

Chapter 1

Introduction

1.1 Background and Motivation

**** This entire section is going to be rewritten ****

In this project we aim to design and develop a system for analyzing medical videos from a camera pill, as seen in Figure 1.1. The pill is swallowed and records video of the entire digestive system. The goal is to be able to detect different irregularities in the patient's digestive system, like a colon polyp, Chron's disease, Colorectal cancer, etc. by using video object tracking, object detection, machine learning or other relevant tools.

Neural networks models that we would like to explore further for this purpose are Convolutional neural networks (CNN), Recurrent neural networks (RNN), Capsule neural networks, Long Short-Term memory networks and more.

The main idea is to go beyond image-based methods and also exploit the time factor of the data. The videos we will be using for this is delivered by Bærum Hospital, and is carefully labeled by using tools such as described in the paper “Expert Driven Semi-Supervised Elucidation Tool for Medical Endoscopic Videos”. In this paper Albisser *et al.* presents a semi-supervised method to gather the annotations in a easy and time saving way [1].



Figure 1.1: Illustration of how such a camera pill could look like¹.

¹Image Credit: Medtronic

1.1.1 Nomenclature used in the field of capsule endoscopy

From a variety of literature searches we have found that there is five different terms in use:

- **VCE**: Video Capsule Endoscopy - capsule endoscopy including an imaging device such as a CCD (the capsule does not have to be wireless);
- **WVE**: Wireless Video Endoscopy (not necessarily a capsule);
- **CE**: Capsule Endoscopy - endoscopic capsule (not necessarily wireless);
- **WCE**: Wireless Capsule Endoscopy (not necessarily containing an image sensor);
- **WVC**: Wireless Video Capsule.

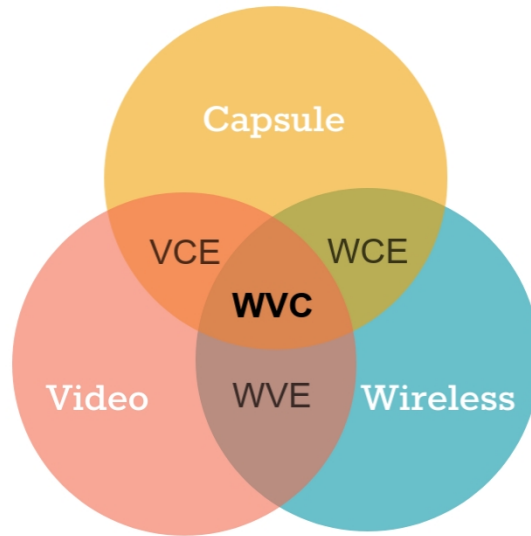


Figure 1.2: Distinction of the nomenclature relating to capsules, wirelessness and video.

The terms are often confused and used interchangeably. Consequently in this thesis they will all be referred to as WCE.

1.2 Problem statement

Colorectal cancer (CRC) is the third most common cause of cancer mortality for both men and women [2], and it is a condition where early detection is of clear value for the ultimate survival of the patient. As statistics show that 15% of male and female above 50 years are at risk, the procedure is recommended on a regular basis (every 3-5 years) for the population over 50, and from an earlier age for high-risk groups.

Colonoscopy is a demanding procedure requiring an significant amount of time by specialized physicians, in addition to the discomfort and risks inherent in the procedure. Traditional methods based on colonoscopy are not cost-effective for population-based screening purposes, so only about 2-3% of the target population is reached at present.

The cost of a population screening program is prohibitively expensive. Colonoscopy is the most expensive cancer screening process in the US, with annual costs of \$10 billion dollars (\$1100 per person). In Norway we have similar costs of around \$1000 per person, with a time consumption of about 1 doctor-hour and 2 nurse-hours per examination.

By researching an automatic system for a camera pill the aim is to greatly increase the number of patients that can be examined, i.e., making the public health care system more scalable and cost effective, while at the same time reducing the need for intrusive procedures like "bottom-up" examinations like colonoscopy.

1.3 Scope and limitations

Based on the described problem statement the scope of this thesis is to compare some prior selected automatic systems with focus on machine learning, for classification tasks in the medical domain. We will test the systems on video and still images taken from a wireless capsule endoscopy system, provided by Bærum Hospital. This data is collected with the Olympus Endocapsule 10 System² and gives us a wide view of the GI-tract. A limitation for our research will therefore be that our dataset lack a great deal of diversity. While we might have a wide range of patients all the data contributions stem from one provider in one location. This might not be much of an issue due to GI-tract being quite homogeneous over large populations. But can presumably cause biasing issues in our deep learning models.

