

الجمهورية الجزائرية الديمقراطية الشعبية  
People's Democratic Republic of Algeria  
وزارة التعليم العالي والبحث العلمي  
Ministry of Higher Education and Scientific Research  
جامعة غرداية  
University of Ghardaia  
كلية العلوم والتكنولوجيا  
Faculty of Science and Technology  
قسم الرياضيات والاعلام الآلي  
Department of Mathematics and Computer Science

## ترخيص بمناقشة مذكرة ماستر Master Thesis Defense Permission

I, the undersigned, Mr(s). **Slimane Oulad-Naoui**. Hereby certify that I have examined the work entitled : **A Deep Learning based Model for Colorectal Cancer Detection**. Presented to the partial fulfillment of the Master degree in Computer Science by: **Mansoura Ballou** and **Asma Harrouzi**

I reviewed the document, and declare it is free from any serious defaults and respects the academic integrity rules. Furthermore, my jury member proposal is the following:

..... Univ.Ghardaia President

..... Univ.Ghardaia Examiner

..... Univ.Ghardaia Supervisor

Issued for all due intents and purposes.

Ghardaia, on June, 9, 2024

Signature



الجمهورية الجزائرية الديمقراطية الشعبية  
People's Democratic Republic of Algeria  
وزارة التعليم العالي والبحث العلمي  
Ministry of Higher Education and Scientific Research



جامعة غرداية  
University of Ghardaia  
كلية العلوم والتكنولوجيا  
Faculty of Science and Technology  
قسم الرياضيات والاعلام الآلي  
Department of Mathematics and Computer Science  
مخبر الرياضيات و العلوم التطبيقية  
Mathematics and Applied Sciences Laboratory

Registration N:

...../...../...../...../.....

## Master Thesis

Presented to obtain the academic Master diploma in Computer Science  
Specialty: Intelligent Systems for Knowledge Extraction

### THEME

# Colorectal Cancer: Segmentation and Detection via Deep Learning Detection

Presented by

Mansoura Ballou & Asma Harrouzi

Defended under the ministerial decree 1275 before the jury Members:

Mr. Slimane BELLAOUAR	MCA	Univ.Ghardaia	President
Mr. ADJILA ABDERRAHMANE	MAA	Univ.Ghardaia	Examiner
Mr. Slimane OULAD-NAOUI	MCB	Univ.Ghardaia	Supervisor
Ms. Farialle LAHRECHE	PhdS	Univ.Ghardaia	Co-Supervisor

University Year: 2023/2024

---

## Acknowledgement

We begin by expressing our deepest gratitude to the Almighty God, whose guidance and strength enabled us to complete this work.

We extend heartfelt thanks to our supervisor, Mr. Oulad Naoui Slimane, and our co-supervisor, Ms. Lahrach Ferial, for their invaluable guidance, unwavering support, and expertise that significantly enriched our research. Their insightful direction and contributions to our project were instrumental in its completion. Additionally, we express our gratitude to both supervisors for their pivotal role in selecting the important thesis title. Delving into this topic was a rewarding experience, and their guidance in this aspect was crucial.

We are grateful to our families for their unwavering support, constant encouragement, and understanding throughout this academic journey.

Lastly, we extend our thanks to all who have assisted us directly or indirectly in completing this work. Their contributions have been invaluable, and we deeply appreciate their support.

## ملخص

وفقاً لمنظمة الصحة العالمية WHO أصبح سرطان القولون والمستقيم (CRC) ثالث أكثر أنواع السرطان شيوعاً في الجزائر. وهو أيضاً السبب الرئيسي الثالث لوفيات السرطان في العالم. يعد سرطان غدي القولون والمستقيم أحد (CRA) أنواع سرطان القولون والمستقيم المعروفة، والذي يمكن تصنيفه إلى الدرجات من الأولى إلى الرابعة وفقاً لمعايير برودر. يعتبر فحص الشرائح تحت المجهر (التحليل النسيجي المرضي) هو أفضل طريقة للأطباء للتمييز بين أنواع سرطان القولون، إلا أنها مهمة صعبة، خاصة بالنسبة للسرطانات في المراحل المتقدمة التي يصعب تشخيصها. هذا التعقيد يمكن أن يجعل التشخيص يستغرق وقتاً طويلاً وقد يؤدي إلى تشخيص خاطئ. ولذلك، فقد أصبحت أجهزة الكمبيوتر مفيدة جداً في تخفيف عبء العمل عن الأطباء. وبما أن الصور الطبية معقدة بطبيعتها، مما يجعل من الصعب تحليلها وتجزئتها. ففي هذه الدراسة، اقترحنا استخدام أحد نماذج التعلم العميق لتجزئة الصور وهو قناع RCNN للتنبؤ بأقنعة لكل غدة في الصور النسيجية. قننا بعد ذلك بتطبيق خوارزمية مستجمعات المياه لفصل أقنعة الغدة المتداخلة المتوقعة خطأً. حقق هذا النهج نتائج جيدة على مجموعة بيانات Glas حيث بلغت درجة F1 87%، ومعامل النرد 86%، و IoU 78%. على الرغم من النتائج الواعدة، واجهت الدراسة تحديات، مثل كون نموذج RCNN ثقيلاً من حيث المتطلبات الحاسوبية، مما جعل التدريب يستغرق وقتاً طويلاً، كما ان هناك نقص في الصور التشريحية المرضية لسرطان القولون والمستقيم.

## كلمات مفتاحية

سرطان القولون والمستقيم، تجزئة الصور الطبية، قناع RCNN، خوارزمية مستجمعات المياه.

---

## Abstract

The World Health Organization (WHO) reports Colorectal Cancer (CRC) as the third most common cancer in Algeria and the third most common cause of cancer-related deaths worldwide. CRC cases commonly involve colorectal adenocarcinoma (CRA), which is categorized into grades I-IV based on Broder criteria. Examining slides under a microscope (histopathological analysis) remains the optimal method for doctors to distinguish between different types of colon cancer. However, this task is particularly demanding for low-grade cancers, leading to potential time-consuming diagnoses and the risk of misdiagnosis. Consequently, computers are increasingly valuable in alleviating the diagnostic burden for medical professionals. Medical images are complex in nature, making them difficult to analyze and segment. In this study, we propose using deep learning image segmentation techniques: Mask R-CNN to predict instance masks for each gland in histological images. We then apply the watershed algorithm to handle the overlapping predicted gland masks. This approach yields good results on the GlaS dataset, with an F1 score of 0.87, a Dice coefficient of 0.86, and an IoU of 0.78. Despite the promising results, We face challenges, such as the heavy computational requirements of the Mask R-CNN model, which makes training time-consuming, and the limited availability of histological CRC images.

**Keywords:** Colorectal Cancer, Medical Image Segmentation, Mask-RCNN, Watershed algorithm.

---

---

## Résumé

L'Organisation Mondiale de la Santé (OMS) reporte le cancer colorectal (CRC) comme le troisième cancer en Algérie et la troisième cause de décès par cancer dans le monde. Les cas de CRC impliquent souvent un adénocarcinome colorectal (ARC), qui peut être classé en grades 1 à 4 selon les critères de Broder.

Bien que l'examen des lames au microscope (analyse histopathologique) soit le meilleur moyen pour les médecins de différencier les types de cancer du côlon, il s'agit d'une tâche difficile, en particulier pour les cancers de bas grade difficiles à diagnostiquer. Cette complexité peut rendre le diagnostic long et conduire à un diagnostic erroné. Par conséquent, les ordinateurs deviennent de plus en plus utiles pour alléger la charge de travail des médecins.

Les images médicales sont de nature complexe, ce qui les rend difficiles à analyser et à segmenter. Dans cette étude, nous proposons d'utiliser des techniques de segmentation d'images par apprentissage profond : Mask R-CNN pour prédire les masques d'instance pour chaque glande dans les images histologiques. Nous appliquons ensuite l'algorithme de bassin versant pour gérer les masques de glandes prédits qui se chevauchent. Cette approche produit de bons résultats sur l'ensemble de données GlaS, avec un score F1 de 0,87, un coefficient Dice de 0,86 et un IoU de 0,78. Malgré les résultats prometteurs, nous avons confronté à des défis, tels que les lourdes exigences de calcul du modèle Mask R-CNN, ce qui a rendu l'entraînement long, et la disponibilité limitée des images histologiques du CRC.

**Mots clés :** Cancer Colorectal, Segmentation d'Images Médicales, Mask-RCNN, Algorithmes des bassins versants.

---

---

# CONTENTS

<b>List of Figures</b>	<b>iv</b>
<b>List of Tables</b>	<b>v</b>
<b>Introduction</b>	<b>1</b>
<b>1 Preliminaries</b>	<b>4</b>
1.1 Introduction . . . . .	4
1.2 Medical Imaging and Cancer Diagnostic . . . . .	5
1.2.1 Imaging Modalities . . . . .	6
1.2.2 Medical Image File Formats . . . . .	12
1.2.3 Histopathology Image Analysis . . . . .	12
1.3 Machine learning, Image Segmentation and Related Tasks . . . . .	14
1.3.1 Preprocessing . . . . .	14
1.3.2 Feature Extraction . . . . .	15
1.3.3 Segmentation . . . . .	15
1.3.4 Object Detection . . . . .	17
1.3.5 Classification . . . . .	18
1.4 Deep Learning . . . . .	19
1.4.1 Convolutional Neural Networks . . . . .	19
1.4.2 Region-Based Convolutional Neural Network . . . . .	23
1.4.3 Fast R-CNN . . . . .	26
1.4.4 Faster R-CNN . . . . .	28

---

1.4.5	Mask R-CNN . . . . .	32
1.5	Conclusion . . . . .	36
<b>2</b>	<b>Colorectal Cancer Recognition: State Of The Art</b>	<b>37</b>
2.1	Introduction . . . . .	37
2.2	Datasets and Evaluation Metrics . . . . .	38
2.2.1	Datasets . . . . .	38
2.2.2	Evaluation Metrics . . . . .	38
2.3	Literature Review . . . . .	40
2.3.1	Classical Techniques . . . . .	41
2.3.2	Machine Learning-based Techniques . . . . .	42
2.3.3	Deep Learning-based Techniques . . . . .	43
2.4	Conclusion . . . . .	45
<b>3</b>	<b>Experiment</b>	<b>47</b>
3.1	Introduction . . . . .	47
3.2	Implementation Setup . . . . .	47
3.2.1	Dataset . . . . .	48
3.2.2	Environment . . . . .	49
3.3	Our Approach . . . . .	50
3.4	Results and Discussion . . . . .	52
3.4.1	Results . . . . .	52
3.4.2	Discussion . . . . .	54
3.5	Conclusion . . . . .	55
	<b>Conclusion</b>	<b>57</b>
	<b>Bibliography</b>	<b>59</b>



---

# LIST OF FIGURES

1.1	(a) Multimodal Discriminative Model: Deep learning architectures integrate image data using convolutional networks and non-image data using generic deep networks. (b) Generative Model: Convolutional neural networks generate images for tasks like picture-to-image regression, super-resolution enhancement, and unique image creation. Esteva et al. (2021).	5
1.2	CT colonographic images in a 64-year-old male patient with true-positive findings (a) axial. (b) sagittal images of a 4 mm sessile polyp (arrow) in the sigmoid colon. (c) virtual colonoscopy image of the same lesion (arrow). (d) Colonoscopy image Kosov et al. (2020).	8
1.3	Normal midline brain MRI. <i>image from: <a href="https://radiopaedia.org/articles/mri-2">https://radiopaedia.org/articles/mri-2</a></i>	9
1.4	Retrorectal cystic mass. Key (2018)	10
1.5	Cancer Histopathologic image of colonic carcinoid. <i>Image from: <a href="https://handwiki.org/wiki/Medicine:Histopathology_of_colorectal_carcinoma">https://handwiki.org/wiki/Medicine:Histopathology_of_colorectal_carcinoma</a></i>	11
1.6	PET scan with persistent and increasing Fluorodeoxyglucose (FDG) uptake in the distal small bowel (arrow). Barge et al. (2021)	11
1.7	A temporal subtraction image created from previous and current bone scan images. The computer correctly identified one cold lesion (white solid circle) and two hot lesions (dark dotted circles) on the subtraction image, highlighting its potential to enhance interval changes between successive whole-body bone scans Doi (2007).	13
1.8	Main Stages in Traditional ML for HI Analysis Tharwat et al. (2022).	14

1.9	Semantic Segmentation vs Instance Segmentation Odemakinde (2023).	16
1.10	Network architecture of CNN Tsai and Tao (2021).	21
1.11	Regions with CNN features. Odemakinde (2023)	24
1.12	Hierarchical grouping algorithm finds objects at different scales, as the girl is contained by the TV van de Sande et al. (2011).	24
1.13	The sample shown in this section depicts, object detection with precise boundary boxes Blue and Brindha (2019).	25
1.14	R-CNN process Huang (2017).	26
1.15	Fast R-CNN process Huang (2017).	27
1.16	Input feature map (top left), output feature map (bottom right), blue box is the ROI (top right) Hui (2018).	28
1.17	The deep network inside the system is used instead of an external region proposal Hui (2018).	29
1.18	The diagram shows the $8 \times 8$ feature maps with a $3 \times 3$ filter, and it outputs a total of $8 \times 8 \times 3$ ROIs (for $k = 3$ ) Hui (2018).	30
1.19	offsets $\delta x$ and $\delta y$ Hui (2018).	31
1.20	a specific anchor but different locations share the same anchor shapes Hui (2018).	32
1.21	Region Proposal Network (RPN)s Ren et al. (2015).	32
1.22	The Mask R-CNN framework for instance segmentation He et al. (2017).	34
1.23	Mask R-CNN framework Anantharaman et al. (2018).	35
3.1	Histological images and their corresponding masks for benign and malignant gland cancer types.	49
3.2	The architecture of our approach	52
3.3	Predicted Masks of Various Gland Grades from Mask R-CNN (Healthy, Moderately Differentiated, Poorly Differentiated)	53
3.4	The predicted masks from mask R-CNN and the Results after applying The Watershed Algorithm	54

---

# LIST OF TABLES

2.1	Some of the publicly available CRC imaging datasets Tamang and Kim (2021) . . . . .	38
3.1	Percentage of each grade type in Test A and Test B Subdatasets . . . . .	48
3.2	Mask R-CNN Model Testing Results . . . . .	53
3.3	The Testing Results After the Post-processing . . . . .	54

---

# INTRODUCTION

## Context and Motivation

Colorectal cancer (CRC) is a major global health issue WHO (2023). Adenocarcinoma is the most common type of CRC. Early and accurate diagnosis, which relies on analyzing colon tissue samples, is critical for effective treatment. Pathologists examine these samples for cancerous features, particularly the morphology of intestinal glands, which is crucial for cancer grading and prognosis. Intestinal glands, located in the epithelial layer of the colon, play key roles in nutrient absorption, mucus secretion, and epithelial cell regeneration. Disruptions in these processes can lead to colorectal adenocarcinoma Sirinukunwattana et al. (2017).

Automated image analysis techniques can detect cancer by analyzing the shape and structure of gland regions, especially gland segmentation techniques that precisely determine the location and shape of the gland. Accurate segmentation of glands in colon histology images is crucial for computer-aided diagnosis (CAD) systems to quantify morphological features associated with cancer progression effectively. Over the years, segmentation methods have evolved from traditional techniques based on separating Hematoxylin and Eosin staining, and shape analysis Cheikh et al. (2016). However, these methods are time-consuming, challenging, and typically effective only for benign glands with static

shapes.

To overcome these limitations, More sophisticated approaches utilizing machine learning algorithms, such as clustering based on the color or the shape Rathore et al. (2019), and the watershed method that segments individual glands Alfonso et al. (2018), have been developed. These methods marked significant improvements but still faced challenges in dealing with the complexity and variability of malignant glands.

Recently, with the availability of histological image datasets Sirinukunwattana et al. (2017) and advancements in deep learning (DL), particularly convolutional neural networks (CNNs), there has been substantial progress in gland segmentation Wang et al. (2022). CNNs have demonstrated remarkable capability in feature detection and extraction, enabling more precise and efficient analysis of complex histopathological images.

## Goals and Approach

To improve gland segmentation in histological images of colorectal cancer, we combined one of the deep learning models, Mask R-CNN, with a traditional watershed algorithm Alfonso et al. (2018). Mask R-CNN is used to segment the glands and generate instance masks for each gland He et al. (2017), leveraging its powerful object detection and instance segmentation capabilities. This allows for the accurate capture of complex features and varying shapes of glands within histopathological images.

