Uma Abordagem para Segmentação de Regiões Tumorais da Cavidade Oral em Imagens Histológicas de Lâmina Inteira Utilizando Redes Neurais Completamente Convolucionais

Dalí Freire Dias dos Santos



Universidade Federal de Uberlândia
Faculdade de Computação
Programa de Pós-Graduação em Ciência da Computação

Uberlândia
2022

Dalí Freire Dias dos Santos

Uma Abordagem para Segmentação de Regiões Tumorais da Cavidade Oral em Imagens Histológicas de Lâmina Inteira Utilizando Redes Neurais Completamente Convolucionais

Tese de doutorado apresentada ao Programa de Pós-graduação da Faculdade de Computação da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Doutor em Ciência da Computação.

Área de concentração: Ciência da Computação

Orientador: Marcelo Zanchetta do Nascimento
Coorientador: Bruno Augusto Nassif Travençolo

Uberlândia
2022

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

S237 Santos, Dali Freire Dias dos, 1985-
2022 Automated segmentation of tumor regions from oral histological whole slide images using fully convolutional neural networks [recurso eletrônico] / Dali Freire Dias dos Santos. - 2022.

Orientador: Marcelo Zanchetta do Nascimento.
Coorientador: Bruno Augusto Nassif Travençolo.
Tese (Doutorado) - Universidade Federal de Uberlândia, Pós-graduação em Ciência da Computação.
Modo de acesso: Internet.
Disponível em: http://doi.org/10.14393/ufu.te.2022.252
Inclui bibliografia.

1. Computação. I. Nascimento, Marcelo Zanchetta do, 1976-, (Orient.). II. Travençolo, Bruno Augusto Nassif, 1981-, (Coorient.). III. Universidade Federal de Uberlândia. Pós-graduação em Ciência da Computação. IV. Título.

CDU: 681.3

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:
Gizele Cristine Nunes do Couto - CRB6/2091
Nelson Marcos Ferreira - CRB6/3074

Agradecimentos

Agradeço à Universidade Federal de Uberlândia (UFU) e à Faculdade de Computação (FACOM) por terem me proporcionado amplo acesso aos laboratórios, espaço físico, equipamentos, suporte administrativo e todo apoio acadêmico necessário para o desenvolvimento deste trabalho. Em especial, agradeço aos professores Marcelo Zanchetta do Nascimento (CNPq #311404/2021-9, FAPEMIG #APQ-00578-18), Bruno Augusto Nassif Travençolo e Paulo Rogério de Faria pela orientação, parceria, ensinamentos e suporte durante todo esse período de doutorado.

Resumo

A segmentação de regiões tumorais em lâminas coradas com hematoxilina e eosina (H&E) é uma tarefa importante para os patologistas no diagnóstico de diferentes tipos de câncer, incluindo câncer derivado da cavidade oral (do inglês, *oral squamous cell carcinoma* - OSCC). Imagens histológicas de toda a lâmina (do inglês, *whole slide image* - WSI) são imagens de alta resolução que podem conter células, estruturas celulares e alguns artefatos derivados do processo de aquisição de imagem. Processar essas imagens enormes para descartar áreas não relevantes, mantendo apenas as regiões clinicamente relevantes, é uma tarefa desafiadora. Nesse trabalho é apresentado uma nova abordagem, baseada em redes neurais completamente convolucionais (do inglês, *fully convolutional network* - FCN), capaz de analisar automaticamente uma WSI para segmentar regiões de tumores da cavidade oral com o objetivo de auxiliar a tomada de decisão do patologista sobre diagnósticos e prognósticos de OSCC. A metodologia proposta utiliza características de cor no espaço de cores HSV para identificar as regiões de tecido em uma etapa de pré-processamento. As regiões do tecido identificadas na WSI são transformadas para o espaço de cores CIE L*a*b* e divididas em subimagens menores. Uma FCN baseada no modelo U-Net, treinada com uma nova estratégia de aumento de dados, é utilizada para localizar e segmentar regiões tumorais a partir das subimagens. Para avaliar a metodologia proposta, uma nova base de imagens com WSIs de tecidos de OSCC corados com H&E foi criada. O método foi testado nesta e em outras bases de imagens disponíveis na literatura, incluindo uma base de WSIs de outro tipo de câncer, alcançando 90% de F1-*score* na nova base de imagens criada. Para as demais bases de imagens de domínio público o método obteve uma média de 83% de F1-*score*, mostrando-se robusto e com o potencial para criar uma ferramenta de apoio ao diagnóstico em casos de OSCC.

Palavras-chave: Redes Neurais Completamente Convolucionais. Aumento de Dados. Análise de Imagens Histológicas. Segmentação de Imagens. Câncer da Cavidade Oral.

Automated Segmentation of Tumor Regions from Oral Histological Whole Slide Images Using Fully Convolutional Neural Networks

Dalí Freire Dias dos Santos



Universidade Federal de Uberlândia
Faculdade de Computação
Programa de Pós-Graduação em Ciência da Computação

Uberlândia
2022

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
Coordenação do Programa de Pós-Graduação em Ciência da Computação
Av. João Naves de Ávila, nº 2121, Bloco 1A, Sala 243 - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902
Telefone: (34) 3239-4470 - www.ppgco.facom.ufu.br - cpgfacom@ufu.br

ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Ciência da Computação				
Defesa de:	Tese, 8/2022, PPGCO				
Data:	4 de maio de 2022	Hora de início:	08:30	Hora de encerramento:	12:50
Matrícula do Discente:	11723CCP003				
Nome do Discente:	Dalí Freire Dias dos Santos				
Título do Trabalho:	Automated Segmentation of Tumor Regions from Oral Histological Whole Slide Images Using Fully Convolutional Neural Networks				
Área de concentração:	Ciência da Computação				
Linha de pesquisa:	Ciência de Dados				
Projeto de Pesquisa de vinculação:	-				

Reuniu-se, por videoconferência, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Ciência da Computação, assim composta: Professores Doutores: André Ricardo Backes - FACOM/UFU, Maurício Cunha Escarpinati - FACOM/UFU, Marcelo Costa Oliveira - IC/UFAL, Ricardo José Ferrari - DC/UFSCAR, Bruno Augusto Nassif Travençolo - FACOM/UFU (Coorientador) e Marcelo Zanchetta do Nascimento - FACOM/UFU orientador do candidato.

Os examinadores participaram desde as seguintes localidades: Marcelo Costa Oliveira - Maceió/AL; Ricardo José Ferrari - São Carlos/SP; Bruno Augusto Nassif Travençolo, André Ricardo Backes, Maurício Cunha Escarpinati e Marcelo Zanchetta do Nascimento - Uberlândia/MG. O discente participou da cidade de Uberlândia/MG.

Iniciando os trabalhos o presidente da mesa, Prof. Dr. Marcelo Zanchetta do Nascimento, apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação do Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir o candidato. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o candidato:

Aprovado.

Esta defesa faz parte dos requisitos necessários à obtenção do título de Doutor.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.

Documento assinado eletronicamente por **Marcelo Zanchetta do Nascimento**, **Professor(a) do Magistério Superior**, em 06/05/2022, às 10:41, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Ricardo José Ferrari**, **Usuário Externo**, em 06/05/2022, às 10:46, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Bruno Augusto Nassif Travençolo**, **Professor(a) do Magistério Superior**, em 06/05/2022, às 10:55, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **André Ricardo Backes**, **Professor(a) do Magistério Superior**, em 06/05/2022, às 14:54, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Marcelo Costa Oliveira**, **Usuário Externo**, em 08/05/2022, às 13:46, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Mauricio Cunha Escarpinati**, **Professor(a) do Magistério Superior**, em 16/05/2022, às 12:02, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3574708** e o código CRC **5F8B99F3**.

Abstract

Segmentation of tumor regions in H&E-stained slides is an important task for pathologists while diagnosing different types of cancer, including oral squamous cell carcinoma (OSCC). Whole slide histological images (WSI) are high-resolution images created from tissue sections that can contain many different types of cells, regions, and some artifacts derived from the image acquisition process. Processing these huge images to discard non-relevant areas while keeping only the clinically relevant regions is a challenging task. This work presents a new approach, based on fully convolutional neural networks (FCN), capable of automatically analyzing a WSI to segment regions of oral cavity tumors to aid and enhance the pathologist's decision making on OSCC diagnoses and prognosis. The proposed methodology uses color characteristics in the HSV color space to identify tissue regions in a pre-processing step. The tissue regions identified in the WSI are transformed to the CIE L*a*b* color space and split into smaller sub-images. An FCN segmentation model, based on the U-Net model and trained using a new data augmentation strategy, is used to locate and segment tumor regions from the sub-images. The evaluation of the proposal was tested using public domain image datasets, and a newly created WSI dataset of H&E-stained OSCC tissues. The results have shown a segmentation F1-score of 97% in the new OSCC dataset created. For the public domain image datasets, including a dataset of WSIs from another type of cancer, the proposed method obtained an average F1-score of 83%, proving to be robust and with the potential to develop a tool to support the diagnosis in cases of OSCC.

Keywords: Fully Convolutional Neural Networks. Data Augmentation. H&E-stained Histological Images Analysis. Image Segmentation. Oral Cavity-Derived Cancer.

List of Figures

Figure 1 – A oral squamous cell carcinoma whole slide image. 15
Figure 2 – Tissue changes can occur before cancer cells are formed. 19
Figure 3 – Anatomy of the oral cavity. 19
Figure 4 – Main steps for the histological analysis process. 20
Figure 5 – Histological tissue preparation. 21
Figure 6 – Artifacts that can lower the quality of histological images. 22
Figure 7 – Digital image representation. 23
Figure 8 – Segmentation of a natural image. 24
Figure 9 – Graphic representation for commonly used activation functions. 27
Figure 10 – Sample pooling operations. 27
Figure 11 – Basic operations in the encoding phase of typical CNNs. 28
Figure 12 – Basic operations in the decoding phase of typical FCNs. 29
Figure 13 – Typical convolutional architectures for image segmentation task. 30
Figure 14 – The architecture of the LeNet model. 31
Figure 15 – The architecture of the AlexNet model. 31
Figure 16 – Original U-Net architecture. 32
Figure 17 – Identification of keratinization and keratin pearl area. 34
Figure 18 – Region of tumor-infiltrating lymphocytes. 34
Figure 19 – WSI segmentation results by Halicek et al. (2019). 35
Figure 20 – Segmentation results by Martino et al. (2020). 36
Figure 21 – Overview of our proposed method. 39
Figure 22 – The labeling process for the OCDC dataset. 41
Figure 23 – Six WSIs from OCDC dataset. 42
Figure 24 – Some image-patches (640×640 pixels) from the OCDC dataset. 43
Figure 25 – Image-patches extracted from the CAMELYON16 dataset. 44
Figure 26 – Sample WSI from the CAMELYON16 dataset. 45
Figure 27 – Images from the ORCA dataset. 46
Figure 28 – H&E-stained histological images of mice oral epithelial tissues. 47

Figure 29 – ROI images from healthy and epithelial dysplasia-harboring tongue mice. 48
Figure 30 – Random Composition Augmentation - RCAug 50
Figure 31 – Augmentation transformations that do not need to be applied in the corresponding label images. 51
Figure 32 – Augmentation transformations applied to both the images and their corresponding labels equally. 52
Figure 33 – Proposed color-based tissue filter 53
Figure 34 – The tissue regions identified by our proposed HSV tissue filter. 54
Figure 35 – FCN architecture used for OSCC segmentation. 56
Figure 36 – FCN architecture used for nuclei segmentation. 58
Figure 37 – General diagram of the performed experiments. 60
Figure 38 – Accuracy and loss learning curves for the OCDC dataset. 65
Figure 39 – Accuracy and loss learning curves for the ORCA dataset. 66
Figure 40 – Accuracy and loss learning curves for the ORCA 512×512 dataset. . . . 67
Figure 41 – One of the WSIs from our OCDC dataset and three corresponding image-patches from the tumor areas. 70
Figure 42 – One of the images from the ORCA dataset and three corresponding image-patches extracted from the tumor areas. 71
Figure 43 – Qualitative segmentation results for the ORCA 512×512 dataset. . . . 72
Figure 44 – One of the WSIs from the CAMELYON16 dataset and three corresponding image-patches extracted from the tumor areas. 73
Figure 45 – Qualitative image-patch segmentation results for OCDC dataset. 74
Figure 46 – Qualitative results for automated nuclei segmentation on oral histological images 76
Figure 47 – Qualitative analysis indicating the nuclei masks obtained by different segmentation methods. 77
Figure 48 – ROI images applied to different color normalization methods. 78
Figure 49 – Accuracy results by nuclei segmentation and color normalization methods. 80
Figure 50 – Gold-standard nuclei segmentation. 81
Figure 51 – Qualitative analysis for RGB color space segmentation. 82
Figure 52 – Qualitative analysis for HSV color space segmentation. 83
Figure 53 – Qualitative analysis for CIE L*a*b* color space segmentation 84

List of Tables

Table 1 – Summary of the image datasets used for the training and test of the proposed method. 40
Table 2 – Average quantitative results for color spaces and image-patch sizes. . . . 63
Table 3 – Average quantitative results for test images from the OCDC dataset. . . 64
Table 4 – Average quantitative results for test images from the ORCA dataset. . . 64
Table 5 – Average quantitative results for test images from the ORCA 512×512 dataset. 64
Table 6 – Average quantitative results for test image-patches. 68
Table 7 – The average quantitative results by color normalization and nuclei segmentation methods applied to test set images. 79
Table 8 – The average quantitative results by each color space transformation applied on the test set. 85

Acronyms list

CAD computer-aided diagnosis

CNN convolutional neural network

CIE International Commission on Illumination

CAMELYON16 CAncer MEtastases in LYmph nOdes challeNge

FCN fully convolutional network

FCL fully connected layer

H&E hematoxylin and eosin

HSV hue, saturation, and value/brightness

INCA *Instituto Nacional do Câncer*

IOU intersection-over-union

ILSVRC ImageNet Large Scale Visual Recognition Challenge

MRI magnetic resonance image

OSCC oral squamous cell carcinoma

OCDC Oral Cavity-Derived Cancer

ORCA ORal Cancer Annotated

ONHI Oral Nuclei Histological Images

RCAug random composition augmentation

ROI region of interest

RGB red, green, and blue

ReLU rectified linear unit

tanh hyperbolic tangent

TIL tumour infiltrating lymphocytes

TMA tissue microarray

TCGA The Cancer Genome Atlas

WHO World Health Organization

WSI whole slide image

Contents

1 INTRODUCTION . . . 12
1.1 Motivation . . . 14
1.2 Research Hypotheses . . . 16
1.3 Goals . . . 16
1.4 Contributions . . . 17
1.5 Thesis Organization . . . 17

2 CONTEXTUALIZATION . . . 18
2.1 Oral Cavity-derived Cancer . . . 18
2.2 Histological Image Analysis . . . 20
2.3 Digital Image Processing . . . 22
2.4 Convolutional Neural Networks . . . 25
2.4.1 Popular Convolutional Models . . . 30
2.5 Related Works . . . 33
2.6 Final Considerations . . . 37

3 PROPOSAL . . . 38
3.1 Method Overview . . . 38
3.2 Image Datasets . . . 40
3.2.1 OCDC dataset . . . 40
3.2.2 CAMELYON16 dataset . . . 44
3.2.3 ORCA dataset . . . 45
3.2.4 ONHI dataset . . . 47
3.3 Dataset Augmentation . . . 48
3.3.1 Data augmentation transformations . . . 50
3.4 Tissue Filter . . . 53
3.5 Histological Image Segmentation . . . 55
3.5.1 Fully Convolutional Network Architecture . . . 55

3.5.2 Nuclei Segmentation on Oral Histological Images 57
3.6 Final Considerations 58

4 EXPERIMENTAL RESULTS 60
4.1 Evaluation Metrics 61
4.2 Experiments for Automated Detection and Segmentation of Tumor Regions 62
4.2.1 Color spaces and image-patch sizes 62
4.2.2 Evaluation of data augmentation strategies 63
4.2.3 External validation 68
4.3 Experiments for Automated Nuclei Segmentation on Oral Histological Images 75
4.4 Final Considerations 86

5 CONCLUSION 87
5.1 Conclusions 87
5.2 List of Publications 88
5.3 Future Works 89

BIBLIOGRAPHY 90

I hereby certify that I have obtained all legal permissions from the owner(s) of each third-party copyrighted matter included in my thesis, and that their permissions allow availability such as being deposited in public digital libraries.

Dalí Freire Dias dos Santos

Chapter 1

Introduction

Cancer can be defined as a group of non-communicable diseases that develop almost anywhere in the human body. The disease is caused when cells begin to divide uncontrollably with the potential to invade other parts of the body and/or spread to other organs and surrounding tissues (NCI, 2022). According to the World Health Organization (WHO), cancer is the second leading cause of death in the world, accounting for 18.1 million cases and 9.6 million deaths in 2018 (WHO, 2022). Cancer cases continue to grow globally and by 2030 the WHO estimates 24.1 million new cancer cases worldwide (BRAY et al., 2018). In Brazil, projections from the *Instituto Nacional do Câncer* (INCA) show that 626,000 cases of cancer were registered in 2020, including 15,000 new cases of oral cavity-derived cancer (INCA, 2022).

Head and neck cancers are the sixth leading group of cancers worldwide. It encompasses a large group of tumors including those from the oral cavity (MAHMOOD et al., 2020). Lesions of the oral cavity can be grouped into two broad categories: benign and malignant lesions. Benign lesions are not considered cancer, as they do not invade other tissues and do not spread to other parts of the body. Precancerous conditions, on the other hand, are still harmless lesions, but they can evolve to cancer over time. The majority of oral cavity-derived cancers are oral squamous cell carcinoma (OSCC) (BOXBERG et al., 2019) and the main risk factors for these malignant tumors are the use of tobacco, including smokeless tobacco, infection with human papillomavirus, and excessive alcohol consumption (HASHIBE et al., 2007; BOFFETTA et al., 2008; CHATURVEDI et al., 2011).

Despite advancements in medical and surgical techniques, the prognosis of head and neck cancers remains poor with the overall 5-year relative survival rates less than 70% (PULTE; BRENNER, 2010). The primary management of tumors from the oral cavity involves surgical resection with safe negative cancer margins (BADDOUR; MAGLIOCCA; CHEN, 2015). Compared with other tumor locations, the treatment of OSCC patients is highly associated with comorbidities including speech problems, oral pain, malnutrition, sexual problems, dysphagia, and loss of appetite, hence leading to poor health-related

quality of life for them (LIAO et al., 2019).

Although the well-established importance of early diagnosis in cancer, the disease continues to be late-diagnosed in most cases, as the diagnosis is an expensive, highly skilled, and time-consuming process. In the traditional cancer diagnosis, after clinical examinations, histological analysis is still one of the most important elements in this process. The histological analysis begins with sample tissue extractions (biopsy) to examine and determine the presence of the disease analyzing a wide variety of features under light microscopic (JUNQUEIRA; CARNEIRO, 2013). To observe visual features under light microscopes, different components of the tissue are stained with different dyes. The hematoxylin and eosin (H&E) stain is one of the most widely used protocol. Hematoxylin stains nuclei in dark blue color, while eosin stains other cell structures with a pink color (KLECZEK; JAWOREK-KORJAKOWSKA; GORGON, 2020). Despite being a gold-standard method and even existing a consensus in observing some characteristics, there is a certain degree of subjectivity in the pathologist analysis, causing some variations in the evaluation among different pathologists and the evaluations of the same case by the same pathologist on different occasions – inter-and intra-observer variability (WARNAKULASURIYA et al., 2008).

Valuable data with diagnostic and prognostic implications can be obtained through the analysis under light microscopic of cellular characteristics and tissue components (ANDREA; BLEGGI-TORRES; ALVES, 2008). Morphological changes in the size and shape of tumor cell nuclei are some of the basic features used to diagnose any kind of cancer (BELL; LAMMERDING, 2016). The presence of keratinization and keratin pearls are also investigated by pathologists to determine OSCC grading (DAS et al., 2015). Another characteristic widely used in the prognosis of the disease is the abundance of tumor-infiltrating lymphocytes (SHABAN et al., 2019). This histological aspect in malignant cells is recognizably different from those observed in non-malignant cells.

The use of computer-aided diagnosis (CAD) can be applied as a key tool to aid and enhance the pathologists' decisions, minimizing human interventions, discovering measurable and traceable clinical information, providing reliable prognostic factors, and reducing inter and intra-observer variability. Fast scanners can be used to capture tissue samples into digital images to obtain the so-called digital histological images, allowing to view tissue samples on a computer rather than through a microscope. Particularly, deep learning techniques, such as convolutional neural network (CNN) and fully convolutional network (FCN), have been successfully applied in medical and biological research (XING et al., 2018). Since CNNs has exceeded the state-of-the-art in many visual recognition tasks (KRIZHEVSKY; SUTSKEVER; HINTON, 2012; LONG; SHELHAMER; DARRELL, 2015; XING et al., 2018), it becomes of great importance to perform further investigations on the application of deep learning techniques in problems as identification of clinically tumor relevant regions, tumor region features extraction, nuclei segmenta-

tion, and cancer grade classification problems in several scenarios. Studies in these fields of knowledge can have a great impact on the improvement of diagnosis, providing more assertive prognoses and better treatment planning for patients around the world.

