Politécnico de Leiria
Escola Superior de Tecnologia e Gestão
Departamento de Engenharia Eletrotécnica
Mestrado em Eng.ª Eletrotécnica

LOSSY COMPRESSION OF BIOMEDICAL IMAGES FOR COMPUTER VISION ANALYSIS

EDGAR DA SILVA PAULO

Leiria, March 2024

Politécnico de Leiria
Escola Superior de Tecnologia e Gestão
Departamento de Engenharia Eletrotécnica
Mestrado em Eng.ª Eletrotécnica

LOSSY COMPRESSION OF BIOMEDICAL IMAGES FOR COMPUTER VISION ANALYSIS

EDGAR DA SILVA PAULO
Number: 2212421

Dissertation performed under the supervision of Professor Sérgio Manuel Maciel de Faria, Professor Luís Miguel de Oliveira Pegado de Noronha e Távora, and Professor Lucas Arrabal Thomaz

Leiria, March 2024

ACKNOWLEDGMENTS

First of all, I would like to express my sincere gratitude to my supervisors Dr. Lucas Arrabal Thomaz, Dr. Luís Miguel de Oliveira Pegado de Noronha e Távora, and Dr. Sérgio Manuel Maciel de Faria. Their guidance, dedication, persistence, patience, and support throughout this research were priceless in my academic and personal growth.

This work was supported by the Fundação para a Ciência e a Tecnologia (FCT), Portugal under project 2022.09914.PTDC (DOI:10.54499/2022.09914.PTDC), by Project CIBME, Programa Operacional Regional do Centro, and by FCT/MCTES through national funds and when applicable co-funded by EU funds under the project UIDB/EEA/50008/2020 and LA/P/0109/2020.

Many thanks to the Instituto de Telecomunicações that hosted the CIBME project, providing resources, and thus contributing to the successful execution of this dissertation. Also, to my colleagues at the Leiria Delegation of the Instituto de Telecomunicações, for sharing knowledge and maintaining good spirits, I would like to thank you.

Thanks also to the Polytechnic of Leiria, namely the Higher School of Technology and Management for the academic training opportunity. Furthermore, for my colleagues, it was a pleasure to walk this path with you.

A special thanks to my sons and my wife for their unconditional support, understanding, and dedication. And to my family for your patience and support. After all, just like in the poem: *...o sonho é uma constante da vida tão concreta e definida como outra coisa qualquer.*

RESUMO

O aumento exponencial da aquisição de dados médicos e biomédicos é impulsionado pelos avanços tecnológicos, nomeadamente na área da imagiologia. No entanto, este crescimento exponencial traz consigo desafios em termos de capacidade de processamento, transmissão e armazenamento de dados. Em resposta a esta procura crescente, têm surgido soluções cada vez mais eficientes, especialmente através da visão computacional para análise automática de imagens e algoritmos de compressão.

Esta dissertação tem como objetivo, por um lado, avaliar o desempenho de sistemas de visão computational em imagens biomedicas previamente comprimidas. Por outro lado, aumentar a gama útil de compressão de imagens, quase sem perdas e com perdas, reduzindo o impacto da distorção adicionada por este método no desempenho de algoritmos de visão computacional na análise de imagens biomédicas.

Neste sentido, utilizam-se os algoritmos YOLO e Detectron2 para avaliar o impacto da distorção da codificação na sua capacidade de detectar mitocôndrias em imagens de microscopia eletrónica. Os resultados revelam que, embora a compressão tenha influência no desempenho da deteção de mitocôndrias, a sua influência é insignificante em taxas de compressão mais baixas.

Adicionalmente, são apresentadas duas propostas para aumento do rácio de compressão util, mantendo as características da imagem que permitem efetuar a deteção automática de mitocôndrias. Por um lado, demostra-se que a metodologia de treino proposta, utilizando versões comprimidas dos dados no treino, permite reduzir o impacto da distorção no desempenho da deteção dos algoritmos de visão computacional; por outro lado, a atribuição de níveis de qualidade mais elevados às regiões de interesse em comparação com o *background* permite manter um desempenho elevado em taxas de compressão onde os algoritmos de visão computacional começam a perder eficácia. Estas abordagens permitem expandir o uso de compressão de imagens com pouco impacto no desempenho de deteção, contribuindo assim para uma maior eficiência no processamento, armazenamento e transmissão de imagens em aplicações biomédicas.

Palavras-chave: Imagens Biomedicas, Imagens de Microscopia Eletrónica, Compressão com e quase sem perdas, YOLO, Detectron2, HEVC, Codificação por regiões, Compressão de imagens médicas

ABSTRACT

The exponential increase in medical and biomedical data acquisition is compelled by technological advances, namely in the imaging field. However, this exponential growth brings with it challenges in terms of processing capacity, transmission, and data storage. In response to this growing demand, increasingly efficient solutions have emerged, especially through computer vision for automatic image analysis and compression algorithms.

This dissertation aims, on the one hand, to evaluate the performance of computer vision systems on previously compressed biomedical images. On the other hand, it increases the useful range of image variations, almost lossless and lossy, decreasing the impact of the change added by this method on the performance of computer vision algorithms in biomedical image analysis.

In this sense, YOLO and Detectron2 are employed to evaluate the impact of coding distortion on their ability to detect mitochondria in electron microscopy images. The results of this study reveal that although the distortion introduced by compression affects their detection performance, it is negligible at lower compression ratios.

Furthermore, two proposals are presented to improve the useful compression ratio, keeping the images characteristics that allow to perform the automatic detection of mitochondria. On the one hand, it is demonstrated that the proposed training methodology, which incorporates compressed versions of the original data during training, mitigates the impact of distortion on the performance of computer vision algorithms; on the other hand, allocating higher quality levels to regions of interest, compared to background elements, helps to sustain high performance at compression rates where computer vision algorithms typically start to lose effectiveness. These approaches allow the extension of the compression range with little impact on detection performance, thus contributing to the improvement of data processing, storage, and transmission in biomedical applications.

Keywords: Biomedical Images, Electron Microscopy Images, Lossy and near Lossless Compression, YOLO, Detectron2, HEVC, Region Coding, Medical Image Compression

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CABAC	Context-adaptive binary arithmetic coding.
CR	Compression Ratio.
CTU	Coding Tree Unit.
CU	Coding Unit.
CUDA	Compute Unified Device Architecture.
DCT	Discrete cosine transform.
DST	Discrete sine transform.
EHRs	Electronic health records.
EM	Electron Microscopy.
FN	False negative.
FP	False positive.
GPU	Graphics processing unit.
HEVC	High Efficiency Video Coding.
JPEG	Joint Photographic Experts Group.
QP	Quantization Parameter.
SEM	Scanning electron microscope.
STM	Scanning tunneling microscope.

TEM	Transmission electron microscopy.
TP	True positive.
VVC	Versatile Video Coding.

1

INTRODUCTION

1.1 CONTEXT AND MOTIVATION

The use of images is widespread in biomedicine, playing an important role in the development of new treatments, improving patient outcomes, and leading to advances in medicine. However, the increase in image data size and number of acquisitions presents itself as a challenge, pressing human resources and infrastructure, namely in terms of processing, storing, and transmitting this type of data [1, 2].

Although in biomedicine the interpretation of the generated data is carried out essentially by specialists, the increasing amount of information in modern imaging and its heterogeneity [3] presents a challenge in identifying relevant regions of the images. The human factor is, thus, a key element in the diagnosis [4], with experience being a non-deterministic but important factor, given the diversity of information to be analyzed (image modalities, patients, and pathologies, among others).

The introduction of computer vision algorithms, despite not aiming to substitute the specialized analysis, has evolved as an auxiliary tool that allows to improve the identification of the object of study. These algorithms resulted in increased performance, with an impact that currently covers different areas, from quality inspection to safety [5].

It can, therefore, be understood that a shift is taking place, where perceptual quality is replaced by the overall performance of machine vision tasks. Since the usefulness of images in decision-making is related to the data analysis capacity [6], identifying areas of interest is a crucial process, as it allows for efficient data analysis and reduces the amount of unnecessary information that needs to be processed.

The large amounts of generated data impose increasing stress on the healthcare infrastructure [7], namely by the huge volumes of storage and transmission bandwidth requirements. On one hand, the capture of relevant high-resolution medical information from different imaging methods is necessary for accurate diagnosis and treatment, on the other hand, compliance with legal obligations, and standards, such as the Digital Imaging and Communications in Medicine (DICOM [8]) standard,

requires the preservation of data for long periods of time. As a result, there is a high demand for storage capacity in the biomedical field.

The use of data compression, namely lossless coding, is a solution that preserves the original information reducing the amount of resources needed to represent the biomedical information. Nevertheless, such coding schemes present limited compression efficiencies when compared to lossy algorithms. In this sense, exploring compression models that permit increased efficiency, without compromising relevant information is therefore essential.

Thus, since higher compression ratios (CR) may lead to higher coding distortion and, consequently, lower performance of computer vision tasks, some relevant research questions arise: i) what are the upper compression boundaries that still preserve the relevant information for detection algorithms?; ii) how can such boundaries be extended in order to achieve the same detection performance with higher compression ratios?; and iii) can the information provided by the computer vision algorithms be used to improve data compression, while keeping the algorithms' performance?

1.2 RESEARCH OUTPUT

The main objective of this research was to investigate and enhance the valuable role of biomedical image lossy compression within computer vision systems. Specifically, the study sought to expand the compression capabilities of electron microscopy (EM) images, demonstrating that a broader compression range could be achieved with minimal adverse effects on the computer vision algorithm's detection capability.

In this sense, the performance of two different computer vision algorithms was assessed in uncompressed and previously compressed images. Based on the results achieved, two strategies are presented to minimize the impacts of compression in this type of biomedical image. As a result of the research work done during this dissertation, the following conference paper was published:

- E. S. Paulo, L. A. Thomaz, L. M. N. Távora, P. A. A. Assuncao and S. M. M. Faria, "Extending the compression range of biomedical images for machine vision analysis," 2022 30th European Signal Processing Conference (EUSIPCO), Belgrade, Serbia, 2022, pp. 1273-1277, doi: 10.23919/EUSIPCO55093.2022.9909663.

1.3 DISSERTATION STRUCTURE

This document is organized as follows: Chapter 2 introduces the diversity of biomedical images, namely the challenges related to their preservation and processing requirements. It also presents some of the most relevant image coding standards and computer vision algorithms used in the scope of this work. Chapter 3 outlines the guidelines followed in this study, aimed at comprehending and enhancing detection performance on previously compressed images. Detailed descriptions of the proposed methods are presented therein. Chapter 4 offers insights into the experimental assessment results, specifically focusing on the effects of compression on the detection performance of YOLO and Detectron2, and the advancements achieved through the proposed methods. Concluding the document, Chapter 5 resumes the findings and provides suggestions for enhancing future endeavors.

2

BACKGROUND

This chapter unfolds into three sections with dual objectives: to describe some types of biomedical images and highlight the significance of ensuring the preservation of vital information; and to introduce the presence of algorithms that facilitate optimization in storage and transmission, as well as the automatic image processing systems for classification and segmentation. The first section resumes the provenance of biomedical images, with highlights for the electron microscopy images and the datasets used in this work. Then, the most relevant image coding standards as well as the use of coding by regions are introduced. Finally, the computer vision algorithms used throughout this work are presented.

2.1 MEDICAL/BIOMEDICAL IMAGING

Imaging within biomedicine encompasses a wide group of techniques for acquiring and processing images. It is pivotal for studying specific elements of interest at both research and diagnostic levels. The diversity of imaging techniques is related to the plurality of elements of interest, ranging from anatomical structures to cellular regions.

Although the term 'anatomy' originates from the Greek 'anatomé' meaning 'incision or cut' and the Latin 'anatomía' referring to the 'dissection of the body' [9], contemporary exploration of the internal morphology of living organisms predominantly employs non-invasive methods, namely through imaging techniques. Of the different branches of medical imaging, the acquisition of images of anatomical structures is carried out essentially using radiology, ultrasound, computerized axial tomography (CT), or magnetic resonance imaging.

The analysis of cellular regions is directly linked to the invention and development of microscopy. Currently, electron microscopy techniques, by transmission, scanning or tunneling, and confocal microscopy techniques (fluorescence) are state-of-the-art methods [10]. This technology allows, on the one hand, the study of detail, and on the other, the accounting of objects that make up our organism.

The remarkable rise in biomedical image data can be attributed to the rise of advanced medical imaging techniques and the widespread adoption of electronic health records (EHRs), including the digitalization of healthcare procedures [11]. These factors have led to an accumulation of extensive data, posing new challenges to the existing infrastructure, namely in terms of processing, interpretation, storage, and transmission capabilities.

The interaction between compression and automatic processing is an advantage that allows efficient handling of biomedical images, enabling more research, less treatment planning, and faster diagnosis. On one hand, data compression reduces storage demands and allows fast communications [12, 13]. On the other hand, automatic processing systems can free human resources from repetitive and time-consuming tasks.

2.1.1 *Electron Microscopy Images*

Electron microscopy is an imaging technique that takes advantage of the shorter wavelength of electrons to inquire about the structure of materials at the nanoscale. The different amplitude signals produced by the electron beams that interact with the sample under study can be combined to create full digital images [14].

Transmission Electron Microscopy (TEM), Scanning Electron Microscopy (SEM), and Scanning Tunneling Microscopy (STM or STEM) stand out as the most prevalent techniques in electron microscopy. The distinction among various techniques arises essentially from the emission and collection of the electrons, as presented in Figure 1 and 2: in TEM the electron beams pass directly through the sample before being collected, offering information about structure and morphology; SEM creates an image by detecting reflected or knocked-off electrons from the surface of the sample [14]; STM picks up electrons that escape from the surface of a sample by the process known as the quantum tunneling effect [15], providing information about the surface and its composition.

In biomedicine, electron microscopy provides information about the elements inside cellular regions, enabling the construction of topographical maps that detail structures and molecular arrangements. The raw images generated by electron microscopy typically contain a large amount of data, often in the form of high-resolution images (e.g. in [18] the images had a size between 1376×1032 and 2048×2048 pixels with pixel sizes vary from 0.26 to 5.57 nm). This means that not only a large amount of storage space is required, but also to extract meaningful

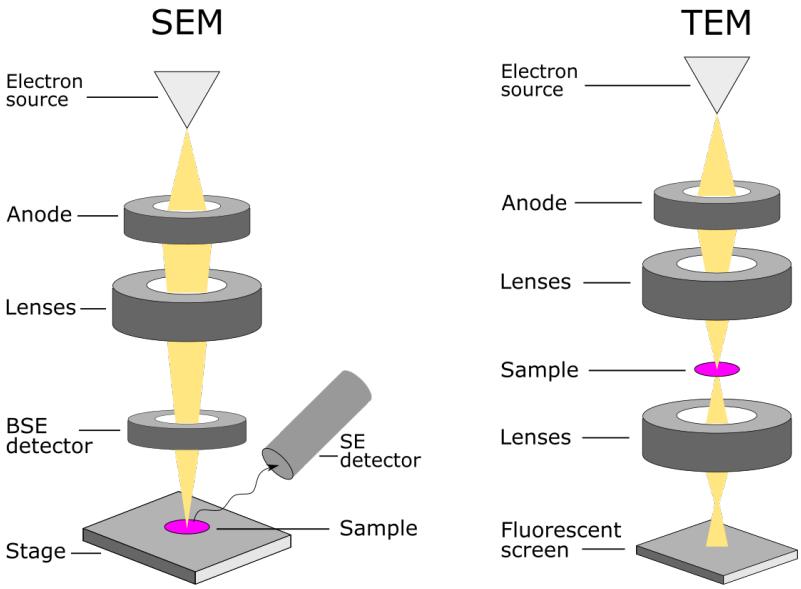


Figure 1: Schematic diagram of SEM and TEM [16].

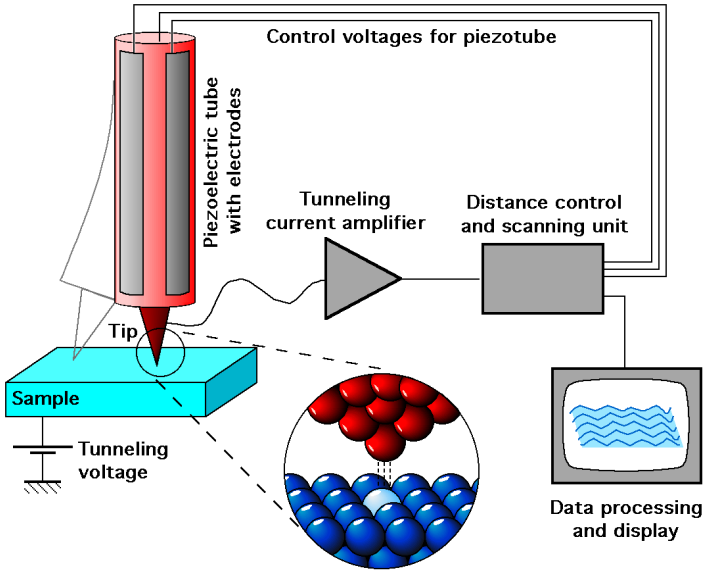


Figure 2: Schematic diagram of STM [17].