We apply the Watershed algorithm to refine these initial results and address the challenge of separating overlapping glands. This algorithm enhances boundary delineation by considering the topographical features of the image, effectively separating closely packed or overlapping glands. By integrating Mask R-CNN and the Watershed algorithm, our approach aims to achieve more precise and reliable gland segmentation. This hybrid method leverages deep learning and traditional image processing strengths, significantly aiding pathologists in cancer diagnosis and grading.

## Organization

This thesis is organized into three chapters, each detailing different aspects of our study on gland segmentation in histological images of colorectal cancer.

**Chapter 1** reviews medical imaging's role in cancer diagnostics and covers key imaging modalities and standardized medical image file formats. It also explores diagnostic systems using histopathological image analysis (HI) and details crucial medical image related machine-learning tasks. Finally, it examines the impact of DL on medical imaging, discussing the evolution of CNN like R-CNN, Fast R-CNN, Faster R-CNN, and Mask R-CNN, emphasizing their contributions to enhanced diagnostic accuracy and efficiency.

**Chapter 2** Chapter 2 examines gland segmentation techniques: traditional methods (color and shape analysis), machine learning advancements, and recent enhancements with deep learning, focusing on histopathology image accuracy and efficiency.

**Chapter 3** Chapter 3 outlines our gland segmentation approach using Mask R-CNN and the watershed algorithm, detailing implementation steps and evaluating its efficacy in improving segmentation accuracy for CRC..

---

---

# CHAPTER 1

---

## PRELIMINARIES

### 1.1 Introduction

Medical imaging is pivotal in modern cancer diagnostics, facilitating the visualization and characterization of pathological conditions. This chapter explores essential concepts in medical imaging modalities, file formats, and diagnostic systems that leverage advanced technologies for cancer diagnosis. Additionally, we delve into image processing tasks, including preprocessing, feature extraction, segmentation, object detection, and classification, highlighting their significance in extracting meaningful information from medical images. Furthermore, this chapter explores deep learning methodologies, particularly convolutional neural networks (CNNs) and region-based architectures like R-CNN variants (Fast R-CNN, Faster R-CNN, and Mask R-CNN), which have revolutionized medical image analysis. Through these discussions, we underscore the transformative impact of deep learning in advancing cancer diagnostics.

## 1.2 Medical Imaging and Cancer Diagnostic

Situating itself at the crossroads of artificial intelligence and computer science, computer vision seeks to provide computers with visual comprehension and interpretation capabilities. Its purview includes developing methods and systems for visual data extraction, analysis, and interpretation, which includes image and video data. All things considered, computer vision's ultimate goal is to have computers see, comprehend, and interpret visual data in the same way that humans can. Computer vision has many different uses in many different industries, such as medicine, transportation, robots, and surveillance Esteva et al. (2021). Figure 1.1 shows an example of medical computer vision tasks.

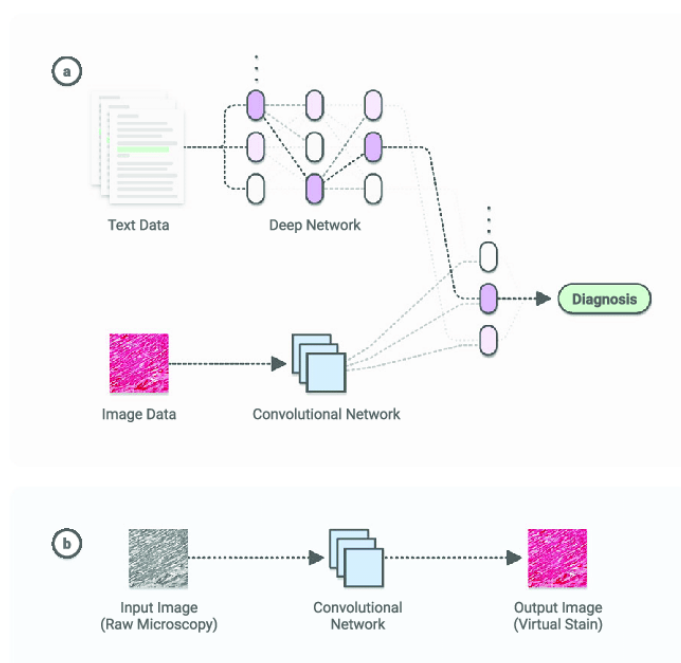


Figure 1.1: (a) Multimodal Discriminative Model: Deep learning architectures integrate image data using convolutional networks and non-image data using generic deep networks. (b) Generative Model: Convolutional neural networks generate images for tasks like picture-to-image regression, super-resolution enhancement, and unique image creation. Esteva et al. (2021).

Medical imaging involves employing a range of technologies and methods to generate visual depictions of the internal aspects of the human body, facilitating clinical analysis and medical interventions. These visual representations are crucial in diagnosing, track-



ing, and addressing various medical conditions and diseases. The array of medical imaging techniques encompasses X-rays, computed tomography (CT), magnetic resonance imaging (MRI), ultrasound, and nuclear medicine imaging. These methodologies empower healthcare practitioners to observe and assess the internal structures and functions of the body, furnishing essential insights for diagnosis, treatment strategizing, and ongoing monitoring of diverse medical conditions Esteva et al. (2021).

Colorectal cancer, affecting either the colon or rectum, is a widespread global malignancy, comprising roughly 10% of all cancer cases and ranking as the second highest cause of cancer-related mortality. The probability of having colorectal cancer grows with age, mostly affecting people over 50 years old. Various lifestyle factors contribute to its development, including high consumption of processed meats, inadequate consumption of fruits and vegetables, a sedentary lifestyle, obesity, smoking, and excessive alcohol consumption. Manifesting symptoms encompass diarrhea, constipation, blood in the stool, abdominal pain, unexplained weight loss, fatigue, and reduced iron levels. Prognosis varies depending on the cancer stage upon diagnosis, with early-stage patients having greater survival rates than advanced-stage ones. Timely diagnosis, adequate treatment, and continuous follow-up care are important for boosting survival rates and overall quality of life WHO (2023). The diagnosis of colon cancer necessitates a comprehensive strategy that combines state-of-the-art medical technologies with various diagnostic methodologies.

### 1.2.1 Imaging Modalities

The primary objective in medical imaging research is to attain precise and automated colon cancer diagnosis without manual intervention. This section thoroughly examines diverse imaging modalities utilized for this objective. The emphasis is placed on five fundamental techniques:

1. Computed Tomography Colonography (CTC)
2. Magnetic Resonance Imaging (MRI)
3. Endorectal UltraSound (ERUS)
4. Histopathology Images (HIs)

## 5. Positron Emission Tomography (PET)

### 1.2.1.1 Computed Tomography Colonography

Computed Tomography Colonography is an advanced imaging technique employed for the non-invasive identification of colorectal neoplasia. This method utilizes thin-section spiral computed tomography (CT) to generate initial two-dimensional axial images, reconstructed into detailed three-dimensional (3D) representations using specialized algorithms. These 3D images offer endoscopic-like perspectives of the colon, enabling the identification of colorectal carcinomas and polyps larger than 6.0 mm in diameter. Beyond neoplasia detection, CTC facilitates the assessment of colonic wall thickness and extracolonic abnormalities, providing crucial staging information. Particularly valuable in cases where traditional colonoscopy is hindered by occlusive colorectal carcinoma, CTC allows for visualization of the proximal colon Laghi et al. (2002). An example of a CT colonographic image is shown in Figure 1.2.

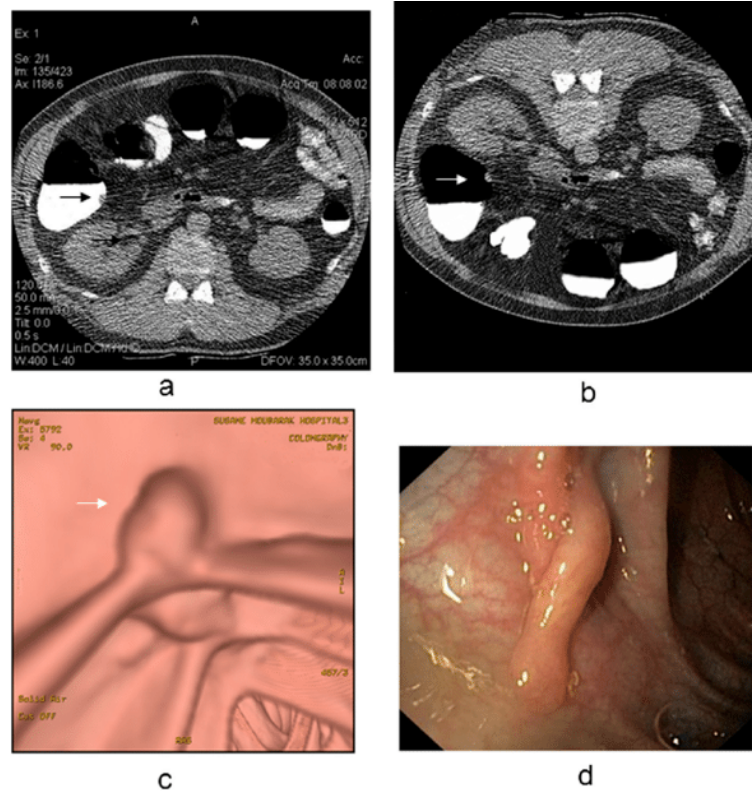


Figure 1.2: CT colonographic images in a 64-year-old male patient with true-positive findings (a) axial. (b) sagittal images of a 4 mm sessile polyp (arrow) in the sigmoid colon. (c) virtual colonoscopy image of the same lesion (arrow). (d) Colonoscopy image Kosov et al. (2020).

### 1.2.1.2 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) uses a strong magnetic field and radio waves to generate detailed images of the colon and rectum without being invasive. In the context of colon and rectal cancer, MRI is essential for staging tumors, evaluating the involvement of adjacent structures, and assessing the response to treatments like chemo-radiation therapy. Additionally, MRI identifies key prognostic indicators, such as the proximity to the anorectal junction, infiltration of the mesorectal fascia, and extramural vascular invasion. These factors are crucial for planning the appropriate treatment strategy and predicting clinical outcomes in patients with colon and rectal cancer Pizzi et al. (2018). The figure 1.3 depicts an example of an MRI image.

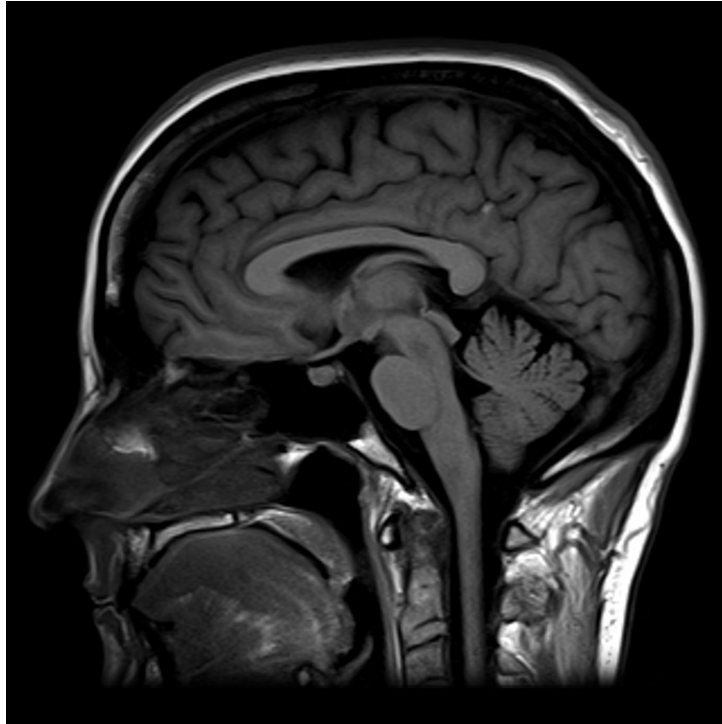


Figure 1.3: Normal midline brain MRI. *image from:* <https://radiopaedia.org/articles/mri-2>

### 1.2.1.3 Endorectal UltraSound

Endorectal ultrasound (ERUS) serves as a diagnostic imaging technique applied in the assessment of colon and rectal cancer. It employs a rotating transducer positioned within the rectal lumen to generate high-resolution images of the rectal wall and surrounding tissues. ERUS proves particularly advantageous for determining the local extent of disease in rectal cancer, offering detailed visualization of rectal wall layers, adjacent structures (prostate, seminal vesicles, vagina, and uterus), as well as the size and number of perirectal lymph nodes. This intricate visualization contributes to the accurate staging of rectal tumors, which is crucial for guiding appropriate treatment decisions and predicting prognosis. Moreover, ERUS has demonstrated high accuracy in the T-staging of rectal tumors, establishing its significance in the overall management of colon and rectal cancer Brook et al. (2013). An example of an ERUS image is illustrated by Figure 1.4.

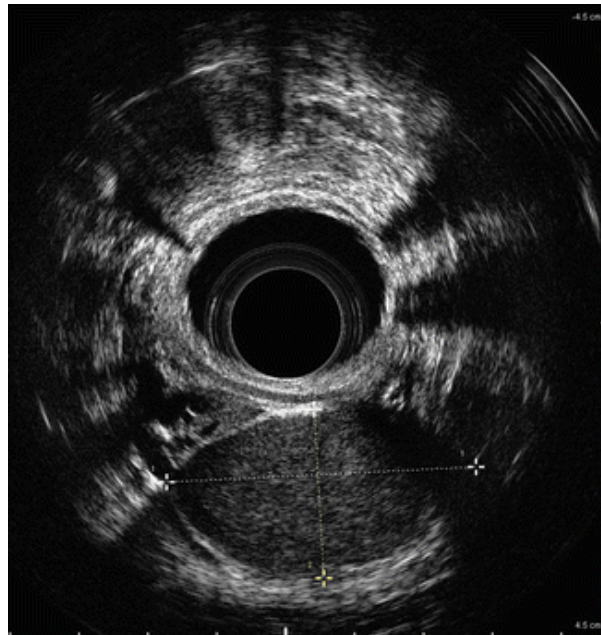


Figure 1.4: Retrorectal cystic mass. Key (2018)

#### 1.2.1.4 Histopathology Images

Histopathology images (HIs) are pivotal in the realm of pathology, offering intricate views of tissue samples at a cellular level. These high-resolution digital images, stained to emphasize specific structures, play a vital role in diagnosing and categorizing diseases, notably cancer. In the context of colon and rectal cancer, it is derived from patients with colorectal cancer and provides essential insights into the microscopic characteristics of the disease. Pathologists leverage these images to discern and classify cancerous tissues accurately Tsai et al. (2023). Figure 1.5 shows an illustration of a histopathology image.

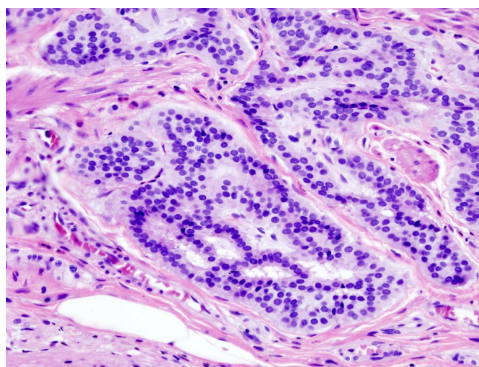


Figure 1.5: Cancer Histopathologic image of colonic carcinoid. *Image from: [https://handwiki.org/wiki/Medicine:Histopathology\\_of\\_colorectal\\_carcinoma](https://handwiki.org/wiki/Medicine:Histopathology_of_colorectal_carcinoma)*

#### 1.2.1.5 Positron Emission Tomography

Positron Emission Tomography (PET) represents a nuclear medicine imaging technique that utilizes a radioactive tracer, commonly  $^{18}\text{F}$ -fluoro-2-deoxy-D-glucose (FDG), to observe metabolic activity in tissues. This non-invasive procedure involves the emission of positrons by the tracer, subsequently detected by a scanner to generate three-dimensional images of the body. In the realm of colon and rectal cancer, PET plays a pivotal role in detecting and staging the disease, evaluating treatment response, and pinpointing potential metastatic sites. By gauging the heightened glucose metabolism characteristic of cancer cells, PET scans offer valuable insights for managing patients with colon and rectal cancer, facilitating treatment planning, and providing ongoing disease monitoring Mukai et al. (2006). Figure 1.6 shows an illustration of a Positron Emission Tomography image.