1.1 Motivation

The analysis of histological slides, i.e., the study of tissue samples of affected regions under the microscope, remains the most reliable way for diagnosing and grade OSCC (BAIK et al., 2014; DAS et al., 2015). This method of microscopic investigation, despite being referred to as gold-standard in cancer diagnosis, is influenced by inter-and intra-observer variability and it is a time-consuming task that requires well-trained and experienced pathologists (BOXBERG et al., 2019).

Digitization of histological slides in the form of whole slide image (WSI) has made it possible to automate tasks that are prohibitively time-consuming for humans. For instance, precisely quantify vast amounts of information about morphological and spatial architectural patterns in tissues within a tumor to discover new disease markers (MCCANN et al., 2015). CAD tools can have a substantial impact on clinical practices, providing repeatable and more precise methods to automatically identify regions of interest (ROIs) on WSIs obtained from OSCC samples. Segment regions of interest to afterward extract and quantify features from those ROIs can substantially reduce the inter-and intra-observer variability and enhance the efficiency and accuracy of pathologist decisions.

WSIs are multi-gigapixel images created from tissue sections that can contain many different types of cells and regions – such as keratin, lymphocytes, blood vessels, glands, muscle, and tumor cells. WSIs may also contain some derived artifacts from the image acquisition process, like tissue folding and blurring, that need to be ignored (SHABAN et al., 2019). Processing billions of pixels from these huge images to discard non-relevant areas, keeping only clinically relevant regions, is a non-trivial problem (GUO et al., 2019). The vast majority of the existing approaches bypass this non-trivial problem by requiring a specialist to previously select accurate small ROI areas from the larger images. Another frequent challenge is the scarcity of image datasets to study different types and tissue regions where OSCC can arise. A single H&E-stained histological WSI, as shown in Figure 1, may contain thousands of nuclei and other cell structures that should be analyzed by pathologists at diagnostic time. This large amount of information requires very skilled professionals and a great effort to label the relevant structures to create these image datasets.

Segmentation of tumor regions in H&E-stained histological slides is an important task for a pathologist while diagnosing OSCC. The traditional feature engineering proposals for segmentation in medical images are based on clustering (FILIPCZUK; KOWAL; OBUCHOWICZ, 2011), morphological operations (HEIJMANS, 1995), watersheds (NAJMAN;

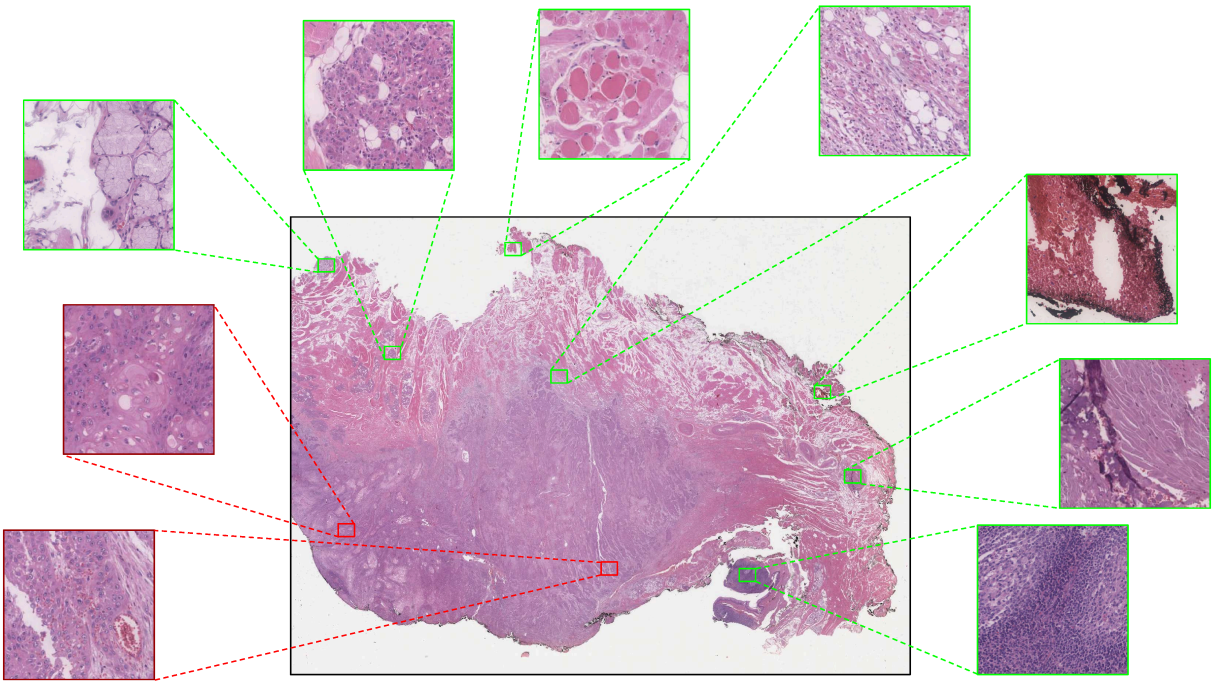


Figure 1 – A oral squamous cell carcinoma whole slide image: some image-patches of size 640×640 pixels are highlighted to show different types of cells and regions that can be present in a single WSI. Green image-patches represent normal areas while red image-patches represent tumor areas. The image has a dimension of 53,783×39,990 pixels and the file size is 483.5 MB.

SCHMITT, 1994), graph-based methods (DAKUA; SAHAMBI, 2011a; DAKUA; SAHAMBI, 2011b; DAKUA; ABI-NAHED, 2013), and thresholding techniques (PHOULADY et al., 2016). These methods have successfully been employed in different types of medical images, like X-ray or magnetic resonance images (MRIs) (DAKUA; ABI-NAHED; AL-ANSARI, 2016; DAKUA; ABI-NAHED; AL-ANSARI, 2018). However, these traditional feature engineering methods have some limitations when applied in histological images, as many parameters have to be set empirically, which will not produce consistent results if there is considerable variability within the image sets (GURCAN et al., 2009). On the other hand, the results achieved by a deep learning model known as AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) provided a major breakthrough in the computer vision community and, in recent years, deep learning methods are outperforming the traditional methods in many image segmentation problems (HU et al., 2018; ALTAF et al., 2019; MAHMOOD et al., 2020). Deep learning techniques traditionally need large amounts of labeled data to be trained. Unfortunately, histological images segmentation is often constrained by the availability of labeled training data once labeling histological images is a highly-skilled, complex, and time-consuming task. One way to surpass this drawback is data augmentation. Data augmentation encompasses a set of data-space transformation techniques that enhance the size and quality

of training datasets, hence deep learning models can be built using them (SHORTEN; KHOSHGOFTAAR, 2019).

Automatic ROI segmentation in OSCC WSIs is challenging due to the high computational costs of pixel-wise segmentation, and to the large regional variations that can be present in a single WSI – besides other factors like the variability that can occur in the image acquisition process (H&E-reagents, staining process, scanner, the specialist who performs the staining, etc). This thesis proposes a novel methodology that can automatically analyze a WSI to segment clinically relevant regions of OSCC to aid and enhance the pathologist decision-making on diagnoses and prognoses of such tumors. To the best of our knowledge, this work is one of the first studies to investigate OSCC segmentation on a WSI level applying FCN methods (LITJENS et al., 2017; XING et al., 2018; MAHMOOD et al., 2020; SRINIDHI; CIGA; MARTEL, 2021).

1.2 Research Hypotheses

Despite the compelling CNN results when applied to image segmentation tasks, hardware limitations still make it prohibitive to use them directly to very large images (as is the case of WSIs). Thus, image segmentation using convolutional networks commonly adopts an approach based on smaller image-patches extracted from the original large images. The main research question discussed in this thesis is that proposing an FCN-based architecture to segment tumor regions in oral cavity-derived H&E-stained WSIs using image-patches with a larger size could produce refined competing results. Small image-patches could perform worse because of the little contextual information while the image-patches with a larger size contain more contextual information to discriminate between different types of cells and regions such as keratin, lymphocytes, blood, glands, muscle, and tumor cells.

1.3 Goals

This thesis aims to develop computational techniques for histological image segmentation to support pathologists in decision-making both in the diagnosis and in the prognosis in cases of oral cavity-derived cancer. The specific goals are:

- ❏ Create a new labeled WSI dataset for studying real OSCC cases;
- ❏ Propose a new method for automated segmentation of tumor regions on WSIs obtained from OSCC samples;
- ❏ Evaluate the systematization of a data augmentation strategy to segment H&E-stained histological images using FCN;

- ❏ Develop and make available to the community an open-source WSI processing tool that implements the methods proposed in this thesis.

1.4 Contributions

The main contributions of this thesis are:

- ❏ A new labeled WSI dataset, fully validated by a pathologist, containing H&E-stained histological images of real OSCC cases;
- ❏ A simple and fast color-based tissue detector for WSIs;
- ❏ An FCN-based method for segmentation of OSCC regions on oral tissue images;
- ❏ Validation of the proposed method across datasets with H&E-stained histological images from different diseases (OSCC and non-OSCC) and image formats (ROI, TMA, and WSI);
- ❏ A data augmentation strategy to train FCN-based methods to segment H&E-stained OSCC histological images;
- ❏ Evaluation of color-stain normalization impacts on automated nuclei segmentation on H&E-stained ROI images of oral tissues;
- ❏ Evaluation of the impact caused by different color space transformations (RGB, HSV, and CIE L*a*b*) on automated nuclei segmentation on H&E-stained ROI images of oral tissues.

1.5 Thesis Organization

To clearly explain the theoretical concepts involved, techniques, and obtained results, this thesis was organized into five chapters. This chapter provides a brief introduction, presents the motivations, goals, main contributions, and organization of this thesis. Chapter 2 presents the theoretical concepts necessary to understand this work along with the main related works. Chapter 3 describes in detail the proposed approaches and the used image datasets. Chapter 4 presents the evaluations and experiments carried out to assess the quality of each proposed approach. Finally, Chapter 5 reports the conclusions and proposals for future work.

Chapter 2

Contextualization

To provide some context and background, this chapter describes some required concepts to better understand the main challenges of our problem such as digital image processing, cancer diagnoses, and the steps to histological evaluation. The main related works also are presented in this chapter.

2.1 Oral Cavity-derived Cancer

Cancer encompasses a group of diseases that can starts in almost any organ or tissue of the body caused by changes to genes that control the way our cells function, especially how they grow and divide. The disease starts when abnormal cells grow uncontrollably, go beyond their usual boundaries to invade adjoining parts of the body, and spread to other organs and surrounding tissues. There are more than 100 different types of cancer and these types are usually named by the organs or tissues where the cancers form. For example, when the disease starts in epithelial tissues, such as skin or mucosa, they are called carcinomas. If the starting point is connective tissues, such as bone, muscle, or cartilage, they are called sarcomas. Other characteristics that differentiate the types of cancer from each other are the speed of cell multiplication and the ability to invade neighboring or distant tissues and organs. The process by which cancer cells spread to other parts of the body is called metastasis (WHO, 2022; NCI, 2022; INCA, 2022).

Not every change in the body's tissues is cancer, but some tissue changes may sometimes become cancer if they are not treated, as shown in Figure 2. Hyperplasia is a soft disorder where there is an increase in the number of cells in an organ or tissue, however, the cells and the way the tissue is organized appear normal under the microscope visualization. Oral epithelial dysplasia is a more serious condition than hyperplasia, being defined as a premalignant disorder of oral mucosa that can be graded as mild, moderate, and severe. Hyperplasia and dysplasia are not considered cancer, once they do not spread into, or invade nearby tissues. Although not all dysplasia becomes cancer over time, this premalignant condition has a significant rate of malignant progression to cancer, and the

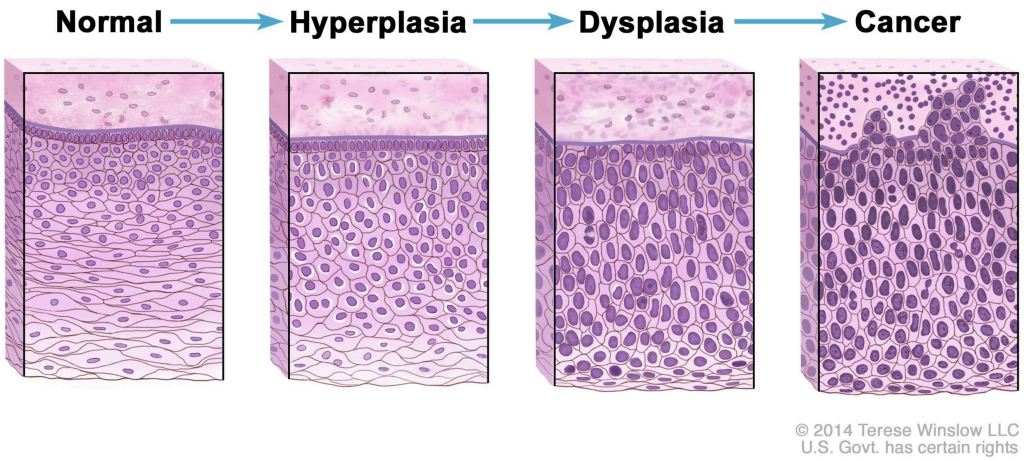
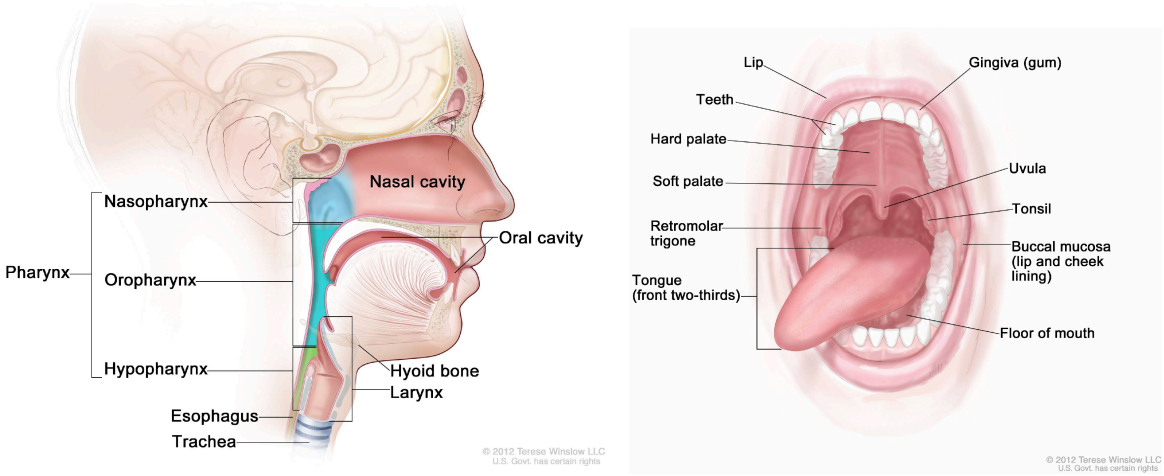


Figure 2 – Some cells may go through abnormal changes called hyperplasia and dysplasia before cancer cells are formed in the body tissues. Normal cells may or may not become cancer cells over time. (Image sourced from (NCI, 2022).)

early treatment is considerably more successful (WILS et al., 2020; NCI, 2022).

Oral cavity-derived cancer, denoted as a category of head and neck cancer that includes the regions highlighted in Figure 3 (b), represents the most common form of head and neck cancer (ETTINGER; GANRY; FERNANDES, 2019). Figure 3 shows the anatomical structure of the oral cavity.



(a) (b)

Figure 3 – Anatomy of the oral cavity: (a) Oral cavity viewed by a head and neck perspective; (b) Details of the oral cavity anatomy. (Images sourced from (NCI, 2022).)

OSCC, the most common subtype of oral cavity-derived cancer, is the sixth most prevalent occurring cancer in the world and is among the twenty leading causes of death in the population. The main risk factors for OSCC development are tobacco use, including smokeless tobacco, human papillomavirus infection, and excessive alcohol drink-

ing (HASHIBE et al., 2007; BOFFETTA et al., 2008; CHATURVEDI et al., 2011). OSCC is also known to have a high recurrence rate and overall 5-year relative survival rates of less than 70%. It has a poor prognosis as most of them are diagnosed in advanced stages (III and IV), hence it has a low survival rate and high-cost treatment (PULTE; BRENNER, 2010). Compared to other tumor locations, patients with this type of cancer may have more problems to address, such as speech problems, oral pain, malnutrition, sexual problems, and reduced ability to swallow along with loss of appetite. These disabilities can lead to poor quality of life for OSCC patients (LIAO et al., 2019).

The study of tissue samples of suspected regions under the microscope (also referred to as histological analysis) remains the most reliable way for diagnosing and grading cancer. The diagnosis and grad of different cancer types are made by pathologists through a complex and time-consuming process where some cytological features, including the nuclei morphology and their spatial arrangement, are analyzed (IRSHAD et al., 2014; MCCANN et al., 2015; XING et al., 2018).

2.2 Histological Image Analysis

Histology is the microscopic study of body tissues and as these tissues are organized to constitute organs. The most common procedure used in the histological analysis is the preparation of histological tissue slides that can be visually examined using a light microscope. The tissue slides must be prepared, with the aim to highlight and preserve some internal features as close as possible it had inside the body before being visualized under the microscope (JUNQUEIRA; CARNEIRO, 2013). Figure 4 shows the basic steps commonly involved in the process of preparing histological tissue slides.

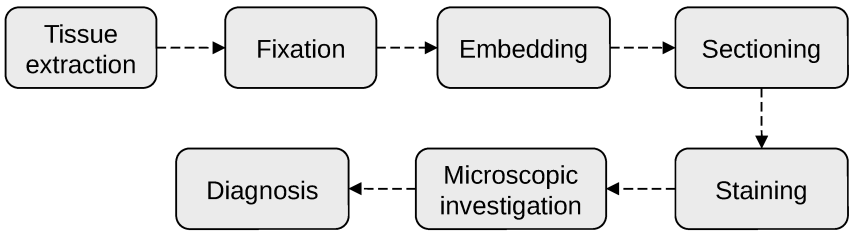


Figure 4 – Main steps for the histological analysis process.

The preparation of histological slides begins with a tissue sample extraction – biopsy. After the biopsy, the extracted tissue samples should be stabilized in the fixation step to preserve cell structure and prevent decomposition. The choice of fixative solution depends on the kind of tissue and the analysis to be carried out. In general, the tissue fragments in the fixation process are immersed for a few hours in fixative chemical solutions such as formalin and glutaraldehyde (MCCANN et al., 2015).

After fixation, the tissue is embedded with other chemical substances in order to eliminate water and substances used in the previous step. The aim is to remove water and impurities and make the tissue samples rigid enough to facilitate the sectioning of the tissue into thin slices for further observation under the microscope. For this, the tissue samples should be infiltrated with substances that provide a rigid consistency. This process is known as embedding. The most used substances for this purpose are melted paraffin and some plastic resins. Once the embedding step has been completed, the paraffin blocks that contain the tissue samples are then cut into sections of 3-10 μm thickness using an instrument called a microtome (see Figure 5(a)). This process is called sectioning or microtomy. After the microtomy, the sections are placed to float on a heated water surface and then are placed on glass slides, where they will later undergo the staining process to facilitate microscopic visualization (MOLINARO; CAPUTO; AMENDOEIRA, 2010; JUNQUEIRA; CARNEIRO, 2013).

When the sectioning process is completed, cells and extracellular material usually become translucent, requiring a staining procedure to differentiate the components of the tissues, cells, and extracellular matrix. The combination of H&E represents the most commonly used dyes, with hematoxylin responsible for pigmenting the cell nuclei in dark blue, while eosin pigments the other structures in pink (IRSHAD et al., 2014; KLECZEK; JAWOREK-KORJAKOWSKA; GORGON, 2020). Once colored, the tissues on the glass slides are ready to be examined under a microscope after receiving a protective transparent coverslip (see Figure 5(c)). According to (JUNQUEIRA; CARNEIRO, 2013), the entire slide preparation procedure, from fixation to the observation of the tissue under the microscope, can take from 12 hours to 2.5 days. Figure 5 shows a paraffin block being cut by a microtome and the resulting tissue slides stained with H&E stain and ready to be visually examined using a light microscope.

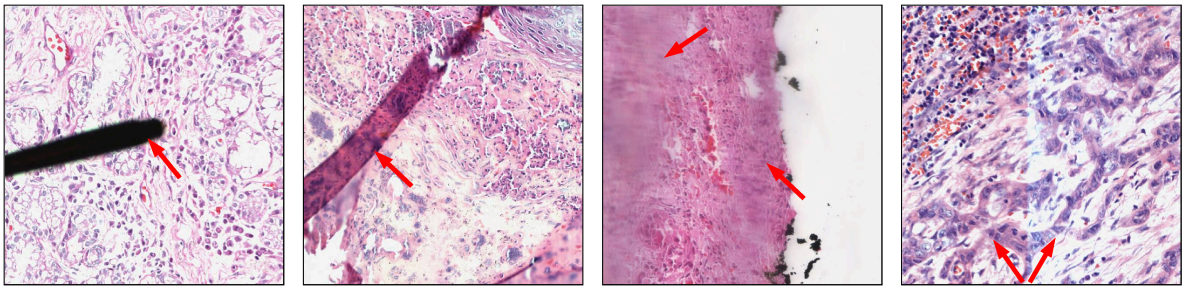


(a) (b) (c) (d)

Figure 5 – Histological tissues produced in the Histology Laboratory of the Federal University of Uberlândia: a) Paraffin block tissue sample positioned for cutting in the microtome; b) Pieces of tissue in a water bath after cutting; c) Tissue slides after staining; d) Tissue slide being visualized by a microscope.