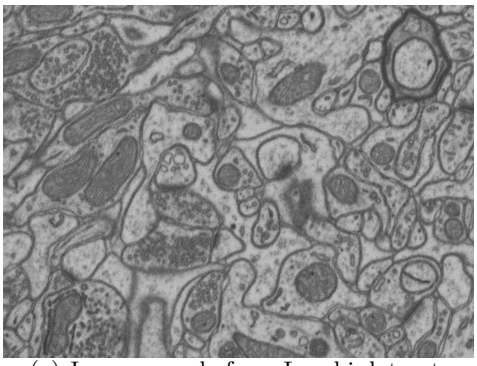
information various steps must be done, including filtering, segmentation, and reconstruction.

The availability of datasets in public archives or scientific articles plays a crucial role in the ongoing advancement of various fields. The development of public archives like the Electron Microscopy Public Image Archive (EMPIAR) [19, 20] and the BioImage Archive [21], among others, are nowadays important resources that provide access to state-of-the-art data.

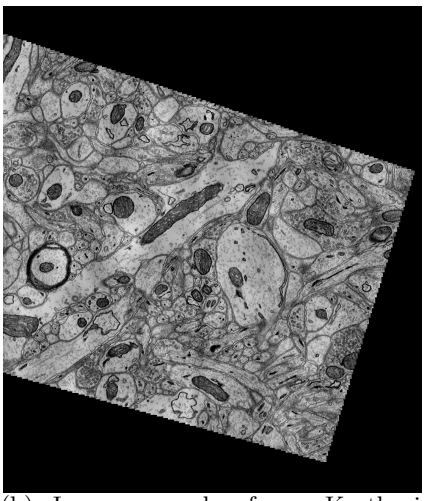
2.1.2 *Datasets*

In the scope of this work two publicly available datasets [22]: Lucchi++ and Kasthuri++ were used. These datasets are re-annotated versions of data originally made available by Lucchi *et al.* [23] and Kasthuri *et al.* [24]. The re-annotations of the masks that accompany each image led to significant changes, and their new versions have ≈20% less mitochondria annotations than the original ones. Also, the overall area of annotations was reported to have increased around 20% and 2% in Lucchi and Kasthuri, respectively.

The Lucchi++ dataset, acquired using a focused ion beam SEM from a 5×5×5µm section of the hippocampus of a mouse brain, is provided in two sets, one for training and one for testing. Each contains 165 images with 1024×768 pixels, in 8-bit grayscale. The Kasthuri++ dataset, acquired using serial SEM from a 3×3×30nm mouse cortex, is also made available in two sets, for training and testing: one with 85 images, 1463×1613 pixels in size, and another containing 75 images (1334×1553 pixels), all in 8-bit grayscale. Figure 3 is presented as a sample of the images provided in the used datasets.



(a) Image sample from Lucchi dataset.



(b) Image sample from Kasthuri dataset.

Figure 3: Illustrative images from the used datasets.

2.2 IMAGE COMPRESSION

Compression encompasses a collection of tools designed to remove redundancies and encode data into more compact forms. This allows the optimization of infrastructures and plays a vital role in storage and transmission efficiency.

Advances in biomedical imaging technology are leading to a significant increase in the amount of generated data. It is estimated that nowadays, the healthcare industry accounts for more than 30% of the data globally produced [25]. Compliance with legal terms very often requires data to be stored for long periods of time, eventually leading to a scarcity of computational and storage resources [7]. In this context, biomedical image data compression is necessary to cope with such increasing demand.

The distinctive nature of biomedical data, which may encompass critical diagnostic information, remains safeguarded under conservative legal norms [26], including adherence to the DICOM standard [8]. These regulations play a vital role in maintaining trust, ethical practices, and interoperability within biomedicine.

The adoption of conventional image coding in lossless mode to ensure the preservation of the data is commonly used. Nevertheless, since lossless compression ratios are far from those obtained from lossy coding schemes, more advanced approaches are required to preserve the relevant biomedical information after lossy coding. The choice, however, depends on the acceptable trade-off between file size and compressed data fidelity.

Currently, DICOM supports various compression standards for medical images, namely by defining guidelines for implementing compression algorithms within DICOM-compliant systems. These compression standards include lossless JPEG, JPEG 2000, and JPEG-LS for image compression; and MPEG2, MPEG-4, and HEVC/H.265 for video compression.

2.2.1 *JPEG2000*

The JPEG2000 [27] is one of the standards widely used in biomedical image coding, allowing lossy or lossless coding. This image encoder, whose International Standard ISO/IEC 15444-1 [28], was developed by *Joint Photographic Experts Group*[1], and

1 Collaboration between *International Standardization Organization* (ISO) and o *International Electrotechnical Commission* (IEC)

can be used free of charge for non-commercial purposes true the reference software OpenJPEG [29].

Based on the wavelet transform [30], it allows high flexibility in accessing information, allowing decoding with resolutions of quality equal to or lower than the initially encoded quality, including lossless. Figure 4 briefly presents the pipeline of a JPEG 2000 encoder.

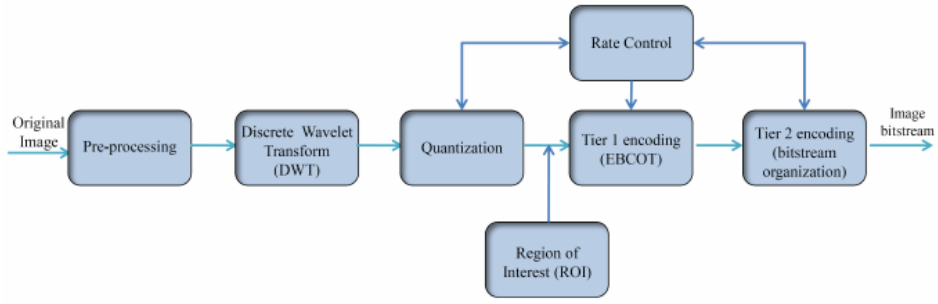


Figure 4: JPEG 2000 simplified pipeline [31].

During the initial phase, pre-processing techniques are used, such as, for example, changing the color space, and subdividing the image into smaller rectangular regions, tiles, or slices. This operation increases processing speed since several tiles can be processed in parallel by the pipeline. In the next phase, the discrete Wavelet transform is applied, horizontally and vertically, to each tile, obtaining the coefficients corresponding to the Wavelet sub-bands. This results into four blocks, as shown in Figure 5, corresponding to the high (H) and low (L) horizontal and vertical frequencies.

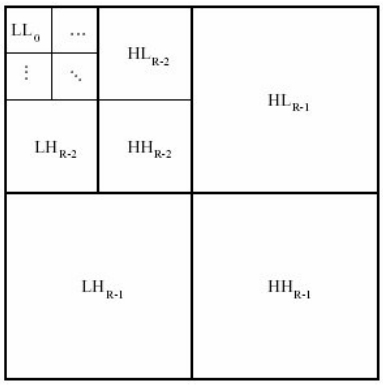


Figure 5: Sub-bands after Wavelet transform [32].

LL blocks can also be decomposed by applying the transform again. Subsequently, the obtained coefficients are quantized, based on the configured compression ratio, between 1 (lossless) and 100. This process will limit the maximum quality of the decoded image, that is, the ability to recover the original information. The higher

the Quantization Parameter (QP), the higher the compression rate, but, the greater the data losses.

Then, the result is encoded using an arithmetic encoder. This process aims to optimize the generated code, not introducing any losses. Finally, relevant information about different parameters, including the original size and compression quality, in the form of headers and trailers is added to the final file. The decoder uses this information and runs in the opposite direction of the pipeline.

Among the innovations introduced, JPEG 2000 [33] includes mechanisms that allow encoding different regions with different qualities, namely through maxshift or scaling techniques. In both methods, a certain level of relevance is applied to the wavelet coefficients of the region of interest, which allows, for example, their encoding to occur before background (maxshift), or as per user control (scaling) [34].

2.2.2 *High Efficiency Video Coding*

The High Efficiency Video Coding (H.265/HEVC) [35]) was approved as a coding standard by the *Telecommunication Standardization Sector* (ITU-T) in 2013, following the work developed jointly by the *ITU-T Video Coding Experts Group* and the *ISO/IEC Moving Picture Experts Group*. Introduced as a solution to enhance the coding efficiency for ultra-high-resolution videos [36, 37], it allowed improvements close to 50% reduction in bit-rate for the same image quality [38] in relation to the previous H.264/AVC standard.

Although it was developed for video compression, exploring Inter and Intra coding, it is also possible to apply it to single images, for lossy or lossless coding, as used in the scope of this dissertation. The algorithm is protected by patents, however it can be used for non-commercial image encoding and decoding purposes, together with the documentation and configuration files, which can be accessed via the HM software [39].

The algorithm starts by partitioning the picture into a block structure known as a Coding Tree Unit (CTU), composed of substructures associated with the color space, namely Luma (associated with luminosity) and Chroma (associated with color components). From this structure, optimized prediction solutions are calculated to exploit redundancies effectively. In the Intra mode, the HEVC internal loop progressively subdivides the image, scrutinizing different intra-prediction modes to determine the optimal correlation between the pixels of a particular Coding Unit

(CU) block and its surrounding area [40], as shown in Figure 6. Subsequently, it applies the DCT transform and the coefficient´s quantization, obtaining the number of bits required per CTU. When QP is modified, namely by the use of a delta QP optimization, a flag is triggered that ensures the recalculated QP is made available to the decoder. Coding only occurs after identifying the best configuration for each CTU (block size, QP, and prediction mode).

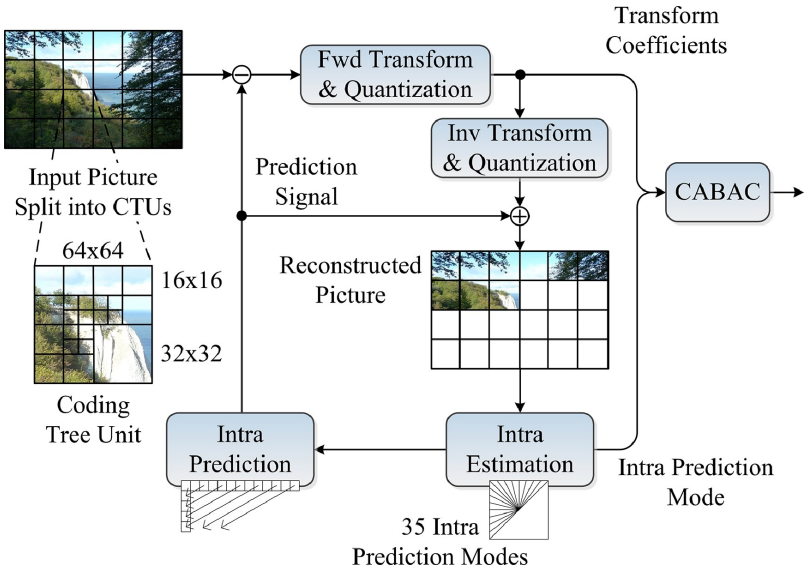


Figure 6: HEVC Intra model pipeline [40].

Depending on the used configuration, blocks ranging from 64×64 to 8×8 pixels, 35 directional intra-prediction modes, and the context-adaptive binary arithmetic coding (CABAC) [41] are added values in lossless coding, achieving highly optimized efficiency. In the case of lossy coding, the intra-prediction is followed by the use of discrete sine (DST) and cosine transforms (DCT) and a quantization process. Similar to other encoders, losses are introduced by the coefficient quantization process, although comparatively, HEVC manages to obtain a reduction of 16% (compared to JPEG 2000 4:4:4) and 43% (compared to JPEG) [42] for the same visual quality. In fact, as demonstrated in [43], when used for lossless compression of medical images, the HEVC encoder is able to achieve compression gains up to 54% higher than its counterparts.

2.2.3 *Versatile Video Coding*

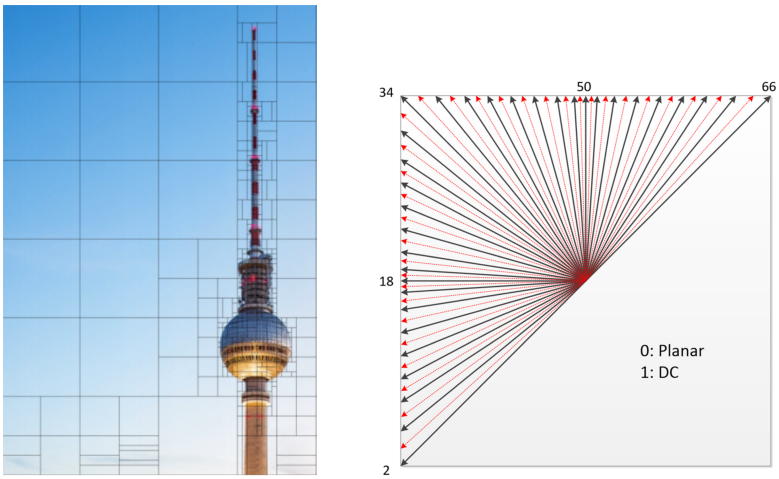
Versatile Video Coding (VVC) [44] is one of the latest image and video coding standards, with the designation H.266 in ITU-T and 23090-3 in ISO/IEC [45].

Developed by *JVET - Joint Video Exploration Team*, this standard was presented as the successor to HEVC.

Like its predecessor, it was designed for video coding, however, it is configurable so that it can be used to encode single images. The reference software *VVC test Model* (VTM) [46] is made available free of charge for non-commercial image encoding and decoding purposes, together with documentation and some configuration files.

The VVC block diagram is identical to that of HEVC, however its modules are more complex. As an example, the greater flexibility in the size and shape of the prediction units (PU) stands out, making it possible to have blocks up to 128x128, and rectangular blocks. Figure 7, on the left, allows us to visualize this innovation with regions that present some uniformity, such as the sky, being the target of prediction in larger regions and regions of the antenna being divided into multiple blocks. This feature increases interpixel redundancy, improving coding.

At the coding level *Intra*[2], there is an improvement in the increase to 67 directional prediction modes, as e.g. in Figure 7 on the right. Since there are rectangular and larger blocks, there are more prediction angles to cover more reference options.



(a) Flexibility in the shape and size of prediction blocks.

(b) 67 intra prediction modes.

Figure 7: Example of *intra perdition* improvements introduced by VVC [47, 48].

Additionally, VVC incorporates a model that allows relating the prediction of *Chroma* elements with those of *Luma*, in what is called a *Cross-component linear model* (CCLM), reducing the redundancy between different color components.

2 The *Inter* coding was not addressed in this report.

VVC also introduces developments in terms of transforms, with new variants of DCT and DST. The algorithm can even identify which transformation best performs to maximize compression efficiency.

The lossless compression profile involves deactivating several tools, including the transforms, as mentioned in the configuration document. In this case, the QP must be zero. Coding ends after the entropic coding, which is CABAC like in HEVC. This guarantees the reduction of image representation without introducing losses.

2.2.4 *ROI Coding*

The integration of computer vision algorithms with coding systems, such as HEVC, has been a subject of study over the last decade, to significantly improve coding performance [49, 50, 51]. This integration offers significant advantages in terms of compression efficiency, enabling the allocation of more bits to regions of interest that, consequently, leads to superior quality of the ROI.

This synergy has given rise to two distinct approaches: a parallel coding system, where the coding of ROIs and background generates two separate bitstreams that are later merged; and an end-to-end system, where ROI identification takes place simultaneously with coding, internally modifying the resulting bitstream, as depicted in Figure 8. However, the integration of such alterations fundamentally reshapes the encoders pipeline [52, 53, 54, 55], making the encoders incompatible with the standards.

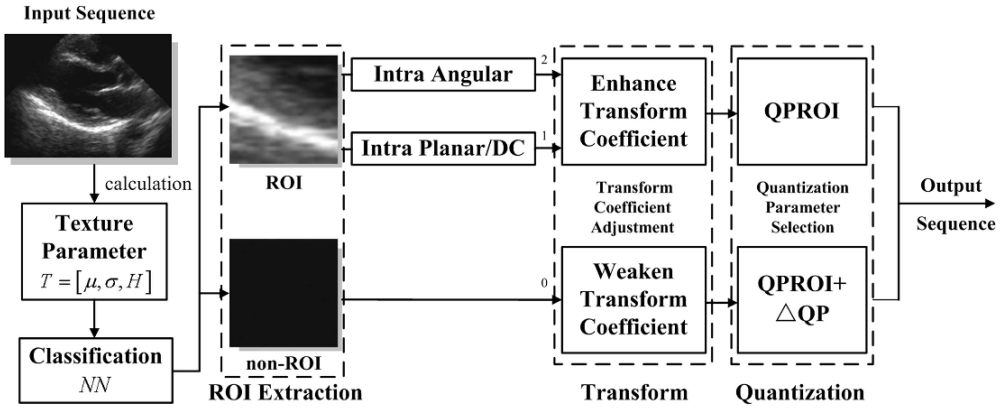


Figure 8: Example of end-to-end ROI compression framework based on HEVC[50].

In this sense, the possibility of using previously identified masks, similar to what can be used in JPEG 2000 [56, 34], serves as a motivating factor for devising a solution grounded in HEVC and/or VVC. Additionally, encoding data of reduced

relevance for a diagnosis, namely the background, increases the difficulty of resource optimization. Therefore, the study of the use of ROI in the compression of medical images is highly relevant [57, 58, 59, 60, 61].