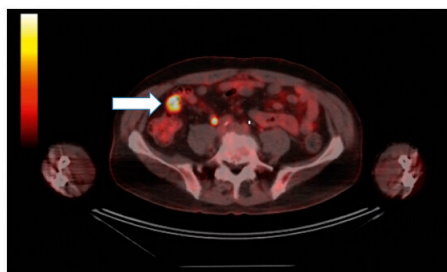


Figure 1.6: PET scan with persistent and increasing Fluorodeoxyglucose (FDG) uptake in the distal small bowel (arrow). Barge et al. (2021)

### 1.2.2 Medical Image File Formats

Medical imaging utilizes various file formats, each tailored to specific needs and applications, including Analyze, Neuroimaging Informatics Technology Initiative (Nifti), Minc, Digital Imaging and Communications in Medicine (DICOM), bitmap, and vector formats Larobina and Murino (2014); Furness (1999):

- **Analyze:** This format employs two different files to contain metadata and picture data, respectively. It is commonly utilized in neuroimaging studies.
- **Nifti:** Building upon the Analyze format, Nifti provides improved support for 3D and 4D imaging data and is also prevalent in neuroimaging.
- **Minc:** Designed for versatility, Minc integrates metadata and picture data in a single file and may be utilized across multiple imaging modalities.
- **DICOM:** A standard in clinical situations, DICOM files not only carry picture data but also include information such as patient details and imaging procedures.
- **Bitmaps:** These pictures comprise a two-dimensional array of pixels, where each pixel represents a dot in the image.
- **Vector diagrams:** These employ mathematical definitions for forms, such as polygons, curves, and their colors.

Choosing the proper format relies on the application's unique needs since each has its benefits and limits.

### 1.2.3 Histopathology Image Analysis

This section discusses common diagnosis systems applied for colon cancer detection based on HI analysis. These systems include Computer-Aided Diagnosis (CAD), and other findings in clinicopathology association systems.

- **Computer-Aided Diagnosis (CAD):** is a concept within medical imaging and diagnostic radiology. It involves the application of computer algorithms to aid

physicians in interpreting medical images and assisting in disease detection, diagnosis, and treatment planning. These systems analyze medical images, emphasizing areas of interest or potential abnormalities, serving as a supplementary tool to support radiologists and clinicians in their decision-making process Doi (2007).

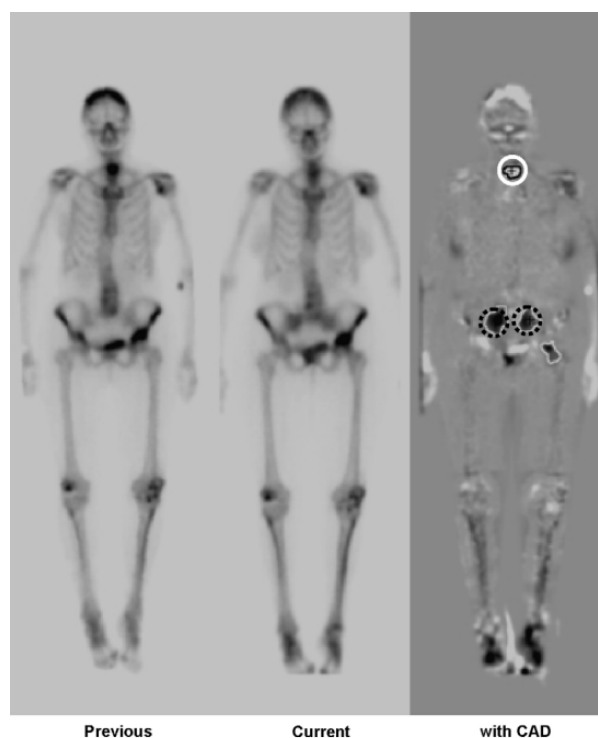


Figure 1.7: A temporal subtraction image created from previous and current bone scan images. The computer correctly identified one cold lesion (white solid circle) and two hot lesions (dark dotted circles) on the subtraction image, highlighting its potential to enhance interval changes between successive whole-body bone scans Doi (2007).

- **Clinicopathological association systems:** Clinicopathological association systems involve exploring the connections between clinical symptoms observed in patients and pathological features associated with a specific condition, like colorectal cancer. These associations help to understand the role of specific genetic mutations, such as the BRAF V600E mutation, in the progression, characteristics, and outcomes of the disease. In the case of colorectal cancer, the known clinicopathological association systems related to the BRAF V600E mutation provide valuable insights into how this mutation might affect the clinical presentation, behavior, and treatment options of the disease Chen et al. (2014).



## 1.3 Machine learning, Image Segmentation and Related Tasks

Machine learning is an area of artificial intelligence that focuses on constructing algorithms and statistical models to assist computer systems improve their efficiency at a certain activity. This is acquired by experience or training with vast datasets. In the medical imaging industry, machine learning algorithms are utilized to analyze and interpret medical pictures. They aid in duties such as risk assessment, detection, diagnosis, prognosis, and therapy response. These methods include both conventional approaches like support vector machines and random forests, as well as more modern techniques like deep learning Giger (2018).

The machine-learning process for histological image (HI) analysis consists of five primary steps, represented in Figure 1.8, which will be further discussed in the next.

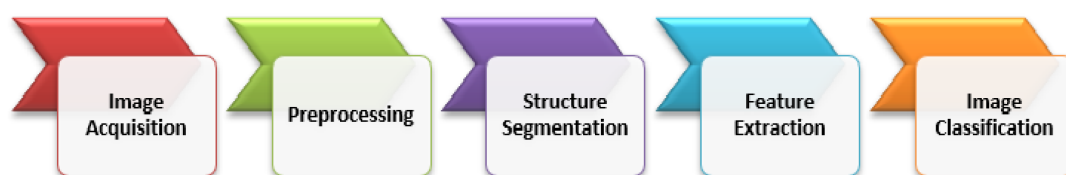


Figure 1.8: Main Stages in Traditional ML for HI Analysis Tharwat et al. (2022).

### 1.3.1 Preprocessing

In the realm of digital pathology, preprocessing involves employing various techniques to prepare data or images for analysis and further processing. Specifically, image preprocessing in digital pathology entails manipulating and enhancing digital images to enhance their quality, diminish noise, and eliminate artifacts. Techniques like image resizing, normalization, contrast enhancement, noise reduction, and image registration may be employed. Preprocessing aims to enhance the accuracy and reliability of subsequent analysis or modeling tasks McCombe et al. (2021).

### 1.3.2 Feature Extraction

Feature extraction is the procedure of recognizing and isolating pertinent information or characteristics from raw data, such as medical images, to produce a more concise and meaningful representation of the data. In the realm of medical imaging, feature extraction entails the identification and quantification of specific attributes within the images that are pertinent to the given task, encompassing aspects like the size, shape, texture, and intensity of particular regions or structures depicted in the image. These extracted features can subsequently serve as inputs for machine learning algorithms, aiding in tasks such as classification, segmentation, and diagnosis. This process holds substantial importance in numerous medical imaging applications, contributing to the reduction of data dimensionality and enhancement of the accuracy and efficiency of subsequent analysis Giger (2018).

### 1.3.3 Segmentation

Segmentation refers to the act of splitting a picture into many segments to simplify and alter its representation into a more comprehensible and analytically accessible form. In the realm of medical imaging, segmentation is widely used to detect and delineate certain structures or regions of interest within the pictures, such as tumors, organs, or anatomical characteristics. This key phrase is vital to many medical imaging applications since it provides the foundation for future analysis and interpretation of the images Giger (2018).

There are two types of Segmentation:

- ***Semantic Segmentation:*** is a computer vision task where every pixel in a picture is labeled with a certain class, such as "person," "car," or "tree." Semantic segmentation seeks to separate a picture into areas with semantically meaningful material, enabling a complete comprehension of the objects and their outlines in the scene. It varies from instance segmentation, which goes beyond providing class labels by discriminating between separate instances of the same class, supplying unique masks for each object as illustrated in the figure 1.9 He et al. (2017).
- ***Instance segmentation:*** presents a computer vision challenge focused on identifying and segmenting individual object instances within an image, assigning distinct

labels and masks to each object. In contrast to semantic segmentation, where all pixels of a specific class receive a single label, instance segmentation differentiates between multiple instances of the same class as shown in the figure 1.9, supplying a unique mask for each object. This task is more intricate than semantic segmentation, demanding the identification of objects and the precise delineation of their boundaries He et al. (2017).

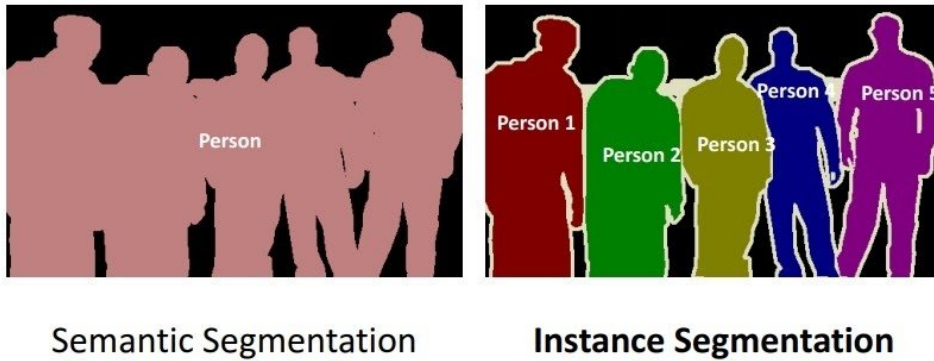


Figure 1.9: Semantic Segmentation vs Instance Segmentation Odemakinde (2023).

### Mathematical Formulation:

Consider an image  $I$  defined on a domain  $\Omega \subset \mathbb{R}^2$ . Each pixel  $p$  of the image  $I$  is in the domain  $\Omega$ . The goal of image segmentation is to determine a function  $S$  that maps each pixel in the domain to a label from a predefined set  $\mathcal{C}$ :

$$S: \Omega \rightarrow \mathcal{C}$$

where  $\mathcal{C} = \{1, 2, \dots, k\}$  represents the set of  $k$  different labels or segments. The function  $S$  assigns a label  $S(p)$  to each pixel  $p$  in  $\Omega$ .

In a supervised learning framework, this function  $S$  is learned from a labeled dataset  $\{(I_i, S_i)\}_{i=1}^n$ , where  $I_i$  denotes the input images and  $S_i$  denotes their corresponding ground truth segmentation maps. The objective is to train a model  $f_\theta$  with parameters  $\theta$  such that the model's output approximates the true segmentation function:

$$f_{\theta}(I) \approx S$$

The training process involves minimizing a loss function  $\mathcal{L}$  that quantifies the discrepancy between the predicted segmentation  $f_{\theta}(I)$  and the ground truth  $S$ . A common loss function for this purpose is the cross-entropy loss, which can be expressed as:

$$\mathcal{L}(\theta) = -\frac{1}{|\Omega|} \sum_{p \in \Omega} \sum_{c \in \mathcal{C}} \mathbf{1}_{\{S(p)=c\}} \log P_c(p; \theta)$$

where  $P_c(p; \theta)$  is the predicted probability that pixel  $p$  belongs to class  $c$ , and  $\mathbf{1}_{\{S(p)=c\}}$  is an indicator function that equals 1 if the true label of pixel  $p$  is  $c$  and 0 otherwise. Ronneberger et al. (2015); Badrinarayanan et al. (2017); Long et al. (2015).

### 1.3.4 Object Detection

Object detection is fundamental in computer vision, and essential across diverse applications such as autonomous driving, surveillance systems, and malady diagnosis. Object detection involves identifying and localizing objects within images or video frames. It encompasses predicting bounding boxes around objects of interest and assigning corresponding class labels to denote the types of objects present Girshick (2015); Ren et al. (2015).

Object detection is a daunting task because of the various natures of things in terms of appearance, size, orientation, and occlusion. Conventional techniques for object recognition depended on manually constructed features and machine learning algorithms like support vector machines (SVMs) and decision trees. However, these methodologies typically battle with the complexity and differences encountered in real-world data. The introduction of deep learning has ushered in major advancements in the accuracy and efficiency of object identification. Deep learning-based solutions employ convolutional neural networks (CNNs) to autonomously extract features from raw picture data, reducing the demand for created features. These models may undergo end-to-end training to concurrently handle object localization and classification. Frameworks for object identi-

fication founded in deep learning, such as Faster R-CNN, YOLO, and SSD, have gained leading-edge performance on benchmark datasets like COCO and PASCAL VOC. These frameworks incorporate varied approaches such as region proposal networks, feature pyramids, and anchor boxes to boost detection accuracy and speed Zhao et al. (2019).

- **Mathematical Formulation**

$I$  denote an input image. The goal of object detection is to predict a set of bounding boxes  $B = \{b_1, b_2, \dots, b_n\}$  and their associated class labels  $C = \{c_1, c_2, \dots, c_n\}$ , where  $n$  represents the number of detected objects.

Each bounding box  $b_i$  is defined by four parameters  $(x_i, y_i, w_i, h_i)$ :

- $(x_i, y_i)$  denotes the coordinates of the top-left corner of the bounding box,
- $w_i$  and  $h_i$  denote the width and height of the bounding box, respectively.

The class label  $c_i$  indicates the category of the object within the bounding box  $b_i$ .

Object detection algorithms typically involve the following steps:

1. **Object Localization:** Predicting accurate bounding boxes  $B$  that tightly enclose each object in the image.
2. **Object Classification:** Assigning appropriate class labels  $C$  to the predicted bounding boxes, identifying the type of objects detected.
3. **Post-processing:** Refining predictions through techniques like non-maximum suppression to eliminate duplicate detections and enhance localization accuracy.

### 1.3.5 Classification

In medical imaging, classification refers to the process of categorizing images based on their visual characteristics or extracted quantitative data. Machine learning algorithms are often used to automate this process, assigning labels to images according to their distinctive features. For example, in diagnosing tumors, a classification algorithm can be trained to identify whether a tumor is benign or malignant by analyzing specific

features in the medical images. This approach is used in various applications, such as diagnosing diseases, characterizing tumors, and planning treatments. The accuracy of these models is essential for providing reliable and effective clinical decision support. Various algorithms, including support vector machines, random forests, and deep learning models, are employed depending on the complexity of the task and the type of data available Giger (2018).

## 1.4 Deep Learning

Deep learning (DL) is a subfield of machine learning that uses neural networks with numerous layers to learn and extract high-level characteristics from datasets. These networks automatically acquire data representations using a hierarchical framework of ideas modeled after the structure and function of the human brain. DL models can do complicated tasks, including picture and audio recognition, natural language processing, and medical imaging analysis. They can learn complicated patterns and representations using numerous layers, making them very useful in a variety of artificial intelligence and data analysis applications. DL models can remove the need for human feature engineering include Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which demonstrate outstanding competence in capturing subtle patterns inside enormous datasets. DL algorithms, trained on large datasets, optimize parameters using methods such as backpropagation and gradient descent, resulting in substantial advances in computer vision, healthcare, finance, and other domains Tsai and Tao (2021).

### 1.4.1 Convolutional Neural Networks

#### 1.4.1.1 Architecture

A CNN stands out as a specialized artificial neural network tailored for image processing and recognition. Specifically optimized for handling pixel data, it excels in tasks related to image recognition and processing. Widely employed in computer vision applications, such as image segmentation, the CNN architecture comprises essential layers that collaborate to analyze and comprehend visual data thoroughly Tsai and Tao (2021):

- **Input layer** Within a neural network, the input layer serves as the point where unprocessed data, such as images or structured information, is introduced into the network for analysis and processing.
- **Convolutional layers** The convolutional layer is a crucial element within a neural network, particularly in tasks related to image recognition. This layer employs filters or kernels to extract features from input data, including edges and textures, and represents them as a feature map. Through the application of these filters, the convolutional layer facilitates the abstraction of input images, enhancing the neural network's ability to identify objects within the image data.
- **The Rectified Linear Unit (ReLU)** stands as a frequently employed activation function in neural networks. It introduces non-linearity by outputting the input as is for positive values and zero for negative values. This straightforward yet efficient function has become popular for its capability to address the vanishing gradient problem and expedite the training of deep learning models.
- **Pooling layers** within a neural network function aim to diminish the spatial dimensions of the input by downsampling the feature maps produced by the convolutional layers. This downsizing results in a reduction of parameters and computations within the network. The role of pooling layers extends to establishing translation invariance, prioritizing crucial features, and lessening sensitivity to minor input alterations. Two prevalent pooling methods are max pooling and average pooling. In these methods, the pooling layer condenses information about feature presence in patches of the feature map to achieve downsampling.
- **Fully connected layers** alternatively known as dense layers, play a pivotal role in neural networks. Within these layers, each node establishes connections with every node in the subsequent layer, facilitating the learning of intricate relationships between features. Typically positioned at the network's conclusion, fully connected layers are frequently employed for classification tasks. Their function involves consolidating the features acquired from earlier layers to formulate final predictions. To encapsulate, fully connected layers create connections between each neuron in one layer and every neuron in another layer.