A growing trend in the area of histological analysis is the digitalization of tissue slides after the staining procedure. In this way, pathologists can make diagnoses aided by CAD tools based on the digital image rather than through a microscope. In addition, storing digital images would be a more attractive alternative than storing glass slides, since glass slides take up considerably more space, can be damaged, lost, or even deteriorate over time (MCCANN et al., 2015).

Similar to other manual tasks, the produced histological tissue slides may suffer high variations between them, once they typically are manually produced by laboratory technicians. The preparation steps can generate some undesired artifacts that can lower the quality of the histological images such as folding of tissue, presence of dust, air bubbles, or other contaminations on the coverslip, and even color differences caused by the staining time. Although these problems only partially limit the interpretation of images by pathologists, they can cause a big impact on automatic image analysis algorithms (SALVI et al., 2021). Figure 6 illustrates some of the most common undesired artifacts in the process of preparing histological tissue slides.

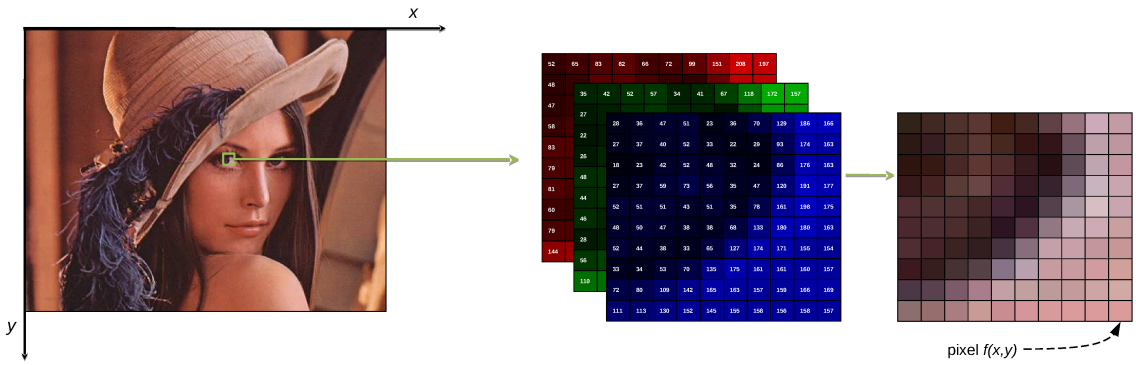


(a) (b) (c) (d)

Figure 6 – Different artifacts, indicated by red arrows, that can lower the quality of histological images: a) Slide contamination; b) Tissue folding; c) Blurring; d) Color and contrast differences.

2.3 Digital Image Processing

An image can be defined as a two-dimensional function, $f(x, y)$, where x and y are spatial coordinates, and the amplitude of f at any point (x, y) is called the intensity or gray level of the image at that point. When x, y, and the amplitude values of f are all finite (discrete values) then this image is a digital image. In this way, the digital image can be interpreted as a matrix whose indexes of lines and columns identify an element in the image, and the value of the element corresponds to the intensity or gray level at that point. Each of the elements in this matrix is called a pixel. In the case of color digital images, as shown in Figure 7, each pixel is described using a set of properties. Thus, the color of each pixel is represented by a point in the space defined by the color space (GONZALEZ; WOODS, 2006).



(a) (b)

Figure 7 – Digital image representation: (a) Color digital image; (b) Matrix representation of a region of size 10×10 pixels using the RGB color space.

Color spaces, also known as color models or color systems, are multidimensional coordinate systems in which each color can be represented by a point in that multidimensional space. There are several color spaces, among those commonly used in image analysis routines are the red, green, and blue (RGB); the hue, saturation, and value/brightness (HSV); and the International Commission on Illumination (CIE) L*a*b*. The purpose of a color space is to facilitate the specification of colors in some standard providing a coordinate system and a subspace in which each color is represented by a single point. There have been numerous different color spaces in use nowadays. Some of these color spaces are ideally suited for hardware implementations (monitors, printers, etc.), others in the form that humans describe and interpret colors and others for computer vision routines (ROBERTSON, 1977; GONZALEZ; WOODS, 2006; FERNáNDEZ-CARROBLES et al., 2015).

The most commonly used color spaces for segmentation purposes in H&E-stained histological images are RGB, HSV, and CIE L*a*b*:

- ❏ RGB (red, green, blue) color space defines each color as a combination of the three primary spectral components: red, green, and blue. The RGB color space is the hardware-oriented color space most widely used for a broad class of video cameras and color monitors;

- ❏ HSV (hue, saturation, value/brightness) color space is a nonlinear transformation of the RGB that describes the pure color (hue) in terms of gray presented in each color (saturation) and how bright the color is (value/brightness). The HSV color space corresponds closely with the way humans describe and interpret color;

- ❏ CIE L*a*b was defined by the International Commission on Illumination and is also based on human perception. The L* channel indicates lightness and a* and b* channels indicate chromaticity directions: a* indicates the color value between green and red and b* indicates the color value between blue and yellow.

Image segmentation is a very important and challenging task in the digital image processing and computer vision field. Segmentation is the process of partitioning an image into its constituent regions or objects of interest based on specific criteria. There are many commonly used image segmentation methods and, generally, they are based on the detection of the following image properties: discontinuity and similarity. The discontinuity-based methods partition an image based on abrupt changes in intensity, such as edges in an image. The similarity-based methods aim to partition an image into regions that are similar according to a set of predefined criteria (GONZALEZ; WOODS, 2006). Thresholding (OTSU, 1979), region-growing (NOCK; NIELSEN, 2004), k-means clustering (DHANACHANDRA; MANGLEM; CHANU, 2015), and watersheds (NAJMAN; SCHMITT, 1994) are examples of classical segmentation methods that use traditional feature engineering to perform image segmentation.

Over the past few years, various algorithms exceeded the classical segmentation methods results by using deep learning methods to perform semantic segmentation. Semantic segmentation associates each pixel of an image with a category label that represents its more meaningful constituent regions (HAO; ZHOU; GUO, 2020). Figure 8 illustrates an original image and its corresponding semantic segmentation output where the meaningful regions of the original image are represented by different colors – dog (yellow), tree (green), grass (dark green), ground (brown), toy (blue), or bicycle (dark gray).



(a) (b)

Figure 8 – Segmentation of a natural image: (a) Original image; (b) Meaningful regions of the image identified in the semantic segmentation output.

The segmentation of nontrivial images, such as histological images, is one of the most difficult tasks in image processing and their accuracy usually determines the success or failure of further image analysis algorithms (GONZALEZ; WOODS, 2006). The subtleties and high diversity of different structures present in histological images are quite complex and sometimes challenging even for well-trained pathologists. A CAD tool that can understand these nuances and identify clinically relevant regions can make a big impact in medical care.

2.4 Convolutional Neural Networks

Deep learning based methods are emerging as a powerful tool that has been successfully applied to microscopy image analysis, object detection, and image segmentation tasks (Xing et al., 2018). Different from algorithms that use classical feature engineering, deep learning methods are able to receive raw input data and automatically learn patterns and features from these input data without having to design handcrafted feature extraction methods (LECUN et al., 1998).

CNN is a specialized kind of deep learning architecture that is designed to process data in the form of multi-dimensional arrays, as is the case of color images, taking into account the spatial structure of these data. The name "convolutional neural network" indicates that the network architecture employs at least one layer with a mathematical operation called convolution[1]. A typical CNN architecture is structured as a sequence of stacked layers, with the convolutional layers being the core building blocks of a CNN where the filter weights are automatically learned. Each layer in a CNN is a three-dimensional volume of size $h \times w \times c$, where h and w are spatial dimensions, and c is the number of feature channels. These layers in a CNN are built using basic components (convolution, pooling, and activation functions) and operation attributes (kernel size, stride, and padding) with the aim to extract a feature map that describes the input image. These extracted features are then made available to be further classified into desired categories by fully connected layer (FCL). While in traditional image processing methods the filters are handcrafted, CNNs, with sufficient training and data samples, have the ability to automatically learn these filters (LECUN; BENGIO; HINTON, 2015; LONG; SHELHAMER; DARRELL, 2015; GOODFELLOW; BENGIO; COURVILLE, 2016).

CNN models traditionally need large amounts of labeled data samples to be trained. Unfortunately, histological images segmentation is often constrained by the availability of labeled training data once labeling histological images are highly-skilled, complex, and time-consuming. One way of surpassing this lack of a sufficient amount of training data is data augmentation. Data augmentation encompasses a set of data-space transformation techniques (as flip and rotation geometric transformations of images) that enhance the size and quality of training datasets, such that deep learning models can be trained using these transformation techniques (SHORTEN; KHOSHGOFTAAR, 2019).

Every state-of-the-art CNN model is composed of encoding (feature extraction) and decoding phases (classification). These two phases are built using the following basic building blocks operations: convolution, activation function, pooling or downsampling, upsampling, and classification. Convolution is a linear operation where multi-dimensional filters are slid across the width and height of the input volume (image) to compute dot

[1] It is important to note that the convolution operation used in a CNN does not correspond precisely to the definition of convolution as used in pure mathematics.

products between the weights of the filter and the input volume at any position. The filters usually have small width and height sizes but extend through the full depth of the input volume. For example, a typical filter employed on the first layer of a CNN might have a size of 3×3×3 (3 pixels width and height, and 3 color channels depth). As the filter is slid over the width and height of the input volume, a 2-dimensional feature map is produced containing the visual feature values activated by the filter (GOODFELLOW; BENGIO; COURVILLE, 2016; DUMOULIN; VISIN, 2018). Each convolutional layer is configured by a set of hyperparameters that will define the size of the output volume:

- ❏ The number of filters used (or the convolutional layer size): as each filter produces a separate 2-dimensional feature map output, this value will define the depth or number of channels of the output volume;
- ❏ The size of the filter kernel: the kernel size will define the width and height sizes of the filter and consequently how many weights will be learned for each filter;
- ❏ The stride to slide the filter over the input volume: the stride value will define how many pixels the filter will be moved at a time (like a sliding window). Larger stride values will produce spatially smaller output volumes;
- ❏ The padding approach: defines if the input volume must be padded with zeros around the border to consider the bordering elements of the input volume in the convolution. This hyperparameter defines if the height and the width of the output volume will be the same or smaller than the input volume;
- ❏ The bias value: represents an additional set of learnable weights that allows the activation function to be shifted to the left or right (analogously to the role of a constant in a linear function).

The feature maps resulting from the convolutional operations are then applied to a nonlinear operation, known as the activation function. Different from the definition of convolution as used in pure mathematics, a learnable bias term is generally added to each convolutional filter result before applying the activation function. There are several activation functions that can be used in CNN models but the most commonly found are the rectified linear unit (ReLU), hyperbolic tangent (tanh), and sigmoid functions (KRIZHEVSKY; SUTSKEVER; HINTON, 2017). Figure 9 shows the graphic representation of these activation functions.

The ReLU activation function replaces all negative values in the feature map by zero and is defined by Equation 1:

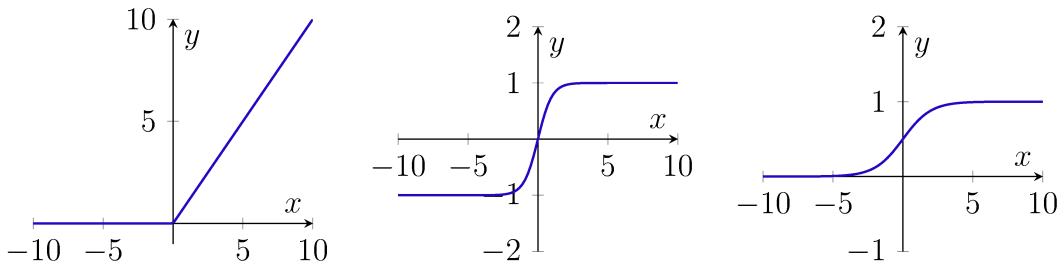
$$ReLU(x) = max(0, x). \tag{1}$$

The tanh activation function transforms the pixel values to the range [-1, 1] and is defined by Equation 2:

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \tag{2}$$

Sigmoid function squashes the pixel values to range between 0 and 1 and is defined by Equation 3:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{3}$$



(a) ReLU (b) tanh (c) Sigmoid

Figure 9 – Graphic representation for commonly used activation functions.

The pooling operation replaces the input pixel value at a certain location with a summary statistic of the nearby pixel values. There are several kinds of pooling operations such as max-pooling and average-pooling. The most common kind of pooling operation is max-pooling using kernels of size 2×2 and stride of 2 (GOODFELLOW; BENGIO; COURVILLE, 2016; DUMOULIN; VISIN, 2018). As shown in Figure 10, the pooling configuration will downsample the input by 2 along both width and height, retaining only the summarized information. Figure 11 shows a single execution of the commented typical layers and operations used to obtain the feature map for an input image in a CNN.

0.00	0.03	0.61	1.11
0.03	0.47	0.91	1.03
0.49	0.90	0.91	0.77
0.97	0.98	0.85	0.83

(a) Original feature map.

0.47	1.11
0.98	0.91

(b) Max-pooling result.

0.13	0.95
0.83	0.83

(c) Average-pooling result.

Figure 10 – Sample pooling operations using kernel of size 2×2 and stride of 2.

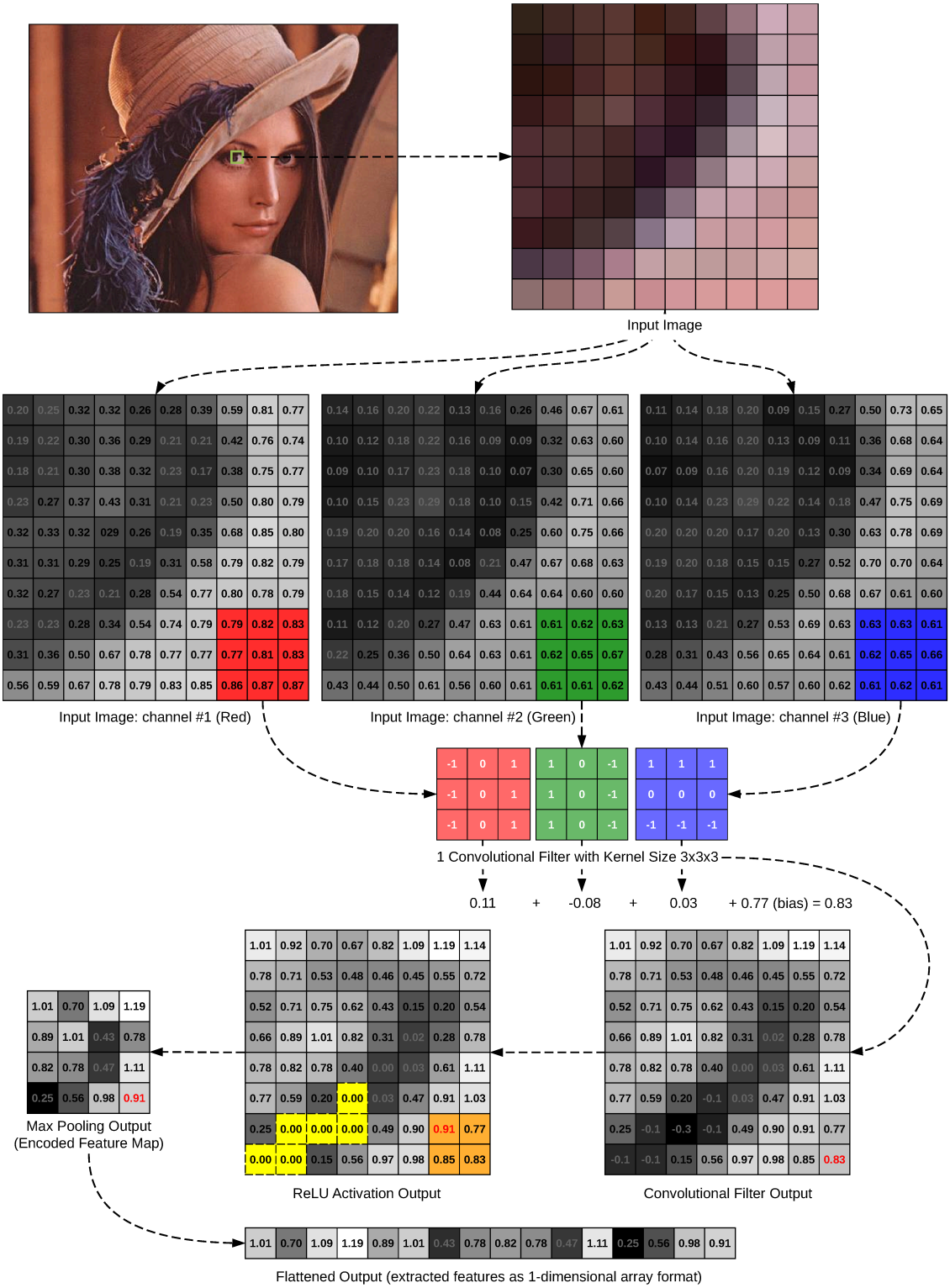


Figure 11 – Basic operations in the encoding phase of typical CNNs. The convolutional filter in this example is composed of three channels with kernel size 3×3, stride 1, no padding, and bias equals 0.77. The learnable weights are represented by numbers inside of each convolutional kernel (red, green, and blue). ReLU activation function transforms negative values to zero (highlighted in yellow). The max-pooling operation chooses the max value present inside each stride in the pooling kernel with size 2×2 and stride 2 (highlighted in orange).

The CNN decoding phase depends on the task that will be addressed. For the image segmentation task, there are two major architectures commonly used: segmentation by image-patches classification (classification network) and segmentation by classification of every pixel present in the image (segmentation network). In the first option, the image is subdivided into several image-patches and FCLs are used to make class label predictions for the input image-patches. The image segmentation result is achieved using a postprocessing operation that associates each image-patch of the original image with the predicted class label value. The second option (known as FCN) performs a pixel-wise prediction, without the need to use FCLs, to directly produces the image segmentation result. For segmentation networks, the resolution of the feature maps obtained in the encoding phase is increased, until the desired output image dimension, replacing the pooling operations by upsampling operations such as transposed convolutions or interpolation methods (LONG; SHELHAMER; DARRELL, 2015; DUMOULIN; VISIN, 2018). Figure 12 illustrates the execution of the common operations used in typical FCNs in the decoding phase considering the feature map encoded in Figure 11.

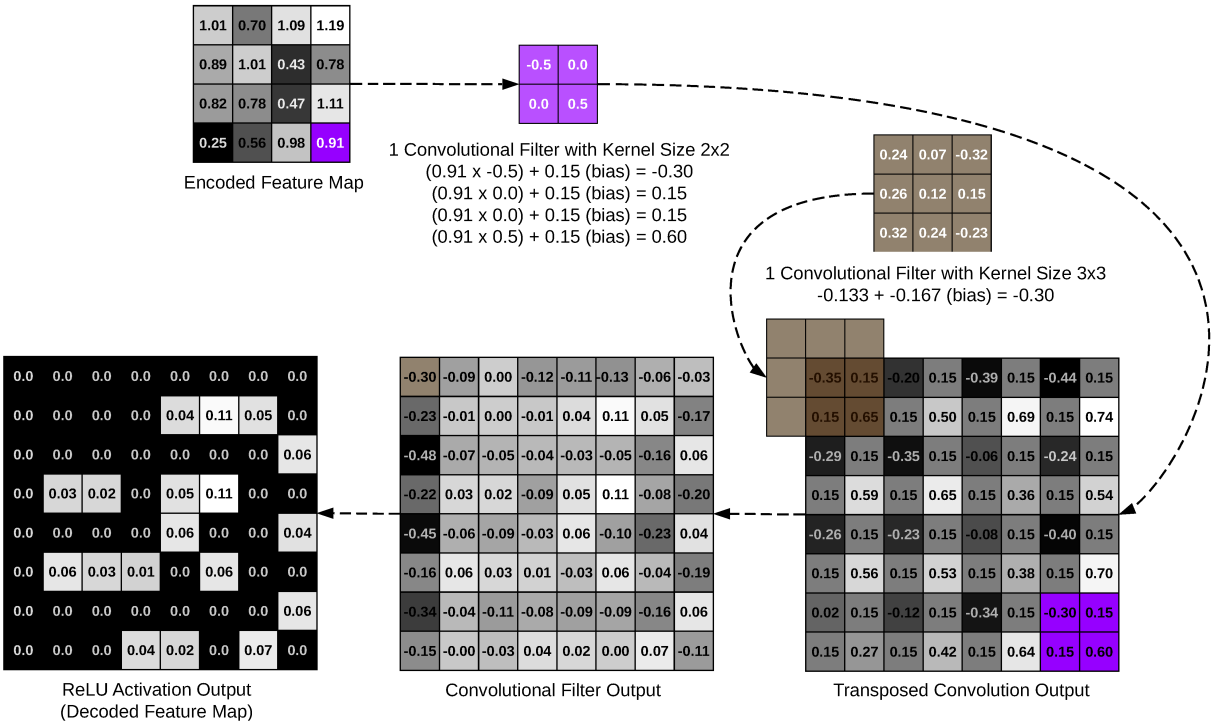


Figure 12 – The basic operations in the decoding phase of typical FCNs. The transposed convolution (purple kernel) is used to increase the feature map dimension (from 4×4 to 8×8). In this example, it is performed by a convolutional one-channel kernel with a size of 2×2, stride 2, and bias equals 0.15. The convolutional filter (brown kernel) is performed using a kernel with size 3×3, stride 1, zero-padding, and bias equals -0.167. The learnable weights are represented by numbers inside of each convolutional kernel (purple and brown kernels). ReLU activation function transforms negative values to zero.