2.3 AUTOMATIC COMPUTER VISION TOOLS

The integration of automated computer vision systems has become present in biomedicine, liberating human resources from repetitive and time-consuming tasks. Tasks like cell counting, as shown in Figure 9, and organelle identification in microscopy images can now be efficiently executed by computer vision algorithms [62]. These systems offer the advantage of being invariant to time and immune to intra- or inter-operator variability [63, 64, 65, 66, 67], driving progress in various medical image analysis applications.

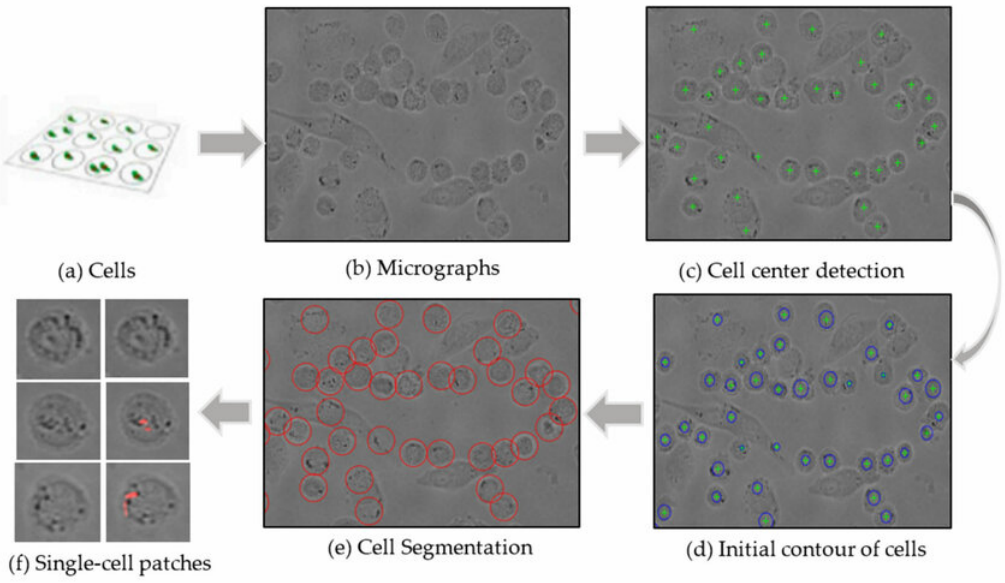


Figure 9: Example of automatic detection process for cell images [68].

The high performance of computer vision algorithms shows its effectiveness as an ROI identification tool, yielding consistent outcomes within significant applications in biomedicine [69, 70]. For instance, in the context of Diabetic Retinopathy detection, the use of computer vision demonstrates remarkable levels of sensitivity and precision (exceeding 95%) [71], aligning with prior research that surpasses even specialist evaluations by over 10% in certain instances [72].

Within convolutional networks designed for object classification and localization via bounding boxes, noteworthy architectures include R-CNN [73], SPP-NET [74],

FAST R-CNN [75], R-FCN, RetinaNet [76], and YOLO [77]. Although initially, the location, identification, and classification processes occurred at different times, leading to excessive time consumption, the emergence of YOLO [78] proved to be a turning point by ensuring that the identification and classification of each object occurred simultaneously.

More recently, Mask R-CNN [79] has made the detection, classification, and segmentation within a single algorithm easier. This innovation served as the foundation for frameworks such as Detectron2 [80] and the Segment Anything Model (SAM) [81], which are examples of state-of-the-art object detection algorithms that were developed by Facebook AI Research (now Fundamental AI Research or FAIR). These frameworks, aside from enabling tasks such as semantic segmentation and object classification or identification, extend their capabilities to incorporate instance segmentation, i.e. a process that involves delineating specific objects within a scene.

Nowadays, automatic segmentation and classification can be found in several areas of biomedicine. For instance, in dermoscopic images, automated systems have achieved remarkable results such as skin lesion segmentation, reporting accuracy rates exceeding 93% and sensitivity of 90% [82]. In ultrasound images, these systems have demonstrated over 85% accurate thyroid gland detection and over 95% precise gland segmentation [83]. In low-dosage CT scans, real-time detection and segmentation of lung nodules have reached 0.89 sensitivity and 93% precision [84]. Other applications include the detection of cholelithiasis and gallstones in CT scans [85].

2.3.1 *YOLO*

One of the most successful deep learning frameworks for image analysis is YOLO, for example its version 4 [78]. This type of learning network is used for object detection and classification of images of various classes. It has also the advantage of being fast, presenting a good performance for new classes (with moderately few samples) after a few epochs of training using transfer learning for large datasets. In medical imaging, specifically for the task of organelles (namely mitochondria) detection and segmentation, the work presented in [86] employs the YOLO v4 with great success.

In YOLO, pinpointing object locations and classifying starts by partitioning the image into a grid of cells, as depicted in Figure 10. Each cell is responsible for detecting objects within its boundaries, thus a set number of bounding boxes are

forecasted, each accompanied by a confidence score indicating the likelihood of accurately localizing an object.

Simultaneously, each bounding box is assigned a probability value that the identified object belongs to one of the pre-trained classes. The final result corresponds only to bounding boxes where the set of location and classification probability is greater than the limit established in the algorithm configuration. Its application in areas such as the segmentation of skin lesions and dermoscopy images [82] showed results greater than 93% accuracy, with 90% sensitivity.

The delimitation of objects by a bounding box, despite clearly identifying its location, presents a total area greater than the object itself due to the difference in geometric shape between the two. Therefore, the delimited region includes a portion of the background, as opposed to semantic segmentation which is pixel-wise.

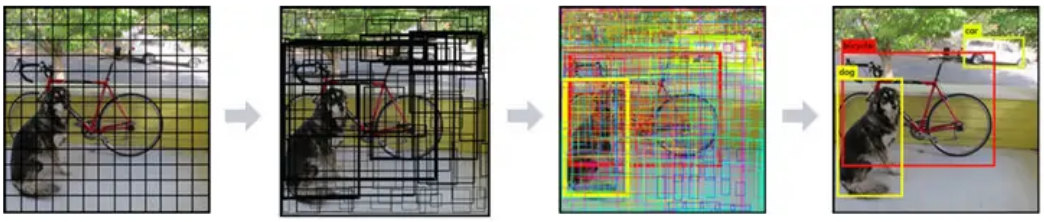


Figure 10: Example of classification by YOLOv4 [87].

2.3.2 *Detectron2*

Detectron2 is a state-of-the-art computer vision model that has been developed by the Facebook AI Research group for object detection and segmentation. Built on top of the Pytorch framework and implementing CUDA-enabled GPU support [88].

Its modular and highly configurable architecture has the advantage of being faster, and highly accurate, presenting support for different object detection techniques. The implementation of several backbone and model options, including Faster R-CNN, Mask R-CNN, RetinaNet, Cascade R-CNN, and Panoptic FPN allows for a choice between detection by bounding box, keypoint detection, and object segmentation: including semantic, instance, and panoptic segmentation, like the ones presented in Figure 11.

The application of Detectron2 in biomedicine has produced impressive results, including achieving accuracy rates of 100% in classifying hemorrhagic stroke and 97% in targeting lung regions [90].

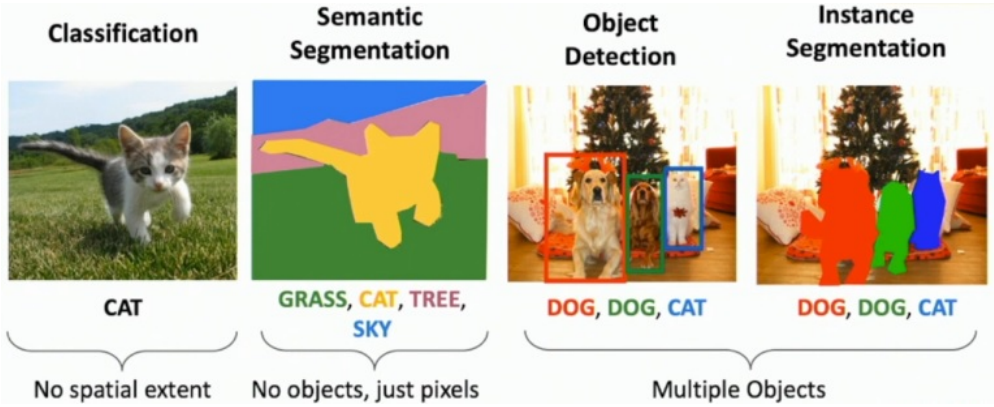


Figure 11: Different Computer Vision detection types [89] available on Detectron2.

2.4 SUMMARY

This chapter presented an introduction to the diversity of biomedical imaging, highlighting the challenges posed by the widespread adoption of electronic health records over the existing infrastructure. In laying the foundations for the work herein described, three distinct sections are introduced: the first explores electron microscopy images and presents the datasets, the second presents some of the most relevant standards for image compression, and the third outlines the automated image processing methods employed throughout this research.

3

EXTENDING THE COMPRESSION RANGE OF BIOMEDICAL IMAGES FOR MACHINE VISION ANALYSIS

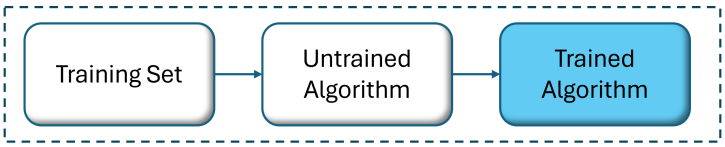
In this chapter, the guidelines followed to enhance the performance of the computer vision algorithms, YOLO and Detectron2, on pre-encoded Electron Microscopy images are described. First, the proposed methods are explained in detail, namely the strategies employed to identify and mitigate the effects of lossy compression on the performance of YOLO and Detectron2. Then, the dataset preparation for both training and performance assessment is described, together with a description of the computer vision algorithms configuration and the compression algorithm used Finally, the metrics used to assess and evaluate the algorithmic performance are explored.

3.1 PROPOSED METHODS

The investigation of how lossy coding affects the detection performance of computer vision algorithms led to the development of novel strategies. These strategies were designed to, not only expand the practical scope of the compression range but also to minimize distortions introduced during the compression process over ROIs.

To attain these objectives, it was necessary to initially assess the algorithms performance in identifying mitochondria within the datasets images and compare them against human classification. As outlined in the pipeline depicted in Figure 12(a), both computer vision algorithms were individually configured and trained using the pre-prepared training set described in 3.2. Subsequently, the trained YOLO and Detectron2 networks were used to detect the mitochondria in the uncompressed images of the test sets (refer to Figure 12(b)), with the inference result being compared with the ground truth. The inference results obtained from uncompressed images were used as reference for subsequent analysis.

Next, in order to establish the upper limits of lossy coding that allows to maintain the detection performance of the computer vision algorithms, the YOLO and



(a) Training.

(b) Predition.

Figure 12: Training and prediction pipelines.

Detectron2 detection performance was evaluated over previously compressed images. The inference results were then compared with those obtained from uncompressed images, as outlined in the pipeline depicted in Figure 13. In this sense, the images of both datasets were individually compressed with different qualities, as explained in section 3.4.

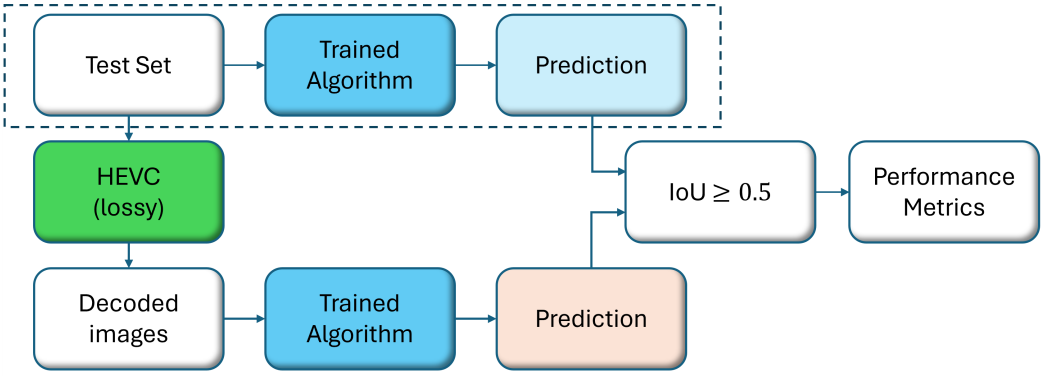


Figure 13: Compression impact assessment on YOLO and Detectron2.

Then, to increase the image compression ratio (CR) with minimum impact on the YOLO and Detectron2 performance, two strategies were followed:

- In the first approach, both detection algorithms were retrained using uncompressed and previously compressed images, intending to enhance their performance at higher CRs;
- In the second approach ROI-based coding is used, with higher quality being assigned to the regions of interest (ROI) that result from the computer vision algorithms inference.

3.1.1 *Enhance the computer vision performance at higher CRs*

In the first case, a transfer learning plus retraining technique that used compressed and uncompressed images here used, as outlined in the pipeline depicted in Figure 14. The process starts by loading the trained algorithm weights from the previous stage involving only non-compressed images; the subsequent training, validation, and testing included both original images and compressed versions. The newly trained models were then employed to detect mitochondria in the compressed images from the test sets, and the results were compared to the reference for evaluation, as depicted in Figure 15.

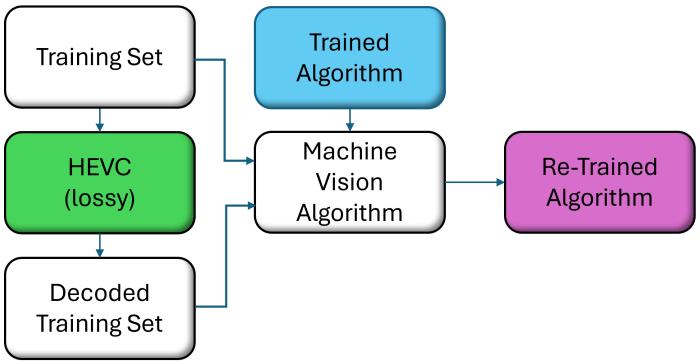


Figure 14: Re-trained pipeline using uncompressed and previously compressed images

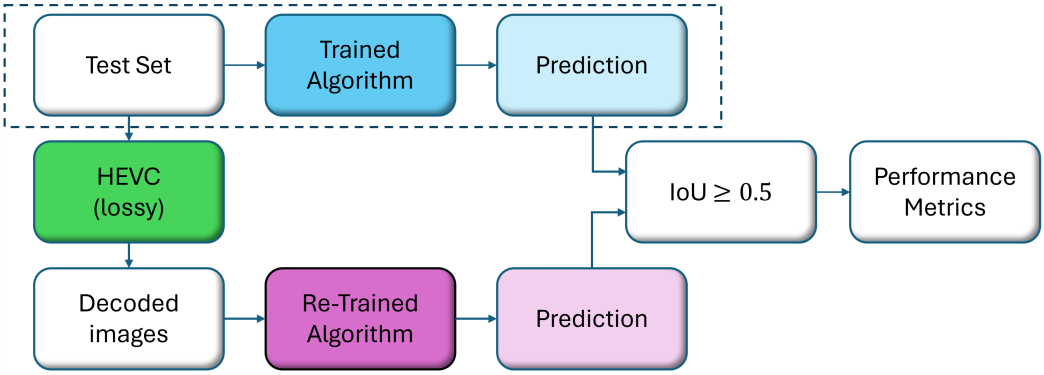


Figure 15: Inference with the re-trained computer vision algorithms.

3.1.2 *ROI-based coding*

In the second case, the information of the ROI, previously identified by the algorithms on non-compressed images, is used on the HEVC encoder to apply different quality configurations between ROI and background. This means that the encoder algorithm

was modified to allow the compression of regions of the same image with different parameters without compromising the standard bitstream format, as explained in section 3.4.

Upon decoding, the performance of YOLO and Detectron2 was re-evaluated using the model trained exclusively with non-compressed images. Once again, the inference results were compared to the reference, as outlined in detail in the pipeline depicted in Figure 16.

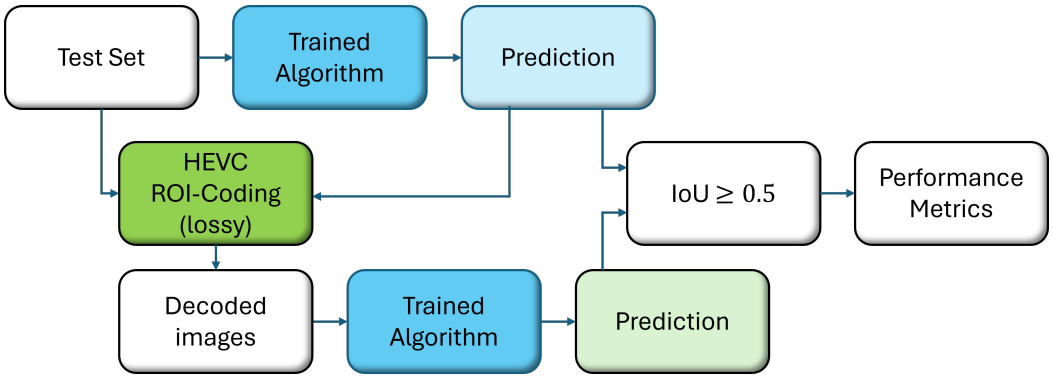


Figure 16: YOLO and Detectron2 performance assessment on images previously coded by regions.

Given the extensive range of configurations and images employed in this study, Python programs were developed to automate both the compression, training, and detection processes. These programs allowed repetitive tasks to be simplified and automated, ensuring consistency in experimental setup across different algorithms and images.