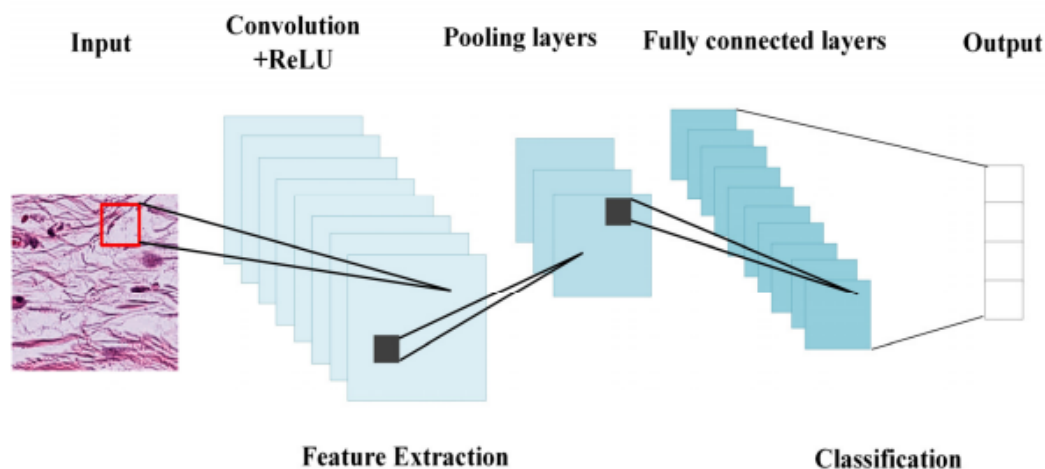


Figure 1.10: Network architecture of CNN Tsai and Tao (2021).

Integrating the layers of a convolutional neural network enables the network to acquire the capability to identify and recognize a specific object within an image. While basic CNNs are tailored for tasks such as image classification and object detection, they are primarily effective for images featuring one object.

#### 1.4.1.2 CNN Models Networks

Diverse CNN models have been created, each distinguished by its unique architecture and strengths. Some well-known CNN models comprise Shin et al. (2016); Tsai and Tao (2021):

- *U-Net* model is a CNN architecture specifically tailored for biomedical image segmentation tasks within medical image analysis. It was introduced by Ronneberger et al. in 2015 Ronneberger et al. (2015) and has since gained popularity for its effectiveness in delineating fine details and boundaries in images.

This architecture features a contracting path to capture contextual information and a symmetric expanding path for precise localization. Unique skip connections concatenate feature maps from the contracting path to the expanding path, maintaining high-resolution details during the upsampling process. This design enhances



U-Net's ability to generate accurate segmentation masks by integrating both local and global information effectively. Due to its capacity to yield high-quality segmentation outputs with modest training data, U-Net has been extensively applied in diverse medical imaging tasks, including tumor segmentation, organ delineation, and cell segmentation.

- *AlexNet* represents a groundbreaking CNN architecture that garnered notable recognition by triumphing in the 2012 Krizhevsky et al. (2012) ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Comprising five convolutional layers, three pooling layers, and three fully connected layers, it boasts around 60 million free parameters. Tailored to handle 224x224 pixel input images, AlexNet demonstrated the capability of deep learning in tasks related to image classification.
- *GoogLeNet* stands out as a sophisticated deep CNN architecture introduced in 2014 Szegedy et al. (2015). It surpasses earlier CNN architectures in complexity, featuring two convolutional layers, two pooling layers, and nine "Inception" layers. The Inception module, a pivotal innovation in GoogLeNet, combines filters of different sizes and dimensions into a single new filter. This enables the network to learn more diverse and representative features from input images. Notably, GoogLeNet achieved state-of-the-art performance in the 2014 ILSVRC, boasting a top-5 classification error rate of 6.67%. Its versatility extends to various computer vision applications, including medical image analysis, where it exhibits high accuracy in tasks such as lesion detection and classification.
- *VGGNet* Simonyan and Zisserman (2015) achieved notable accuracy on the ImageNet dataset in 2014. Developed by the Visual Geometry Group at the University of Oxford, it comprises 16-19 layers, all of which are convolutional. VGGNet employs compact 3x3 filters with a stride of 1, along with max pooling, to distill features from input images. Renowned for its depth, surpassing earlier CNN models in the number of layers, VGGNet is distinguished by the use of the ReLU activation function and dropout regularization to mitigate overfitting. The VGGNet model culminates in a Softmax layer for classification. Beyond its success on ImageNet, VGGNet has found applications in diverse fields, including object recognition, image segmentation, and medical image analysis.
- *ResNet* CNN model launched in 2015 He et al. (2016), was meant to overcome the vanishing gradient issue widespread in deep neural networks. The vanishing gradient issue develops when gradients become excessively low, preventing their propagation over several layers and inhibiting the training of deep networks. ResNet

introduces the use of residual connections, enabling gradients to flow straight across the network by including the input of a layer into its output. This allows the network to learn residual functions, which are more susceptible to optimization than the original functions. Featuring a wide layer count, with some versions approaching 100 layers, ResNet has proven cutting-edge performance across different computer vision applications, spanning picture categorization, object identification, and segmentation.

- *SqueezeNet Iandola et al. (2016)* is a CNN model strategically crafted to attain superior accuracy with a reduced parameter count when juxtaposed with alternative models. Its development focused on facilitating efficient inference on devices constrained by resources, like smartphones and embedded systems. Employing a fusion of 1x1 and 3x3 filters, SqueezeNet efficiently reduces the number of parameters while maintaining high accuracy. Furthermore, leveraging model compression techniques has allowed SqueezeNet to be compressed to under 0.5 MB, a notably smaller size compared to other CNN models such as AlexNet. Despite its compact dimensions, SqueezeNet delivers competitive performance and finds utility in applications with limited computational and memory resources.

### 1.4.2 Region-Based Convolutional Neural Network

Region-Based Convolutional Neural Network (R-CNN) Chakraborty et al. (2022) is a DL model employed for object detection in computer vision tasks.

The RCNN architecture is specifically crafted for solving image detection tasks and is the foundation for other architectures such as Mask R-CNN. It initiates the process by generating region proposals for bounding boxes through a selective search. These proposals are then converted into standardized squares and sent through a CNN to produce a feature vector map as the output.

The output dense layer contains features extracted from the image, subsequently fitting into a classification algorithm responsible for categorizing the objects in the region proposal network. The algorithm also predicts offset values to enhance the precision level of the region proposal or bounding box.

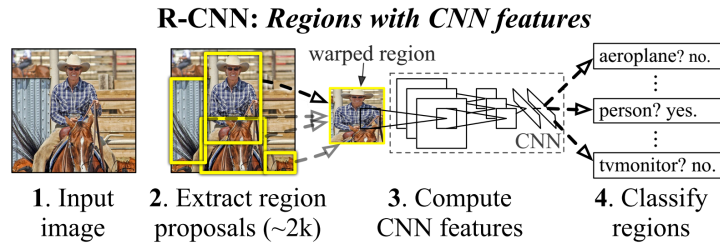


Figure 1.11: Regions with CNN features. Odemakinde (2023)

As illustrated by Figure 1.11, the initial step involves applying a region-extraction algorithm that suggests around 2,000 object boundaries. For each proposed region, the regions are resized to suit the convolutional neural network (CNN), and their features are calculated. Ultimately, the classification of the object within each region is determined.

#### 1.4.2.1 Selective Search

Selective Search represents a technique for producing object proposals or candidate regions within an image for object recognition as shown below in the figure 1.12. The approach relies on segmentation to create a varied set of regions encompassing the entire image at different scales and resolutions. Next, these regions are ranked according to their resemblance to object instances, and the top-ranked regions serve as object proposals for subsequent stages. Noteworthy for object recognition, selective search efficiently minimizes the number of regions for processing compared to exhaustive search, all while preserving high recall rates van de Sande et al. (2011).



Figure 1.12: Hierarchical grouping algorithm finds objects at different scales, as the girl is contained by the TV van de Sande et al. (2011).

### 1.4.2.2 Bounding Box Regressor

A bounding box regressor refers to a machine-learning model utilized for predicting the coordinates of a bounding box encircling an object within an image or video. This regressor processes an input image or a video frame and produces the coordinates of the bounding box that optimally encompasses the object of interest. Typically trained on a substantial dataset of labeled images, the regressor learns during training to forecast the coordinates of the bounding box, aligning with the object in the image as shown in Figure 1.13. This learning process relies on features extracted from the image by a CNN or another feature extractor. The objective is to minimize the disparity between the predicted bounding box and the ground truth bounding box, employing a loss function like Mean Squared Error (MSE) or smooth L1 loss. Once trained, the boundary box regressor can be applied to predict bounding box coordinates for objects in new images or video frames, facilitating efficient and accurate object detection and tracking Blue and Brindha (2019).



Figure 1.13: The sample shown in this section depicts, object detection with precise boundary boxes Blue and Brindha (2019).

Combining the components, the R-CNN identifies objects within an image and generates the bounding box.

The figure 1.14 explains the R-CNN process in a concise and understandable manner

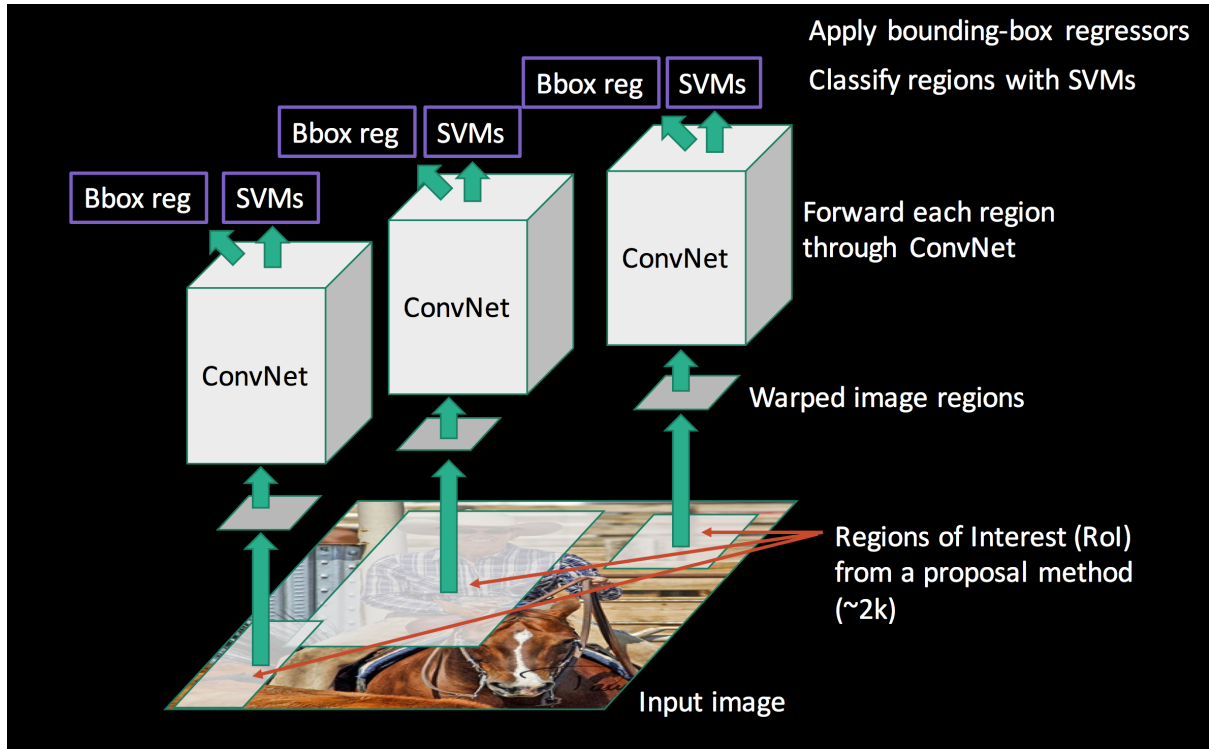


Figure 1.14: R-CNN process Huang (2017).

### 1.4.3 Fast R-CNN

Fast R-CNN is a CNN architecture designed for object detection and localization in computer vision. It overcomes the computational inefficiencies of its predecessor, R-CNN, by employing a single convolutional network for both region proposals and classification tasks. This eliminates redundant CNN computations on overlapping proposed regions, enhancing computational efficiency. Additionally, Fast R-CNN incorporates an external region proposal method, like selective search, to generate regions of interest (ROIs) as shown in the figure 1.15. These ROIs and corresponding feature maps create patches for object detection. After warping these patches to a fixed size using ROI pooling, they are processed by fully connected layers for classification and localization. By avoiding repeated feature extractions, Fast R-CNN significantly reduces processing time compared to R-CNN Chakraborty et al. (2022).

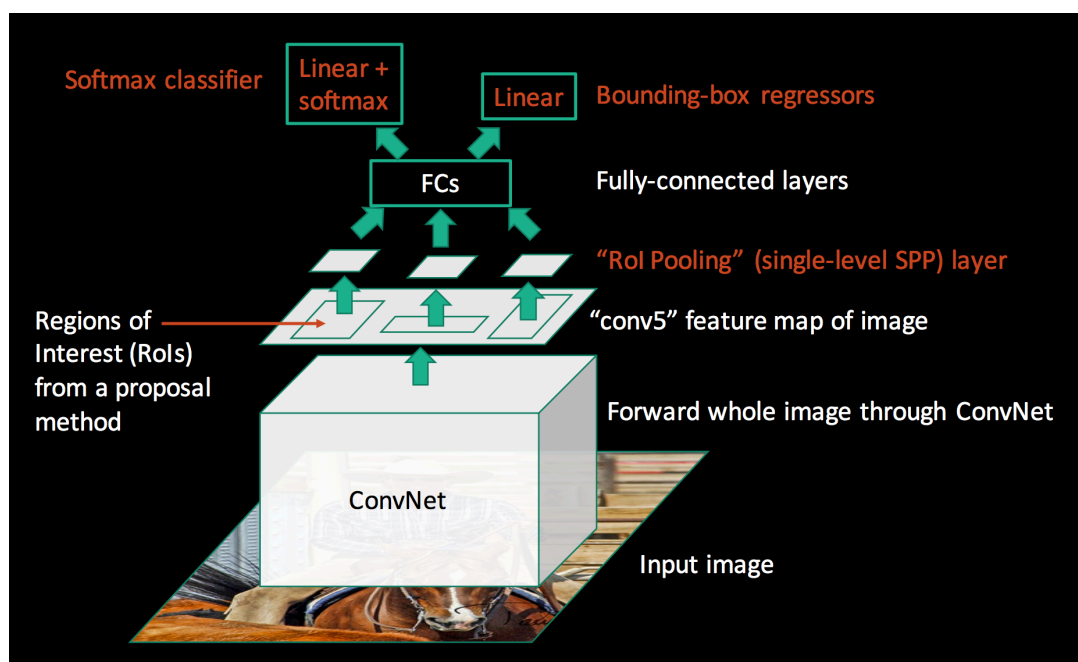


Figure 1.15: Fast R-CNN process Huang (2017).

#### 1.4.3.1 Region Of Interest Pooling

Region of Interest (ROI) pooling is a technique applied in computer vision for object detection and localization tasks. It extracts a fixed-size feature map from an input feature map of variable size, focusing on a specified region of interest defined by a bounding box indicating the object's location. The process involves dividing the region of interest into a fixed number of sub-windows or bins, rounding each bin to the nearest integer, and mapping it to the corresponding location in the output feature map. The output feature map is generated by max-pooling the values in each bin, ensuring a consistent size irrespective of the input region's size Chakraborty et al. (2022).

Used in architectures like Fast R-CNN, ROI pooling extracts features from the region of interest for input into fully connected layers, aiding in classification and localization. To illustrate, let's simplify the concept by transforming  $8 \times 8$  feature maps into a predefined  $2 \times 2$  shape as shown in Figure 1.16 Hui (2018):

- *Top left:* Represents the feature maps.

- *Top right*: Overlaps the ROI (blue) with the feature maps.
- *Bottom left*: Divides ROIs into the target dimension (e.g.,  $2 \times 2$ ).
- *Bottom right*: Determines the maximum for each section, resulting in our warped feature maps.

