to the professors and the staff at Research Institute for Nanodevice and Bio Systems, Hiroshima University.

I would like to show my grateful appreciation to Professor Kazufumi Kaneda, and Associate Professor Bisser Raytchev, Graduate School of Engineering, Hiroshima University for cooperation of research. I would like to show my grateful appreciation to Associate Professor Toru Tamaki, Department of Computer Science, Nagoya Institute of Technology. I would like to show my grateful appreciation to Professor Shigeto Yoshida, and Dr. Hiroshi Mineno, JR Hiroshima Hospital for cooperation of research. I would like to show my grateful appreciation to Professor Shinji Tanaka, Department of Endoscopy, Hiroshima University for cooperation of research.

I'm deeply grateful to Mr. Takumi Okamoto, Cadence Design Systems (Japan) B.V., for many discussions, advices and supports since he was a doctoral student in Hiroshima University.

I would like to express my special thanks to Mr. Toshifumi Kaneko,

President of Cadence Design Systems (Japan) B.V., for permission to study in a doctoral course and assistance in a study.

I would like to express deepest thanks to Dr. Nobuo Tamba, CEO of T2Laboratory, for an encouragement to enter a doctoral course and enormous suggestions and advices since he was my boss in Cadence Design Systems (Japan) B.V.

I would like to show grateful appreciation to Mr. Hiroshi Toishi for enormous cooperation of hardware implementation of the computer-aided diagnosis system in Chapter 4 when he was my colleague in Cadence Design Systems (Japan) B.V.

I would like to express deepest appreciation to my colleagues, Mr. Takayuki Sugawara and Dr. Masayuki Tsuji, Cadence Design Systems (Japan) B.V., for grateful cooperation of implementation and optimization onto a customizable DSP cores in Chapter 4.

I would like to show my special thanks to R&D teams in Cadence Design Systems Inc. and colleagues in Cadence Design Systems (Japan) B.V. for many supports and cooperation.

Finally, I would like to express deepest appreciation to my family for their cooperation and patience. It took long time to write this thesis since I decided to enter a doctoral course. I would like to apologize my family for placing large strain.

Masayuki Odagawa

Yokohama, Kanagawa, Japan

July, 2021

公表論文 (Articles)

- (1) A Hardware Implementation on Customizable Embedded DSP Core for Colorectal Tumor Classification with Endoscopic Video toward Real-Time Computer-Aided Diagnosis System.
Masayuki Odagawa, Takumi Okamoto, Tetsuhi Koide, Toru Tamaki, Bisser Raytchev, Kazufumi Kaneda, Shigeto Yoshida, Hiroshi Mieno, Shinji Tanaka, Takayuki Sugawara, Hiroshi Toishi, Masayuki Tsuji and Nobuo Tamba
IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences, Vol.E104.A, No.4, pp.691-701 (2021).
DOI: 10.1587/transfun.2020EAP1069

- (2) Feasibility Study for Computer-Aided Diagnosis System with Navigation Function of Clear Region for Real-Time Endoscopic Video Image on Customizable Embedded DSP Cores.
Masayuki Odagawa, Tetsushi Koide, Toru Tamaki, Shigeto Yoshida, Hiroshi Mieno, Shinji Tanaka
IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences, Vol.E105-A, No.1, pp.-, In press.
DOI: 10.1587/transfun.2021EAL2044

- (3) Classification with CNN features and SVM on Embedded DSP Core for Colorectal Magnified NBI Endoscopic Video Image.
Masayuki Odagawa, Takumi Okamoto, Tetsushi Koide, Toru Tamaki, Shigeto Yoshida, Hiroshi Mieno, Shinji Tanaka
IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences, Vol.E105-A, No.1, pp.-, In press.
DOI: 10.1587/transfun.2021EAP1036

参 考 論 文

(Thesis Supplements)

- (1) A Hardware Implementation of Colorectal Tumor Classification for Endoscopic Video on Customizable DSP toward Real-time Computer-Aided Diagnosis System.

Masayuki Odagawa, Takumi Okamoto, Tetsuhi Koide, Toru Tamaki, Bisser Raytchev, Kazufumi Kaneda, Shigeto Yoshida, Hiroshi Mieno, Shinji Tanaka, Takayuki Sugawara, Hiroshi Toishi, Masayuki Tsuji and Nobuo Tamba
Proceedings of 2019 IEEE International Symposium on Circuits and Systems, Sapporo, Japan, pp.1-5 (2019).

DOI: 10.1109/ISCAS.2019.8702379

- (2) Classification Method with CNN features and SVM for Computer-Aided Diagnosis System in Colorectal Magnified NBI Endoscopy.

Masayuki Odagawa, Takumi Okamoto, Tetsuhi Koide, Toru Tamaki, Bisser Raytchev, Kazufumi Kaneda, Shigeto Yoshida, Hiroshi Mieno and Shinji Tanaka
Proceedings of IEEE Regional 10 Conference, Online, Japan, pp.1-6 (2020)

DOI : 10.1109/TENCON50793.2020.9293709