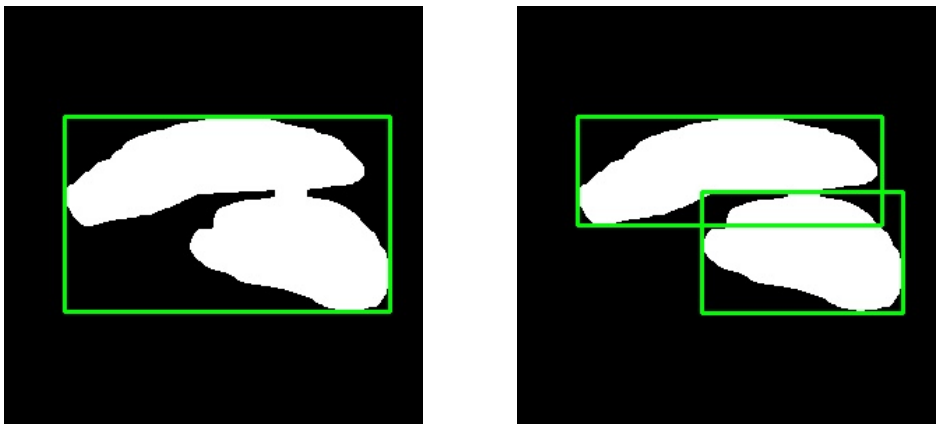
3.2 DATASET PREPARATION

In order to adequately evaluate the undergoing training performance, the Lucchi++ training set underwent additional subdivision, randomly allocating 125 images for training and 40 for validation. Similarly, the Kasthuri++ training set was also randomly divided, with 61 images designated for training and 24 for validation. Table 1 resumes the number of training images in each dataset.

		Training	Validation	Test
Original	Lucchi++	125	40	165
	Kasthuri++	61	24	75

Table 1: Number of images in the two datasets

In order to train and assess the computer vision algorithms, it was necessary to provide the images and the location of each study object (mitochondrion) herein, in the form of a list of bounding boxes. This involved extracting precise coordinates and dimensions from the provided masks using OpenCV2 *connectedComponentsWithStats* function.



(a) Two ROI in one bonding box. (b) Splitting the bounding box

Figure 17: Creating one bounding box for each ROI.

A careful inspection carried out in the scope of this work led to the identification of a few cases where the produced bounding boxes contained more than one mitochondrion, as shown in Figure 17. It was, therefore, necessary to manually correct the bounding boxes in order to properly train the algorithms and avoid future misclassifications; Table 2 resumes the total number of bounding boxes before and after the reannotation, with a total increase of 2.9% and 2.1% for the Lucchi++ and Kasthuri++ datasets respectively.

		Training	Validation	Test
Lucchi++	Before	2026	668	2641
	After	2087	690	2715
Kasthuri++	Before	2049	821	2360
	After	2107	843	2389

Table 2: Number of bounding boxes in the two datasets before and after the reannotation.

To avoid spurious "noisy" detections, both networks were configured to ignore mitochondria with less than 50 pixels in area. An analysis of the ground truth shows that this led the algorithms to ignore 261 objects with an average area of less than 5 pixels in the Kasthuri++ training dataset and 10 objects with an average of 9 pixels in the Lucchi++ training dataset.

3.3 ALGORITHM TRAINING

The training process of a YOLO and Detectron2 was essential to optimize its efficiency in the detection of the mitochondria. In this process, different configurations were tested so it could be possible to identify the setup that can achieve high performance without occurring in under- or over-fitting.

3.3.1 *YOLO*

The YOLO version 4 was published in April 2020 and uses the Darknet framework, which is a Open Source Neural Network written in C and optimized for GPU with CUDA. During the scope of this work, the training and detection with YOLO were carried out on a computer equipped with a CPU Intel(R) Xeon(R) Gold 6336Y 22.4GHz and two NVIDIA GeForce RTX 3090.

After the identification of each mitochondrion, as described in 3.2, a text file (.txt) in YOLO format has been created for each image, containing the list of mitochondria present in that image, as depicted in Figure 18. In this specific format, each line corresponds to one object (bounding box), and must included in this order: the number of is class (start in 0); normalized X-center position; normalized Y-center position; normalized width; normalized height. This normalization is achieved, e.g. by dividing the value of X or Y by the width or height of the image.

```
0 0.2275390625 0.037109375 0.0390625 0.07421875
0 0.94677734375 0.06640625 0.0830078125 0.1171875
0 0.330078125 0.08268229166666667 0.060546875 0.08984375
0 0.6171875 0.11002604166666667 0.080078125 0.13151041666666666
0 0.98779296875 0.138671875 0.0244140625 0.08463541666666667
0 0.46142578125 0.13736979166666666 0.0400390625 0.04296875
0 0.8125 0.20833333333333334 0.09375 0.10416666666666667
0 0.4140625 0.24609375 0.04296875 0.057291666666666664
0 0.19287109375 0.3138020833333333 0.0849609375 0.08854166666666667
0 0.40478515625 0.353515625 0.0810546875 0.06901041666666667
0 0.8037109375 0.5104166666666666 0.072265625 0.140625
0 0.97998046875 0.5026041666666666 0.0400390625 0.08854166666666667
0 0.44873046875 0.5846354166666666 0.2275390625 0.18229166666666666
0 0.923828125 0.5592447916666666 0.0703125 0.08463541666666667
0 0.17431640625 0.5755208333333334 0.0595703125 0.06510416666666667
0 0.69140625 0.6360677083333334 0.056640625 0.08984375
0 0.916015625 0.720703125 0.0390625 0.07161458333333333
0 0.81591796875 0.8040364583333334 0.1337890625 0.10286458333333333
0 0.84912109375 0.9563802083333334 0.0791015625 0.08723958333333333
0 0.4111328125 0.9713541666666666 0.068359375 0.057291666666666664
0 0.10400390625 0.9915364583333334 0.0869140625 0.016927083333333332
```

Figure 18: Example of the YOLO format annotation of image one of Lucchi++ train set.

Next, four new files need to be created for each dataset:

- a file named *obj.name* containing the list of classes;

- files named *train.txt* and *valid.txt* containing the location of each image for training and validation;
- a file named *obj.data* containing the total number of classes, the location of the previously mentioned files, and where the training results must be stored.

As mentioned in section *Proposed Methods* 3.1, initially these files only accommodated information about the dataset with uncompressed images. Later on, these files were changed to include both uncompressed and previously compressed images.

Although YOLO is provided with generic configuration files, which include network parameters, a data augmentation process, train optimization, and layer configuration, they must be modified to achieve better results [91]. This has been made empirically, namely by changing the following parameters:

- batch: number of images and labels used in the forward pass to compute a gradient and update the weights via backpropagation;
- subdivisions: subdivision of the batch that ran in parallel on the GPU;
- width, height, and channels: network size, meaning that every image is converted/resized with these values during Training and Detection;
- angle, saturation, exposure, and hue: for data augmentation by randomly changing these values on images during training;
- policy: use of steps or/and scales to adjust the learning rate during training;
- steps: list of the batch number when the learning rate is adjusted by the scale factor;
- scales: list of scales that are multiplied by the current learning rate at the current batch;
- max_batches: number of iterations after which the train is stopped;
- learning_rate: initial learning rate for training;
- anchors: initial set of predefined bounding boxes;

Other important parameters used are:

- mosaic: used for data augmentation, by mixing 4 images in one;
- momentum: how much of the history affects the further change of weights;
- burn_in: number of initial interactions where the learning rate is gradually increased. Also used as an interval after which the training progress has evaluations;

- filters and num_filters: type and number of convolutional kernels used by layer;
- activation: activation function after convolution.

Since during training, YOLO processes images in a random way, including augmentation on the fly, these training sessions are non-deterministic. This means that even with identical parameters, consecutive training sessions yield varying detection performances. For this reason, each setting was trained multiple times (6, in the case).

The analysis and monitoring of the training sessions were obtained essentially considering the metric mean average precision (mAP), presented in red in Figure 19. The best weights attained in each training session were subsequently used to detect and classify the test sets.

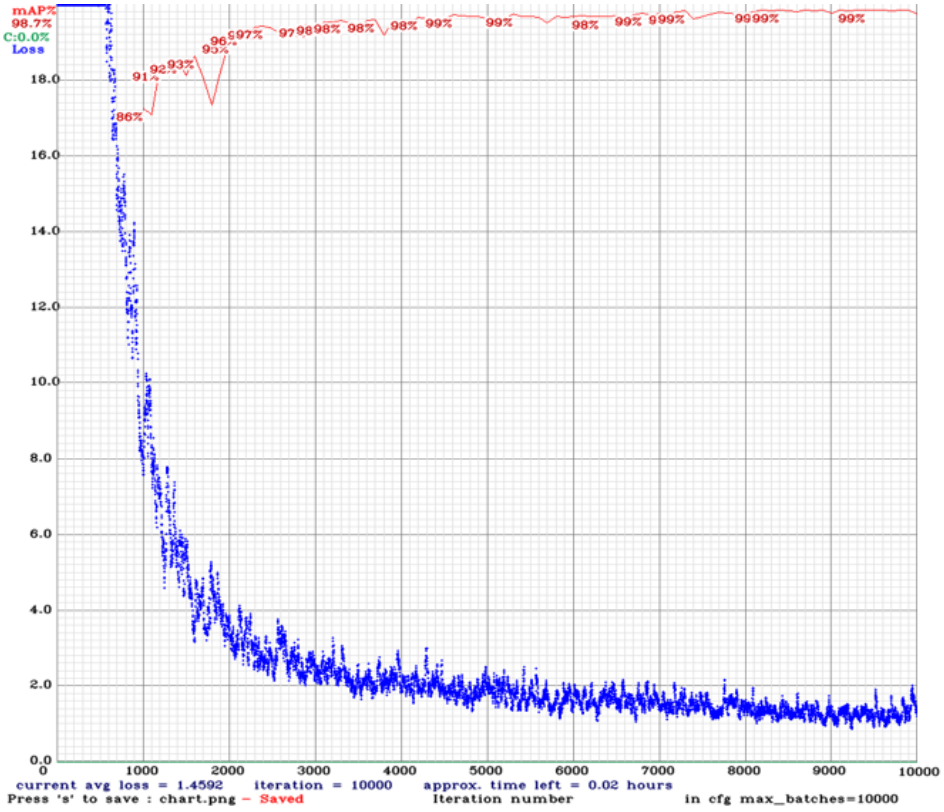


Figure 19: Evolution of a training session of YOLO: mAP (red line) and loss function (blue line).

The configuration that achieved better results had a maximum of 10000 iterations, and a batch size of 16 split by 8 (subdivisions). The stochastic gradient descent optimizer was used with a momentum term of 0.949 and a learning rate of 0.5×10^{-4}. A step learning rate policy was adopted, so after 80% and 90% of the maximum

number of batches, the learning rate was decreased by 90%. In the first layer, each image was resized to a resolution of 416×416 pixels. The augmentation done by YOLO during training includes saturation and exposure up to 1.5, and hue up to 1.

Although it was possible to retrain YOLO starting from the provided pre-weights, it was observed during the initial studies, that better performance was achieved when training from scratch. The retrained process, using the best weights from the initial train, was later used to improve the performance of the algorithm over previously compressed images, as mentioned in section 3.1.

3.3.2 *Detectron2*

The built-in support for GPU acceleration using CUDA and Python in Detectron2 facilitates efficient training and detection processes, making it feasible to use the same computer that was previously used for YOLO 3.3.1.

As with YOLO, training Detectron2 requires access to images along with their corresponding annotations. However, since Detectron2 lacks native support for YOLO format annotations, the initial step involved translating them into COCO JSON format [92].

In this format, each dataset information is consolidated into a single file, adhering to a structure as exemplified in Figure 20. This file encompasses a list of images, classes, and objects, each assigned a unique ID starting from zero, which facilitates establishing relationships among them, enabling efficient processing and analysis. The information about each of the bounding boxes is presented in the following order: top left X position, top left Y position, width, and height. Unlike YOLO, values are not normalized.

The next step was the configuration of the algorithm. Although Detectron2 allows different types of segmentation, in the scope of this work, and for comparison with YOLO, only object detection was considered. In this sense, the COCO-Detection model *faster_rcnn_R_101_FPN_3x.yaml* configuration was as used, which involves loading pre-trained weights and configuring the network accordingly.

While some parameters are similar to those of YOLO, such as learning rate or anchors, Detectron2 offers a broader range of options for resizing and augmentation of the images (e.g. list for minimum resize, flipping, or cropping the images). Such flexibility enables an approach to fine-tune the model for better results [80].

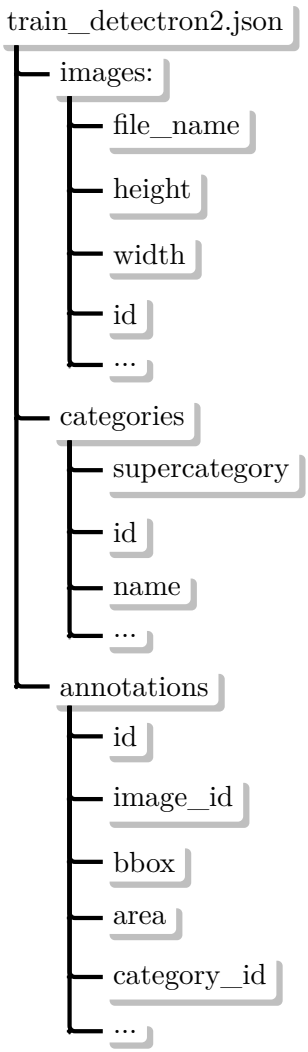


Figure 20: COCO JSON format.

The Detectron2 network was therefore trained with a maximum of 15000 iterations, and a *IMS_PER_BATCH* of 8 (with a *BATCH_SIZE_PER_IMAGE* of 16). The gradient descent optimizer was configured with the same momentum and learning rate used with YOLO, but the step learning rate policy decreases the learning rate by 80% at iterations 7000 and 10000 (46.7% and 66.7% of the maximum number of iterations).

Concerning augmentation, Detectron2 training included the possibility of horizontal flip, crop (0.5 of relative range), and resize of the images. The limits used for the length of the short edge of the images in the Lucchi++ resize were 416, 512, 640, and 768, and for the Kasthuri++ 384, 512, 640, 768, 896, 1024, and 1334. In addition, the initial anchor boxes have been changed to sizes 8, 16, 32, 64, 128, and 256, and the aspect ratio to 0.25, 0.5, 1.0, 2.0, and 4.0. These values were obtained empirically, by monitoring the training sessions being made with TensorBoard (a

tool that provides the measurements and visualizations of measurements during the computer vision training workflow). Figure 21 presents the training evolution, in terms of mAP, of different configuration setups.

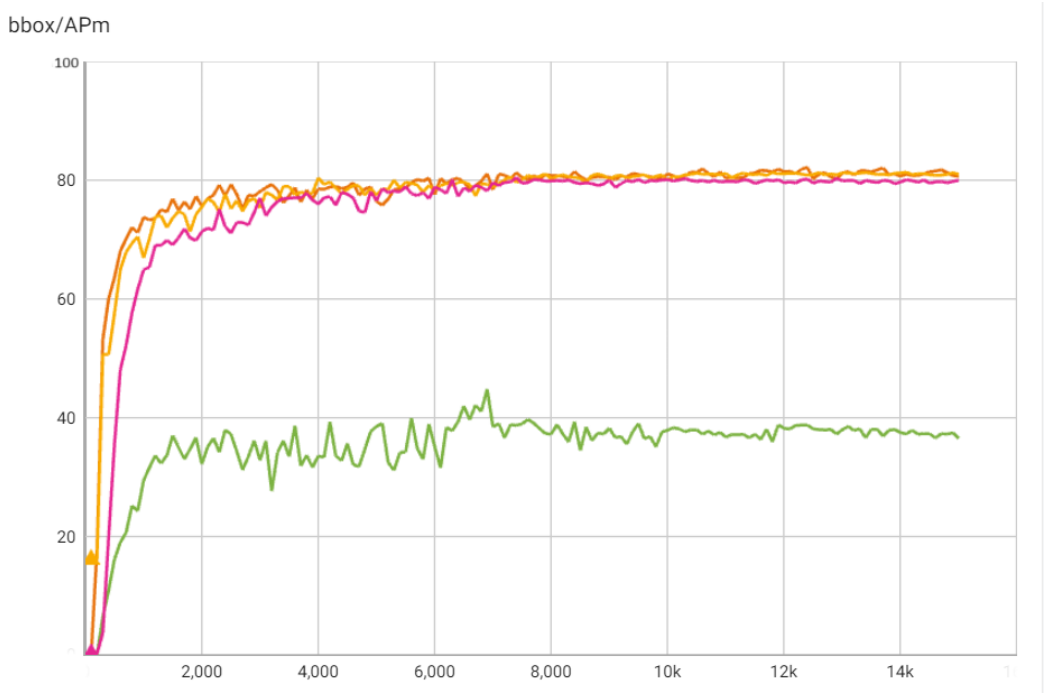


Figure 21: Evolution of different training setups of Detectron2 (each line represents one trained with 15000 interactions).

3.4 HEVC CONFIGURATION

To investigate the effect of compression on the performance of YOLO and Detectron2, and subsequently mitigate any distortion on their performance, the HEVC reference software was used, namely HM version 16.9 [93]. This software offers various configuration files that can be customized to achieve the desired levels of image quality.