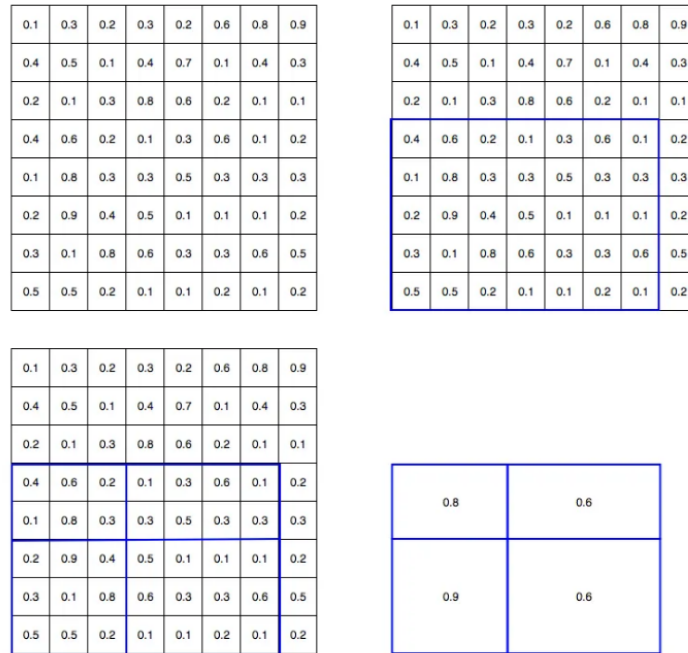


Figure 1.16: Input feature map (top left), output feature map (bottom right), blue box is the ROI (top right) Hui (2018).

#### 1.4.4 Faster R-CNN

Faster R-CNN Chakraborty et al. (2022) Builds upon the Fast R-CNN architecture by seamlessly incorporating a Region Proposal Network (RPN) directly into the model. This integration streamlines the process of generating region proposals within the network, eliminating the necessity for external methods like selective search. Through this combination of RPN with Fast R-CNN, Faster R-CNN attains notable advancements in computational efficiency and achieves enhanced accuracy in object detection.

Within the Faster R-CNN framework, a unified convolutional network handles both

region proposals and object classification tasks. The RPN autonomously generates bounding boxes of diverse shapes and sizes as shown in the figure 1.17, forming regions of interest for subsequent object detection. These proposed regions undergo further processing in the Fast R-CNN for precise classification and localization.

The pivotal inclusion of RPN facilitates end-to-end training of the complete object detection system, leading to superior performance and expedited inference compared to its predecessors. Faster R-CNN stands as a widely adopted and influential model in the realm of object detection, laying the groundwork for subsequent advancements in this field.

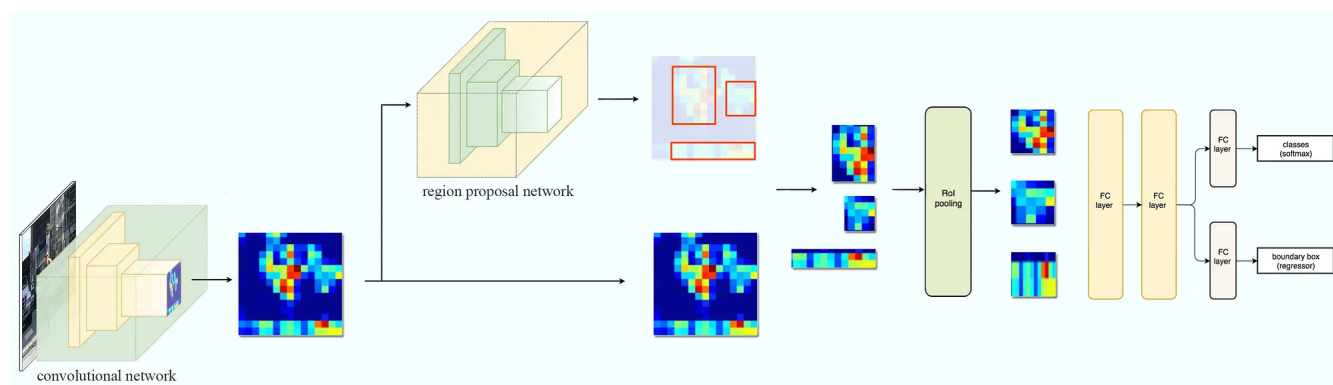


Figure 1.17: The deep network inside the system is used instead of an external region proposal Hui (2018).

#### 1.4.4.1 Region Proposal Network (RPN)

The Region Proposal Network (RPN) stands as a pivotal element within the Faster R-CNN architecture, specifically crafted for efficient and precise object detection in images. Its primary function involves generating region proposals, which represent potential bounding boxes containing objects of interest Chakraborty et al. (2022).

Functioning as a fully convolutional network, the RPN traverses the convolutional feature map of the input image. At each position of the sliding window, the RPN adeptly predicts multiple region proposals and their corresponding objectness scores simultaneously. These region proposals play a crucial role in suggesting regions of interest for subsequent stages of object detection and classification.



The RPN's distinctive capability to generate region proposals within the network itself, without resorting to external methods like selective search, significantly enhances the overall efficiency and effectiveness of the Faster R-CNN architecture. The integration of the RPN with subsequent stages of object detection results in improved performance and faster inference times, solidifying Faster R-CNN as a potent tool for object detection tasks in the realm of computer vision.

In detail, the region proposal network (RPN) takes the output feature maps from the initial convolutional network as input. Employing  $3 \times 3$  filters sliding over the feature maps, it formulates class-agnostic region proposals using a convolutional network like the ZF network. In contrast, alternative deep networks such as VGG or ResNet can be employed for more intricate feature extraction, this comes at the cost of processing speed Hui (2018).

The Region Proposal Network (RPN) generates multiple hypotheses for each location within the feature maps. Specifically, the RPN yields  $4 \times k$  coordinates and  $2 \times k$  scores for each location. To illustrate, consider  $8 \times 8$  feature maps with a  $3 \times 3$  filter, and for a given value of  $k$  (let's say  $k = 3$ ), the RPN produces a total of  $8 \times 8 \times 3$  Region of Interest (ROI) proposals. The right side of the diagram of Figure 1.18 visually represents the three proposals emanating from a single location Hui (2018).

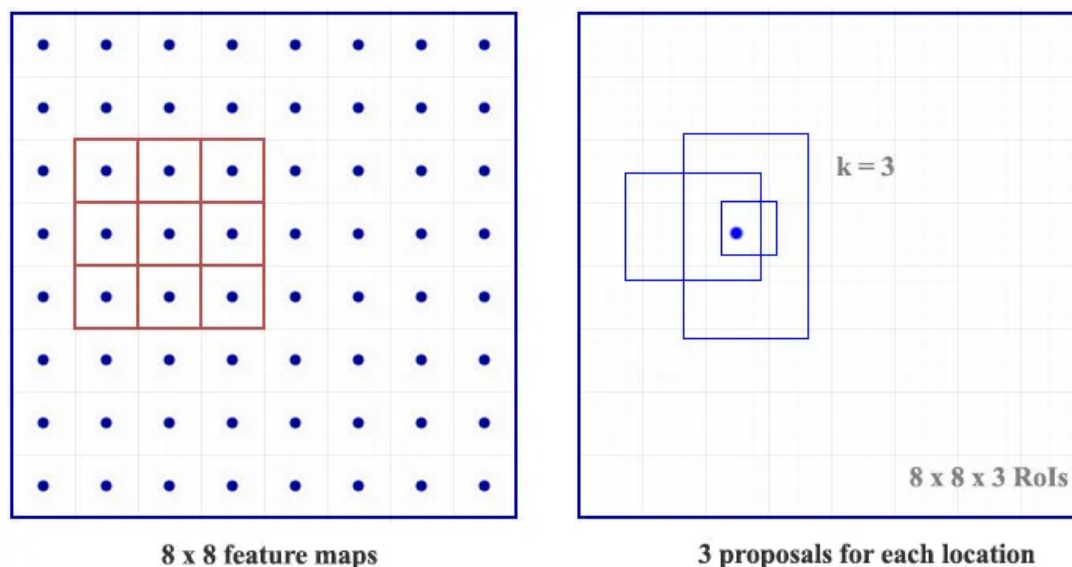


Figure 1.18: The diagram shows the  $8 \times 8$  feature maps with a  $3 \times 3$  filter, and it outputs a total of  $8 \times 8 \times 3$  ROIs (for  $k = 3$ ) Hui (2018).

There are a total of three opportunities to make accurate guesses, with the option to refine assumptions afterward. It is recommended that we begin with diverse guesses in terms of shape and size to maximize our likelihood of success. This is precisely why Faster R-CNN avoids utilizing arbitrary boundary box proposals, and instead predicts offsets like  $\delta x$  and  $\delta y$ , which are related to the top-left corner of reference boxes known as anchors as shown in the figure 1.19. Constraints are imposed on the magnitude of these offsets to ensure that the assumptions resemble the anchors Hui (2018).

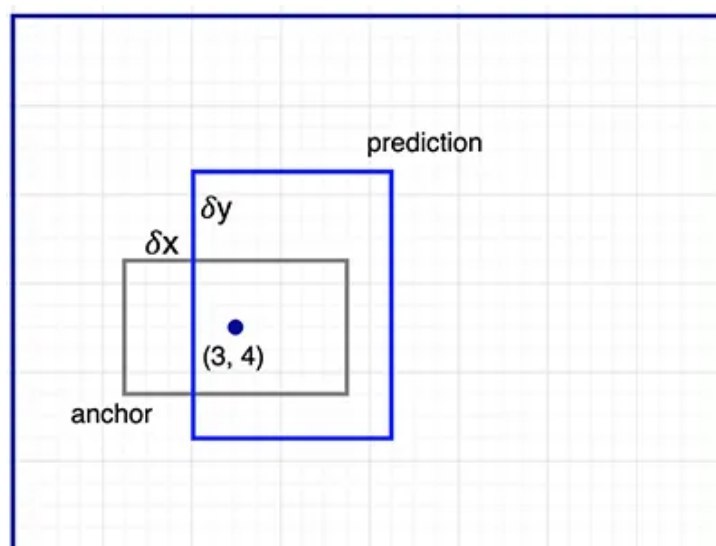


Figure 1.19: offsets  $\delta x$  and  $\delta y$  Hui (2018).

To generate  $k$  predictions for each location, it is essential to have  $k$  anchors positioned at the center of each location as shown in the figure 1.20 below. Each prediction is linked to a distinct anchor, yet various locations share identical anchor shapes. These anchors are thoughtfully pre-determined to ensure diversity, effectively covering real-world objects with varying scales and aspect ratios. This approach facilitates more informed initial training, provides better estimations, and allows each prediction to specialize in a particular shape. Consequently, this strategy enhances the stability and ease of the early training process.

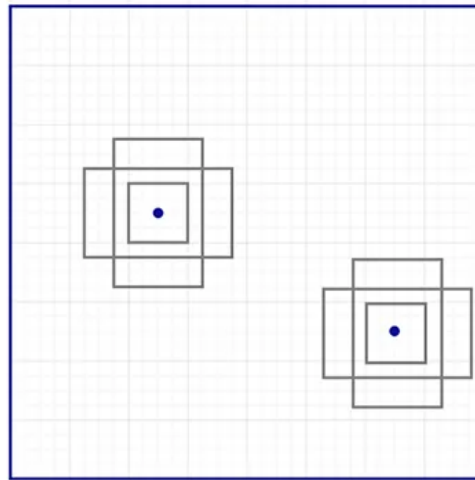


Figure 1.20: a specific anchor but different locations share the same anchor shapes Hui (2018).

Faster R-CNN employs a larger number of anchors, utilizing 9 anchor boxes that encompass three distinct scales and three different aspect ratios. With 9 anchors assigned to each location, the model produces  $2 \times 9$  objectness scores and  $4 \times 9$  coordinates for every location as depicted in the figure 1.21 Ren et al. (2015).

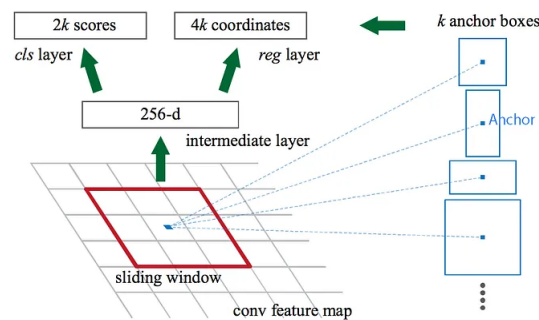


Figure 1.21: Region Proposal Network (RPN)s Ren et al. (2015).

### 1.4.5 Mask R-CNN

Mask R-CNN, represents an advanced method that employs a convolutional neural network to perform image segmentation and instance segmentation. It is constructed on the

foundation of Faster R-CNN Odemakinde (2023).

Mask R-CNN Anantharaman et al. (2018) extends the Faster R-CNN algorithm, which employs a RPN to generate object proposals and a separate network for proposal classification. The distinguishing feature of Mask R-CNN is the addition of a third branch to the network, dedicated to creating a binary mask for each object proposal, outlining the pixels associated with the object.

The Mask R-CNN architecture encompasses various modules as shown in the figure 1.22, such as a backbone network (typically ResNet50 or ResNet101), a region proposal network, an ROI (region of interest) align layer, a classification head, a bounding box regression head, and a mask head. The backbone network extracts features from the input image, utilized by the region proposal network to formulate object proposals. The ROI align layer then captures features from each object proposal, aligning them to a fixed size before forwarding them to the classification and bounding box regression heads. Simultaneously, the mask head generates a binary mask for each object proposal, identifying the pixels corresponding to the object.

During the training phase, the network undergoes end-to-end training, employing a multi-task loss function that combines classification, bounding box regression, and mask prediction losses. The training dataset comprises many annotated images, enabling the network to learn the detection and segmentation of objects in new images.

Mask R-CNN proves its effectiveness as a robust algorithm for object detection and instance segmentation, finding successful applications across diverse fields, including medical imaging, autonomous driving, and robotics.

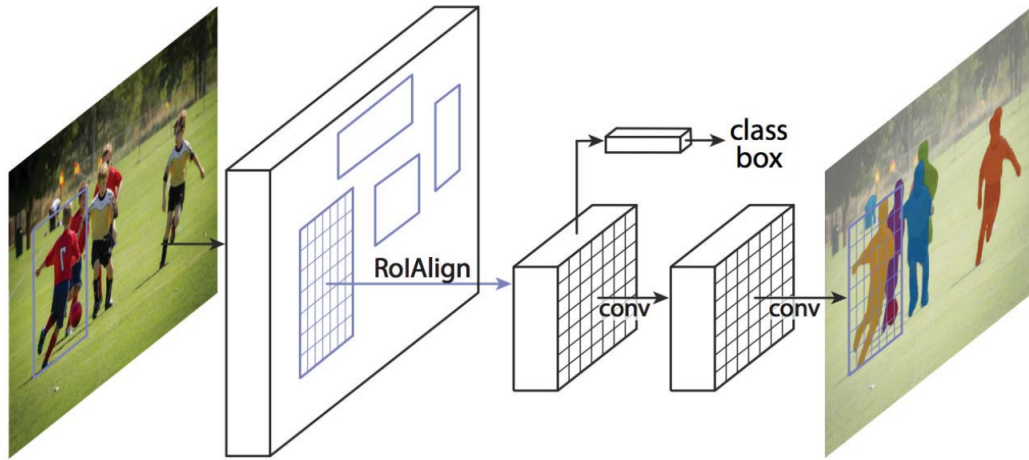


Figure 1.22: The Mask R-CNN framework for instance segmentation He et al. (2017).

#### 1.4.5.1 RoIAlign

RoIAlign is a pivotal element within the Mask R-CNN framework, strategically designed to counteract the alignment challenges associated with RoIPool, the conventional method for extracting a compact feature map from each Region of Interest (RoI). RoIPool introduces coarse spatial quantization during feature extraction, leading to potential misalignments between the RoI and the extracted features, particularly impacting the precision of pixel-level mask predictions.

In contrast, RoIAlign mitigates the drawbacks of RoIPool by eliminating its stringent quantization. Instead, RoIAlign ensures accurate alignment by faithfully preserving exact spatial locations without quantization. This approach guarantees that the features extracted from RoIs, which inherently constitute small feature maps, maintain precise alignment to retain explicit per-pixel spatial correspondence as shown in figure 1.23. Such alignment proves indispensable for achieving accuracy in mask prediction within instance segmentation tasks and empirical evidence has demonstrated that RoIAlign markedly enhances performance compared to RoIPool in Mask R-CNN experiments He et al. (2017).