By using a tool provided by Augere Medical, we have sequentially classified each frame in XXX WCE videos, assisted by doctor Thomas de Lange, with many years of experience. To the best of our efforts we managed to divide the data into XXX classes which we later will train our networks on.

Considering the scope of this thesis, we will limit ourselves to use some of the more common deep convolutional neural networks and a more novel stacked hourglass network. There are many other more common network used in the field of object classification like NN, RNN, statistical methods, but these methods are widely tested and we wish to push the envelope of machine learning systems on medical videos and also have a wide variety of ways to score our model against previously tested algorithms.

1.4 Research methods

We have based this paper on Association for Computing Machinerys (ACMs) research methodology. In the spring of 1986 ACM president Adele Goldberg and ACM Education Board Chairman Robert Aiken appointed a task force with the prime objective of describing the core fundamentals of computer science and computer engineering [3]. They introduced the phrase "discipline of computing" to embrace the two fields. The task force were given three objectives to complete.

1. Present a description of computer science that emphasizes fundamental questions and significant accomplishments. The definition should recognize that the field is constantly changing and that what is said is merely a snapshot of an ongoing process of growth.
2. Propose a teaching paradigm for computer science that conforms to traditional scientific standards, emphasizes the development of competence in the field, and harmoniously integrates theory, experimentation, and design.
3. Give a detailed example of an introductory course sequence in computer science based on the curriculum model and the disciplinary description.

To fully elaborate the core fundamentals of computer science and computer engineering they agreed upon no core fundamentals is different in the two fields, but the difference is manifested in the way the two fields elaborate the core; computer science focuses mainly on analysis and abstraction and computer engineering focuses mainly on abstraction and design.

Furthermore the task force appointed three major paradigms which represent different areas of competence in the field. Some will argue that the different paradigms are implicitly based on an assumption that one of the three processes is the most fundamental, but as we will see, the three paradigms are so intricately intertwined that it is irrational to say that one is the most fundamental. The three paradigms, or cultural styles, are listed below.

1.4.1 Theory

The first paradigm, Theory, is deeply rooted in mathematics and is concerned with the ability to describe and prove relationships among object. The paradigm consist of the following four steps.

1. Characterize objects of study (definition);

²<https://www.olympus-europa.com/medical/en/Products-and-Solutions/Products/Product/ENDOCAPSULE-10-System.html>

2. Hypothesize possible relations among them (theorem);
3. Determine whether the relationships are true (proof);
4. Interpret the results found.

When working in this paradigm it is expected to follow this reasoning and iterate the steps when errors or inconsistencies are discovered, until they can be explained and interpreted.

1.4.2 Abstraction

The second paradigm is abstraction. Abstraction is a form of modeling and is rooted in the experimental scientific method. This paradigm is concerned with the ability to use the relationships found in the theory paradigm to make predictions that then can be compared to the real world. It has four steps for which a scientist is expected to follow.

- Form a hypothesis;
- Construct a model and make a prediction;
- Design an experiment and collect data;
- Analyze the results.

1.4.3 Design

The third paradigm, design, is rooted in engineering and consist, like the others, of four main steps listed below. The design paradigm is concerned with the ability to implement specific instances of those relationships and use them to perform useful actions. The following four steps will help an engineer to construct a device or system to complete a given task.

- State requirements;
- State specifications;
- Design and implement the system;
- Test the system.

An engineer is to iterate these steps until the results from the testing satisfy the requirements and specifications of the system.

1.5 Main Contributions

Write a section on how our work contributes to the field... TBW (to be written)

1.6 Thesis Outline

This thesis is split into five chapters. Chapter one and two are mostly to introduce the reader to the topic and to fill in the necessary knowledge to understand the rest of the thesis. In the last chapter we conclude on our findings and discuss our findings and propose further work. The papers that have been referenced in the thesis is added in the bibliography at the very end. The chapters in the thesis is summarized below:

- In Chapter ?? we discuss the literature that focus on the topic of automated lesion detection in computer systems.
- In Chapter ?? we present the details of design, implementation of system and the processing and collection of data.
- In Chapter ?? we present the experiments we have conducted and..
- In Chapter ?? we provide a comprehensive overview of the results found and discuss what that contributes to the field and propose some further work.