The images presented in the Lucchi++ and Kasthuri++ datasets were individually encoded exclusively in intra-mode, following the *main_intra* profile. As all images are in 8-bit grayscale, the 4:0:0 chroma format was used, with 8 *InputBitDepth*. Also, due to the size of the images, it was necessary to apply a level of 6.2 and activate the compliance window.

On the one hand, the levels employ a set of constraints on the bitrate in accordance with the standard [94]. On the other hand, the compliance window allows that

during image partitioning, there is an automatic fill to the next minimum CU size in case the image size does not fill this CU in full.

Other configuration parameters included max CU of 64 by 64 pixel with a *maxpartition* depth of 4. Also, quadtree-based TU coding depth (QuadtreeTUMaxDepth) of 3 with a Log2 of minimum transform size between 2 and 5. This allows the HEVC to exploit various subdivisions and fine-tune the compression process. This allows the encoder to optimize compression efficiency by adapting the partitioning of the image and adjusting the transform size according to the content characteristics.

As mentioned in section 3.1, during the initial phase of this study, each image was individually compressed with different quality levels. This was made by using the fixed QP: 22, 27, 32, 37, 41, 43, 45, 47, and 51, i.e., suppressing possible variations of delta QPs. At a later stage of this study, namely to compress unequally different regions, the following QP were used for the background: 41, 43, 45, 47, 49, and 51, while the ROIs were coded with QPs: 22, 27, 32, and 37.

The assignment of QP based on regions implies that, during the coding process, the image structure in HEVC must be adapted according to the ROIs position. Furthermore, not only the encoded bitstreams need to include the QP used in each subdivision, but the decoder also needs to be able to correctly extract this information.

3.4.1 *HEVC ROI-based adaptation*

To achieve a selective compression ratio in the ROI and in the background, a novel function was developed to establish a relation between each structural block and an external source containing the ROI positions. This connection enables the assignment of different QP values to be used at the CU level, as shown by Figure 22.

When an intersection occurs between the ROI and a block, the QP is modified to a predefined value, thereby distinguishing the rate-distortion characteristics of the encoded block based on whether it belongs to the relevant region (ROI) or the background. Also, a flag is triggered, ensuring the recalculated QP is made available to the decoder at the CU level. It should be noted that the QP only affects the value of the coefficients other than zero that result from the prediction and transform, being irrelevant in the remaining cases.

To enhance the accuracy of the mitochondria detection, an 8-pixel frame was added around the ROI, as shown by Figure 23. This allows a margin between

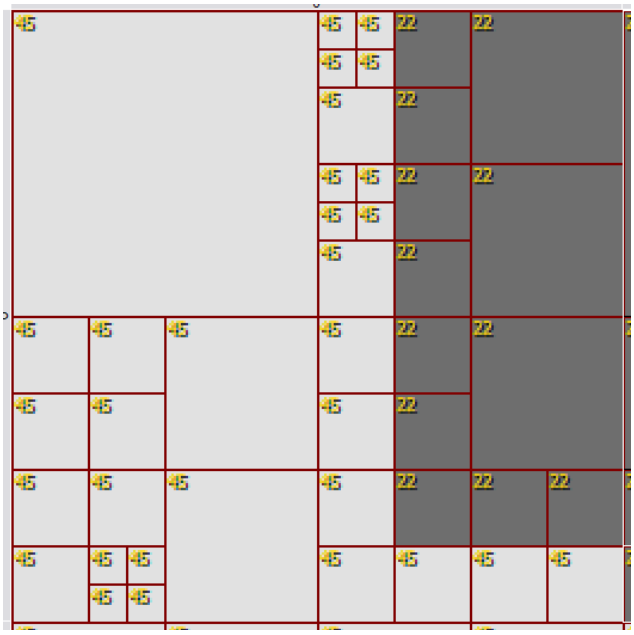


Figure 22: Coded CTU with different levels of quality at the CU level (QP 45 and QP 22).

the mitochondrion membrane and the limit of the bounding box for compression. While this adjustment resulted in a minor increase in required compression ratios, it significantly improved the overall detection performance of the algorithms.

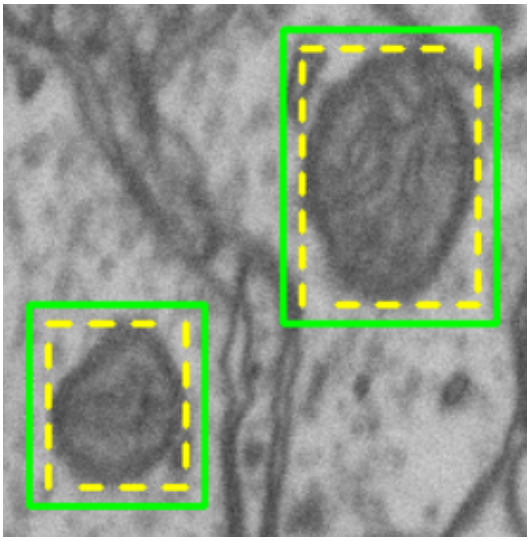


Figure 23: Delimitation of mitochondria by bounding box (yellow dashed line); Extension of the same bounding box for ROI compression (solid green line).

It is noteworthy that all images compressed throughout this study were decoded using an unmodified HEVC, ensuring that the bitstreams generated by the proposed modified encoder was still compatible with the H.265 standard and its decoder.

3.5 METRICS

To evaluate the performance of YOLO and Detectron2 in the detection of mitochondria, the intersection over union (IoU) indicator was used. A successful mitochondria detection, or true positive (TP) occurs when:

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \geq 0.5, \tag{1}$$

i.e. the algorithms' output and the annotated reference intersection area should represent at least 50% of the union. Accordingly, a proposed detection with IoU bellow this value is considered false positive (FP). Also, in this type of studies the false negative rate (FN) can be then simply estimated from $FN = 1 - TP$ (values in percentage).

The assessment indicators TP, FP, and FN are usually combined to establish other performance metrics such as Precision, Recall, and F_1 Score.

Precision quantifies, among objects classified as positive, which percentage is actually correct or relevant, according to

$$Precision = \frac{TP}{TP + FP}. \tag{2}$$

Recall, on the other hand, indicates the ability to identify the sample object on the dataset:

$$Recall = \frac{TP}{TP + FN}. \tag{3}$$

The F_1 Score combines this previous metrics, in terms of the harmonic mean of Precision and Recall. This gives equal weight to both metrics, as show in

$$F_1\ Score = 2 \times \frac{Recall \times Precision}{Recall + Precision}. \tag{4}$$

As previously described for the training of YOLO and Detectron2, the well-known mean Average Precision metrics were used to analyze the ongoing performance. It computes the area under the precision-recall curve for the detected mitochondria (only one class was used) given an IoU of 0.5.

The evaluation of compression performance employed the compression ratio (CR) metric according to

$$\mathrm{CR} = \frac{\sum_{i=1}^{N} \mathrm{Size}(\mathrm{ORIG(i)})}{\sum_{i=1}^{N} \mathrm{Size}(\mathrm{COMP(i)})}, \tag{5}$$

where N is the total number of images in the dataset, Size(ORIG(i)) is the number of bits of the uncompressed image i, and Size(COMP(i)) is the number of bits representing the compressed image.

3.6 SUMMARY

This chapter outlines the strategies used to enhance the performance of computer vision algorithms, specifically YOLO and Detectron2, on pre-encoded Electron Microscopy images. Firstly, a description of the pipeline used to assess the impact of lossy compression is provided. This is followed by an outline of the approaches employed to extend the usable CRs, while maintaining good algorithm performance. On one hand, the possibility of training the computer vision algorithm using previously compressed images is exploited. On the other hand, the strategy of compression by regions enables the encoding of ROI with higher quality. Also, it is detailed the description of the dataset preparation, the configuration of the computer vision algorithms, and the used compression algorithm. Finally, the metrics employed to comprehensively assess and evaluate the algorithmic performance are presented.

4

EXPERIMENTAL ASSESSMENT

This chapter delves into assessing the impact of compression on the detection performance of computer vision algorithms within the scope of this dissertation. Initially, the detection performance of YOLO and Detectron2 against the provided ground truth is presented. Subsequently, the influence of compression on the detection performance of YOLO and Detectron2, along with the results of the devised strategies aimed at enhancing the compression's contribution, are detailed.

4.1 DETECTION PERFORMANCE OF YOLO AND DETECTRON2 AGAINST THE GROUND TRUTH

The performance of both algorithms in detecting mitochondria in electron microscopy images was evaluated by directly comparing their predictions with the ground truth. As mentioned in Chapter 2.1.2, the ground truth provided in the dataset resulted from a reannotation process conducted by specialists.

The evaluation results, summarized in Table 3, reveal that both algorithms achieved similar F_1 Score values, averaging 92%.

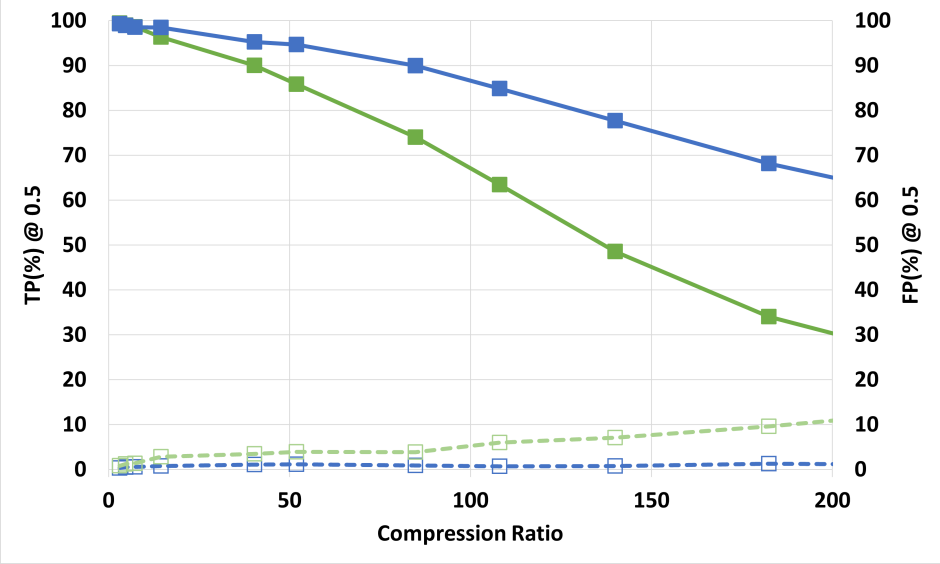
Lucchi++	**TP**	**FP**	**Precision**	**Recall**	F_1 **Score**
YOLO	89.5	5.1	94.6	85.0	91.9
Detectron2	91.7	8.0	92.0	91.7	91.9

Kasthuri++	**TP**	**FP**	**Precision**	**Recall**	F_1 **Score**
YOLO	92.8	9.9	90.3	92.8	91.6
Detectron2	92.2	6.4	93.5	92.2	92.9

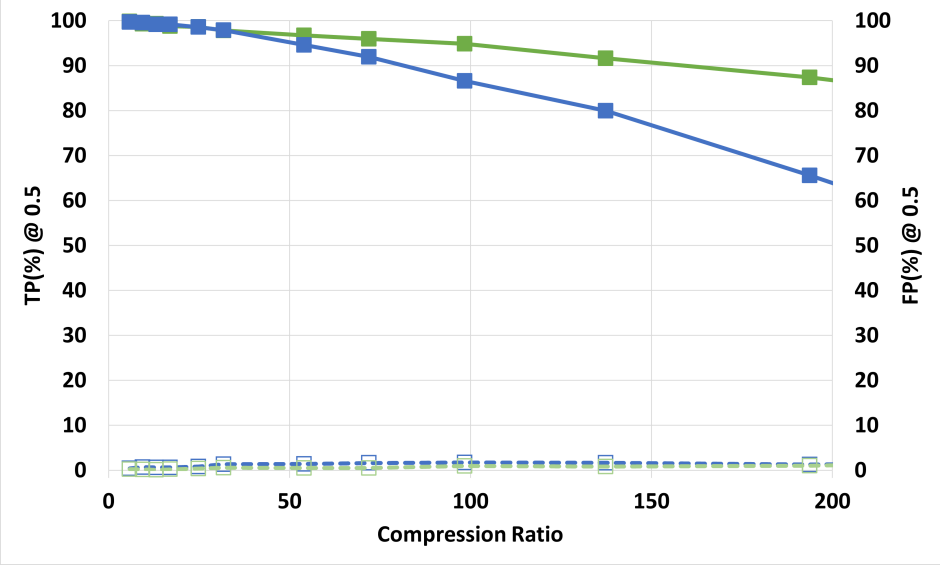
Table 3: Detection performance (%) of YOLO and Detectron2 versus ground truth.

4.2 IMPACT OF IMAGE COMPRESSION ON DETECTION PERFORMANCE

To evaluate the performance of the detection algorithms with the images affected by compression-related distortions, the test datasets were compressed using HEVC, as detailed in Section 3.4. Subsequently, YOLO and Detectron2 were employed for mitochondria detection, and their results were compared with those obtained in the preceding section (TP and FP).

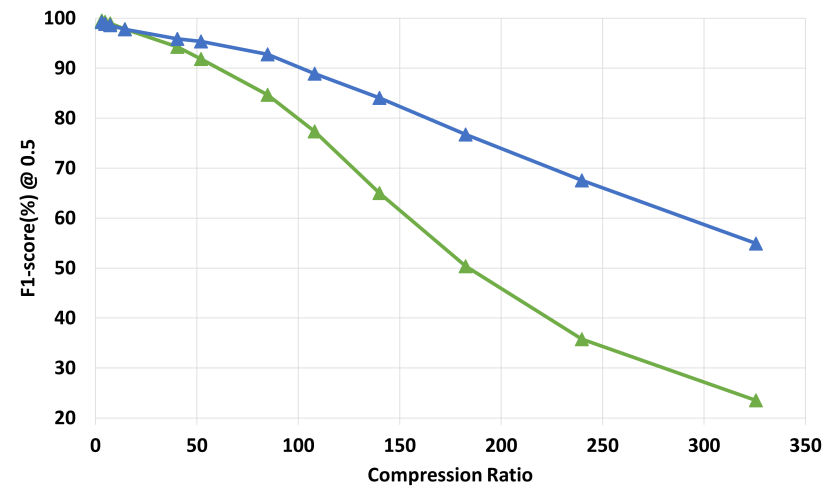


(a) Lucchi++ dataset.

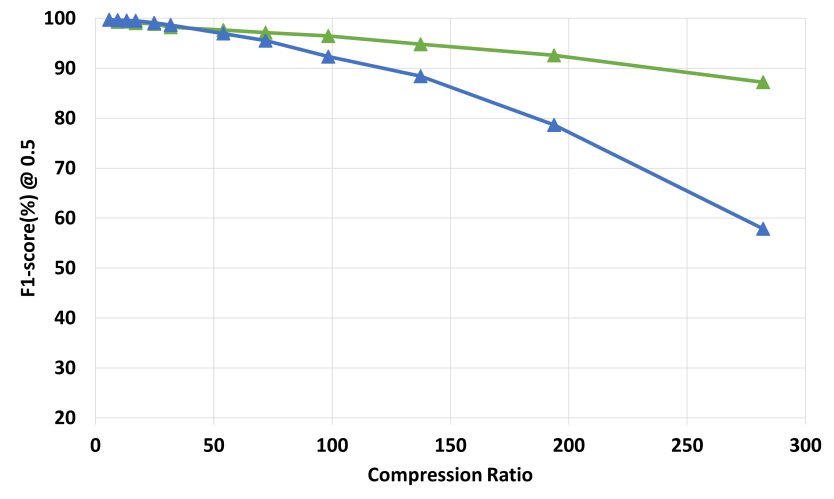


(b) Kasthuri++ dataset.

Figure 24: Mitochondria detection performance in compressed images for YOLO (green lines) and Detectron2 (blue lines): TP (full square) and FP (empty square).



(a) Lucchi++ dataset.



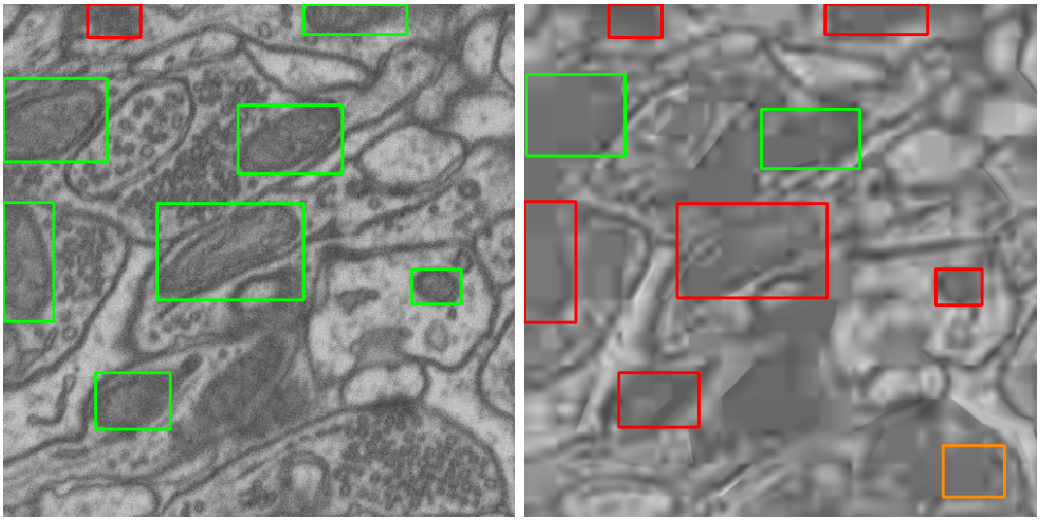
(b) Kasthuri++ dataset.