Despite the attractive results of the CNNs applied to image segmentation tasks, it is still prohibitive to apply them directly to very large images (as is the case of WSIs) mainly because of the hardware limitations to processing this kind of image. Thus, the task of image segmentation using CNNs commonly adopts an approach based on smaller image-patches extracted from the original large images. Figure 13 illustrates the two major architectures commonly used to address the image segmentation task using CNNs. The classification network option has the advantage that obtaining training annotations at patch-level is less laborious but has the disadvantage that it needs to use very small image-patches to obtain a reasonable resolution of the resulting segmented image (image aliasing effect). On the other hand, the segmentation network option has the advantage that for obtaining much more refined pixel-level segmentation results, but as a drawback, producing pixel-level training annotations requires a lot of labor.

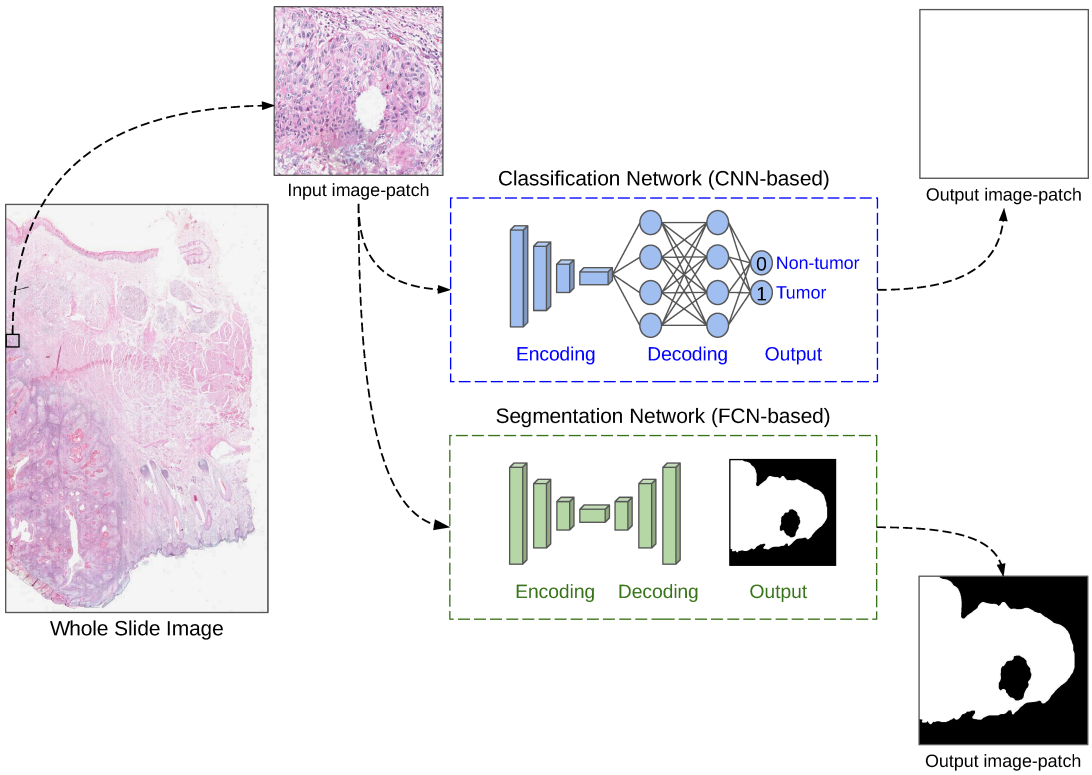


Figure 13 – Typical convolutional architectures used to the image segmentation task: classification network (CNN-based) and segmentation network (FCN-based).

2.4.1 Popular Convolutional Models

LeNet (LECUN et al., 1990; LECUN et al., 1998) is considered the pioneering CNN model to be successfully applied in image recognition tasks (ALOYSIUS; GEETHA, 2017). The model comprises seven layers and was specifically designed to classify handwritten digits using input images of 32×32 pixels size. However, due to a lack of sufficient training

data and computing power available at the time, this architecture failed to perform well in more complex problems. Figure 14 illustrates the proposed LeNet architecture.

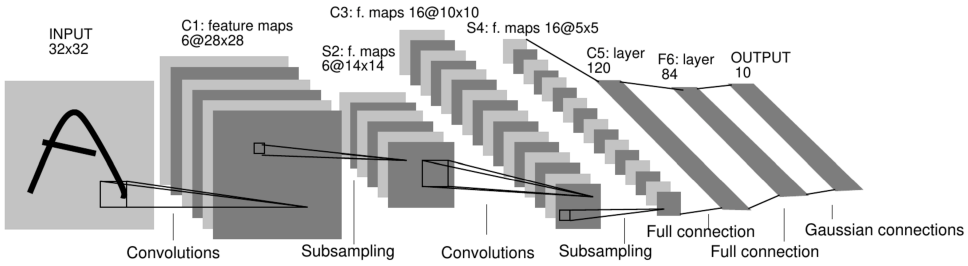


Figure 14 – The architecture of the LeNet model, the first CNN successfully employed in the handwritten character recognition tasks. (Image sourced from (LECUN et al., 1998).)

Almost twenty years later, the deep learning model known as AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) provided a major breakthrough in the computer vision community. The AlexNet model succeeded in bringing down the error rate in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competition (RUSSAKOVSKY et al., 2015) and has become one of the most influential CNN models in the computer vision field. The AlexNet architecture, shown in Figure 15, can be considered as a major variant of LeNet having five convolutional layers followed by three FCLs (ALOYSIUS; GEETHA, 2017).

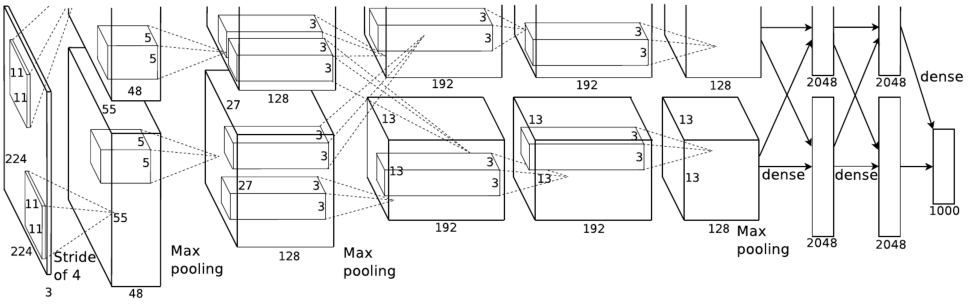


Figure 15 – The architecture of the AlexNet model. The AlexNet model is credited to popularize CNN models in the computer vision field. (Image sourced from (KRIZHEVSKY; SUTSKEVER; HINTON, 2012).)

Over the years some more and more deep convolutional architectures emerged inspired by AlexNet success (SIMONYAN; ZISSERMAN, 2015; SZEGEDY et al., 2015; HE et al., 2016). In the last few years, some deep architectures have become popular for medical image segmentation, and the architecture most well-recognized applied to medical image analysis is U-Net (LITJENS et al., 2017; SALVI et al., 2021). The U-Net model is an FCN-based architecture that was specifically designed for the segmentation of medical microscopy images (RONNEBERGER; FISCHER; BROX, 2015).

As illustrated in Figure 16, the U-Net model has a u-shaped architecture composed of an encoding phase followed by a decoding phase. The encoding phase follows the typical convolutional network architecture (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) using blocks of two no-padding convolutions with 3×3 kernels followed by ReLU activation function (blue arrows) and max-pooling operations with 2×2 kernel and stride of 2 for downsampling (red arrows). After each downsampling step, the next convolution block doubles the number of feature channels. The decoding phase upsamples the feature map resolution using transposed convolutions with 2×2 kernels, stride of 2, and halves the number of feature channels (green arrows). Then a concatenation of the resulting feature map with the corresponding cropped feature map from the encoding phase (grey arrows) is used as the input feature map to the next convolution blocks (decoding blue arrows). This concatenation reinforces that the features learned while the encoding phase will be used to reconstruct the resulting image. At the final layer (another green arrow), a convolutional filter with a 1×1 kernel followed by a Sigmoid activation function is used to provide the final pixel-prediction for each label class (each resulting feature channel predicts the pixel probability to each class label).

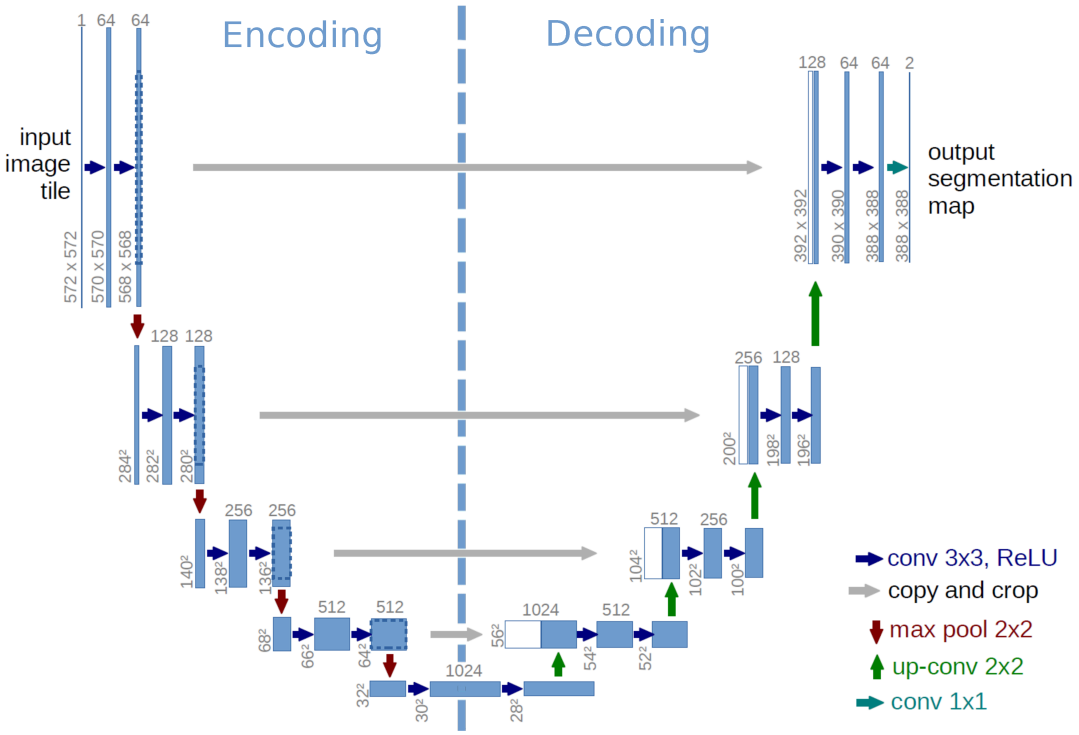


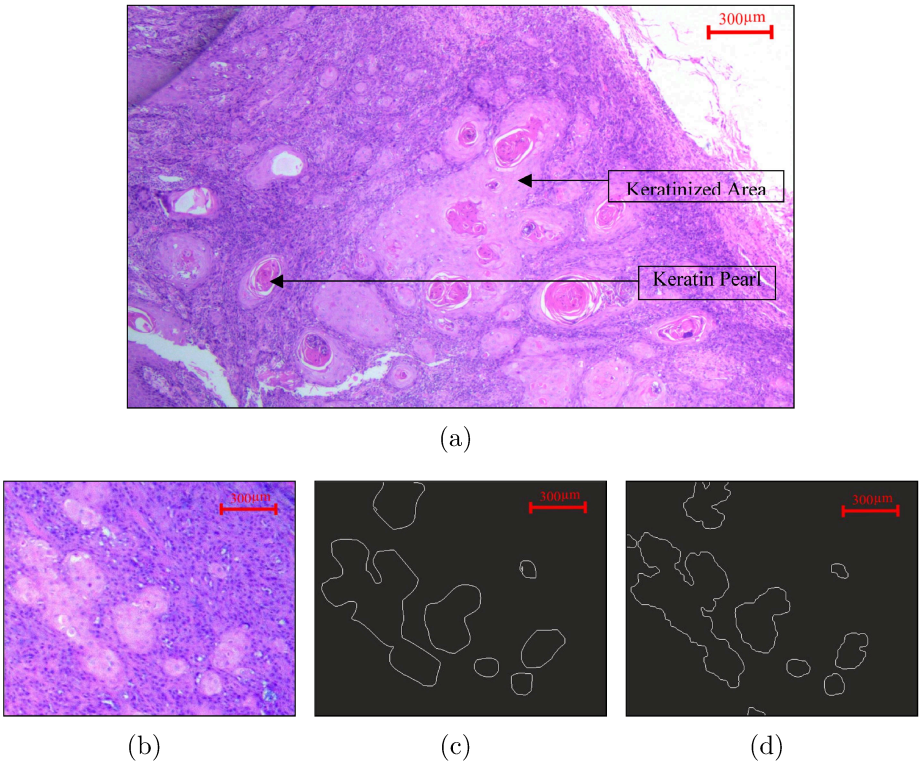
Figure 16 – The original U-Net architecture highlighting the encoding and decoding phases. Each blue box represents multi-channel feature maps, while white boxes represent copied feature maps (from the dashed blue boxes). The number of feature channels is denoted on top of the box and the spatial dimensions are provided at the lower-left edge of the box. The arrows of different colors represent different operations. (Image sourced from (RONNEBERGER; FISCHER; BROX, 2015).)

2.5 Related Works

Automated analysis of H&E-stained histological images covers a wide variety of potential applications, such as the segmentation of regions and cellular structures of interest, the counting of normal and cancerous cells, the recognition of tissue structures, grading, and prognoses of cancer, etc. The progress of image processing and pattern recognition techniques has been made CAD systems an important tool to assist pathologists in the interpretation of histological images. However, most studies are conducted mainly on the most common tumors (i.e., prostate or breast cancer) while OSCC using histological WSIs has been scarcely studied (SRINIDHI; CIGA; MARTEL, 2021).

Das et al. (2015) propose an automated method to localize, segment, and quantify keratinization and keratin pearl areas from H&E-stained histological images for detection of OSCC using traditional feature engineering methods. A keratinization index measure was proposed for grading OSCC cases (poorly, moderately, and well-differentiated) through selected ROI images containing the keratinized area. According to the authors, these keratin areas are an important histological feature for histopathological evaluation and clinical grading of OSCC. The study uses 90 ROI image-patches (selected and labeled by pathologists) with different sizes extracted from 10 histological images with an effective magnification of 40×. The method achieved 95.08% segmentation accuracy of keratin areas in comparison with pathologists' ground truths. Figure 17 shows a sample of the segmentation results obtained. Three years later Das et al. (2018) propose a two-stage approach for detection and segmentation of keratin pearls in H&E-stained histological OSCC images. In the first stage, a 12-layered CNN is used to classify very small image-patches (7×7 pixels) into four different classes (background, epithelial, subepithelial, and keratin) to afterward obtain a segmented version of the images. In the second stage, the keratin pearls are detected from the segmented keratin regions using texture-based features (Gabor filters) with a random forest classifier. The proposed texture-based random forest classifier has achieved 96.88% detection accuracy for the detection of keratin pearls using a proprietary dataset.

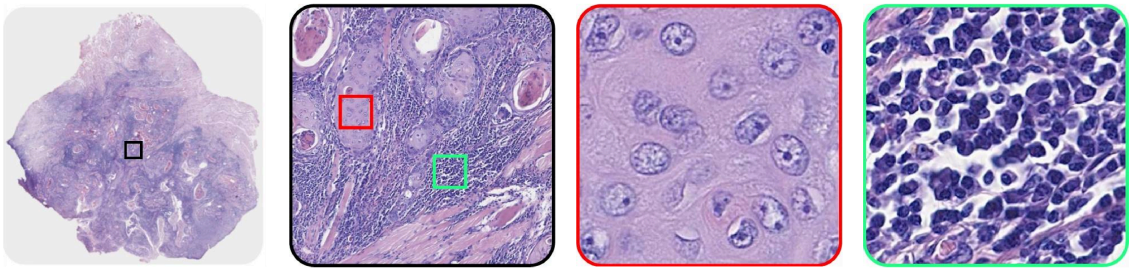
Few previous papers have been published in developing automated methods to localize and segment tumor regions using oral cavity-derived histological WSIs. The vast majority of the existing approaches use CNN-based methods to classify small image-patches from the original WSI between tumor and normal patches. Shaban et al. (2019) propose an automated approach to quantify the abundance of tumour infiltrating lymphocytes (TIL) in OSCC H&E-stained WSIs. There is a correlation of TIL density with improved overall survival and longer disease-free survival. The proposed TIL abundance score is calculated by first performing an indirect segmentation of the WSI into four different classes (tumor, lymphocytes, stroma, and non-ROI) and then quantifying the co-localization of lymphocytes and tumor areas in a novel fashion. To segment the WSI, a state-of-the-art CNN was used to classify image-patches at 20× effective magnification between one of the



(a) (b) (c) (d)

Figure 17 – Identification of keratinization and keratin pearl areas: (a) H&E-stained histological image of OSCC; (b) ROI image-patch; (c) Contour of the ground-truth image; (d) Contour of the segmented image obtained from the Das et al. (2015) methodology. (Images sourced from Das et al. (2015).)

four possible classes, and this information was used to rebuild a segmented version of the original image. The study used a total of 70 WSIs and 500,000 image-patches (400,000 for training and 100,000 for the test) of size 128×128 pixels – the image-patches were chosen so that each image-patch contains only a single class. The reported results achieve 96.31% image-patch segmentation accuracy on a proprietary test dataset. Figure 18 shows a TIL region on a WSI.



(a) WSI. (b) TIL region. (c) Tumor. (d) Lymphocytes.

Figure 18 – Region of tumor-infiltrating lymphocytes on a WSI of OSCC: (a) The WSI; (b) The TIL region highlighting the tumor (red) and lymphocytes (green) regions; (c) Tumor region; (d) Lymphocytes region. (Images sourced from Shaban et al. (2019).)

Halicek et al. (2019) used a state-of-the-art CNN model to detect and localize primary OSCC and to classify WSIs into cancer or normal tissues. The proposed methodology uses image-patches of 101×101 pixels size in the HSV color space with an effective magnification of 10×. The WSI is also segmented using a patch-based CNN approach to classify image-patches into the tumor or non-tumor classes to rebuild a segmented version of the original image. To achieve a reasonable resolution for the resulting segmented image, an overlapping image-patch sliding approach was used with a stride of 50 pixels (the results of overlapping image patches were achieved as averaged in the overlapping area). The authors tested the proposed method on a proprietary dataset containing 228 WSIs from OSCC cases and also on a public dataset of WSIs from sentinel lymph nodes with breast cancer metastases to obtain patch-based cancer localization and slide-level cancer diagnoses. The results achieved 84.8% patch-based segmentation accuracy in the proprietary dataset and 93.4% in the public dataset. Figure 18 shows WSI segmentation results obtained by the method proposed by Halicek et al. (2019).