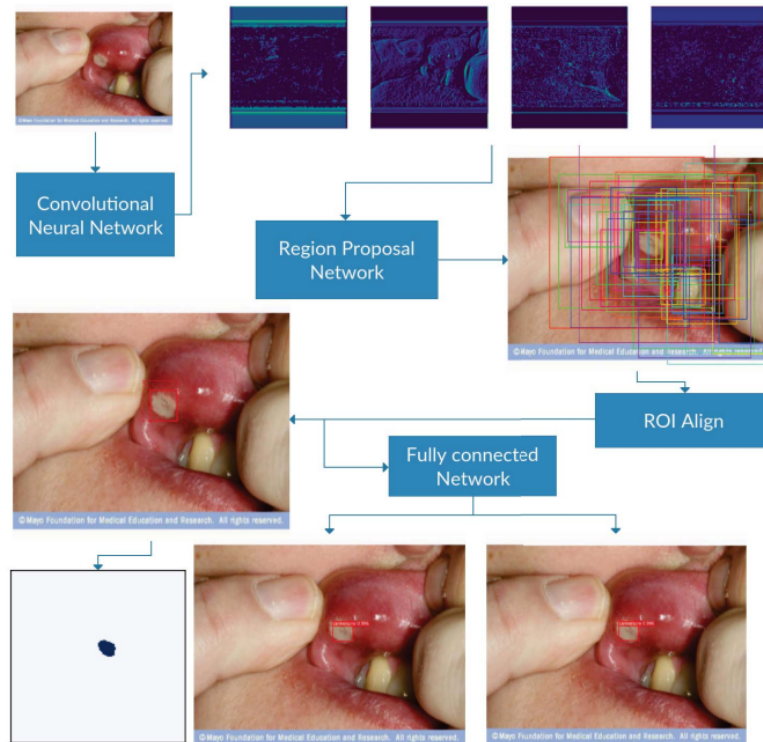


Figure 1.23: Mask R-CNN framework Anantharaman et al. (2018).

#### 1.4.5.2 Benefits of Mask R-CNN

Mask R-CNN offers several notable advantages that make it a standout model in the field of computer vision:

- *Exceptional performance:* Mask R-CNN surpasses all current single-model competitors, such as Faster R-CNN, YOLO, and SSD, in each task.
- *High efficiency:* This method is incredibly efficient and only requires a small amount of additional time to Faster R-CNN.
- *Versatility:* Mask R-CNN is not confined to a single task and can be easily adapted for other purposes.

## 1.5 Conclusion

This chapter reviewed the integral role of medical imaging in cancer diagnostics, emphasizing key imaging modalities, file formats, and diagnostic systems utilizing histopathological image analysis. It highlighted essential machine learning tasks such as preprocessing, feature extraction, segmentation, object detection, and classification, which are fundamental in interpreting medical images.

The chapter also explored the significant advancements in deep learning, particularly through Convolutional Neural Networks (CNNs) and their derivatives like R-CNN, Fast R-CNN, Faster R-CNN, and Mask R-CNN. These advancements have significantly enhanced diagnostic accuracy and efficiency, showcasing the transformative impact of deep learning in medical imaging. Overall, the integration of these technologies holds great promise for improving cancer diagnostics and patient outcomes.

---

---

# CHAPTER 2

---

## COLORECTAL CANCER RECOGNITION: STATE OF THE ART

### 2.1 Introduction

This chapter explores the evolution of gland segmentation techniques, tracing their development from early approaches that relied on substructure identification (nuclei, lumens) or hematoxylin intensity features, to the current landscape dominated by advanced deep learning architectures. We will delve into classic methods like thresholding, explore the advancements brought by machine learning algorithms, and culminate in the discussion of powerful deep learning techniques for this task. Each approach offers valuable insights and contributes to the foundation upon which we build our proposed solution a method for robust, precise, and clinically significant segmentation of colon gland images in the context of colorectal cancer.



## 2.2 Datasets and Evaluation Metrics

To improve the analysis of colorectal cancer (CRC) images, two important things are needed: good datasets and effective evaluation metrics. Quality datasets help train machine learning algorithms, making them more accurate and reliable. Evaluation metrics measure how well these algorithms perform, ensuring the results are trustworthy. Together, these elements help advance CRC image analysis.

### 2.2.1 Datasets

ML and DL algorithms depend on meticulously curated datasets to uncover hidden patterns and facilitate accurate image segmentation of CRC. In the realm of CRC image segmentation, a few datasets have garnered widespread recognition, each contributing unique strengths and collectively advancing the frontiers of research and clinical practice. Table 2.1 presents some datasets used in the CRC image segmentation. In our study, we

Table 2.1: Some of the publicly available CRC imaging datasets Tamang and Kim (2021)

Name	Size	Image Space	Ground Truth	Origin
KVASIR Pogorelov et al. (2017)	1000 iamges	720x576 to 1280x1024	dataset manually masks annotated by a medical doctor	Endoscopy
Colorectal adenocarcinoma gland(CRAG) Graham et al. (2019)	213 images	1512 x 1512	Instance-level annotations for gland locations	38 WSI at 20x
Gland Segmentation (Glas) Sirinukunwatana et al. (2017)	165 images	775 x 522	Precise annotations for gland boundaries and lumen structures	16 WSI at 20x

choose the glas dataset because it suits our approach and is available in Kaggle.

### 2.2.2 Evaluation Metrics

While datasets provide the raw material for learning, evaluation metrics serve as the discerning compass, guiding researchers toward the most effective segmentation strategies.

These metrics, acting as impartial arbiters, illuminate the strengths and weaknesses of various approaches, ensuring continuous progress toward clinical excellence.

As we train our model to detect colon cancer from gland images, careful evaluation is paramount to assess its performance and guide improvements. To this end, we employ a robust set of metrics, each offering a unique perspective on model quality Huynh (2021).

- **Accuracy:** This metric reflects the overall success rate of the model, capturing the proportion of correctly classified instances. The formula for accuracy is:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (2.1)$$

- **Recall:** measures the percentage of actual positive instances that were correctly identified by the model out of the total number of actual positive instances. The formula for the recall is:

$$Recall = \frac{TP}{TP + FN} \quad (2.2)$$

- **Precision:** measures the percentage of instances predicted as positive by the model that is actually positive out of the total number of instances predicted as positive. The formula for precision is:

$$Precision = \frac{TP}{TP + FP} \quad (2.3)$$

where

- TP: True Positives represent the cases where the model correctly predicted the positive class. In other words, these are instances where both the actual value and the predicted value are positive.
- TN: True Negatives represent the cases where the model correctly predicted the negative class. These are instances where both the actual value and the predicted value are negative.
- FP: False Positives occur when the model incorrectly predicts the positive class when the actual class is negative. In other words, these are instances where the model falsely predicts the presence of the condition.
- FN: False Negatives happen when the model incorrectly predicts the negative class when the actual class is positive. These are instances where the model fails to detect the presence of the condition.

- **F-Measure:** also known as F1 score is the harmonic mean of precision and recall. The formula for precision is:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (2.4)$$

- **IoU score:** also known as the Jaccard Index, measures the area of overlap between the predicted gland segmentation and the ground truth (actual gland mask) in the image. The formula for IoU score is:

$$IoU = \frac{TP}{TP + FP + FN} \quad (2.5)$$

A higher IoU score indicates a better match between the predicted and actual gland segmentation. A perfect score of 1 signifies complete overlap, while 0 suggests no overlap.

- **Dice Coefficient:** also known as the Dice Similarity Coefficient or Dice Score, is a measure used to assess the similarity between two sets of data, often represented as binary arrays. It is commonly employed to evaluate the overlap between predicted and ground truth segmentation masks. The formula of the Dice Coefficient is:

$$Dice = \frac{2 \times |X \cap Y|}{|X| + |Y|} \quad (2.6)$$

Where:

- $X$  represents the predicted segmentation.
- $Y$  represents the ground truth.
- $|X|$  and  $|Y|$  denote the cardinality (number of elements) of sets  $X$  and  $Y$ .

The coefficient ranges from 0 to 1, where 1 signifies perfect overlap and 0 indicates no overlap.

## 2.3 Literature Review

In this section, we succinctly review the main techniques used for colorectal cancer segmentation and detection. We categorize these works into three classes: classical tech-

niques, ML-based techniques, and DL-based techniques.

### 2.3.1 Classical Techniques

The initial efforts in gland segmentation from histological images involved extracting features characteristic of gland structures. These features typically focused on identifying substructures like cell nuclei or lumens, or on intensity variations associated with hematoxylin staining. For instance, Paul and Mukherjee (2016) propose a three-step method for analyzing gland morphology in Hematoxylin and Eosin (H&E) stained histology images. First, they apply an edge-preserving filter to smooth the image while maintaining the integrity of the epithelial layer (EL) boundaries. These boundaries hold crucial information about gland morphology. Next, they segment the EL using informative morphological scale space, a technique that analyzes the image at various scales to extract features of different sizes. The *informative* aspect likely refers to selecting scales most suitable for capturing the EL structures. Finally, they fill the regions within the segmented EL to obtain the complete gland. The method was evaluated on the Glas dataset, achieving a moderate F1 score of 0.68 and a good Dice coefficient of 0.75, indicating successful segmentation with good overlap between the automatically identified glands and the actual glands.

Cheikh et al. (2016) utilized color classification techniques to pinpoint the locations of cell nuclei. They employed advanced morphological operations on segmented nuclei objects to analyze their spatial arrangement, thereby discerning glandular regions. These methods were integrated to generate glandular structures. Their approach was tested on a Warwick-QU dataset as part of the Glas2015, resulting in an F-score of 0.8663 and an object dice of 0.9113. Notably, this approach offers scalability across various histology stains, such as immunohistochemistry and fluorescence, as well as different types of glandular tissues like prostatic glands and mammary glands. These classical techniques achieved promising results, laying the groundwork for more advanced segmentation methods. Their reliance on handcrafted features, however, often limited their robustness to image variations and complex gland shapes.

### 2.3.2 Machine Learning-based Techniques

Classical methods for gland segmentation have laid the groundwork for more sophisticated approaches. Recent advancements in this field have seen a significant shift towards ML-based methodologies, capable of autonomously learning discriminative features from data. These features, such as intensity variations, textures, and spatial relationships, can then be used to differentiate gland pixels from background pixels, facilitating accurate segmentation.

Segmentation, a crucial initial step in this process, involves dividing the image into meaningful regions, facilitating precise delineation of gland boundaries. Subsequently, ML models engage in classification, employing techniques such as *Support Vector Machines* (SVM), *decision trees*, *k-nearest neighbors*, etc. Tamang and Kim (2021), where learned features aid in identifying glandular tissues. The following paragraphs will provide a comprehensive review of prevalent ML-based research focused on gland segmentation.

Khatun and Chatterjee (2018) introduced an algorithm designed for the automated segmentation of glands in colon histology. The method relies on local intensity and texture features, utilizing images from the GlaS dataset. The images are partitioned into patches of varying window sizes, from which the intensity and texture-based features are extracted. Classification of these patches into different labels is achieved through the application of a *Random Forest classifier*. Notably, the proposed algorithm employs a multilevel Random Forest technique in a hierarchical manner, ensuring both speed and accuracy. Its applicability in a clinical setup is emphasized. The algorithm demonstrates an average accuracy ranging from 69% to 84%, contingent on the specific test image utilized, affirming its effectiveness in automated gland segmentation for practical clinical applications.

Alfonso et al. (2018) proposed a method for gland segmentation that utilizes Hematoxylin and Eosin (H&E) staining information. The method first separates the H&E image into its constituent channels and then focuses on the hematoxylin channel. Smoothing and thresholding are applied to this channel to detect likely gland candidate regions based on their characteristic white coloration. Further refinement steps are employed to achieve a more precise set of candidate glands. Surrounding these candidates, the method segments the nuclei using a *watershed algorithm*. Finally, a feature extraction step is performed to obtain information about both the candidate glands and their neighboring

nuclei. These features encompass shape, color, and spatial relationships. The authors validated their method on a dataset of 1330 fields of view obtained from H&E whole slide images of patients with various conditions. The dataset was manually annotated for gland locations, resulting in 2372 labeled glands. Their method achieved an accuracy of around 90% and an F-score of approximately 72%.

Rathore et al. (2019) aimed to develop an accurate technique for diagnosing colon cancer using machine learning on digital pathology images, focusing on distinguishing between benign and malignant tumors. They utilized the RMC and GlaS datasets and employed a gland segmentation algorithm that involved several steps. These steps included clustering tissue components using K-means, extracting features from both the overall image and individual glands, and employing the Radial Basis Function (RBF) kernel of SVM in the initial layer of detection. Subsequently, a second layer of SVM classification utilized Linear, RBF, and sigmoid kernels. The final classification was determined through majority voting based on predictions from these kernels. Cross-evaluates the intersection of two data sets, detection accuracy achieved 94.5% (Train RMC – Test GlaS) and 93.7% (Train GlaS – Test RMC), while classification accuracy reached 95% for both Train RMC – Test GlaS and Train GlaS – Test RMC scenarios. However, the study noted a limitation: the segmentation method was specifically designed for glandular regions and may not be suitable for accurately separating benign and malignant tissue across broader tissue regions. This highlights a potential area for future refinement in their approach.

### 2.3.3 Deep Learning-based Techniques

The DL, particularly convolutional neural networks (CNNs), has significantly advanced gland segmentation by automatically learning complex features from images, such as spatial relationships, textures, and intricate gland boundaries. These features are then leveraged to differentiate gland pixels from background pixels, leading to highly accurate segmentation. Unlike traditional methods, CNNs excel at discerning intricate glandular structures, distinguishing between healthy and malignant tissues. Their scalability and adaptability make them well-suited for the challenges of colon gland image segmentation. The subsequent paragraphs will delve into a comprehensive review of prevalent DL-based research focused on the diagnosis of colon cancer.

The approach proposed by Kosov et al. (2020) for gland segmentation employs a two-step strategy that effectively integrates traditional image processing techniques with DL. The first step constructs a probability map for gland positions based on shape and intensity features. A grid is overlaid on the image, and at each grid point, a weighted graph is created to analyze local image properties. This analysis leads to the identification of potential gland contours. However, this reliance on shape analysis might limit the effectiveness for segmenting malignant glands with irregular shapes. The second step employs a pre-trained CNN to classify these candidate regions, refining the segmentation by removing those unlikely to be glands. While this method achieves good results with an accuracy of 0.825 and a Dice score of 0.790 on the Glas dataset, it may be particularly suited for benign glands due to their well-defined shapes.

Mei et al. (2020) introduced the Dense Contour-Imbalance Aware (DCIA) framework for precise object segmentation in images. Leveraging DenseNet for feature extraction and Focal Loss to address challenging edges, DCIA significantly outperformed existing methods. The framework polished its predictions with two refinements, leading to superior performance in visual quality and Hausdorff distance metrics. Compared to a similar method, DCIA demonstrated accuracy improvements of up to 12%, promising precise object separation even for intricate outlines. These results highlight DCIA's potential for advanced image segmentation tasks.

Dabass et al. (2021) introduced a modified U-Net architecture. This study presents a comprehensive methodology for gland segmentation in medical imaging, characterized by meticulous pre-processing steps including manual ground truth delineation, standardized patch selection, stain normalization, data augmentation, and quality control. At the core of the method lies a sophisticated symmetrical Encoder-Decoder Network Architecture. This architecture not only boasts symmetrical design but also integrates Atrous-Residual Units for robust feature extraction, Attention Units for gland-specific emphasis, and Transitional Atrous Units to address resolution issues. A weighted cross-entropy loss function guides the training process, ensuring a balanced ratio of foreground and background classes and enhancing segmentation accuracy. Post-processing steps involving thresholding and morphological operations refine the final segmentation. This methodology achieves impressive results, with an F-score of 0.93 and Object-dice coefficient of 0.94 on the Glas dataset, and an F-score of 0.91 and object-dice coefficient of 0.94 on the CREG dataset, underscoring its robustness and effectiveness across varied datasets and scenarios.

Wang et al. (2022) propose a method, called Topology-Aware Network (TA-Net), that aims to improve gland segmentation using a specialized neural network. TA-Net has two main parts: an Encoder and two decoder branches. It uses a framework called SegNet for the Encoder, which is like the backbone of the network, and this helps with capturing important features. The decoder branches focus on two tasks: separating glands from the background (INST segmentation) and understanding the arrangement of glands (TOP estimation). To enhance accuracy, the network considers two aspects: the shape of glands and specific markers within them. This helps prevent errors like over-segmentation or under-segmentation. After the initial segmentation, the network applies a technique called the Watershed algorithm along with some adjustments to refine the results further. Evaluation on the GlaS and CRAG datasets, TA-Net achieved significant improvements in accuracy, with F1 scores of 90.5 and 84.2, and object Dice scores of 90.2 and 89.3, respectively.

Sun et al. (2023). introduced a new approach for image segmentation using diffusion models. Their method incorporates a ResNet and Feature Pyramid Network (FPN) within an image Encoder to extract multi-scale features from input images. A diffusion model Decoder then applies noise progressively to the image in a forward process, learning to reverse this to reconstruct the original image with conditional encoding and Instance-aware Filters. During this reconstruction, segmentation information is integrated into the Decoder. Additionally, a separate mask branch with a dynamic mask head utilizes multi-scale data to predict precise instance masks, addressing potential detail loss during diffusion. Training utilizes a combination of two loss functions with a balancing parameter for optimization. Their method achieves notable results on medical image datasets: an F1 score of 0.98 and object Dice score of 0.93 on the Glas dataset, and an F1 score of 0.85 and object Dice score of 0.83 on the CRAG dataset. However, challenges remain in handling image quality issues such as artifacts and noise. Overall, these findings underscore the method's significant potential for automated gland segmentation, while also identifying avenues for enhancing its practical utility.