Figure 25: F_1 Score results for Mitochondria detection in compressed images: YOLO (green lines) and Detectron2 (blue lines).

As can be observed in Figures 24 and 25, although encoding algorithms allow for the reduction of data required to represent biomedical information, the inherent distortion introduced by lossy algorithms can raise concerns regarding data fidelity, potentially affecting the usefulness of the encoded data. Nevertheless, lossy compression can be considered up to a certain limit, still maintaining the high-performance benchmarks set by YOLO and Detectron2.

The presented behavior can be understood on the grounds of the coding distortions incurred by compressed images, especially the well-known smoothing due to the suppression of higher spatial frequencies. This effect can be observed in Figure 26, where it is clear that the loss of image detail leads to a higher number of undetected

mitochondria. It is important to observe that the algorithm's ability to identify the mitochondria, although dependent on the overall image quality, is not as sensible as the human vision, as it can be noted by the accurate detection of some mitochondria, even in a blurry image such as the one in Figure 26(b).



(a) Image compressed with QP22. (b) Image compressed with QP51.

Figure 26: Detail of image 18 of test subset of dataset Lucchi++: TP (Green bounding boxes), FN (red bounding boxes), and FP (orange bounding boxes).

For comparison, visual quality metrics related to the human visual system (PSNR and SSIM) obtained from the compression of both datasets with the chosen QPs are shown in Table 4. For this type of content, usually PSNR below 35dB produces visually poor images. In the proposed experiment the detection network was able to detect with very good results even when the PSNR was close to 30dB (QP 32 in Lucci++).

Among the two computer vision algorithms under consideration, Detectron2's performance is more consistent across both datasets. In turn, YOLO's performance exhibits a rapid degradation beyond certain limits within the Lucchi++ dataset, evidenced by a decline in TP and an increase in FP. On the contrary, its performance is remarkable in the Kasthuri++ dataset, achieving a F_1 Score higher than 87% with a CR of 282.

In essence, these results demonstrate that automatic mitochondria detection may perform equally well in both uncompressed and compressed images up to a certain limit. This finding suggests the potential for infrastructure optimization at the storage and transmission level without compromising vital information.

	Lucchi++						
QP	22	27	32	37	43	47	51
CR	3.0	4.7	14.5	52.0	108.1	182.5	325.6
YOLO F_1 Score (%)	99.6	99.3	97.8	91.8	77.3	50.3	23.5
Detectron2 F_1 Score (%)	99.2	98.9	97.8	95.3	88.9	76.7	54.9
PSNR (dB)	41.2	35.8	30.0	28.2	27.0	25.8	24.4
SSIM	0.996	0.986	0.943	0.913	0.885	0.853	0.800

	Kasthuri++						
QP	22	27	32	37	43	47	51
CR	5.8	9.4	17.0	31.8	71.9	137.4	282.1
YOLO F_1 Score (%)	99.7	99.3	99.1	98.2	97.1	94.8	97.2
Detectron2 F_1 Score (%)	99.7	99.7	99.5	98.7	95.5	88.4	57.9
PSNR (dB)	43.2	38.2	34.2	31.1	27.7	25.6	23.7
SSIM	0.999	0.997	0.993	0.986	0.963	0.904	0.893

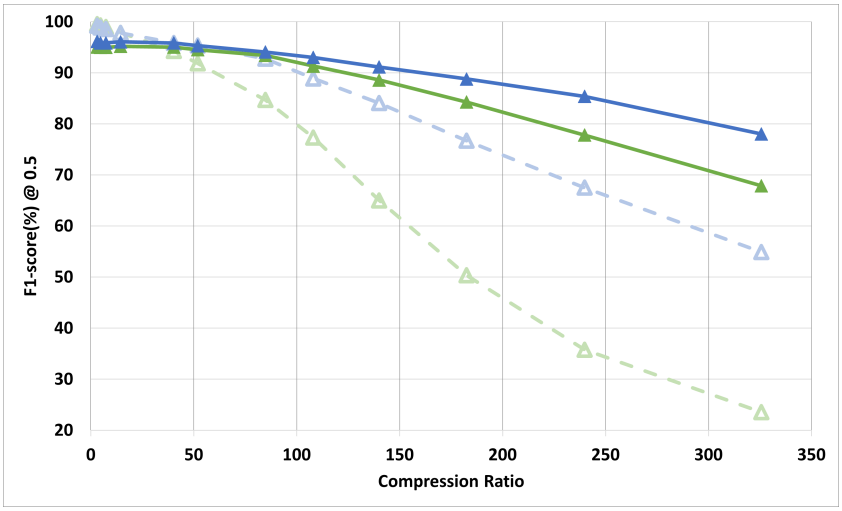
Table 4: Objective image quality per QP

4.3 EXTENDING THE PERFORMANCE FOR HIGHER CR

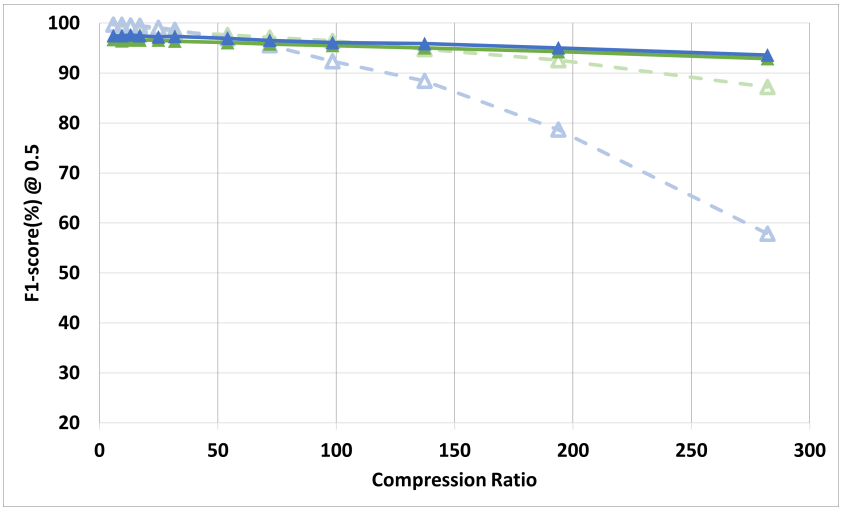
A second set of experiments was carried out, with algorithms now re-trained with both compressed and uncompressed images. The rationale is that exposing the algorithms to lower quality images at the training stage could increase their performance when later considering images coded with higher CRs.

The results, using the prediction over uncompressed images as reference (section 4.2), obtained with this new training approach are presented in Figures 27(a) and 27(b), alongside the previous ones (dashed lines). As it can be observed, in both algorithms this new training approach results in enhanced performances, enabling the extension of the useful range to higher compression rates.

With the Kasthuri++ dataset, both detectors exhibit now similar performances, achieving F_1 Score results exceeding 92% for compression ratios up to 282; i.e. an increase of 5 and (over) 50 percentual points in YOLO and Detectron2, respectively, at this CR. The enhanced performance is also evident in the Lucchi++ dataset, surpassing the previous results: at a CR of 325, Detectron2 has a substantial increase in the F_1 Score, rising from 55% to 78%, while YOLO exhibits a remarkable improvement from 23% to 68%.



(a) Lucchi++ dataset.

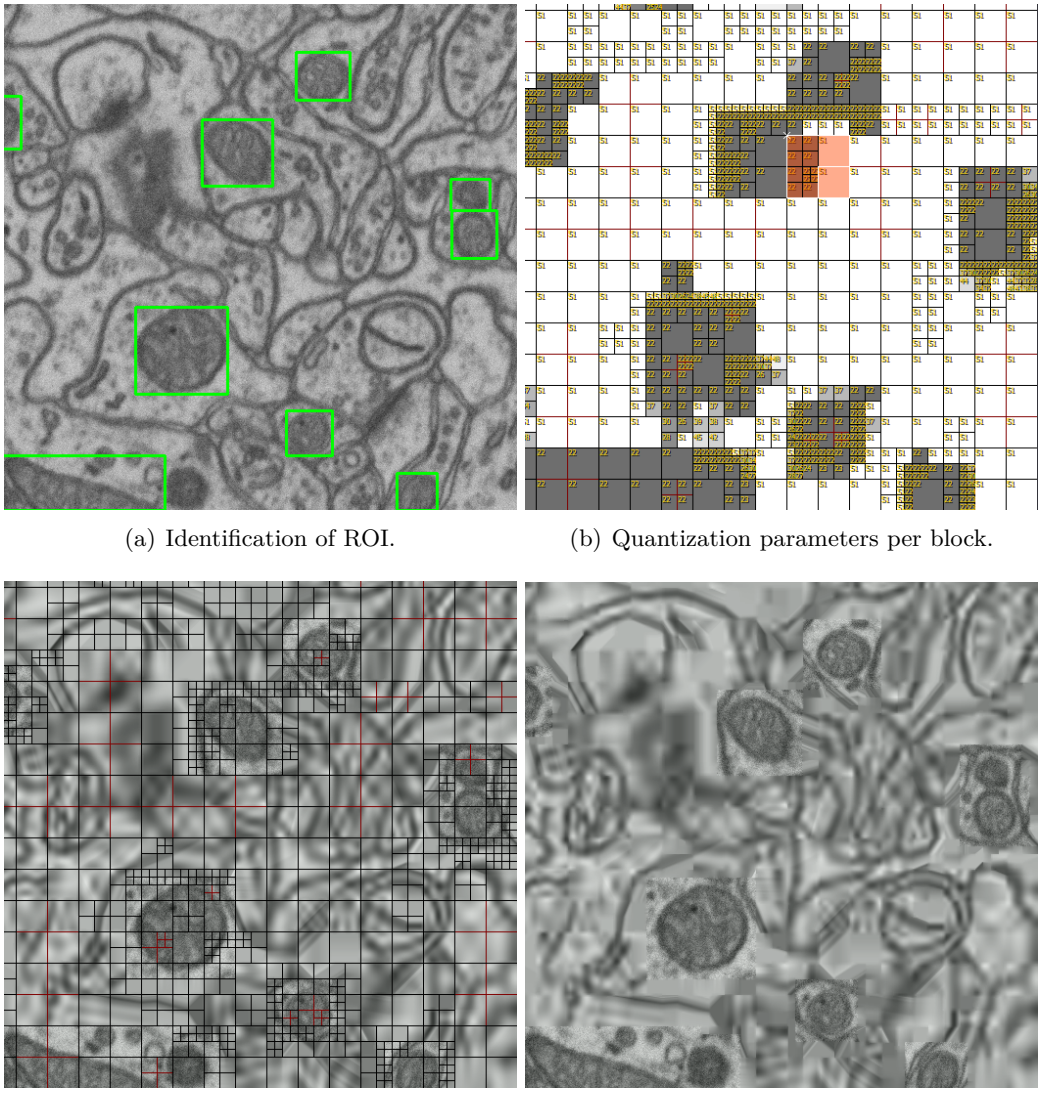


(b) Kasthuri++ dataset.

Figure 27: F_1 Score results of the proposed method for Mitochondria detection in compressed images: YOLO (green lines) and Detectron2 (blue lines); proposed method (solid lines), and previous results (dashed lines).

4.4 REGION-SPECIFIC CODING WITH IMPROVED QUALITY ALLOCATION TO ROI

A further set of studies was carried out, based on the underlying idea that as mitochondria are the truly informative elements, the spatial regions where they are located should be better preserved in terms of image quality. Accordingly, the detection performance of YOLO and Detectron2 was assessed in images where the ROI underwent encoding with higher quality than the background.



(a) Identification of ROI.

(b) Quantization parameters per block.

(c) Partition structure.

(d) Image coded by regions.

Figure 28: Example of unequal quality compression of background (QP 51) and ROI (QP 22).

The predicted mitochondria from the algorithms in non-encoded images served as ROIs for this process, as outlined in Section 3.4. As demonstrative, Figure 28 displays a cropped section of the first image of the Lucchi++ test dataset, allowing an analysis of the region-specific coding:

- In image 28(a) the bounding box inferred from YOLO, and used as ROI, are delineated in green;
- Image 28(b) illustrates the QP assigned to each CU, and highlights one of the CTU;
- Image 28(c) presents the quadtree partition made by the HEVC;

- In image 28(d) the resulting compression of the ROIs with QP 27 and the background with QP 51 can be observed.

A comparison between Figures 28(a) and 28(d) allows to observe that the background exhibits more distortion, as expected, since higher quality has been applied to the ROI. The proposed approach allows for variations in quality between zones to be made at the CU level, as depicted in Figure 28(b). It is clear that within a single CTU (highlighted in red), there are CUs with different QPs, such as QP 22 and QP 45, depending upon their association with either the ROI or the background. This allows an optimization of the bitstream since less area of the background is encoded with higher quality, allowing a reduction in image storage data.

As can be observed in Figure 28(b), and detail in Figure 29, in some background regions there are QPs with different values. This is expected, as the quantization solely affects the coefficients after post-prediction and transforms. In instances without coefficients, the encoder determines the QP to transmit based on the preceding CU as mentioned in the standard, with no impact on the compression.

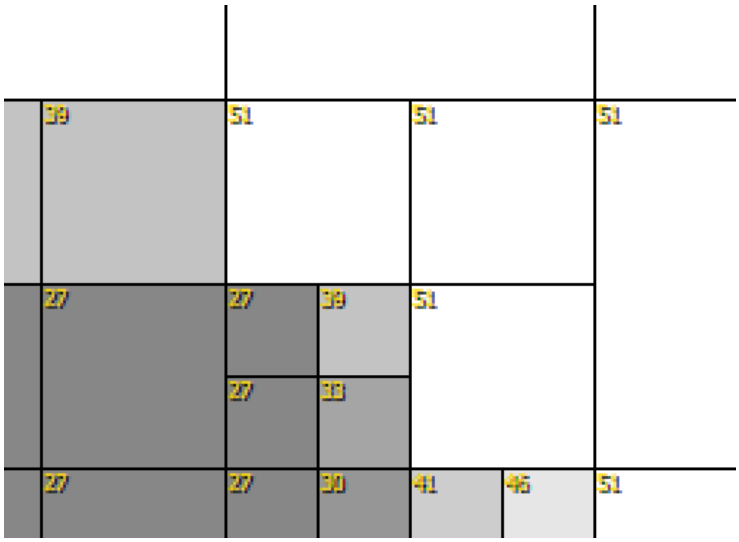


Figure 29: Detail of QPs with different values in neighboring CUs.

Furthermore, by examining the CTU structure generated by HEVC in Figure 28(c), it is possible to see that the subdivision in different sizes of CU was not bounded and still relies on the inner loop of the HEVC to find the best intra-compression. This aspect holds significance, as it ensures that the final bitstream can be decoded by any software supporting standard HEVC.

As detailed in Section 3.4, the bounding box area was extended by 8 pixels in all directions. This extension ensures that the compression margins of each QP do not align directly with the boundaries of the mitochondria. A closer examination of this

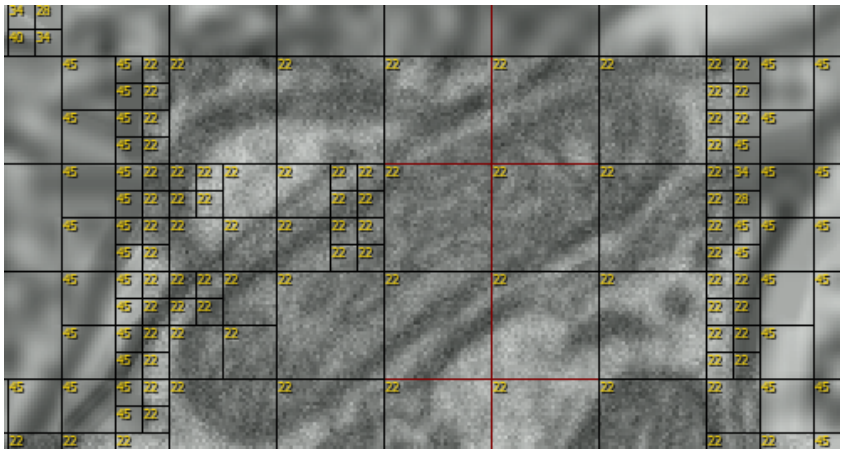


Figure 30: Detail of the QP used in the mitochondrion region (QP 22) and background (QP 45).

phenomenon can be found in Figure 30, where it is evident that the membrane of the mitochondria falls well within QP 22 (high quality - ROI).

The performance results obtained by YOLO and Detectron2, using the prediction over uncompressed images as reference, are presented in Figure 31 along with previous results. The yellow highlighted areas, delineated below by the previous best results, point out to configurations with improved performance, meaning there is potential for further improvement in detection performance at higher CR, especially in cases where the previously trained algorithms, with and without coded images, exhibited a decline in performance.