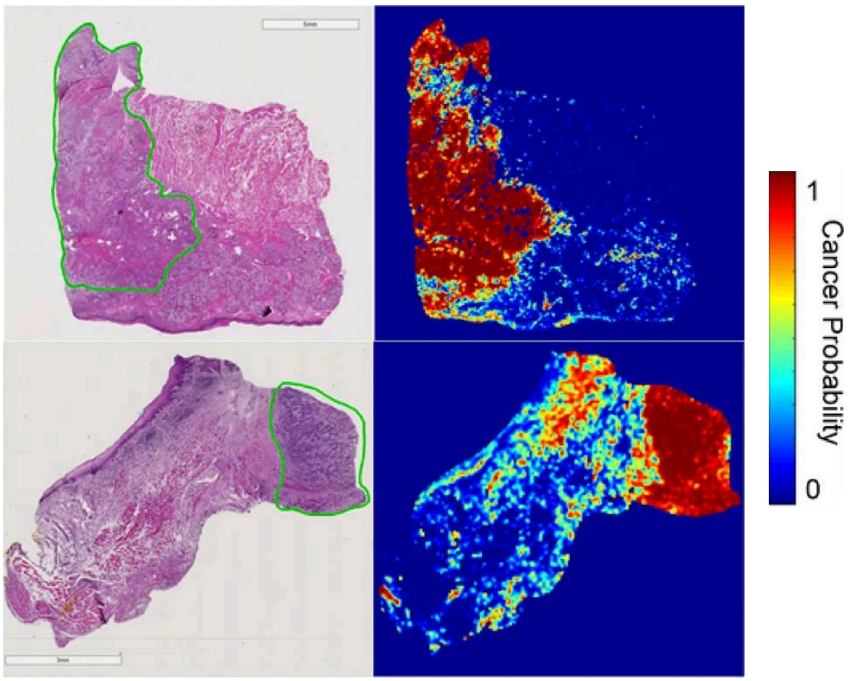
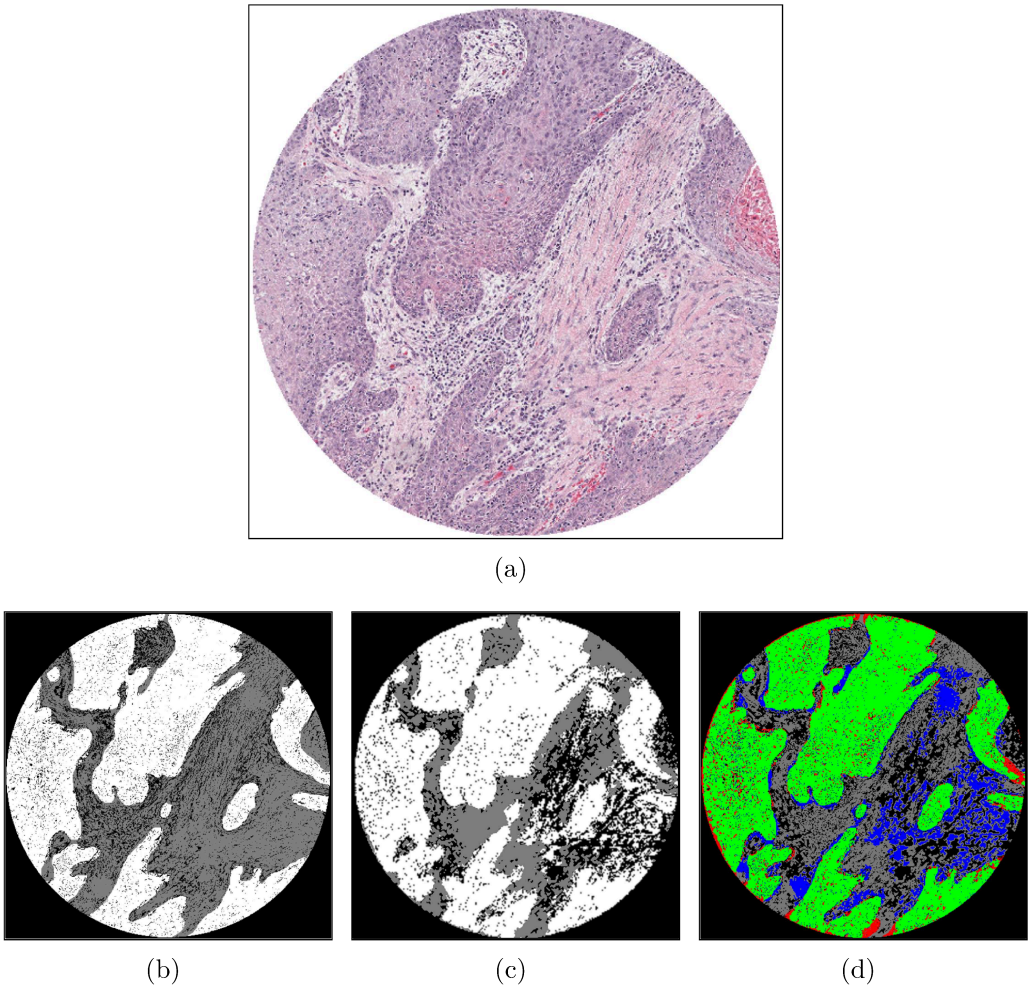


Figure 19 – WSI segmentation results by Halicek et al. (2019): the ground-truth cancer area is outlined in green on the left images, and the heat maps (right images) represent the obtained cancer probability. (Images sourced from Halicek et al. (2019).)

Martino et al. (2020) recently published the ORal Cancer Annotated (ORCA) dataset. ORCA is a labeled dataset containing 200 H&E-stained histological OSCC tissue microarray (TMA) images extracted from The Cancer Genome Atlas (TCGA) dataset (TCGA, 2022). TMA is a tool to extract small tissue cores from regions of interest from paraffin-embedded tissues. Unlike a WSI, where the entire histological slide is digitized into a large single image, TMA images represent small pieces of tumor regions of interest orga-

nized into a microarray. To the best of our knowledge, this is the first publicly available segmentation-proper labeled dataset containing OSCC tumor sample images. The authors used the ORCA dataset to compare the segmentation results achieved by four different FCN-based architectures inspired by the U-Net model. The segmentation performance was measured using a down-sampled version of the original image. All the original input images were resized to 512×512 pixels before being segmented by the FCN models into three different class labels: 1) carcinoma pixels (white); 2) tissue pixels not belonging to a carcinoma (grey); 3) non-tissue pixels (black). The authors also evaluated the use of the red channel of the RGB color space in combination with the hue channel of the HSV color space. The FCN model with the best configuration achieved 0.56 of intersection-over-union (IOU) for the carcinoma class. Figure 20 shows an example of one image from the ORCA dataset and its corresponding segmentation result.



(a)

(b) (c) (d)

Figure 20 – Segmentation results by Martino et al. (2020): (a) Original input image; (b) Corresponding ground-truth image where carcinoma pixels are colored in white; (c) The predicted segmented image obtained; (d) The colored error mask, where green pixels are true-positive carcinoma pixels, grey are true-negatives, blue are false-positives, and red are false-negatives. (Images sourced from Martino et al. (2020).)

From these papers, we can note that the segmentation of OSCC tumor regions is achieved mostly by rebuilding the original WSI using the classification result for each image patch (classification network). Thus, these methods have to use very small image patches to obtain cellular-level segmentation results and need an extra effort to rebuild a segmented version of the original image. Although FCN-based methods have been applied on the most common tumors (i.e., prostate or breast cancer) using H&E-stained histological images or other image types like magnetic resonance images, recent systematic reviews show that OSCC segmentation on a WSI level applying FCN methods (segmentation network) has been scarcely studied (LITJENS et al., 2017; HU et al., 2018; XING et al., 2018; MAHMOOD et al., 2020; SRINIDHI; CIGA; MARTEL, 2021).

2.6 Final Considerations

In this chapter, the theoretical foundations necessary for the understanding of this thesis were presented. In each of the topics presented, we address the fundamental concepts that will be useful throughout this thesis. In addition, this chapter also presented some of the main related works.

Localization and segmentation of malignant tumor regions in H&E-stained histological tissues are crucial tasks for the diagnosis of OSCC. The research activity related to histological WSIs of OSCC tissues is still limited compared to other diagnostic imaging branches mainly because of the following important limitations:

- ❏ WSIs are extremely large images with significant variability, heterogeneity, and complexity of features;
- ❏ There are a small number of available H&E-stained histological image datasets of OSCC tissues and most of them are not annotated;
- ❏ To our knowledge, there are no public datasets with H&E-stained WSIs of OSCC tissues available containing pixel-level annotations.

The next chapters present the image datasets created and used for validation, the proposed methodology to address the task of segmenting oral tumor regions in H&E-stained histological whole slide images, and the performed experimental evaluations and their results.

Chapter 3

Proposal

This chapter details our proposed contributions and the used image datasets. First, an overview of the proposed methods is presented. Then, the four image datasets used in the experimental evaluations are described. Next, we present the proposed data augmentation strategy and the methodology to segment OSCC tumor regions in H&E-stained histological images is detailed.

3.1 Method Overview

The proposed method for automated detection and segmentation of OSCC tumor region-derived H&E-stained WSI is established on seven pivotal steps. Figure 21 illustrates an overview of our proposed method according to the pivotal steps described below:

1. The WSI is scaled down at a factor of 32× to perform background removal and relevant tissue detection using a proposed color-based tissue region filter;
2. The tissue regions identified on the scaled-down image are mapped back into the original WSI;
3. Image-patches of size 640×640 pixels are extracted from the WSI on the original magnification, with no overlapping, considering only the identified tissue regions;
4. An FCN model, implemented in our own way based on U-Net model, is proposed to segment the identified image-patches;
5. A data augmentation strategy, using a pipeline with a random composition of different image transformations on the fly, is used to train our proposed FCN model;
6. A pixel-level cancer probability (a value between 0.0 and 1.0) is computed using the trained proposed FCN model;
7. The final tumor region segmentation is achieved by thresholding the cancer probability image.

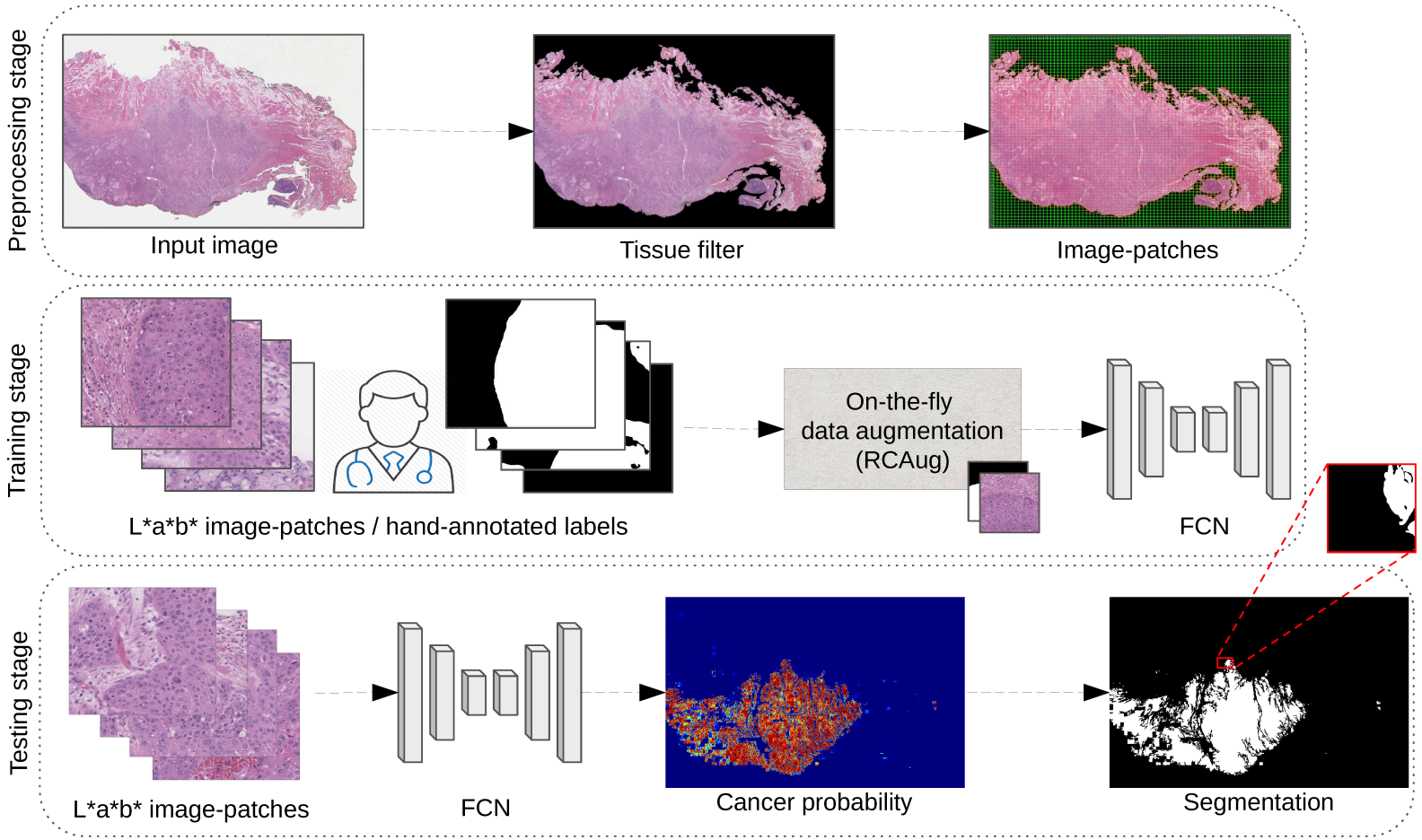


Figure 21 – Overview of our proposed method. Preprocessing stage: the identified tissue regions of the histological images are split into image-patches of size 640×640 pixels; Traning stage: a patch-based FCN is trained using data augmentation operations; Testing stage: the cancer probability heatmap is computed using the trained FCN; the tumor regions are localized and segmented by thresholding the heatmap image. An image-patch sample is highlighted to show that our proposed method is able to obtain a refined segmentation of tumor regions.

3.2 Image Datasets

Four different datasets of H&E-stained histological images were employed: Oral Cavity-Derived Cancer (OCDC), ORal Cancer Annotated (ORCA), CAncer MEtastases in LYmph nOdes challeNge (CAMELYON16), and Oral Nuclei Histological Images (ONHI). The OCDC and ORCA datasets contain digitized images from OSCC tumor samples. The OCDC dataset was used to develop and evaluate the data augmentation strategy and the proposed segmentation method to process WSIs obtained from OSCC tumor samples. The ORCA dataset was used to evaluate our proposed methods on a publicly available dataset of TMA images with different resolutions. As no similar OSCC labeled WSI dataset is publicly available to date, for comparison, the proposed segmentation method was evaluated on a well-known WSI dataset from breast cancer-derived lymph node metastases: the CAMELYON16 dataset. The ONHI dataset was employed to evaluate how the proposed segmentation method performs on the nuclei segmentation task using ROI images.

The original images of OCDC (WSI), ORCA (TMA), and CAMELYON16 (WSI) datasets were employed extracting image patches of 640×640 pixels without any resize or resolution modifications from the original images. In addition, the OCDC dataset was used to evaluate different image-patch sizes (160×160, 320×320, and 640×640) and the ORCA was also evaluated using resized images of 512×512 pixels. Table 1 summarizes information from the image datasets.

Table 1 – Summary of the image datasets used for the training and test of the proposed methods. Considering all experiments, a total of 90,628 image-patches were used to train and 114,093 to test the proposed FCN-based segmentation model.

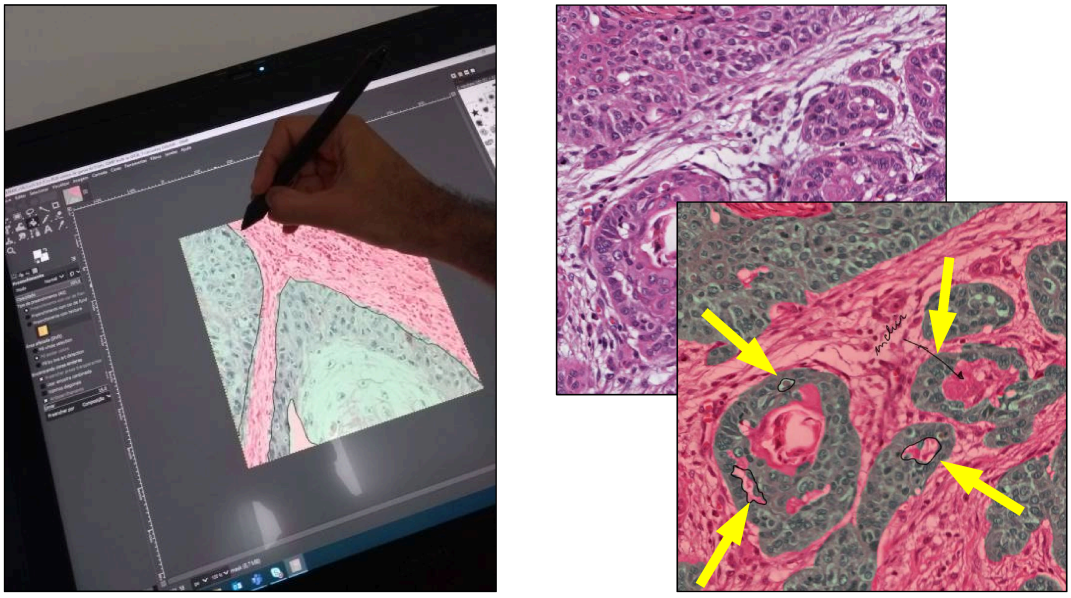
Dataset	Disease	Pixel-accurate	TMAs/WSIs	Test	Training
OCDC	OSCC	Yes	15	210	840
ORCA	OSCC	Yes	200	4,481	4,181
ORCA 512x512	OSCC	Yes	200	100	100
CAMELYON16	Breast cancer	No	399	85,411	109,278
ONHI	Oral dysplasia	Yes	-	24	96
		Total	814	90,226	114,495

3.2.1 OCDC dataset

The H&E-stained Oral Cavity-Derived Cancer (OCDC) dataset consists of a couple of images and their corresponding tumor regions annotation masks. The OCDC dataset was built using tissue specimens collected from human patients diagnosed with OSCC. All OSCC cases were retrieved from the Department of Oral and Medical Pathology Services archives in the Federal University of Uberlândia between 2006 and 2013 under

approval by the Committee on Research and Ethics of the institution (CAAE number: 15188713.9.0000.5152).

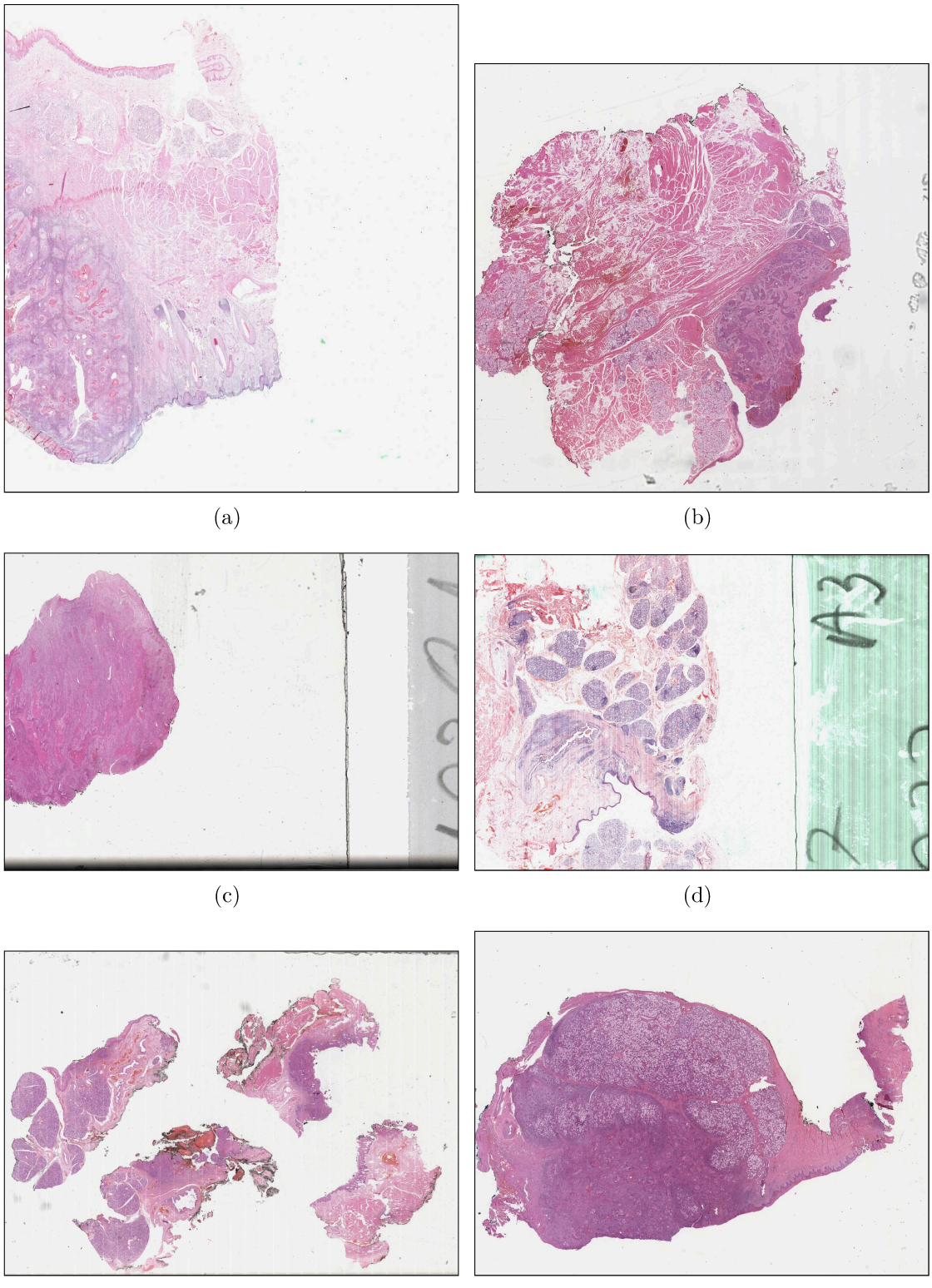
A total of 15 whole slide images were digitized using the Slide Scanner Aperio AT2 (Leica Biosystems Imaging, Inc., Nussloch, Germany) coupled to a computer (Dell Precision T3600) at 20× magnification and pixel-level resolution of 0.5025 μm × 0.5025 μm. The digitized images have different sizes, the larger one has almost three billion pixels (63,743×45,472 pixels), and were stored in SVS (OPENSLIDE, 2022) format using the RGB color space. A total of 1,050 image-patches of size 640×640 pixels were randomly extracted from these 15 WSIs and the tumor regions present in each image-patch were hand-annotated. The resulting annotated images were validated by a pathologist. This kind of task and a large amount of information present in the images requires very skilled professionals and a great effort to label the relevant structures to create this kind of image dataset. Despite involving hard work, relevant learning to better understand the main features involved in the OSCC diagnosis was acquired creating this labeled dataset. Figure 22 illustrates the OSCC regions labeling process for the OCDC dataset.



(a) (b)

Figure 22 – The labeling process for the OCDC dataset: (a) Hand-annotating image-patches (green areas indicate tumor regions while red areas are normal tissue); (b) Yellow arrows indicate some adjustments requested by the pathologist.

The image-patch dataset was split into two subsets: the training set with 840 image-patches and the test set with 210 image-patches. Samples of the 15 WSI from our dataset are illustrated in Figure 23. Representative examples from the 1,050 produced image-patches are shown in Figure 24. Figure 24 (m-r) also shows the corresponding hand-annotated regions produced from the image-patches presented in Figure 24 (g-l).



(a) (b)

(c) (d)

(e) (f)

Figure 23 – Six WSIs from OCDC dataset. The images have different sizes.