## 2.4 Conclusion

In summary, the state-of-the-art exploration reveals the evolution of image segmentation techniques in the context of colorectal cancer diagnosis. From classic methods to



cutting-edge deep learning, each approach contributes valuable lessons, guiding the development of our proposed solution. As we transition to the experimental phase, the insights gained from this section will inform and validate the practical application of our method, promising advancements in the precision and clinical impact of colorectal cancer image segmentation.

---

---

# CHAPTER 3

---

## EXPERIMENT

### 3.1 Introduction

In this chapter, we present our approach to training the Mask R-CNN model to predict an instance mask for each gland. We then combine this model with the Watershed algorithm for enhanced segmentation and to address overlapping glands. Following this, we present the results obtained from our study to evaluate whether the combination of the watershed algorithm and Mask R-CNN improves the segmentation. Additionally, we discuss the challenges associated with this approach.

### 3.2 Implementation Setup

We introduce the dataset used in our experiment, then the environment setups that we worked on.

### 3.2.1 Dataset

We utilized a dataset obtained from the Glas challenge, The dataset consists of 165 histological images derived from 16 H&E stained histological sections of stage T3 or T4 colorectal adenocarcinoma. Each section originates from a different patient.

The histological sections were digitized into whole-slide images (WSIs) using a Zeiss MIRAX MIDI Slide Scanner, with a pixel resolution of 0.465  $\mu\text{m}$ , which were later rescaled to a pixel resolution of 0.620  $\mu\text{m}$ , equivalent to a 20 $\times$  objective magnification. The dataset contains masks for each image, delineating the glandular structures, created by an expert pathologist. These masks serve as valuable ground truth data for automated segmentation algorithms.

The dataset is divided into 85 images for training and 80 for testing divided into two parts: test A with 60 images and test B with 20 images.

The Test A sub-dataset contains a mix of benign and malignant glands, including healthy, normal-shaped glands and various stages of abnormal growths, contributing to more balance.

While, the Test B sub-dataset consists entirely of abnormal glands, with no glands of normal shape.

Table 3.1 representing the different grades of the glands and the percentage of each grade in the subtest A and B:

Table 3.1: Percentage of each grade type in Test A and Test B Subdatasets

<b>Grade Type</b>	<b>test A(%)</b>	<b>test B(%)</b>
<b>Adenomatous</b>	<b>20.00</b>	<b>20.00</b>
<b>Healthy</b>	<b>35.00</b>	<b>0.00</b>
<b>Moderately Differentiated</b>	<b>25.00</b>	<b>40.00</b>
<b>Moderately-to-Poorly Differentiated</b>	<b>6.67</b>	<b>20.00</b>
<b>Poorly Differentiated</b>	<b>13.33</b>	<b>20.00</b>

This dataset served as a cornerstone for our investigation into gland segmentation methodologies tailored specifically to the analysis of Colorectal Cancer. Here, we showcase

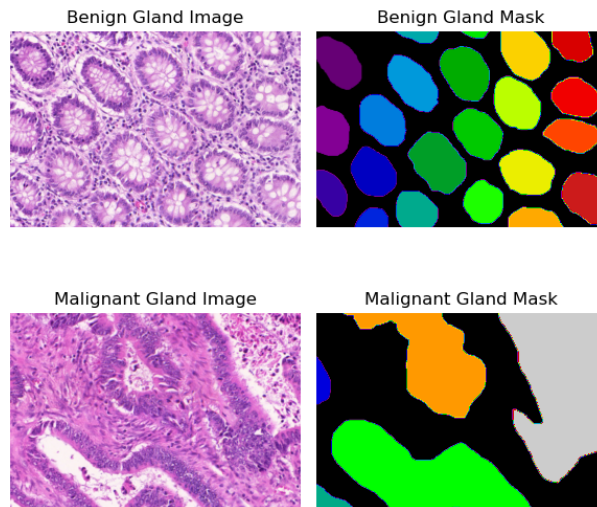


Figure 3.1: Histological images and their corresponding masks for benign and malignant gland cancer types.

examples from the Glas dataset in Figure 3.1, featuring both malignant and benign gland images along with their corresponding masks. It's noteworthy that the images maintain dimensions of  $775 \times 522$  pixels.

**Data Augmentation:** to enhance the generalization capability of our model, we employed data augmentation techniques. This process involved applying a series of geometric transformations to both the images and their corresponding segmentation masks. These transformations included horizontal flipping, vertical flipping, and rotation by 120 degrees.

### 3.2.2 Environment

For this experiment, we chose Anaconda as our platform to manage software environments. It includes Jupyter Notebook, a user-friendly interface for writing and running Python code.

Due to the computationally demanding nature of Mask R-CNN training, which requires significant time and powerful hardware, we trained our model on a high-performance server equipped with an Intel(R) Xeon(R) Silver 4216 CPU (2.10GHz) processor with 32 cores and 128GB of RAM running Ubuntu 22.04 LTS.

Tests were performed on an Intel(R) Core(TM) i5-6200U CPU (2.30 GHz), 8 GB of RAM, and Windows 10 operating system.

Python: Python is a high-level programming language that is easy to use. It's flexible and boasts a rich collection of libraries and frameworks. Below are listed some used libraries.

Torchvision <sup>1</sup>: is a library for PyTorch for computer vision. It streamlines workflows with pre-built datasets, image transformations, and pre-trained models (like Mask R-CNN) for faster development and better results.

Numerical Python <sup>2</sup> (NumPy): It provides effective multidimensional arrays along with an extensive library of mathematical functions for manipulating and analyzing data.

OpenCV <sup>3</sup>: is a powerful library for computer vision tasks. It offers tools for various tasks like object detection, feature extraction, and picture processing.

Matplotlib <sup>4</sup>: is an excellent tool for creating a variety of data visualizations, including efficient picture charting. It comes with the 'imshow' feature, which lets one easily incorporate photos with other plots. This gives the ability to thoroughly examine and evaluate visual data. The library is utilized to produce diverse data displays, encompassing visuals.

### 3.3 Our Approach

Our approach is divided into two parts:

#### Part 1: Mask R-CNN Model

We leverage transfer learning to tackle gland segmentation in CRC images. A pre-trained Mask R-CNN model with a ResNet-50 backbone serves as the foundation. This

---

<sup>1</sup>Torchvision, [https://pytorch.org/tutorials/intermediate/torchvision\\_tutorial.html](https://pytorch.org/tutorials/intermediate/torchvision_tutorial.html)

<sup>2</sup>NumPy, <https://numpy.org/>

<sup>3</sup>OpenCV, <https://opencv.org/>

<sup>4</sup>Matplotlib, <https://matplotlib.org/>

model was originally trained for the detection and classification of daily objects.

The Torchvision library streamlines the application of transfer learning of pre-trained models like Mask R-CNN. To fine-tune the model we freeze the bounding box layers, and we essentially transfer this valuable knowledge to the domain of gland segmentation. We trained the model on different experiments as we will show. This focuses learning on predicting segmentation masks through fine-tuning with the SGD optimizer.

Our training process involved several experiments to optimize the model's performance. Initially, we trained the model for 5 epochs with a batch size of 2. In the second experiment, we extended the training to 15 epochs while maintaining the batch size of 2. The third experiment also used 15 epochs but increased the batch size to 3. In the fourth and final experiment, we trained the model for 15 epochs with a batch size of 16. Throughout these experiments, we observed a consistent improvement in the Intersection over Union (IoU) scores, indicating better segmentation accuracy with each iteration. This iterative training approach allowed us to fine-tune the model effectively, enhancing its ability to generalize and perform precise gland segmentation.

### **Part 2: Post-processing with Morphological Methods and Watershed Algorithm**

Despite the robustness of the Mask R-CNN model, some of its gland outlines sometimes overlapped in the images, a problem that occurred more frequently in malignant glands than in benign ones. This is because malignant glands often have irregular, dynamic shapes, whereas benign glands have more static, regular shapes. To fix this and get perfectly separated glands, we applied some additional techniques. First, we got rid of any tiny objects the model might have mistakenly identified as glands. This cleaning step reduced errors and made the overall gland identification more accurate. Next, we used a method called erosion to slightly shrink the outlined gland areas, which helped separate glands that were practically touching. Then, to bring the glands back to their original size, we used dilation. Finally, we used the Watershed algorithm, along with the distance transform, to further improve the separation. This last step imagines the image as a landscape with hills (glands) and valleys (background). The distance transform calculates how far each point in the image is from the nearest valley. By finding the highest points in this distance map, we could identify the best places to create boundaries and separate the overlapping glands.

The architecture of our approach is illustrated in Figure 3.2.

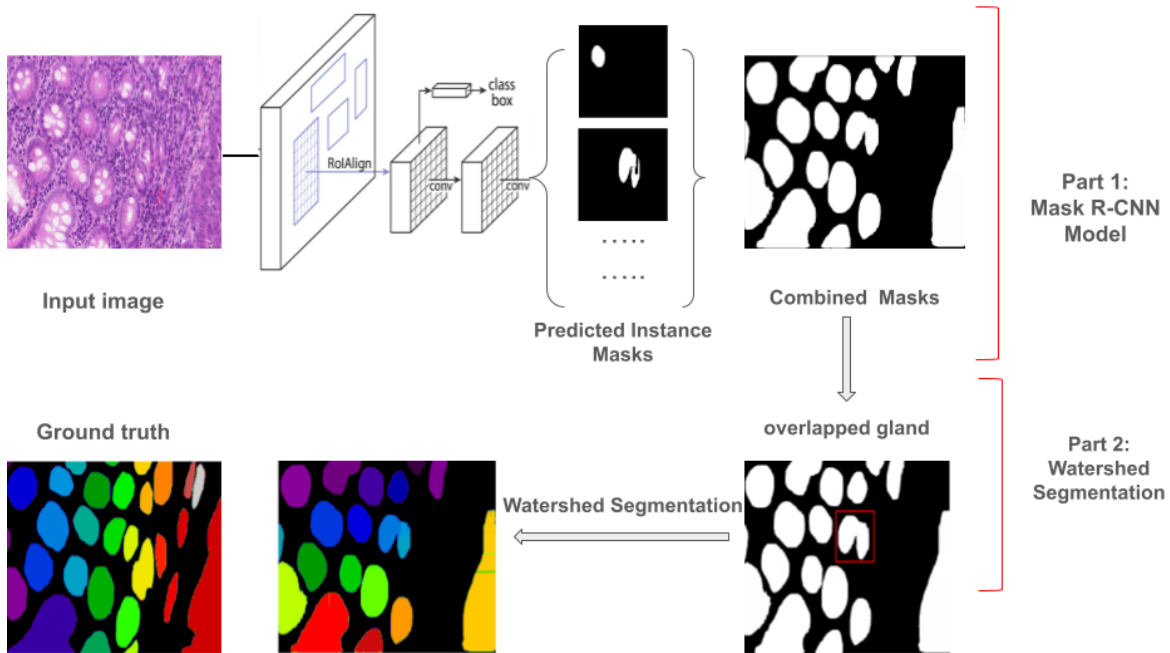


Figure 3.2: The architecture of our approach

## 3.4 Results and Discussion

We present the results obtained from our model and discuss these findings. We divide this section into two parts: first, the results obtained from testing the Mask R-CNN model on Test A and Test B datasets, and second, the results of combining the Mask R-CNN model with post-processing techniques, including morphological methods and the watershed algorithm.

### 3.4.1 Results

#### Part 1: Mask R-CNN Model Testing

In this part, we evaluate the performance of the Mask R-CNN model using standard metrics such as Intersection over Union (IoU), Dice coefficient, Precision, Recall, and F1

score. The model was tested on two datasets, Test A and Test B. The result obtained are presented in Table 3.2 as follows:

Table 3.2: Mask R-CNN Model Testing Results

Metric	Test A	Test B
Average IoU	0.7672	0.7260
Average Dice Coefficient (object dice)	0.8566	0.8304
Average Precision	0.8285	0.7722
Average Recall	0.9082	0.9285
Average F1 Score	0.8665	0.8431

The Mask R-CNN model gives good results in predicting instance segmentation. It shows particularly strong performance in Test A as shown in figure 3.3, likely because this test set is balanced, containing both malignant and benign glands. In contrast, Test B is composed of glands with more difficult shapes, primarily malignant, which presents greater challenges for the model.

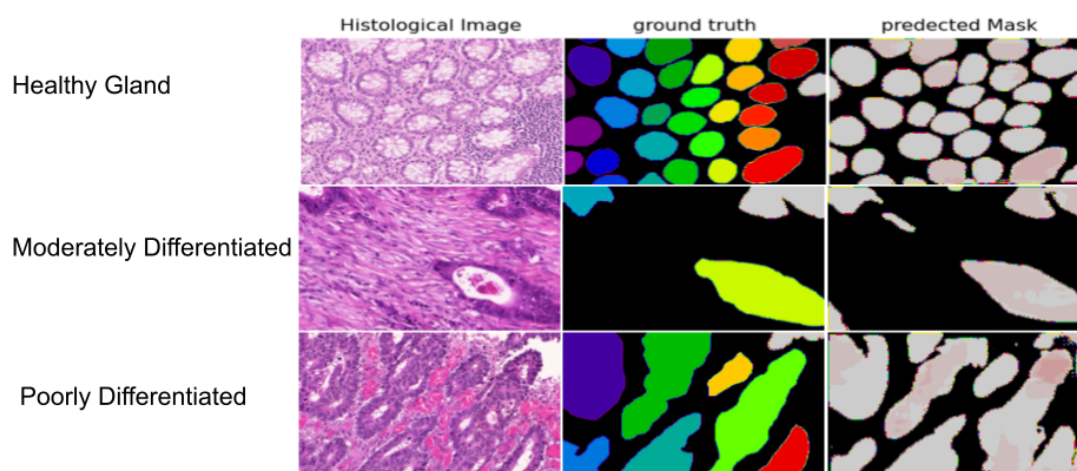


Figure 3.3: Predicted Masks of Various Gland Grades from Mask R-CNN (Healthy, Moderately Differentiated, Poorly Differentiated)

## Part 2: Post-Processing with Morphological Methods and Watershed Algorithm

While the Mask R-CNN model provided good segmentation results, we observed that it had difficulties in accurately predicting some small objects identified as glands and in



separating overlapping glands, especially the malignant glands. To address these issues, we applied additional post-processing steps, including morphological operations (erosion and dilation) and the watershed algorithm. These techniques aimed to improve the segmentation masks produced by the Mask R-CNN model.

The results after applying the post-processing steps are summarized in the table 3.3 and shown in figure 3.4 below:

Table 3.3: The Testing Results After the Post-processing

Metric	Test A	Test B
Average IoU	<b>0.7833</b>	<b>0.7270</b>
Average Dice Coefficient (object dice)	<b>0.8678</b>	<b>0.8304</b>
Average Precision	<b>0.8378</b>	<b>0.7722</b>
Average Recall	<b>0.9206</b>	<b>0.9285</b>
Average F1 Score	<b>0.8772</b>	<b>0.8431</b>

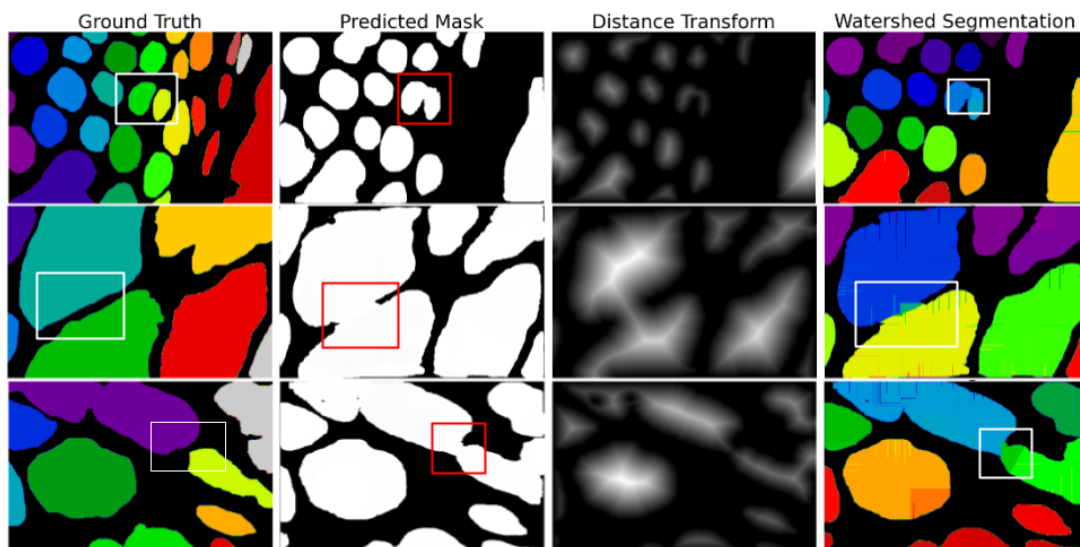


Figure 3.4: The predicted masks from mask R-CNN and the Results after applying The Watershed Algorithm

### 3.4.2 Discussion

After presenting the results, this section will discuss these findings and evaluate the strengths and challenges of our approach.