While the desired CR depends on the configuration of image quality (materialized by QP) between ROI and background, it is worth noting that several F_1 score-to-CR relationships surpass those observed in earlier findings. Even in the Kasthuri++ dataset, the performance of YOLO trained only with uncompressed images can also be matched in some configurations of this technique, as shown for compression ratios of 63 (F_1 Score 97,4%).

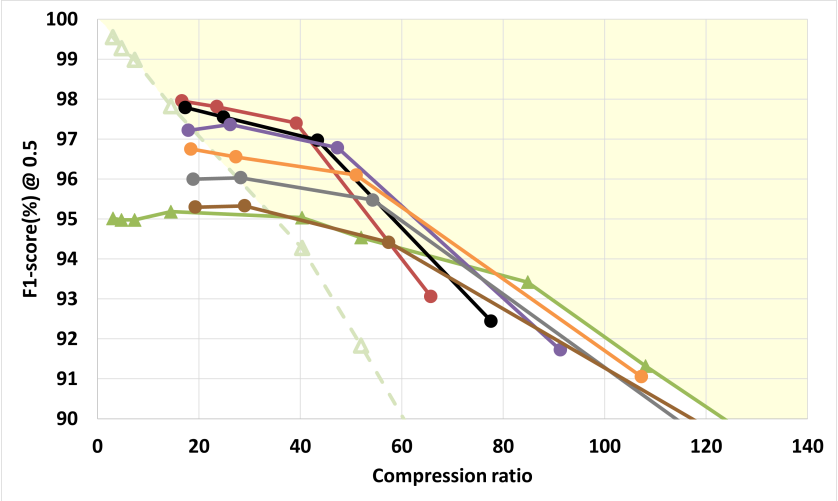
Although, the CR range of this method is upper and lower bounded by the QP of the background and foreground regions, employing region-specific coding with improved quality allocation to ROI can yield superior compression results for some configurations.

For example for the Lucchi++ dataset, the configurations of QP 41 for background and 32 for ROI resulted in an F_1 Score above 97% for CR=40 in both algorithms. As a reference, the HEVC encoder exhibits an average lossless CR of 1.32 and 2.18 on Lucchi++ and Kasthuri++ datasets, respectively. This means that an image of

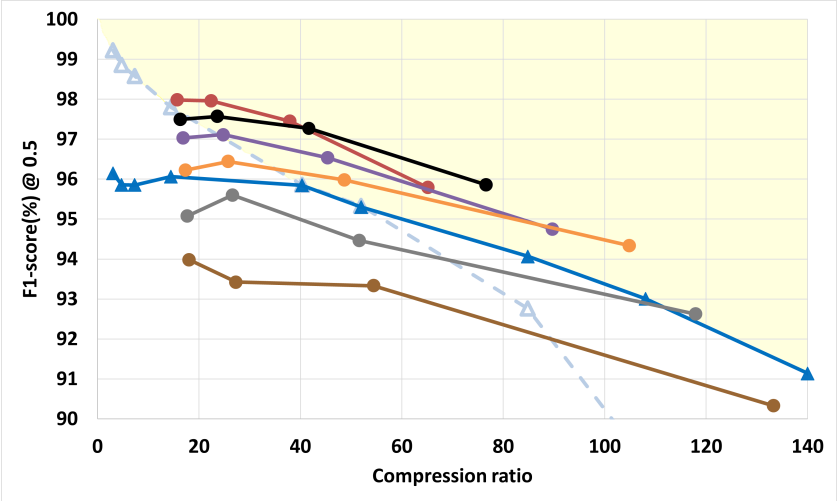
1024×768 pixels in grayscale could be stored in ≈ 20 kB, with little impact on the detection.

4.5 SUMMARY

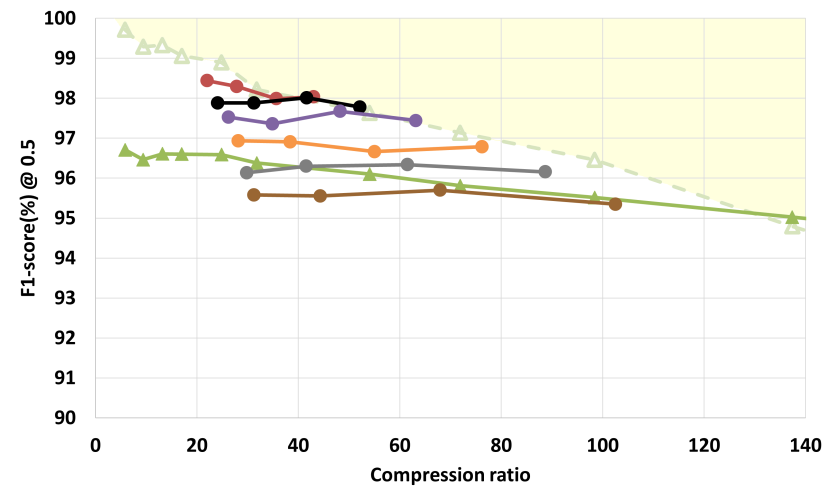
In this chapter, the results of the experimental assessment are discussed. Initially, the detection performance of YOLO and Detectron2 is compared against the provided ground truth (human-based), with the algorithm predictions being used as the reference for subsequent analysis. Following this, the impact of compression on the detection performance is assessed, namely with the performance drop with the increase of CRs being quantitatively shown.



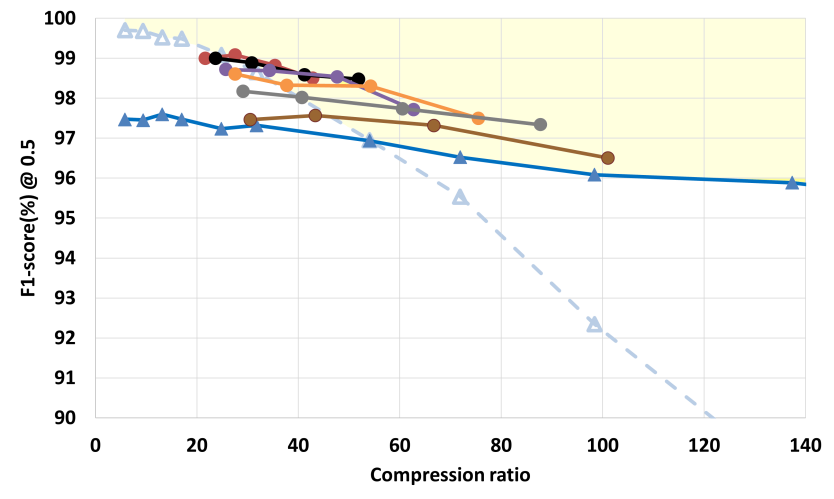
(a) YOLO.



(b) Detectron2.



(c) YOLO.



(d) Detectron2.

Figure 31: F_1 Score results for Mitochondria detection of the Lucchi++ (*a* and *b*) and Kasthuri++ (*c* and *d*) dataset; when ROI and background are coded with different QP values: different color lines (solid circles) are associated to background compression with QP 41 (red), 43 (black), 45 (purple), 47 (orange), 49 (gray), and 51 (brown); in each line, from right to left, the solid circles are associated to ROI QPs of 22, 27, 32, and 37. The green and blue, solid and dashed, lines (triangles) represent previous results (Figure 27(a) and 27(b)).

The results of the approaches devised to enhance the performance of both computer vision algorithms were presented. Firstly, it was observed that using a training method also based on compressed images leads to a substantial increase in performance at higher CRs. Afterward, the implementation of region-based coding with different qualities between ROI and background surpasses previous results for certain configurations. This suggests the possibility of increasing the compression range to higher values, with minimal impact on the detection performance of these

two computer vision algorithms, consequently reducing the amount of data required to represent such images.

5

CONCLUSIONS AND FUTURE WORK

5.1 CONCLUSION

The widespread adoption of electronic health records alongside with the rise of advanced medical imaging techniques has brought both opportunities and challenges to the healthcare infrastructure. The increase of digital data generated by these systems requires robust processing, storage, and transmission capabilities. In this sense, computer vision algorithms and compression techniques emerge as crucial tools to ease the management of this data.

Nowadays, if on one hand, computer vision is used with great success freeing human resources from repetitive tasks, on the other hand, a trade-off between CR and image fidelity is a critical consideration in the implementation of lossy compression techniques for medical imaging data.

In this sense, the main objective of this research was to investigate and enhance the valuable role of biomedical image compression within computer vision systems. Specifically, the goal of the study was to establish compression ratio boundaries for lossy coding, with minimal negative impact on the computer vision detection algorithm's, as well as to develop mechanisms that allow to increase the compression ratio for electron microscopy (EM) images.

In this context, three main contributions were outlined: Firstly, to expand the ongoing research on automatic detection, in order to understand the compression boundaries that have minimal impact on its performance. Secondly, to exploit training methodologies for learning-based computer vision algorithms, that incorporate images previously compressed at different quality. Finally, adapt a standard image encoder (i.e. HEVC) to compress the ROI, identified by the computer vision algorithm, and the background with different quality, while keeping the bitstream standard.

The results show that while the performance of YOLO and Detectron2 may be influenced by the distortion resulting from lossy compression, the impact is negligible at lower CR. Moreover, it is possible to significantly enhance the performance of

detection algorithms at higher CR by including images previously compressed during the training process. For instance, a gain exceeding 20 percentual points was achieved for a CR of 325 on the Lucchi++ dataset, for both detectors.

Furthermore, unequal quality region coding allows to maintain high performance at compression rates where computer vision algorithms start to lose efficiency, meaning that it is possible to significantly reduce the amount of data required to represent these types of images. For example, at CR of 40, different configurations of quality between ROI and background achieve F_1 Score above 97% on both detectors and datasets. In this sense, priority and higher quality access to regions of interest (ROI) becomes an added value.

In conclusion, with a shift taking place, where perceptual quality from humans is being replaced by the overall performance of computer vision in repetitive tasks, the use of lossy compression may significantly improve the efficiency of imagiology infrastructures. Although a critical balance between CR and image fidelity must be carefully managed, the results and proposed methods can land the foundations for future advances in biomedical data storage and transmission.

5.2 FUTURE WORK

The evolution of computer vision has been marked by significant advancements in technology, algorithms, and applications over the years. Nowadays, there is a noticeable transition from single object detection techniques to new approaches such as semantic segmentation, instance segmentation, and panoptic segmentation, meaning further studies must the made to assess the impact of distortion in these cases. Moreover, the extension of the presented studies to the VVC/H.266 needs to be assessed since this standard can achieve higher CR for the same quality when compared to previous compression methods.

BIBLIOGRAPHY

[1] Alex A. T. Bui et al. “Envisioning the future of ’big data’ biomedicine.” In: *Journal of biomedical informatics* 69 (May 2017), pp. 115–117. DOI: 10.1016/j.jbi.2017.03.017.

[2] Sabyasachi Dash et al. “Big data in healthcare: management, analysis and future prospects”. In: *Journal of Big Data* 6 (1 Dec. 2019), p. 54. DOI: 10.1186/s40537-019-0217-0.

[3] Erik Meijering. “Cell segmentation: 50 Years down the road [life Sciences]”. In: *IEEE Signal Processing Magazine* 29.5 (2012), pp. 140–145. DOI: 10.1109/MSP.2012.2204190.

[4] Rebecca J. Mitchell et al. “Application of a human factors classification framework for patient safety to identify precursor and contributing factors to adverse clinical incidents in hospital”. In: *Applied Ergonomics* 52 (2016), pp. 185–195. DOI: 10.1016/j.apergo.2015.07.018.

[5] Melvyn L. Smith et al. “The quiet revolution in machine vision - a state-of-the-art survey paper, including historical review, perspectives, and future directions”. In: *Computers in Industry* 130 (2021), p. 103472. DOI: 10.1016/j.compind.2021.103472.

[6] Edward Shortliffe and G. Barnett. “Biomedical Data: Their Acquisition, Storage, and Use”. In: *Biomedical informatics* (Jan. 2006), pp. 46–79. DOI: 10.1007/0-387-36278-9_2.

[7] European Commission et al. *Assessment of the EU Member States’ rules on health data in the light of GDPR*. Publications Office, 2021. DOI: 10.2818/546193.

[8] Medical Imaging & Technology Alliance. *DICOM PS3.10 2022a - Media Storage and File Format for Media Interchange*. Tech. rep. Medical Imaging & Technology Alliance, NEMA, 2022. URL: https://dicom.nema.org/medical/dicom/current/output/pdf/part10.pdfl (visited on 02/27/2022).

[9] “anatomia in Dicionário infopédia da Língua Portuguesa”. In: *Porto: Porto Editora* (2021). URL: https://www.infopedia.pt/dicionarios/lingua-portuguesa/anatomia.

[10] Yimei Zhu and Hermann Durr. "The future of electron microscopy". In: *Physics Today* 68 (Apr. 2015), pp. 32–38. DOI: 10.1063/PT.3.2747.

[11] R. Scott Evans and MS R. Scott Evans. "Electronic Health Records: Then, Now, and in the Future". In: *IMIA Yearbook* 25 (May 2016), S48–S61. DOI: 10.15265/IYS-2016-s006.

[12] Jean-Yves Tanguy et al. "Relevance of Biomedical Data Compression". In: *Compression of Biomedical Images and Signals*. John Wiley & Sons, Ltd, 2008. Chap. 1, pp. 1–14. DOI: 10.1002/9780470611159.ch1.

[13] Mostafa Hanoune and Mezui Eya'a Guy Lysmos. "Data Compression mechanisms in an intelligent E-Health gateway for medical monitoring applications". In: *Procedia Computer Science* 175 (2020), pp. 578–584. DOI: 10.1016/j.procs.2020.07.083.

[14] Nicolas Brodusch et al. "Scanning Electron Microscopy versus Transmission Electron Microscopy for Material Characterization: A Comparative Study on High-Strength Steels". In: *Scanning* 2021 (May 2021), p. 5511618. DOI: 10.1155/2021/5511618.

[15] G. Binnig and H. Rohrer. "Scanning tunneling microscopy". In: *Surface Science* 126.1 (Mar. 1983), pp. 236–244. DOI: 10.1016/0039-6028(83)90716-1.

[16] *Electron Microscopy*. 2023. URL: https://anapath.ch/electron-microscopy-2/.

[17] Michael Schmid. *The Scanning Tunneling Microscope*. TU Wien; adapted from the IAP/TU Wien STM Gallery. 2005. URL: https://www2.iap.tuwien.ac.at/www/surface/stm_gallery/stm_schematic.

[18] Damian J. Matuszewski and Ida-Maria Sintorn. "TEM virus images: Benchmark dataset and deep learning classification". In: *Computer Methods and Programs in Biomedicine* 209 (Sept. 2021), p. 106318. DOI: 10.1016/j.cmpb.2021.106318.

[19] Andrii Iudin et al. "EMPIAR: the Electron Microscopy Public Image Archive". In: *Nucleic Acids Research* 51.D1 (Nov. 2022), pp. D1503–D1511. DOI: 10.1093/nar/gkac1062.

[20] Andrii Iudin et al. "EMPIAR: a public archive for raw electron microscopy image data". In: *Nature Methods* 13 (May 2016), pp. 387–388. DOI: 10.1038/nmeth.3806.

[21] Eleanor Williams et al. “Image Data Resource: A bioimage data integration and publication platform”. In: *Nature Methods* 14 (Aug. 2017), pp. 775–781. DOI: 10.1038/nmeth.4326.

[22] Vincent Casser et al. “Fast Mitochondria Detection for Connectomics”. In: *Proceedings of the Third Conference on Medical Imaging with Deep Learning*. Ed. by Tal Arbel et al. Vol. 121. Proceedings of Machine Learning Research. Amsterdam,Netherlands: PMLR, July 2020, pp. 111–120.

[23] Aurélien Lucchi et al. *Structured Image Segmentation Using Kernelized Features*. Vol. 7573 LNCS. Springer, Berlin, Heidelberg, 2012, pp. 400–413. DOI: 10.1007/978-3-642-33709-3_29.

[24] Narayanan Kasthuri et al. “Saturated Reconstruction of a Volume of Neocortex”. In: *Cell* 162 (3 July 2015), pp. 648–661. DOI: 10.1016/j.cell.2015.06.054.

[25] S. Coughlin et al. “Looking to tomorrow’s healthcare today: a participatory health perspective”. In: *Internal Medicine Journal* 48.1 (Jan. 2018), pp. 92–96. DOI: 10.1111/imj.13661.

[26] Amine Nait-Ali and Christine Cavaro-Menard. *Compression of Biomedical Images and Signals*. 1st ed. Wiley-IEEE Press, 2008. ISBN: 1848210280.

[27] David S. Taubman. “JPEG2000: Image Compression Fundamentals, Standards and Practice”. In: *Journal of Electronic Imaging* apr.2 (Aug. 2002), pp. 286–287. DOI: 10.1117/1.1469618.

[28] *ISO - ISO/IEC* 15444 – 1 : 2000 - *Information technology — JPEG 2000 image coding system — Part 1: Core coding system*. Dec. 2000. URL: https://www.iso.org/standard/27687.html (visited on 08/14/2021).

[29] The Joint Photographic Experts Group. *Jpeg - Jpeg 2000*. Tech. rep. The Joint Photographic Experts Group, 2017. URL: https://jpeg.org/jpeg2000/index.html (visited on 08/17/2021).

[30] M. Adams and R. Ward. “Wavelet transforms in the JPEG-2000 standard”. In: *2001 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (IEEE Cat. No.01CH37233)*. Vol. 1. Victoria, BC, Canada, Aug. 2001, 160–163 vol.1.