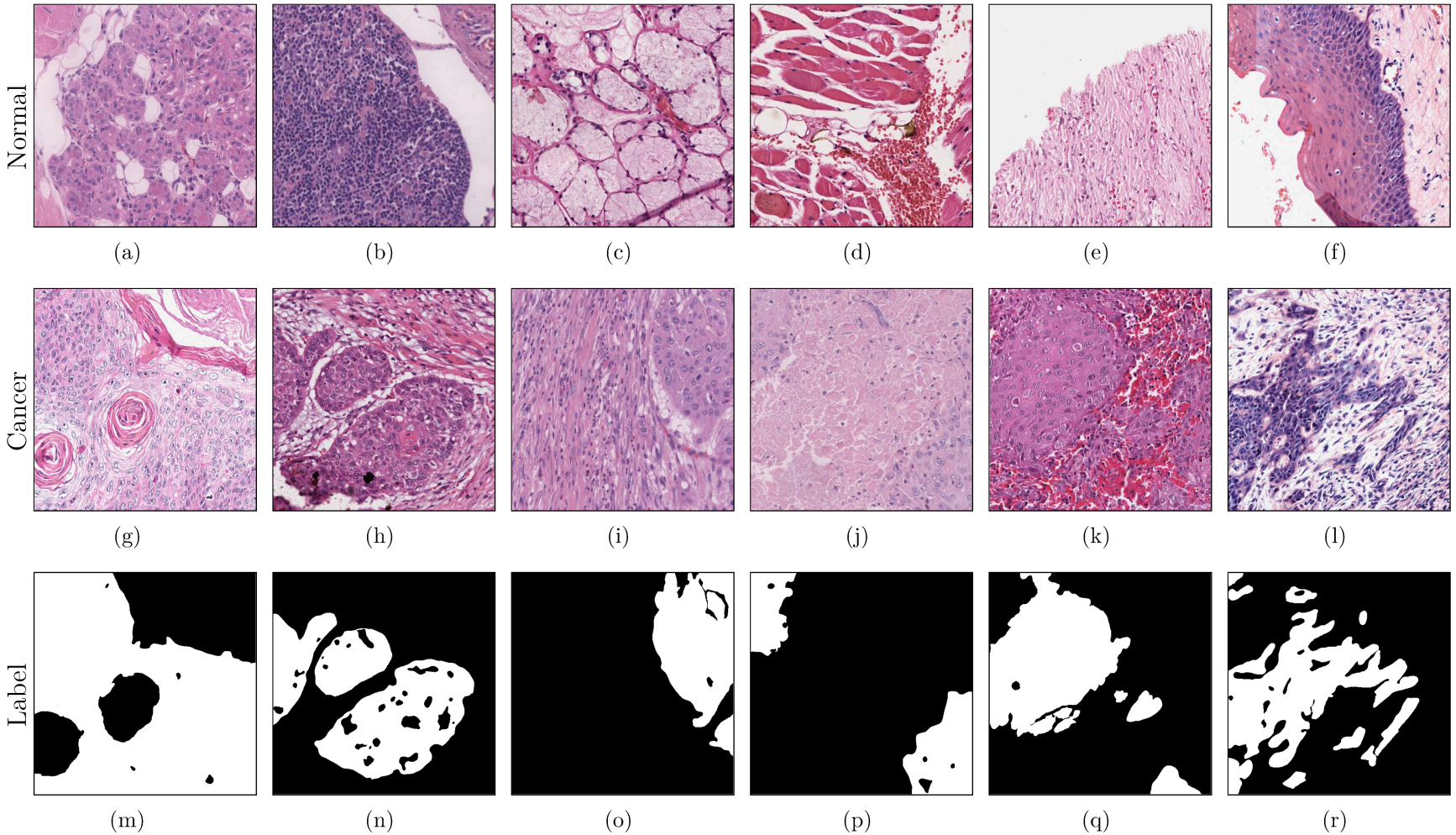


Figure 24 – Some image-patches (640×640 pixels) selected from our dataset to show the anatomic diversity present in whole slide images: 1st row (a-f) shows normal image-patches (serous and mucous salivary gland lobules, tonsil, striated muscle, and connective and epithelial tissues); 2nd row (g-l) shows OSCC image-patches with and without keratin pearl-containing areas; 3rd row (m-r) shows hand-annotated regions (in white) produced by a pathologist from the image-patches presented in the 2nd row (g-l).

3.2.2 CAMELYON16 dataset

The CAncer MEtastases in LYmph nOdes challeNge (CAMELYON16) dataset (LITJENS et al., 2018) is a publicly available dataset with a combination of 399 WSIs of sentinel lymph node with breast cancer metastases tissue sections collected from two medical centers in the Netherlands for the CAMELYON16 challenge (BEJNORDI et al., 2017). These 399 WSIs are split into 270 for training and 129 for testing.

The ground-truth data for training and testing are provided as XML files containing the vertices of the delineation of tumor regions at the WSI level. Unfortunately, the provided annotations for the CAMELYON16 dataset are not pixel-accurate annotations (the tumor annotations can include both normal and tumor cells as indicated by the red arrows in Figure 25). A training dataset was built extracting 109,278 image-patches of size 640×640 pixels from the 270 WSI training images. The normal image-patches were extracted only from the 159 normal-WSIs and the normal parts of the tissues from the 111 tumor-WSIs were not used. The testing dataset was built extracting 85,411 image-patches from the 129 WSI testing images. Only valid tissue areas were considered and no image-patches from the background areas were used in this testing dataset. These image-patch training and testing datasets were approximately balanced between tumor and normal image-patch areas. Figure 25 shows some image-patches extracted from the training and test datasets and Figure 26 shows a WSI from the CAMELYON16 dataset.

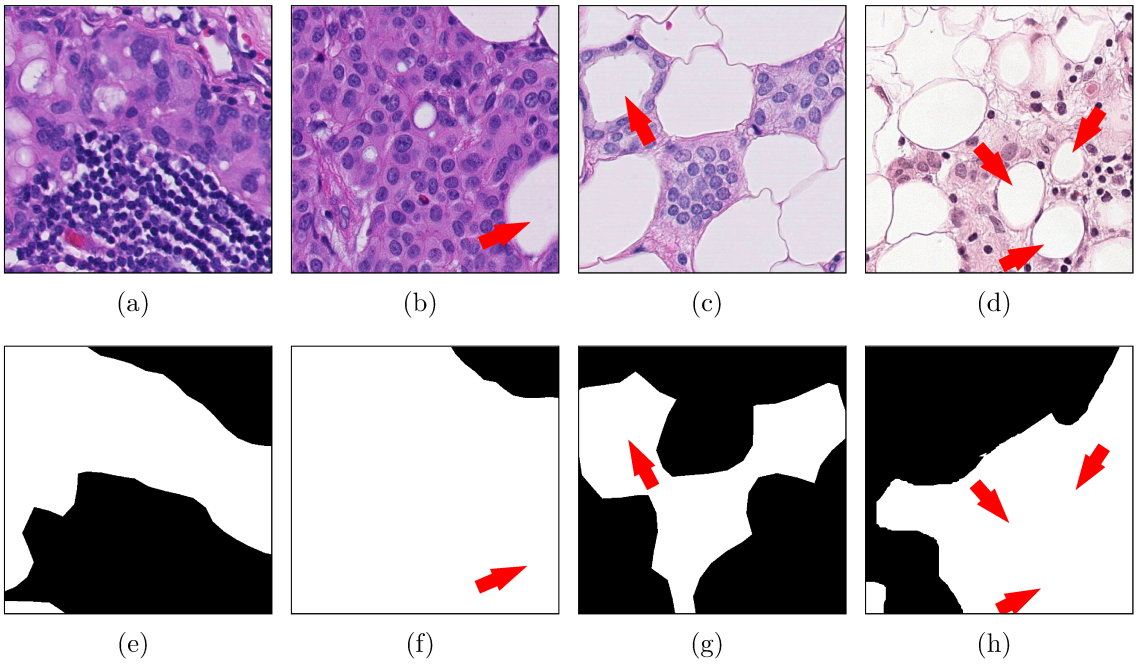
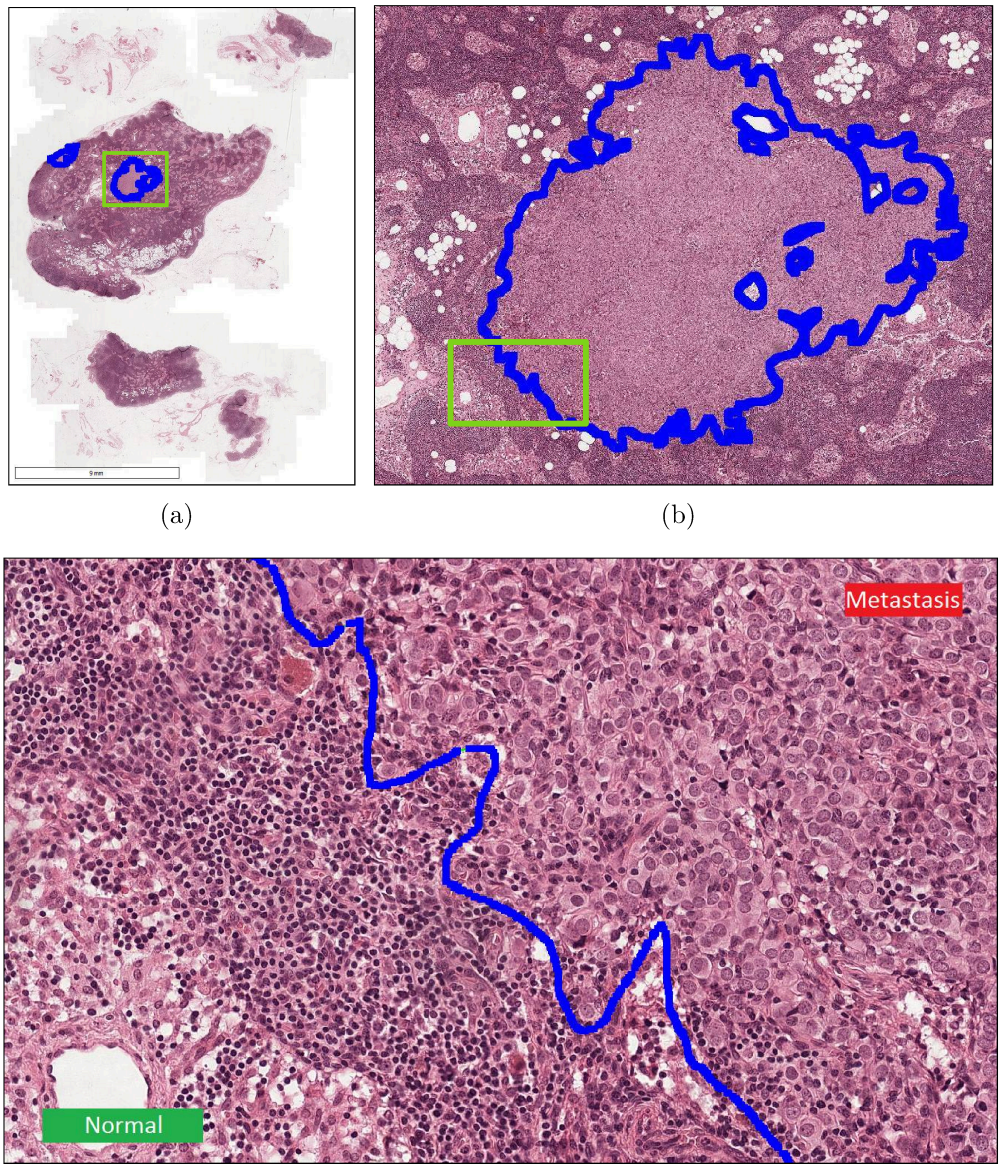


Figure 25 – Image-patches extracted from the CAMELYON16 dataset: 1st row (a-d) shows original images-patches (640×640 pixels); 2nd row (e-h) shows the annotated tumor regions (in white) for each corresponding original image-patch. Red arrows indicate some normal areas that are included in the tumor areas markdown.



(a) (b)

(c)

Figure 26 – WSI from the CAMELYON16 dataset: (a) The WSI at low resolution; (b) Selected ROI area; (c) Details of the ground-truth marking in the selected ROI area. (Images sourced from (LITJENS et al., 2018).)

3.2.3 ORCA dataset

The ORal Cancer Annotated (ORCA) dataset is a public labeled dataset containing 200 TMA images from representative OSCC tumor areas extracted from WSIs of the TCGA dataset (MARTINO et al., 2020; TCGA, 2022). The 200 labeled images were split into 100 images to train and 100 to test and contain images with two different dimensions: 4,500×4,500 or 2,250×2,250 pixels. The ground-truth data for training and testing are provided as mask images containing three different class labels: carcinoma pixels (white), tissue pixels not belonging to carcinoma (grey), and non-tissue pixels (black). In our experiments, a training dataset was built using 4,181 image-patches of

size 640×640 pixels extracted from the original 100 images of the training dataset. The testing dataset was built using 4,481 image-patches extracted from the 100 original images of the testing dataset. Figure 27 illustrates an original image from the ORCA dataset, some image-patches extracted, and their corresponding ground-truth.

The 200 original ORCA TMA images were also used to build a derived dataset in the same way as published by ORCA dataset authors. The 100 images from the training dataset and the 100 images from the testing dataset were resized to 512×512 pixels to compose the dataset identified as ORCA 512×512. This dataset was chosen for external validation and it is used in the same way as published by (MARTINO et al., 2020).

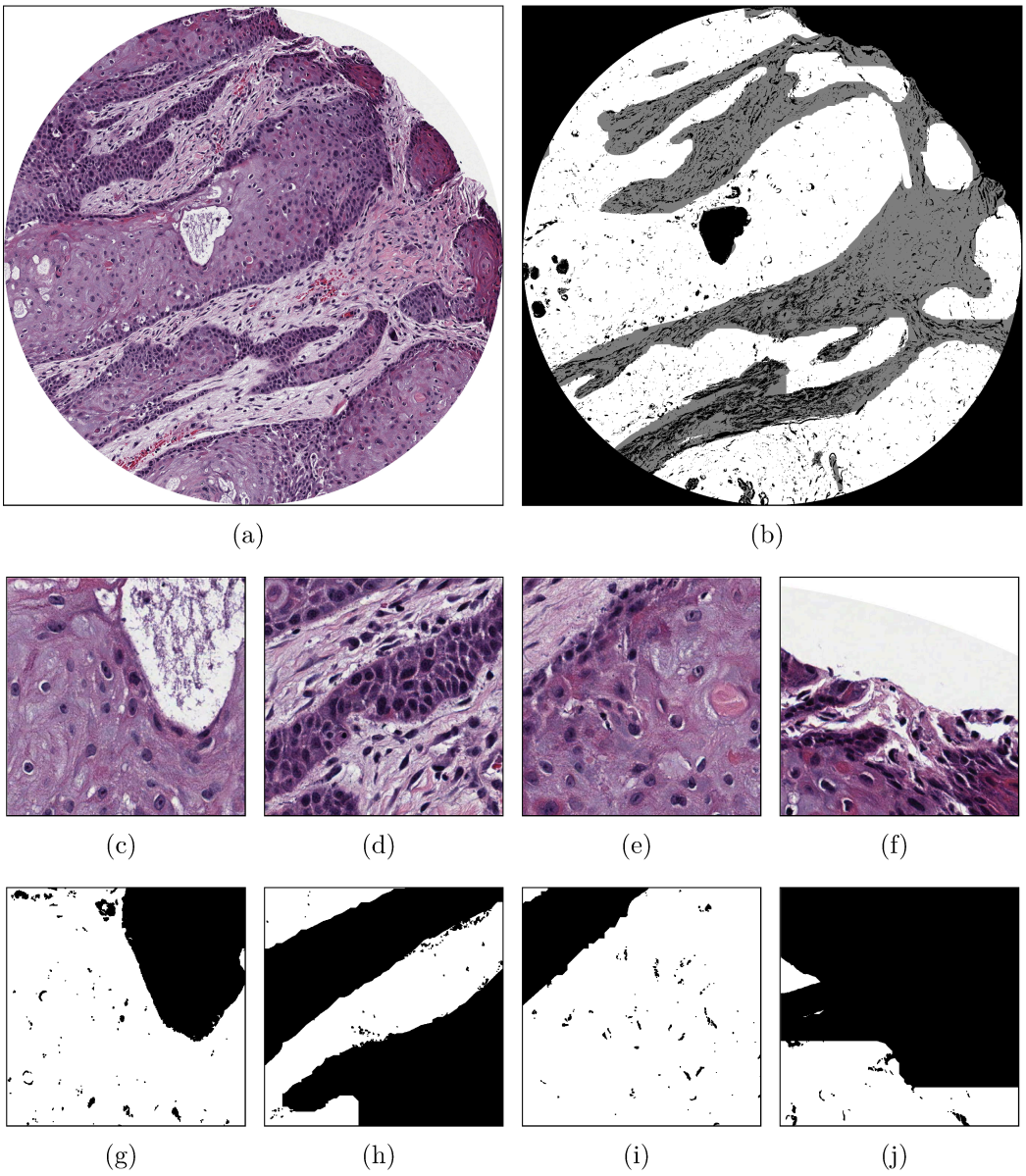
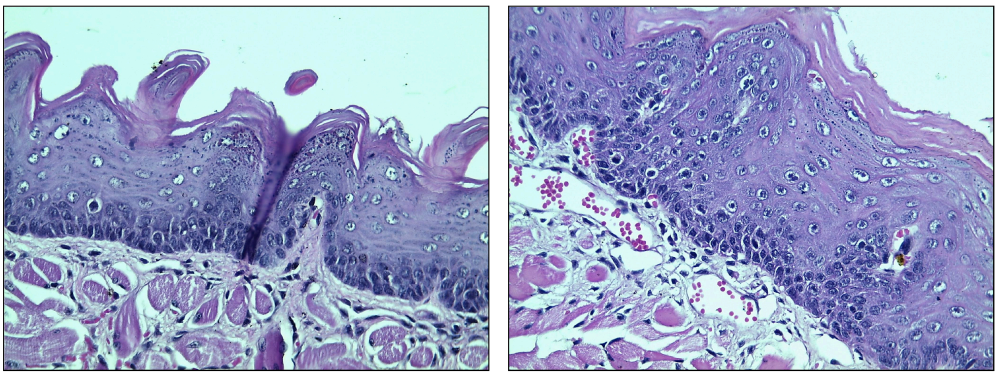


Figure 27 – Images from the ORCA dataset: 1st row (a-b) shows an original image and its corresponding ground-truth; 2nd row (c-f) shows image-patches (640×640 pixels) extracted from the original images; 3rd row (g-j) shows the annotated tumor regions for each corresponding image-patch from the 2nd row.

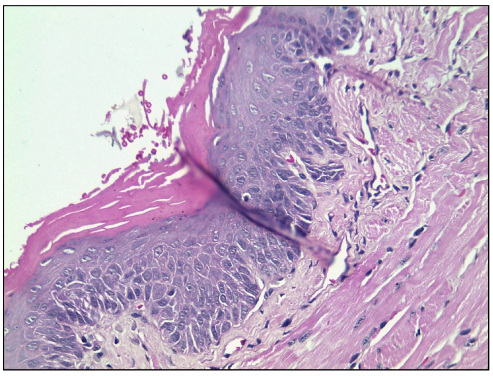
3.2.4 ONHI dataset

The Oral Nuclei Histological Images (ONHI) dataset (SILVA et al., 2019; SILVA et al., 2020) is a private dataset built using tongue tissue slides extracted from 30 dysplasia-harboring mice. The histological images were obtained using a Leica DM500 light microscope with a magnification of 40×. A total of 66 images were digitized and saved in TIFF format with a resolution of 2048×1536 pixels. A pathologist used the criteria described by (LUMERMAN; FREEDMAN; KERPEL, 1995) to classify each image into four predominant classes: healthy mucosa, mild, moderate, or severe dysplasia. Some of the obtained images for each predominant dysplasia grade are shown in Figure 28.

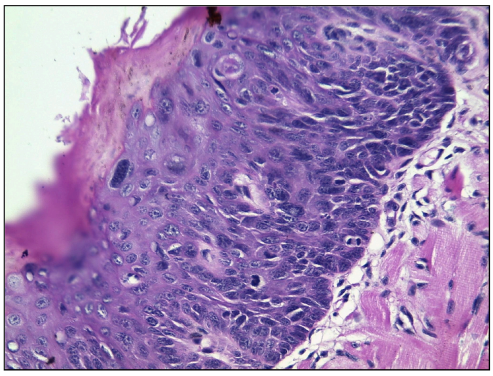


(a) Healthy mucosa.

(b) Mild dysplasia.