In the first part, the Mask R-CNN model gives good results in predicting instance segmentation. It shows particularly strong performance in Test A, This is likely because Test A is balanced. In contrast, Test B is composed of glands with more difficult shapes, primarily malignant, which presents greater challenges for segmentation.

Although the Mask R-CNN model provides a solid starting point for gland segmentation, applying the watershed algorithm to handle overlapping glands and employing morphological methods significantly improves performance. This combination of the Mask R-CNN and the watershed algorithm enhances segmentation accuracy, especially in Test A. Specifically, the Dice coefficient and F1 score increased by approximately 1.18% after applying morphological methods, while the IoU increased by approximately 2.63%.

The improvement wasn't as dramatic in Test B because the oddly shaped glands were inherently harder to segment precisely. Overall, the study suggests that combining Mask R-CNN's power with the classic watershed technique is a promising approach for gland segmentation.

One limitation of our approach is the long training time required for the Mask R-CNN model, which prevented us from experimenting extensively with different parameters. Additionally, the scarcity of histological datasets with gland annotations, especially those representing colorectal cancer, poses a challenge. Nevertheless, the model demonstrated good results, indicating its potential for effective gland segmentation.

## 3.5 Conclusion

In this section, we presented our experiment. We trained the Mask R-CNN model to predict the instance masks of glands. This first step yielded good results, especially on Test A. The second step involved applying the watershed algorithm to handle the overlapping predicted masks. This step improved segmentation performance, particularly in Test A, where the Dice coefficient and F1 score increased by approximately 1.18%, and the IoU increased by approximately 2.63%.

Overall, the combination of Mask R-CNN and the watershed algorithm proved effective in enhancing gland segmentation, particularly in balanced datasets. However, the

challenges posed by more complex gland shapes in Test B highlight the need for further optimization and refinement of our approach.

---

# CONCLUSION

Colorectal gland cancer diagnosis remains a challenging task for pathologists, often requiring extensive time, expertise, and can be particularly difficult when the cancer is located deeper within tissues. The identification and accurate localization of glands within complex histological images are crucial for effective cancer classification.

Our study aims to assist doctors in the analysis of colorectal gland cancer by leveraging computer vision, specifically through image segmentation. We focus on segmenting glands in histological images, accurately identifying their locations among various other tissues, and analyzing their shapes using AI. This approach can significantly aid in the accurate classification of glands as cancerous or non-cancerous, ultimately contributing to more efficient and precise diagnosis.

We began by reviewing the role of medical imaging in cancer diagnostics and defining some types of medical images. Additionally, we discussed some deep learning architectures used in image segmentation. To gain insight into the studies that employ this approach, we reviewed several studies that use machine learning and deep learning to segment glands in medical images.

Based on this foundation, we presented our approach, which combines the powerful deep learning architecture Mask R-CNN with the watershed algorithm for segmentation. We applied the watershed algorithm on the predicted masks by the Mask R-CNN model.

The data used in our study was the GlaS dataset.

The Mask R-CNN alone produced good results, achieving a Dice coefficient of 0.85, an F1 score of 0.86, and an IoU of 0.76. However, we sought to enhance these outcomes. By applying the watershed algorithm and some morphological methods to the predicted masks, we achieved a Dice coefficient and F1 score increase of approximately 1.18%, and the IoU increased by approximately 2.63%.

Despite the promising results, our study faced some challenges. We found that Mask R-CNN is computationally heavy and requires significant time and powerful hardware to train. Additionally, the limited availability of histological images of CRC posed constraints on generating accurate segmentations.

Looking towards the future, medical image segmentation holds immense promise. By leveraging deep learning architectures for object segmentation in medical images and by combining these with established segmentation methods for evaluation, significant advancements can be made. The inherent complexity of medical images, particularly histological ones with diverse tissue types, necessitates exploring preprocessing techniques to isolate the objects of interest from surrounding tissues. This can significantly aid deep learning techniques in feature extraction, ultimately improving segmentation performance.

---

# BIBLIOGRAPHY

- Alfonso, S., Corredor, G., Moncayo, R., Barrera, C. R., Sanchez, A. Y., Toro, P., and Romero, E. (2018). A method to detect glands in histological gastric cancer images. In Romero, E., Lepore, N., and Brieva, J., editors, *14th International Symposium on Medical Information Processing and Analysis*, volume 10975, page 109750X. International Society for Optics and Photonics, SPIE. (Cited in pages 2 et 42.)
- Anantharaman, R., Velazquez, M., and Lee, Y. (2018). Utilizing mask r-cnn for detection and segmentation of oral diseases. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2197–2204. (Cited in pages iv, 33 et 35.)
- Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. (Cited in page 17.)
- Barge, W., Albertini, M. R., Cold, C., and Abbott, D. (2021). Retrograde balloon-assisted deep enteroscopy in the diagnosis of metastatic melanoma. *Case Reports in Gastrointestinal Medicine*, 2021. (Cited in pages iii et 11.)
- Blue, S. T. and Brindha, M. (2019). Edge detection based boundary box construction algorithm for improving the precision of object detection in yolov3. In *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pages 1–5. (Cited in pages iv et 25.)
- Brook, O. R., Sun, M., Kane, R. A., and Kruskal, J. B. (2013). Endorectal ultrasound. In *Abdominal Imaging*, pages 759–768. Springer. (Cited in page 9.)

- Chakraborty, C., Malhotra, P., Gupta, S., Koundal, D., Zaguia, A., and Enbeyle, W. (2022). Deep neural networks for medical image segmentation. *Journal of Healthcare Engineering*, 2022:9580991. (Cited in pages 23, 26, 27, 28 et 29.)
- Cheikh, B. B., Bertheau, P., and Racoceanu, D. (2016). A structure-based approach for colon gland segmentation in digital pathology. In Gurcan, M. N. and Madabhushi, A., editors, *Medical Imaging 2016: Digital Pathology*, volume 9791, page 97910J. International Society for Optics and Photonics, SPIE. (Cited in pages 1 et 41.)
- Chen, D., Huang, J.-F., Liu, K., Zhang, L.-Q., Yang, Z., Chuai, Z.-R., Wang, Y.-X., Shi, D.-C., Huang, Q., and Fu, W.-L. (2014). BRAF V600E mutation and its association with clinicopathological features of colorectal cancer: A systematic review and meta-analysis. *PLOS ONE*, 9(3):e90607. (Cited in page 13.)
- Dabass, M., Vashisth, S., and Vig, R. (2021). Attention-guided deep atrous-residual u-net architecture for automated gland segmentation in colon histopathology images. *Informatics in Medicine Unlocked*, 27:100784. (Cited in page 44.)
- Doi, K. (2007). Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Computerized Medical Imaging and Graphics*, 31(4):198–211. Computer-aided Diagnosis (CAD) and Image-guided Decision Support. (Cited in pages iii et 13.)
- Esteva, A., Chou, K., Yeung, S., and et al. (2021). Deep learning-enabled medical computer vision. *npj Digital Medicine*, 4(1):5. (Cited in pages iii, 5 et 6.)
- Furness, P. N. (1999). The use of digital images in pathology. *Pathology*, 183(3):253. (Cited in page 12.)
- Giger, M. L. (2018). Machine learning in medical imaging. *Journal of the American College of Radiology*, 15(3, Part B):512–520. Data Science: Big Data Machine Learning and Artificial Intelligence. (Cited in pages 14, 15 et 19.)
- Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. (Cited in page 17.)
- Graham, S., Chen, H., Gamper, J., Dou, Q., Heng, P.-A., Snead, D., Tsang, Y. W., and Rajpoot, N. (2019). Mild-net: Minimal information loss dilated network for gland instance segmentation in colon histology images. *Medical Image Analysis*, 52:199–211. (Cited in page 38.)
- He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask r-cnn. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988. (Cited in pages iv, 2, 15, 16 et 34.)
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778. (Cited in page 22.)

- Huang, J. (2017). Fast r-cnn and faster r-cnn: Demystified. (Cited in pages iv, 26 et 27.)
- Hui, J. (Mar 28, 2018). What do we learn from region-based object detectors (faster r-cnn, r-fcn, fpn)? (Cited in pages iv, 27, 28, 29, 30, 31 et 32.)
- Huynh, N. (2021). Understanding evaluation metrics in medical image segmentation. (Cited in page 39.)
- Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., and Keutzer, K. (2016). Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <0.5mb model size. In *arXiv preprint arXiv:1602.07360*. (Cited in page 23.)
- Key, O. (2018). Endorectal ultrasound. Accessed: 18-12-2023. (Cited in pages iii et 10.)
- Khatun, R. and Chatterjee, S. (2018). Machine learning approach for segmenting glands in colon histology images using local intensity and texture features. In *2018 IEEE 8th International Advance Computing Conference (IACC)*, pages 314–320. (Cited in page 42.)
- Kosov, A., Khvostikov, A., and Krylov, A. (2020). Adaptive method of glands segmentation on histological images. *Proceedings of the 30th International Conference on Computer Graphics and Machine Vision (GraphiCon 2020). Part 2*, pages paper39–1. (Cited in pages iii, 8 et 44.)
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1097–1105. (Cited in page 22.)
- Laghi, A., Iannaccone, R., Carbone, I., Catalano, C., Panebianco, V., Di Giulio, E., Schillaci, A., and Passariello, R. (2002). Computed tomographic colonography (virtual colonoscopy): blinded prospective comparison with conventional colonoscopy for the detection of colorectal neoplasia. *Endoscopy*, 34(06):441–446. (Cited in page 7.)
- Larobina, M. and Murino, L. (2014). Medical image file formats. *Journal of Digital Imaging*, 27:200–206. (Cited in page 12.)
- Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (Cited in page 17.)
- McCombe, K. D., Craig, S. G., Viratham Pulsawatdi, A., Quezada-Marín, J. I., Hagan, M., Rajendran, S., Humphries, M. P., Bingham, V., Salto-Tellez, M., Gault, R., and James, J. A. (2021). Histoclean: Open-source software for histological image pre-processing and augmentation to improve development of robust convolutional neural networks. *Computational and Structural Biotechnology Journal*, 19:4840–4853. (Cited in page 14.)



- Mei, L., Guo, X., Huang, X., Weng, Y., Liu, S., and Lei, C. (2020). Dense contour-imbalance aware framework for colon gland instance segmentation. *Biomedical Signal Processing and Control*, 60:101988. (Cited in page 44.)
- Mukai, K., Ishida, Y., Okajima, K., et al. (2006). Usefulness of preoperative fdg-pet for detection of gastric cancer. *Gastric Cancer*, 9:192–196. (Cited in page 11.)
- Odemakinde, E. (2023). Mask r-cnn - deep learning for object detection and segmentation. (Cited in pages iv, 16, 24 et 33.)
- Paul, A. and Mukherjee, D. P. (2016). Gland segmentation from histology images using informative morphological scale space. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 4121–4125. (Cited in page 41.)
- Pizzi, A. D., Basilico, R., Cianci, R., et al. (2018). Rectal cancer mri: protocols, signs and future perspectives radiologists should consider in everyday clinical practice. *Insights Imaging*, 9:405–412. (Cited in page 8.)
- Pogorelov, K., Randel, K. R., Griwodz, C., Eskeland, S. L., de Lange, T., Johansen, D., Spampinato, C., Dang-Nguyen, D.-T., Lux, M., Schmidt, P. T., Riegler, M., and Halvorsen, P. (2017). Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection. *MMSys'17*, page 164–169, New York, NY, USA. Association for Computing Machinery. (Cited in page 38.)
- Rathore, S., Iftikhar, M. A., Chaddad, A., Niazi, T., Karasic, T., and Bilello, M. (2019). Segmentation and grade prediction of colon cancer digital pathology images across multiple institutions. *Cancers*, 11(11). (Cited in pages 2 et 43.)
- Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc. (Cited in pages iv, 17 et 32.)
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241. Springer. (Cited in pages 17 et 21.)
- Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., and Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, 35(5):1285–1298. (Cited in page 21.)
- Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations (ICLR)*. (Cited in page 22.)

- Sirinukunwattana, K., Pluim, J. P., Chen, H., Qi, X., Heng, P.-A., Guo, Y. B., Wang, L. Y., Matuszewski, B. J., Bruni, E., Sanchez, U., Böhm, A., Ronneberger, O., Cheikh, B. B., Racoceanu, D., Kainz, P., Pfeiffer, M., Urschler, M., Snead, D. R., and Rajpoot, N. M. (2017). Gland segmentation in colon histology images: The glas challenge contest. *Medical Image Analysis*, 35:489–502. (Cited in pages 1, 2 et 38.)
- Sun, M., Wang, J., Gong, Q., and Huang, W. (2023). Enhancing gland segmentation in colon histology images using an instance-aware diffusion model. *Computational Biology and Medicine*, 166:107527. Epub ahead of print. (Cited in page 45.)
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9. (Cited in page 22.)
- Tamang, L. D. and Kim, B. W. (2021). Deep learning approaches to colorectal cancer diagnosis: A review. *Applied Sciences*, 11(22). (Cited in pages v, 38 et 42.)
- Tharwat, M., Sakr, N. A., El-Sappagh, S., Soliman, H., Kwak, K. S., and Elmogy, M. (2022). Colon cancer diagnosis based on machine learning and deep learning: Modalities and analysis techniques. *Sensors*, 22(23):9250. (Cited in pages iii et 14.)
- Tsai, M.-J. and Tao, Y.-H. (2021). Deep learning techniques for the classification of colorectal cancer tissue. *Electronics*, 10(14). (Cited in pages iv, 19 et 21.)
- Tsai, P. C., Lee, T. H., Kuo, K. C., et al. (2023). Histopathology images predict multi-omics aberrations and prognoses in colorectal cancer patients. *Nature Communications*, 14:2102. (Cited in page 10.)
- van de Sande, K. E. A., Uijlings, J. R. R., Gevers, T., and Smeulders, A. W. M. (2011). Segmentation as selective search for object recognition. In *2011 International Conference on Computer Vision*, pages 1879–1886. (Cited in pages iv et 24.)
- Wang, H., Xian, M., and Vakanski, A. (2022). Ta-net: Topology-aware network for gland segmentation. In *IEEE Winter Conf Appl Comput Vis*, pages 3241–3249. (Cited in pages 2 et 44.)
- WHO (2023). Colorectal cancer. <https://www.who.int/fr/news-room/fact-sheets/detail/colorectal-cancer>. Accessed: 15-12-2023. (Cited in pages 1 et 6.)
- Zhao, Z.-Q., Zheng, P., Xu, S.-T., and Wu, X. (2019). Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*, 30(11):3212–3232. (Cited in page 18.)

République Algérienne Démocratique et Populaire  
وزارة التعليم العالي و البحث العلمي  
Ministère de l'Enseignement Supérieur et de La Recherche Scientifique  
كلية العلوم والتكنولوجيا  
Faculté des Sciences et de la Technologie  
قسم الرياضيات و الإعلام الآلي  
Département des Mathématiques & de l'Informatique  
جامعة غرداية  
Université de Ghardaia



## شهادة الترخيص بالإيداع

أنا الأستاذ : سليمان بالاعور

بصفتي رئيس لجنة المناقشة والمسؤول عن تصحيح مذكرة الماستر الموسومة ب:

### Colorectal Cancer: Segmentation and Detection via Deep Learning Detection

من إنجاز الطالبة: بالو منصور

والطالبة: حروزي أسماء

الكلية: العلوم والتكنولوجيا  
القسم: الرياضيات والإعلام الآلي  
الشعبة: إعلام آلي

التخصص: الأنظمة الذكية لاستخراج المعارف  
تاريخ المناقشة: 2024/07/01

أشهد أن الطالبة قاموا بالتصحيات المطلوبة من طرف لجنة المناقشة وأن المطابقة بين النسخة الورقية والالكترونية استوفيت جميع شروطها.

مصادقة رئيس القسم

إمضاء المسؤول عن التصحيح

رئيس قسم الرياضيات والإعلام الآلي  
الحاج موسى ياسين



سليمان بالاعور