[31] Dalila Goudia et al. “Joint Trellis Coded Quantization Watermarking for JPEG2000 images”. In: *Annals of telecommunications* 67 (Dec. 2011), pp. 1–15. DOI: 10.1007/s12243-011-0280-9.

[32] Farzad Zargari and Omid Fatemi. “Error concealment of damaged LL subband in motion JPEG 2000”. In: *Proceedings of the Fifth IEEE International Symposium on Signal Processing and Information Technology* 2005 (Dec. 2005), pp. 606–611. DOI: 10.1109/ISSPIT.2005.1577166.

[33] George Anastassopoulos and Athanassios Skodras. “JPEG2000 ROI coding in medical imaging applications”. In: *Proc. 2nd IASTED International Conference on Visualisation, Imaging and Image Processing (VIIP2002)*. Marbella, Spain, Sept. 2002, pp. 783–788.

[34] Joan Bartrina-Rapesta et al. “JPEG2000 ROI coding method with perfect fine-grain accuracy and lossless recovery”. In: *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*. Pacific Grove, CA, USA, 2009, pp. 558–562. DOI: 10.1109/ACSSC.2009.5469892.

[35] Gary J. Sullivan, Jens Rainer Ohm, et al. “Overview of the high efficiency video coding (HEVC) standard”. In: *IEEE Transactions on Circuits and Systems for Video Technology* 22.12 (Sept. 2012), pp. 1649–1668. DOI: 10.1109/TCSVT.2012.2221191.

[36] K. Wegner et al. “Multi-generation encoding using HEVC All Intra versus JPEG 2000”. In: *2015 57th International Symposium ELMAR (ELMAR)*. Zadar, Croatia, 2015, pp. 41–44. DOI: 10.1109/ELMAR.2015.7334491.

[37] Francisco Cunha et al. “Data compression algorithms for biomedical images”. In: *Imaging Modalities for Biological and Preclinical Research: A Compendium, Volume 2*. 2053-2563. IOP Publishing, May 2021, III.4.f-1 to III.4.f–11. DOI: 10.1088/978-0-7503-3747-2ch29.

[38] Jens-Rainer Ohm et al. “Comparison of the Coding Efficiency of Video Coding Standards—Including High Efficiency Video Coding (HEVC)”. In: *IEEE Transactions on Circuits and Systems for Video Technology* 22.12 (2012), pp. 1669–1684. DOI: 10.1109/TCSVT.2012.2221192.

[39] Joint Video ExplorationTeam. *jvet / HM · GitLab*. (last access 17/08/2021). URL: https://vcgit.hhi.fraunhofer.de/jvet/HM.

[40] Jens Brandenburg and Benno Stabernack. “Simulation-based HW/SW co-exploration of the concurrent execution of HEVC intra encoding algorithms for heterogeneous multi-core architectures”. In: *Journal of Systems Architecture* 77 (June 2017), pp. 26–42. DOI: 10.1016/j.sysarc.2016.12.009.

[41] D. Marpe, H. Schwarz, and T. Wiegand. "Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard". In: *IEEE Transactions on Circuits and Systems for Video Technology* 13.7 (Aug. 2003), pp. 620–636. DOI: 10.1109/TCSVT.2003.815173.

[42] Philippe Hanhart et al. "Subjective evaluation of HEVC intra coding for still image compression". In: *Proceedings of Seventh International Workshop on Video Processing and Quality Metrics for Consumer Electronics*. Scottsdale, Arizona, USA, Jan. 2013.

[43] Saurin S. Parikh et al. "High Bit-Depth Medical Image Compression With HEVC". In: *IEEE Journal of Biomedical and Health Informatics* 22 (Mar. 2018), pp. 552–560. DOI: 10.1109/JBHI.2017.2660482.

[44] Benjamin Bross et al. "Overview of the Versatile Video Coding (VVC) Standard and its Applications". In: *IEEE Transactions on Circuits and Systems for Video Technology* 31.10 (Oct. 2021), pp. 3736–3764. DOI: 10.1109/TCSVT.2021.3101953.

[45] ISO/IEC JTC 1/SC 29. *Information technology — Coded representation of immersive media — Part 3: Versatile video coding*. Tech. rep. International Standardization Organization, Feb. 2021, p. 504. URL: https://www.iso.org/standard/73022.html (visited on 08/13/2021).

[46] Joint Video ExplorationTeam. *jvet / VVCSoftware_VTM · GitLab*. (last access 22/08/2021). URL: https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM (visited on 08/22/2021).

[47] Benjamin Bross. "Versatile Video Coding (VVC)". In: *ITU Workshop on "The future of media"*. Geneva, Switzerland, Oct. 2019.

[48] Xiaozhong Xu and Shan Liu. "Recent advances in video coding beyond the HEVC standard". In: *APSIPA Transactions on Signal and Information Processing* 8 (June 2019), e18. DOI: 10.1017/ATSIP.2019.11.

[49] Mai Xu et al. "Region-of-Interest Based Conversational HEVC Coding with Hierarchical Perception Model of Face". In: *IEEE Journal of Selected Topics in Signal Processing* 8.3 (June 2014), pp. 475–489. DOI: 10.1109/JSTSP.2014.2314864.

[50] Yueying Wu et al. "Medical Ultrasound Video Coding with H.265/HEVC Based on ROI Extraction". In: *PLOS ONE* 11.11 (Nov. 2016), pp. 1–13. DOI: 10.1371/journal.pone.0165698.

[51] Sun Xuebin et al. “Content-aware Rate Control Scheme for HEVC Based on Static and Dynamic Saliency Detection”. In: *Neurocomputing* 411 (June 2020). DOI: 10.1016/j.neucom.2020.06.003.

[52] Huaying Xue et al. “Fast ROI-based HEVC coding for surveillance videos”. In: *2016 19th International Symposium on Wireless Personal Multimedia Communications (WPMC)*. Shenzhen, China, 2016, pp. 299–304.

[53] Ling Yang et al. “A ROI quality adjustable rate control scheme for low bitrate video coding”. In: *2009 Picture Coding Symposium*. Chicago, Illinois, USA, 2009, pp. 1–4. DOI: 10.1109/PCS.2009.5167372.

[54] Xiaoyi He et al. “Enhancing HEVC Compressed Videos with a Partition-Masked Convolutional Neural Network”. In: *2018 25th IEEE International Conference on Image Processing (ICIP)*. Athens, Greece: IEEE, Oct. 2018. DOI: 10.1109/icip.2018.8451086.

[55] Jean-Pierre Henot et al. “High efficiency video coding (HEVC): Replacing or complementing existing compression standards?” In: *2013 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. London, UK, 2013, pp. 1–6. DOI: 10.1109/BMSB.2013.6621675.

[56] Charilaos Christopoulos. “The jpeg2000 still image coding system: an overview”. In: *IEEE Transactions on Consumer Electronics* 46.4 (Nov. 2000), pp. 1103–1127. DOI: 10.1109/30.920468.

[57] E. G. C. Arapa et al. “Métodos de codificação ROI para compressão de imagens DICOM”. In: *Ensaios e Ciência: Ciências Agrárias, Biológicas e da Saúde*. Vol. 15. 6. Anhanguera Educacional Ltda, Sept. 2011, pp. 105–120.

[58] S. Urvashi et al. “Region of interest based selective coding technique for volumetric MR image sequence”. In: *Multimedia Tools and Applications* 80 (Mar. 2021), pp. 1–23. DOI: 10.1007/s11042-020-10396-5.

[59] Bindu P.V. and Jabeena Afthab. “Region of Interest Based Medical Image Compression Using DCT and Capsule Autoencoder for Telemedicine Applications”. In: *2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*. Erode, India, Sept. 2021, pp. 1–7. DOI: 10.1109/ICECCT52121.2021.9616748.

[60] Vinayak Bairagi and Ashokm Sapkal. “ROI-based DICOM image compression for telemedicine”. In: *Sadhana* (Feb. 2013), pp. 1–9. DOI: 10.1007/s12046-013-0126-4.

[61] Sin Ting Lim and Nurulfajar Bin Abd Manap. "A Region-based Compression Technique for Medical Image Compression using Principal Component Analysis (PCA)". In: *International Journal of Advanced Computer Science and Applications* 13.2 (Jan. 2022). DOI: 10.14569/IJACSA.2022.0130229.

[62] Cheolsoo Park et al. "Machine learning in biomedical engineering". In: *Biomedical Engineering Letters* 8 (1 Feb. 2018), pp. 1–3. DOI: 10.1007/s13534-018-0058-3.

[63] H. Bahurel-Barrera et al. "Inter- and intra-observer variability in detection and progression assessment with MRI of microadenoma in Cushing's disease patients followed up after bilateral adrenalectomy". In: *Pituitary* 11.3 (Sept. 2008), pp. 263–269. DOI: doi:10.1007/s11102-008-0123-5.

[64] Ashleigh Storr et al. "Inter-observer and intra-observer agreement between embryologists during selection of a single Day 5 embryo for transfer: a multicenter study". In: *Human Reproduction* 32.2 (Jan. 2017), pp. 307–314. DOI: 10.1093/humrep/dew330.

[65] Riitta Salonen et al. "Measurement of intima-media thickness of common carotid arteries with high-resolution B-mode ultrasonography: Inter- and intra-observer variability". In: *Ultrasound in Medicine & Biology* 17.3 (1991), pp. 225–230. DOI: 10.1016/0301-5629(91)90043-v.

[66] Martijn P.A. Starmans et al. "Chapter 18 - Radiomics: Data mining using quantitative medical image features". In: *Handbook of Medical Image Computing and Computer Assisted Intervention*. Ed. by S. Kevin Zhou, Daniel Rueckert, and Gabor Fichtinger. The Elsevier and MICCAI Society Book Series. Academic Press, 2020, pp. 429–456. DOI: 10.1016/B978-0-12-816176-0.00023-5.

[67] F. Cunha et al. "Lossy Image Compression in a Preclinical Multimodal Imaging Study". In: *Journal of Digital Imaging* 36.4 (Apr. 2023), pp. 1826–1850. ISSN: 1618-727X. DOI: 10.1007/s10278-023-00800-5.

[68] Yan Zhuang et al. "AutoCellANLS: An Automated Analysis System for Mycobacteria-Infected Cells Based on Unstained Micrograph". In: *Biomolecules* 12 (Feb. 2022), p. 240. DOI: 10.3390/biom12020240.

[69] Saad Albawi et al. "Understanding of a convolutional neural network". In: *2017 International Conference on Engineering and Technology (ICET)*. Antalya, Turkey, Aug. 2017, pp. 1–6. DOI: 10.1109/ICEngTechnol.2017.8308186.

[70] Shujian Yu et al. “Understanding Convolutional Neural Networks With Information Theory: An Initial Exploration”. In: *IEEE Transactions on Neural Networks and Learning Systems* 32.1 (Jan. 2021), pp. 435–442. DOI: 10.1109/TNNLS.2020.2968509.

[71] Jo-Hsuan Wu et al. “Performance and Limitation of Machine Learning Algorithms for Diabetic Retinopathy Screening: Meta-analysis”. In: *Journal of medical Internet research* 23.7 (July 2021), e23863. DOI: 10.2196/23863.

[72] Varun Gulshan et al. “Performance of a Deep-Learning Algorithm vs Manual Grading for Detecting Diabetic Retinopathy in India”. In: *JAMA Ophthalmology* 137.9 (Sept. 2019), pp. 987–993. DOI: 10.1001/jamaophthalmol.2019.2004.

[73] Ross Girshick et al. “Rich feature hierarchies for accurate object detection and semantic segmentation”. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Nov. 2013), pp. 580–587. DOI: 10.1109/CVPR.2014.81.

[74] Kaiming He et al. *Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition*. Vol. 8691 LNCS. Springer Verlag, June 2014, pp. 346–361. DOI: 10.1007/978-3-319-10578-9_23.

[75] Ross Girshick. “Fast R-CNN”. In: *2015 IEEE International Conference on Computer Vision (ICCV)*. Santiago, Chile, Dec. 2015, pp. 1440–1448. DOI: 10.1109/ICCV.2015.169.

[76] Tsung-Yi Lin et al. “Focal Loss for Dense Object Detection”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42 (Aug. 2017), pp. 318–327. DOI: 10.1109/TPAMI.2018.2858826.

[77] Joseph Redmon et al. “You only look once: Unified, real-time object detection”. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (June 2016), pp. 779–788. DOI: 10.1109/CVPR.2016.91.

[78] Alexey Bochkovskiy et al. “YOLOv4: Optimal Speed and Accuracy of Object Detection”. In: *CoRR* abs/2004.10934 (Apr. 2020). DOI: 10.48550/arXiv.2004.10934.

[79] Kaiming He et al. “Mask R-CNN.” In: *IEEE transactions on pattern analysis and machine intelligence* 42 (2 Mar. 2020), pp. 386–397. DOI: 10.1109/TPAMI.2018.2844175.

[80] Yuxin Wu et al. *Detectron2*. (last access 16/02/2024). 2019. URL: https://github.com/facebookresearch/detectron2.

[81] Alexander Kirillov et al. “Segment Anything”. In: *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. Paris, France, Nov. 2023, pp. 3992–4003. DOI: 10.1109/ICCV51070.2023.00371.

[82] Halil Murat Ünver and Enes Ayan. “Skin Lesion Segmentation in Dermoscopic Images with Combination of YOLO and GrabCut Algorithm”. In: *Diagnostics* 9 (July 2019), p. 72. DOI: 10.3390/diagnostics9030072.

[83] Xiangqiong Wu et al. “CacheTrack-YOLO: Real-Time Detection and Tracking for Thyroid Nodules and Surrounding Tissues in Ultrasound Videos”. In: *IEEE Journal of Biomedical and Health Informatics* 25 (Oct. 2021), pp. 3812–3823. DOI: 10.1109/JBHI.2021.3084962.

[84] Sindhu Ramachandran S. et al. “Using YOLO based deep learning network for real time detection and localization of lung nodules from low dose CT scans”. In: *Medical Imaging 2018: Computer-Aided Diagnosis*. Ed. by Nicholas Petrick and Kensaku Mori. Vol. 10575. International Society for Optics and Photonics. Houston, Texas, USA: SPIE, Feb. 2018, pp. 347–355. DOI: 10.1117/12.2293699.

[85] Shanchen Pang et al. “A novel YOLOv3-arch model for identifying cholelithiasis and classifying gallstones on CT images.” In: *PloS one* 14 (6 2019), e0217647. DOI: 10.1371/journal.pone.0217647.

[86] Asmaa Haja et al. *A Fully Automated End-to-End Process for Fluorescence Microscopy Images of Yeast Cells: From Segmentation to Detection and Classification*. Vol. 784 LNEE. Springer Science and Business Media Deutschland GmbH, Apr. 2022, pp. 37–46. DOI: 10.1007/978-981-16-3880-0_5.

[87] Gabriel Alves. *Detecção de Objetos com YOLO – Uma abordagem moderna – IA Expert Academy*. (last access 06/09/2021). Oct. 2020. URL: https://iaexpert.academy/2020/10/13/deteccao-de-objetos-com-yolo-uma-abordagem-moderna/ (visited on 08/27/2021).

[88] Van Vung Pham and Tommy Dang. Packt Publishing, 2023. ISBN: 1800561628.

[89] Justin Johnson et al. *Deep Learning for Computer Vision — Lecture 15*. (last access 26/03/2024). May 2017. URL: http://cs231n.stanford.edu/slides/2017/.

[90] Tao Han et al. “Internet of Medical Things-based Deep Learning Techniques for Segmentation of Lung and Stroke Regions in CT Scans”. In: *IEEE Access* PP (Apr. 2020), pp. 1–1. DOI: 10.1109/ACCESS.2020.2987932.

[91] Alexey Bochkovskiy. *YOLOv4.weights*. URL: https://github.com/AlexeyAB/darknet/ (visited on 05/04/2021).

[92] Tsung-Yi Lin et al. “Microsoft COCO: Common Objects in Context”. In: *Computer Vision – ECCV* 2014. Ed. by David Fleet et al. Vol. 8693. Springer International Publishing, Apr. 2014, pp. 740–755. DOI: 10.1007/978-3-319-10602-1_48.

[93] ISO/IEC JTC 1/SC 29. *ISO - ISO/IEC 23008-5:2017/Amd 1:2017 - Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 5: Reference software for high efficiency video coding — Amendment 1: Reference software for screen content coding extensions.* Tech. rep. International Standardization Organization, 2017. URL: https://www.iso.org/standard/72289.html (visited on 08/17/2021).

[94] David Flynn et al. “Overview of the Range Extensions for the HEVC Standard: Tools, Profiles, and Performance”. In: *IEEE Transactions on Circuits and Systems for Video Technology* 26.1 (Jan. 2016), pp. 4–19. DOI: 10.1109/TCSVT.2015.2478707.

D E C L A R A Ç Ã O

Declaro, sob compromisso de honra, que o trabalho apresentado nesta dissertação, com o título *" Lossy compression of biomedical images for computer vision analysis "*, é original e foi realizado por Edgar da Silva Paulo (2212421) sob orientação do Professor Sérgio Manuel Maciel de Faria, Professor Luís Miguel de Oliveira Pegado de Noronha e Távora, e Professor Lucas Arrabal Thomaz.

Edgar da Silva Paulo