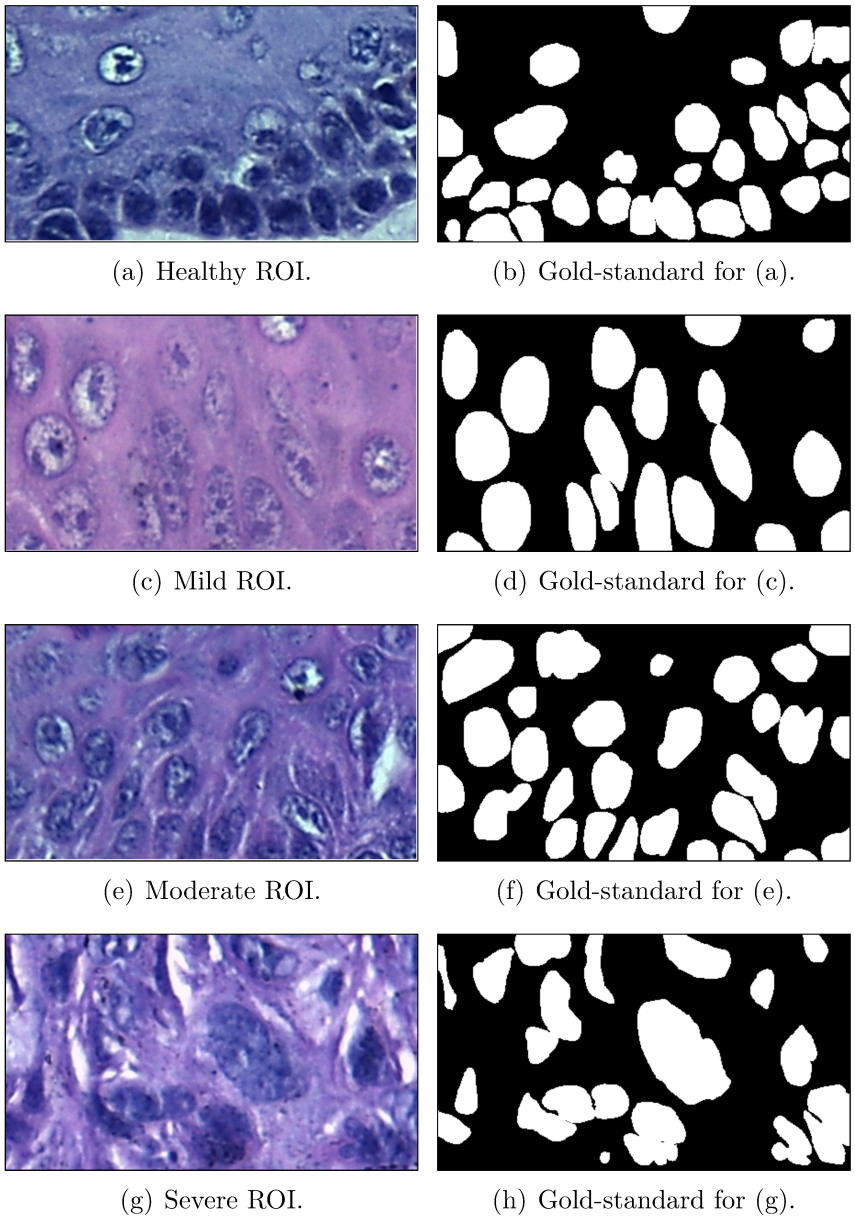
(c) Moderate dysplasia.



(d) Severe dysplasia.

Figure 28 – H&E-stained histological images of mice oral epithelial tissues. The images have a magnification of 40× and a resolution of 2048×1536 pixels.

After digitization, the images were cropped into ROIs with a fixed size of 448×256 pixels, resulting in 120 ROI images (30 ROI images for each dysplasia class). The 120 ROI images were randomly split into two subsets: the training set with 96 images (24 images for each class) and the test set with 24 images (6 images for each class). Examples of the digitized ROI images are illustrated in the first column in Figure 29 (a, c, e, g). The nuclei for the 120 ROI images were then manually marked by a specialist and used as a gold-standard. The second column in Figure 29 (b, d, f, h) shows the nuclei identified by the specialist for each corresponding ROI image.



(a) Healthy ROI. (b) Gold-standard for (a).

(c) Mild ROI. (d) Gold-standard for (c).

(e) Moderate ROI. (f) Gold-standard for (e).

(g) Severe ROI. (h) Gold-standard for (g).

Figure 29 – ROI images with resolution of 448×256 pixels from healthy and epithelial dysplasia-harboring tongue mice: 1st column shows original ROI images; 2nd column shows the annotated nuclei for each corresponding ROI image.

3.3 Dataset Augmentation

Even though our datasets provides a high number of training images (114,495 images as shown in Table 1), it is common knowledge that the more data a deep learning method has access to, the more effective and least prone to overfitting it will be. Overfitting occurs when the gap between the training error and test error is too large such as the network perfectly models the training data, causing poor performance on the testing set relative to the training set. Data augmentation is the most used approach to avoid overfitting problems in deep learning models (GOODFELLOW et al., 2014; SHORTEN; KHOSHGOFTAAR, 2019).

There are several data augmentation techniques and different approaches can be more or less suitable when applied to specific problems. For example, geometric transformations such as rotation and translation of images could be very useful to obtain an augmented dataset version that avoids positional biases present in the training data. On the other hand, rotation and flip operations could be not suitable, as, for example, in digit recognition tasks such as number 6 versus number 9. Most of the existing data augmentation operations focus on improving image classification models (CNN-based). For segmentation models (FCN-based), there is an extra effort to address, once data augmentation operations typically need to be done with transformations that are applied to both the images and their corresponding labels equally (SHORTEN; KHOSHGOFTAAR, 2019).

Standard data augmentation approaches consist of combining transformations to generate augmented data before training the model. For instance, geometric transformations are applied to create a larger dataset, thus, for each original image present in the dataset, a new image is generated after applying flip, rotation, distortion, and other operations. For a dataset of size N and T different transformations, an extended dataset of size $N \times T$ is generated. The extended image dataset containing the augmented and original images is then fed into the model at the training stage.

Despite on the fly augmentation had been employed in other types of images, like MRIs (POLONI; FERRARI, 2022), an on the fly data augmentation strategy, named random composition augmentation (RCAug), is proposed to train FCN-based segmentation models using H&E-stained histological images. RCAug is based on a random composition of different transformations applied to the input images on the fly. Given the input image and their corresponding label, a pipeline is executed with different geometric, distortion, color transfer, and generative image transformations. At each pipeline step, a roll dice with a probability of applying the transform equal to 0.5 is performed to choose if the corresponding transformation will or will not be executed. This parameter was chosen empirically on the investigated histological image datasets. After each step, the resulting image and their corresponding label, that may or may not have been augmented, are then used as the input to the next pipeline step.

The proposed strategy performs image augmentations on-the-fly during the model training and does not generate the augmented training image dataset in advance. This provides a much more flexible manner to increase the training dataset size. Since image datasets to train deep learning models are typically large in size, this approach allows overcoming the limitations of the storage space without penalizing the diversity and amount of image transformations. This strategy also generates an undefined amount of new training images using unique arrangements from different augmentation transformations on the fly. A total of eight transformations were used: horizontal flip, vertical flip, rotation, transpose, grid distortion, optical distortion, color transfer, and inpainting. Each used image transformation is detailed in Section 3.3.1 and Figure 30 illustrates an overview of

the proposed data augmentation strategy.

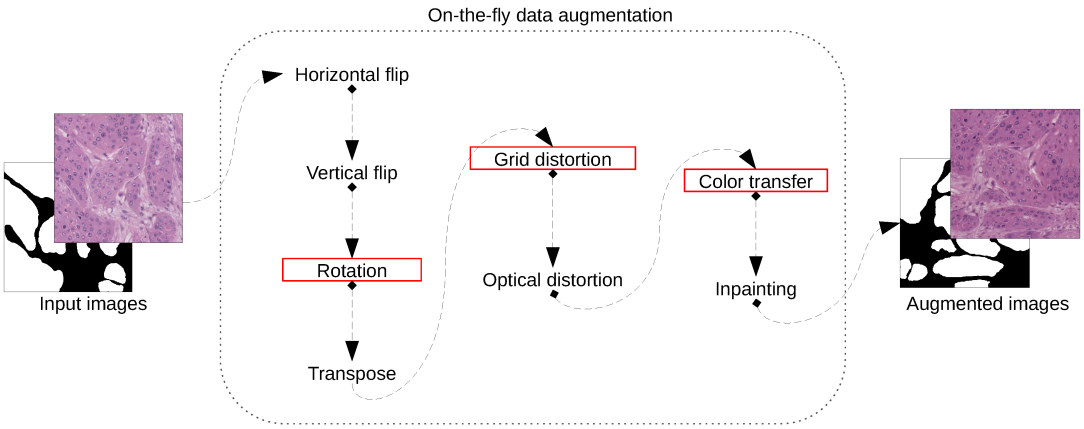


Figure 30 – Random Composition Augmentation - RCAug: before each image and label be used for training, a random composition of eight transformations is defined on the fly. In this example only the transformations highlighted in red were randomly chosen and applied to the input image and their corresponding label: Rotation, Grid distortion, and Color transfer.

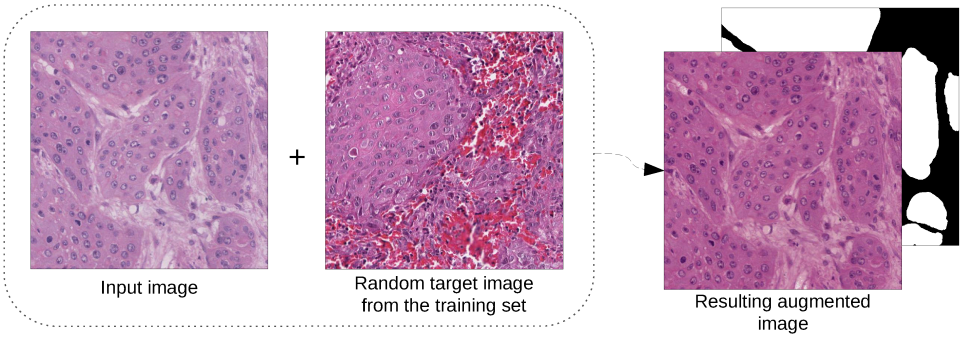
3.3.1 Data augmentation transformations

Digitized images from histological slides must be invariant to positioning and orientation. In this way, for histological images, geometric transformations are crucial to train deep learning models that avoid positional biases present in the training data. Furthermore, geometric image transformations are some of the most common and easiest transformations to implement. Four different augmentations based on geometric transformations were used: rotation, transpose, and horizontal and vertical axis flipping (BUSLAEV et al., 2020). These image transformations are applied to both images and their corresponding labels. Figure 32 (c/i), (d/j), (e/k), and (f/l) exemplifies the geometric transformations used as data augmentation.

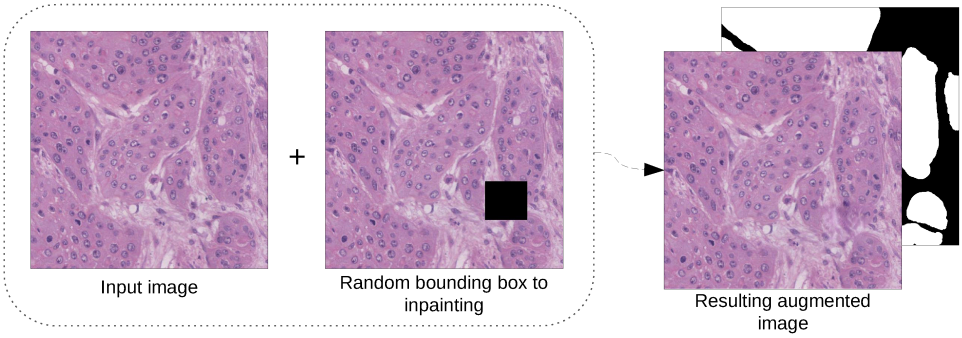
Tissue folding, image distortions, blurring, and even color differences caused by the staining process and lightning conditions are common undesired artifacts that reduces the quality of the histological images. To minimize problems related to distortions and color biases, augmentation transformations considering grid distortions, optical distortions, and color appearances were used. The distortion transformations used as data augmentation were implemented using the Albumentation (BUSLAEV et al., 2020) open-source Python library and are illustrated in Figure 32 (g/m) and (h/n). The used color augmentation transformation was proposed by (XIAO et al., 2019). This technique combines the structure of one image with the color appearance of another image to construct augmented images on-the-fly for each training iteration. More specifically, for each training image, a target image is randomly selected from the training set and a new augmented image is

generated through a color transfer procedure. Using the color transfer augmentation, the augmented image takes the color appearance from the target image while its structure remains the same as in the original image (XIAO et al., 2019). Figure 31(a) illustrates the used color transfer transformation.

To make the training process more robust to noise and occlusion problems, a combination of random erasing (ZHONG et al., 2017) with image inpainting (YU et al., 2018) transformations is also proposed as data augmentation. Random erasing was implemented by randomly selecting a 128×128 pixels patch on the training image and using a GAN-based image inpainting transformation to restore the missing pixels of the selected patch. The size of 128×128 pixels was chosen because this is the largest size with good results that our available hardware supports. Figure 31(b) illustrates the proposed GAN-based image inpainting transformation. The used image inpainting technique was proposed by (YU et al., 2018). We use the python implementation for the GAN-based image inpainting provided by the authors with the model pre-trained on ImageNet dataset (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) for 430k iterations. All the augmentation transformations were used as specified by their respective authors with the provided default configurations (no adjustment modifications were applied).

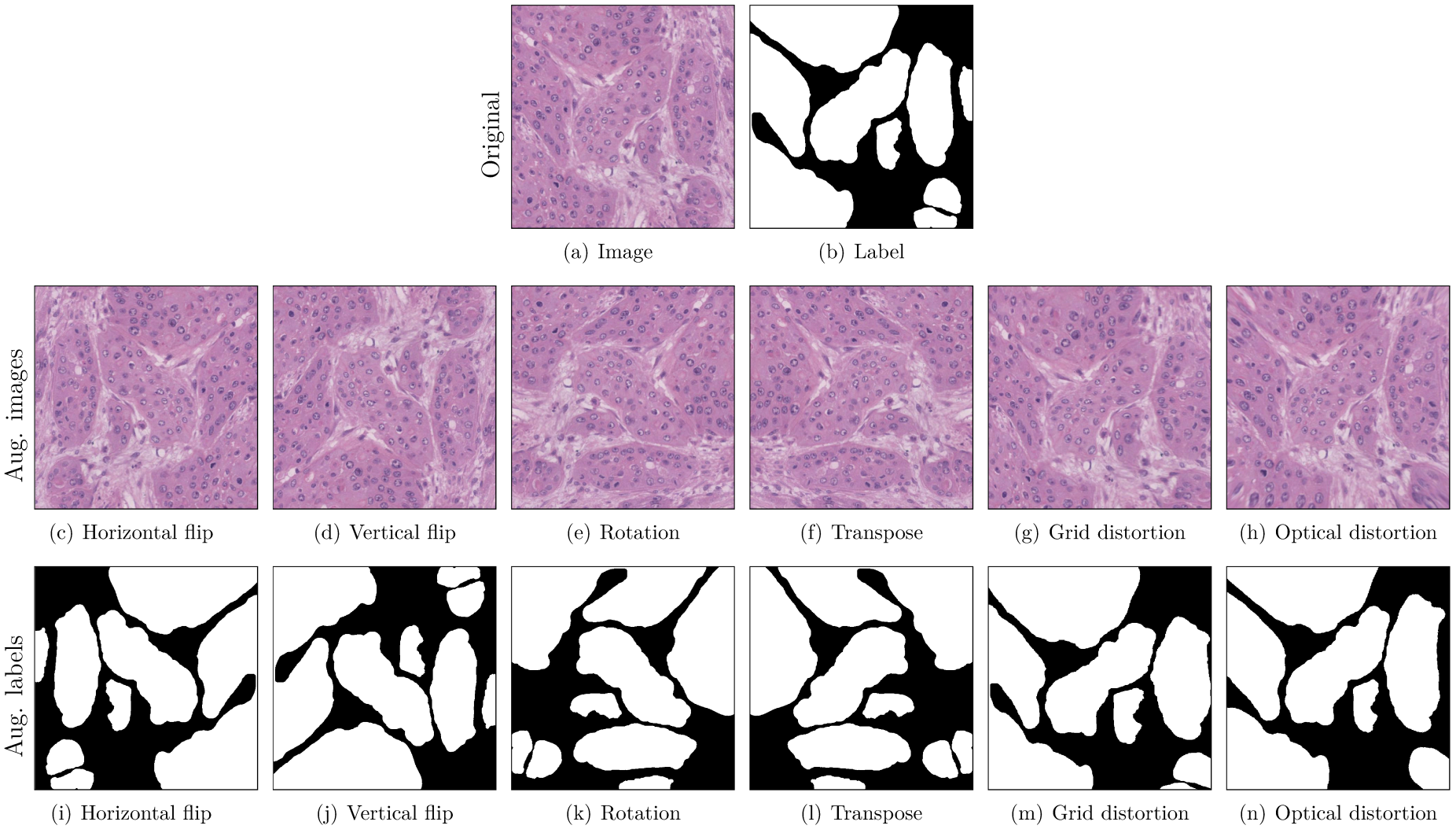


(a) Color transfer augmentation.



(b) Image inpainting augmentation.

Figure 31 – The used augmentation transformations that do not need to be applied in the corresponding label images: (a) Color transfer augmentation (XIAO et al., 2019); (b) GAN-based image inpainting augmentation (ZHONG et al., 2017; YU et al., 2018). The original corresponding labels remain unmodified for these transformations.



(a) Image (b) Label

(c) Horizontal flip (d) Vertical flip (e) Rotation (f) Transpose (g) Grid distortion (h) Optical distortion

(i) Horizontal flip (j) Vertical flip (k) Rotation (l) Transpose (m) Grid distortion (n) Optical distortion

Figure 32 – Augmentation transformations applied to both the images and their corresponding labels equally: (a/b) original images; (c/i) horizontal flip; (d/j) vertical flip; (e/k) rotation; (f/l) transpose; (g/m) grid distortion; (h/n) optical distortion.

3.4 Tissue Filter

Identification and segmentation of malignant tumor regions in H&E-stained histological images are crucial tasks for the diagnosis of OSCC. Typical H&E-stained WSIs, as shown in Figure 23, generally have large background areas that must be discarded. To reduce computation time, a simple and fast color-based tissue filter is performed on a 32× scaled-down WSI version to remove non-relevant areas while preserving the major tumor regions inside the tissue. Some scale-down values were evaluated and the 32× factor achieved good results, removing non-relevant areas while preserving the relevant tumor regions. The non-relevant tissue regions are removed from the scaled-down image by discarding pixels whose values are not in the blue/pink color interval using the HSV color space. In other words, pixel values in the scaled-down image that are not in the range (260°, 0.11, 0.19) and (340°, 1.0, 1.0) in the HSV color space are discarded. Figure 33 illustrates the proposed color-based tissue filter and Figure 34 shows some resulting tissue regions extracted by our tissue filter from the WSIs presented in Figure 23. The HSV tissue filter is applied whenever non-ROI images (WSI/TMA) are processed by our proposed segmentation method. For datasets in which the full resolution images are smaller than 640×640 pixels, as, in the case of the ORCA 512×512 dataset, the scale-down process is not necessary.

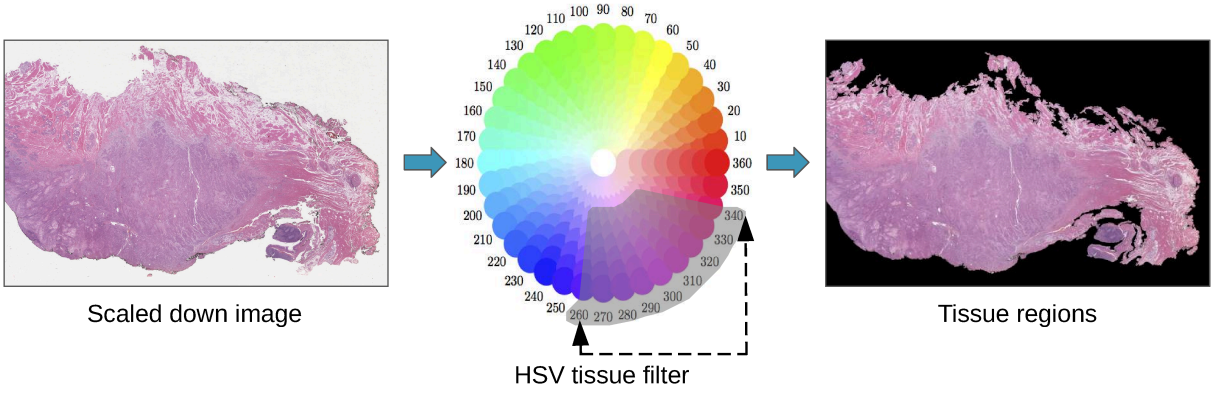
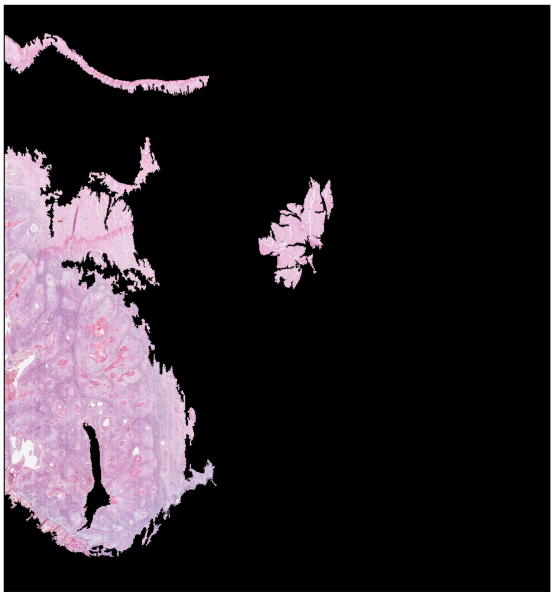


Figure 33 – The proposed color-based tissue filter: pixel values that are not in the highlighted range of the HSV color space are discarded. In addition to discarding background and some not relevant tissue areas, the proposed tissue filter preserves the major tumor regions inside the tissue.



(a) Tissue region for Figure 23 (a).



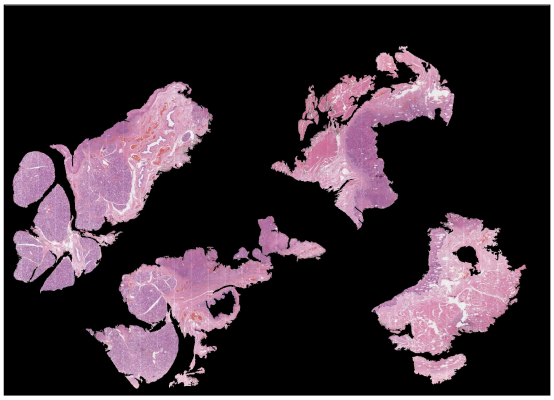
(b) Tissue region for Figure 23 (b).



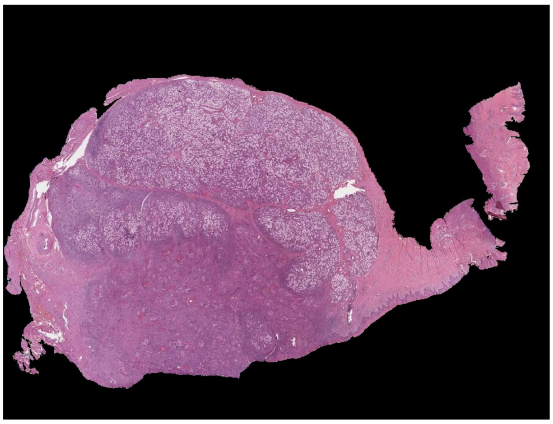
(c) Tissue region for Figure 23 (c).



(d) Tissue region for Figure 23 (d).



(e) Tissue region for Figure 23 (e).



(f) Tissue region for Figure 23 (f).

Figure 34 – The tissue regions identified from Figure 23 by our proposed HSV tissue filter. Besides being simple and fast, the proposed tissue filter removes the most background, irrelevant structures, scanning artifacts, and remains all the major tumor regions inside the tissue. Only the image-patches that contain the remaining tissue regions are processed by our FCN model.

3.5 Histological Image Segmentation

The resulting tissue regions identified by our proposed tissue filter are then split from the original WSI at 20× magnification into image-patches of size 640×640 pixels. These image-patches are converted from the RGB to the CIE L*a*b* color space before being submitted to our proposed FCN model. The L* channel from the CIE L*a*b* color space is independent of chromatic information. It ranges from 0 to 100, and only indicates lightness, while the other two channels are chromatic yellow-blue and red-green opponent channels and range from -128 to 127. The pixel values for each CIE L*a*b* channel are normalized to the interval between 0 and 1 (using min-max normalization considering the range of each channel) before FCN processing. The influence of different color spaces and different image-patch sizes were explored and the image-patch with 640×640 size using CIE L*a*b* color space achieved the best results (see Section 4.2).

Each of the image-patches is then submitted to our trained FCN-based segmentation model, with no overlapping, to predict the cancer probability (a value between 0.0 and 1.0) for each pixel present in the image-patch. Image patches containing only removed background areas are not processed by our FCN model and have all pixel values predicted with the lowest possible cancer probability value (0.0). The final probability image for the WSI is built based on the cancer probability value achieved for each pixel. A heatmap using a blue-to-red color map is shown for better visualization of the probabilities. For the final tumor regions segmentation, a threshold at 0.5 is performed in the probability image to create a first binary version of the segmented tumor regions. Different threshold values have been tested (0.5, 0.7, and 0.9) and the 0.5 value achieved the best results. The final version is achieved by identifying the connected components with 8-adjacency and removing small resulting regions that are smaller than 200 μm.

3.5.1 Fully Convolutional Network Architecture

We propose to use an FCN to address the challenge of patch-based segmentation of tumor regions in oral cavity-derived H&E-stained images at a pixel level. The proposed FCN architecture, which consists of a couple of encoding and decoding layers, is based on U-Net model (RONNEBERGER; FISCHER; BROX, 2015) and is illustrated in Figure 35. The weights for all convolutional layers are initialized by a normal distribution with an average of zero and the standard-deviation value calculated according to Equation 4:

$$std = \sqrt{\frac{2}{kernel_size \cdot in_channels}}, \tag{4}$$

where *kernel_size* (height × width) is the size of the convolutional kernel and *in_channels* is the number of feature channels used in the previous convolutional layer. The weight initialization process, the number of feature channels, kernel size and stride parameters

for convolutional and pooling operations used are the same as defined in the original U-Net model (RONNEBERGER; FISCHER; BROX, 2015). All the other parameters were chosen empirically.

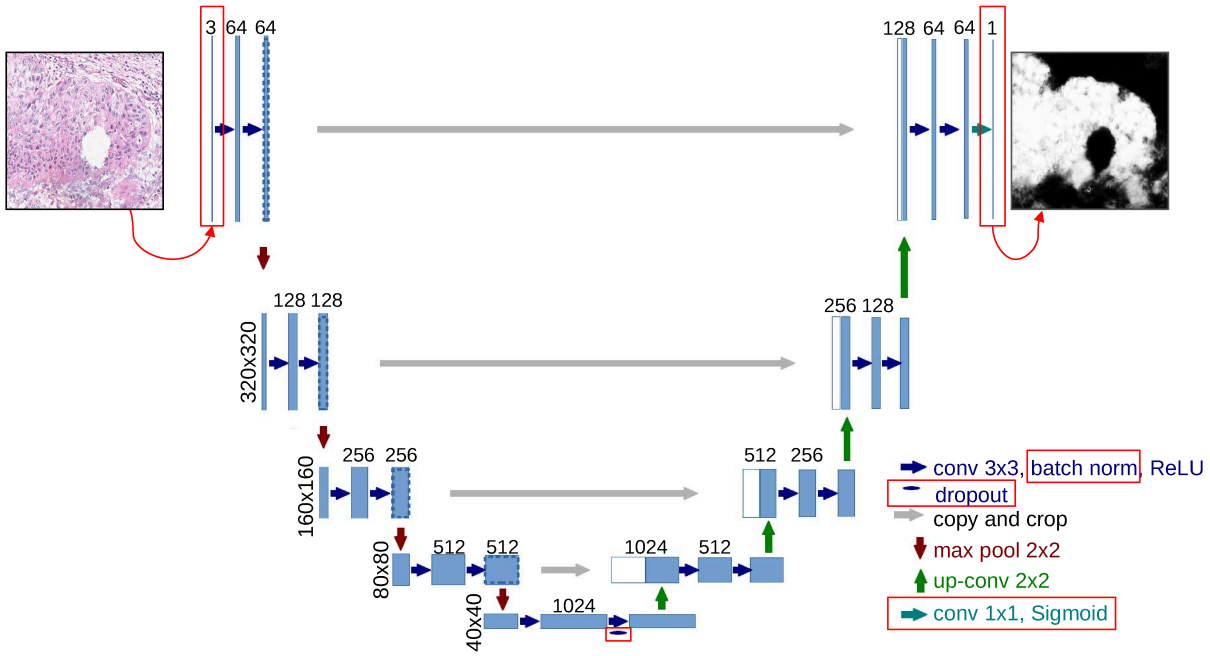


Figure 35 – Our FCN architecture is based on U-Net model (RONNEBERGER; FISCHER; BROX, 2015). The main proposed changes are highlighted in red, the number of convolutional feature channels and the height×width of images for each layer is denoted on their corresponding boxes. The model outputs a gray-level image with the cancer probability for the input image.

Batch normalization and the ReLU activation function are used after all convolutional layers. Except for the last convolutional layer, where the kernel size 1×1 is used, all other convolutional layers use kernel size 3×3 with zero-padding to ensure it keeps the same image size of its previous layer. The same pooling operations as proposed in the original U-Net model (max-pooling operations with kernel size 2×2 and stride 2) are performed in all encoding layers for downsampling. After each downsampling step, the number of feature channels is multiplied by two in the next convolution step (64 to 128, 128 to 256, 256 to 512, 512 to 1024). A dropout layer with a 0.5 drop rate is used after the last encoding layer to help prevent over-fitting and make the model more robust and generalizable. For all decoding layers, an upsampling using transposed convolution operations with kernel size 2×2 and stride 2 is performed. Each upsampling step halves the number of feature channels for the next step (1024 to 512, 512 to 256, 256 to 128, 128 to 64). At the final layer, a 1×1 convolution with a Sigmoid activation function is used to squash the values to the range between 0 and 1, resulting in a gray-level image that predicts the cancer probability for the input image at the pixel level.

The image-patches created according to Section 3.2 were used to train and validate our FCN model. The model was trained using the image-patches from the training sets, using Adam optimizer (KINGMA; BA, 2014), RCAug as data augmentation strategy, and a batch size of one image-patch. The configuration for the Adam optimizer used the value provided by the PyTorch framework (default learning rate of 1×10^{-3}) (PASZKE et al., 2017). This optimizer automatically controls the learning rate for each parameter during the training process, so it is not necessary to manually set momentum and decay.

3.5.2 Nuclei Segmentation on Oral Histological Images

The diagnosis and grading of several types of cancer are made by pathologists through a complex and time-consuming process where some cytological features, mostly the nuclei morphology and their spatial arrangement, are analyzed. Automated methods capable to perform accurate and effective nuclei segmentation can directly influence the precision of the analyzed features and remains a complex and challenging task (IRSHAD et al., 2014; MCCANN et al., 2015; XING et al., 2018). We propose to use our FCN-based segmentation method to segment nuclei from the oral tumor regions previously identified aiming to help further works to extract and quantify features from those nuclei regions.

Due to the complexities of the WSIs, and the fact that we do not have the gold standard for the nuclei regions in the WSIs, the performance of the nuclei segmentation task for our proposed segmentation method was evaluated using a dataset that is best suited for the nuclei segmentation task – the ONHI dataset (see subsection 3.2.4). The images from the ONHI dataset are preselected ROIs with the nuclei marked by a pathologist. The ONHI dataset also was used to evaluate one of the most prevalent artifacts that can lower the quality of the digitized histological images: the color and contrast differences caused by the staining and/or digitization process.

Color variations could play a high influence on the automated analysis of histological images leading to the poor performance of the segmentation algorithms (KHAN et al., 2014). Color normalization methods have been used to minimize the influence of color variations and improve the performance of the automated segmentation methods (TOSTA et al., 2019). This improvement may contribute to increasing the sensitivity of the segmentation process, feature extraction, and classification of CAD systems. The impact of color normalization techniques in H&E-stained histological images applied to the automated nuclei segmentation task was evaluated. For this, we investigate the proposed method and other segmentation methods under normalized-color and non-normalized-color images to evaluate the influence of normalization techniques in the nuclei segmentation step. An evaluation of the impact in different color space transformations (RGB, HSV, and CIE L*a*b*) also was performed.

Our proposed approach to address the challenge of nuclei segmentation on oral tissue images at a pixel-level is shown in Figure 36 and is the same detailed in subsection 3.5.1

with slight modifications. In the training step, we directly apply the histological ROI images and their corresponding masks (see Figure 29) to train our FCN model. After the last convolution step is performed, the Otsu threshold (OTSU, 1979) is applied to the resulting gray-level image to binarize the resulting mask image that predicts the nuclei for the input ROI image. The Otsu threshold was employed as an alternative to the fixed threshold value previously evaluated (0.5, 0.7, 0.9). However, as the last convolutional layer of our proposed FCN model outputs a probability map between 0.0 and 1.0, we conclude that a fixed threshold value of 0.5 would be the best choice once produces similar results while needing low computational resources. The ROI images created according the subsection 3.2.4 were used to train and validate our FCN model. The training was performed using the Adam optimizer, batch size of one ROI image, and the same random composition data augmentation.

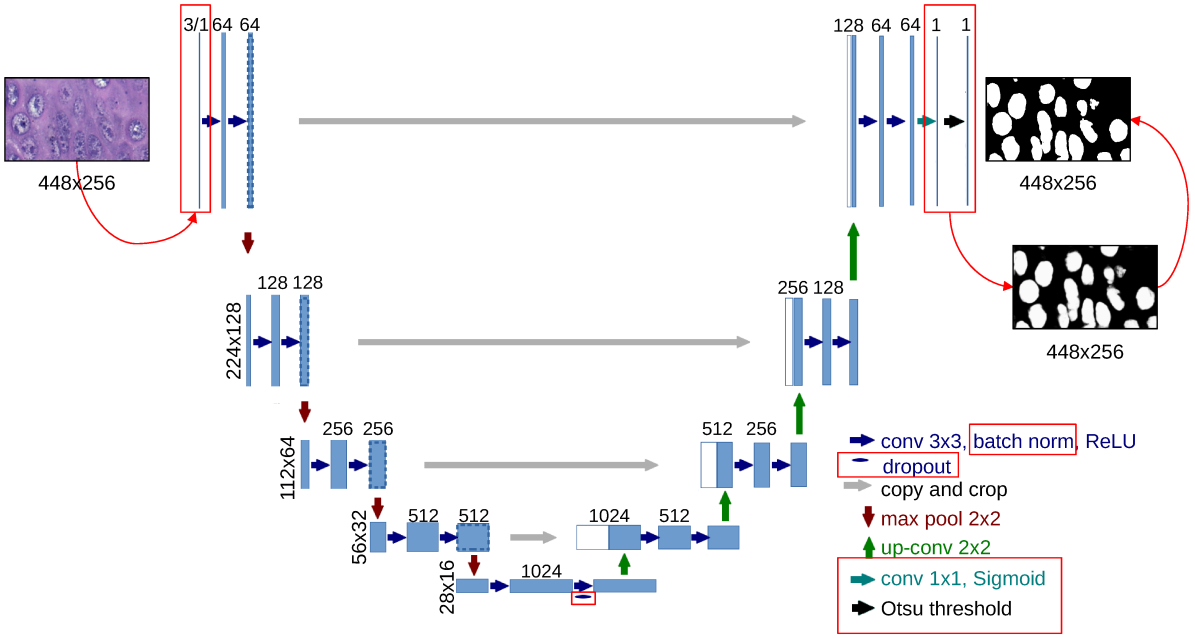


Figure 36 – The adaptation of the FCN architecture illustrated in Figure 35. The main changes of the original U-Net model are highlighted in red. The model outputs a binary image with the nuclei prediction after an Otsu threshold be applied. The first layer support images with one or three color channels as input.

3.6 Final Considerations

This chapter presented our proposed method to address the challenge of detection and segmentation of the OSCC tumor regions in WSIs. The proposed method is based on a fully convolutional network, CIE L*a*b* color space, and image-patches subdivision that localizes and performs refined segmentation of oral cavity-derived tumor regions in H&E-

stained histological whole slide images. Once the tumor regions of interest have been identified, the same FCN architecture with minor modifications is proposed to perform nuclei segmentation from those identified tumor regions.

A data augmentation strategy, named RCAug, was also presented to train FCNs to segment OSCC tumor regions in H&E-stained histological images. Given the input image and their corresponding label, a pipeline with a random composition of geometric, distortion, color transfer, and generative image transformations is executed on the fly.

This chapter also described the four image datasets used to evaluate the performance of our proposed method for the segmentation of tumor and nuclei regions in histological images. The image datasets have images from different types of disease: OSCC and non-OSCC. The next chapter presents the performed experimental evaluations for our proposed method using the image datasets presented in this chapter.

Chapter 4

Experimental Results

This chapter describes the experimental evaluations and the metrics used to evaluate the proposed methods. First, the most commonly used metrics for image segmentation are described. The quantitative evaluation of our proposed methods is performed using these metrics. Next, the experimental evaluation results are presented and discussed. Image datasets of different diseases were used in the experimental evaluation and Figure 37 presents the general diagram of the performed experiments.

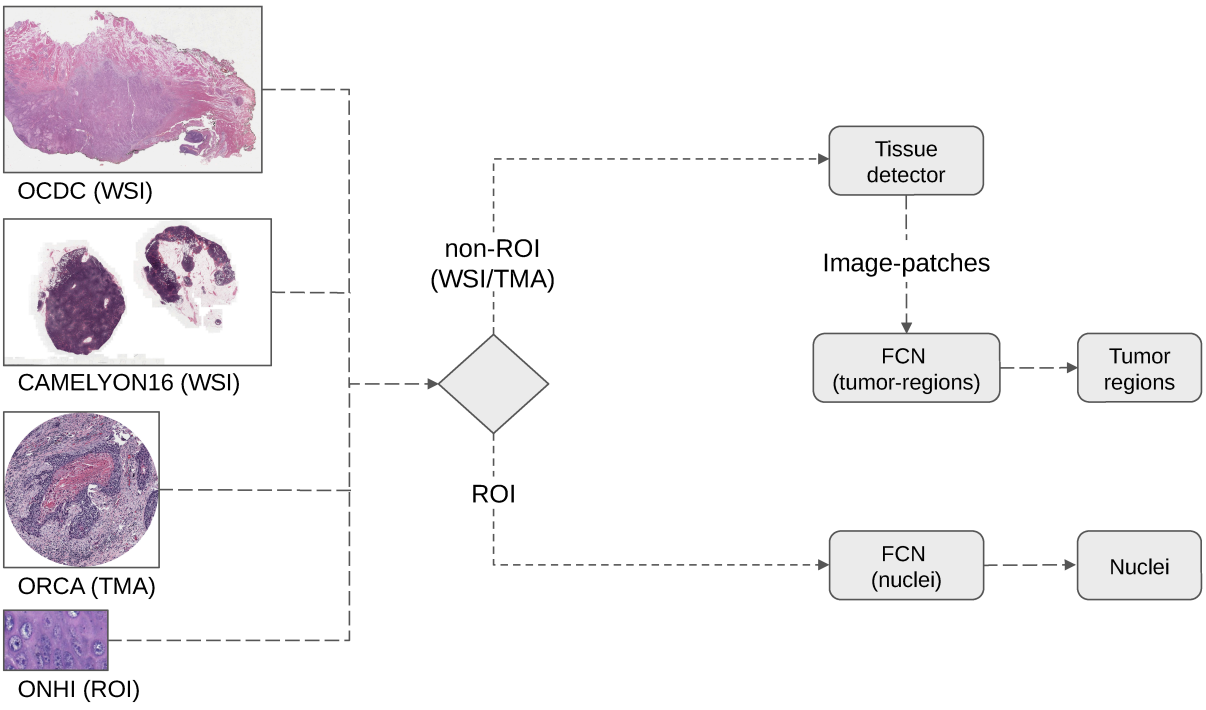


Figure 37 – General diagram for the experimental evaluations performed using different image datasets. The tissue detector process is applied to non-ROI images. The influence of color normalization techniques in nuclei segmentation and the impact of different color space transformations are evaluated in the ROI images. Data augmentation was used to train the FCN model in all experiments.

4.1 Evaluation Metrics

The performance of the proposed segmentation method was measured by calculating the overlapping regions of the resulting automated segmented image and the regions of a reference image segmented by the specialist (gold-standard). The average performance was measured by some of the most commonly used quantitative criteria concerning pixel-level classification.

Accuracy indicates the overall effectiveness of a classifier and is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \tag{5}$$

Precision indicates if the segmentation results bring only desired areas (positive class labels) and do not bring any non-desired areas (negative class labels):

$$Precision = \frac{TP}{TP + FP}. \tag{6}$$

Sensitivity (or Recall, or True Positive Rate) indicates the effectiveness of the segmentation results to identify positive class labels and is defined as:

$$Sensitivity = \frac{TP}{TP + FN}. \tag{7}$$

F1-score (or Dice Coefficient) is the harmonic mean of the precision and recall and can be defined as:

$$F1 - score = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}. \tag{8}$$

IOU (or Jaccard index) emphasizes the similarity between gold-standard and segmentation results for both positive and negative class labels and is defined as:

$$IOU = \frac{TP}{TP + FP + FN}. \tag{9}$$

Specificity measures the proportion of negative class labels correctly identified and is defined as:

$$Specificity = \frac{TN}{TN + FP}, \tag{10}$$

where TP means true positives (the amount of correctly detected pixels, i.e. cancer detected as cancer), TN means true negatives (the amount of correctly undetected pixels, i.e. non-cancer detected as non-cancer), FP means false positives (the number of incorrectly detected pixels, i.e. non-cancer detected as cancer) and FN means false negatives (the number of incorrectly undetected pixels, i.e. cancer detected as non-cancer) (SOKOLOVA; LAPALME, 2009; WANG; WANG; ZHU, 2020; XUN et al., 2021).

4.2 Experiments for Automated Detection and Segmentation of Tumor Regions

To measure the segmentation quality of our proposed method for detection and segmentation of the OSCC tumor regions, experimental evaluations were performed using the image-patches from the test subsets detailed in Section 3.2. Experiments with different color spaces, different image-patch sizes, and different data augmentation strategies were evaluated.

The results were compared quantitatively and qualitatively with the tumor regions manually segmented by the specialist. The detection and segmentation of tumor regions at the WSI, TMA, and image-patch levels were visually analyzed by a specialist as a qualitative evaluation. As there are no other similar public datasets (with pixel-level labeled OSCC tumor region-derived H&E-stained WSIs) available for comparison, for further validation and comparison with other proposals, we also applied our proposed method in a dataset of sentinel lymph nodes with breast cancer metastases (CAMELYON16) and in a dataset of TMA images from OSCC tissues (ORCA).

The segmentation method, the data augmentation transformations, and the experimental evaluations were implemented using the PyTorch framework (PASZKE et al., 2017) with the Albumentation (BUSLAEV et al., 2020) open-source Python library. The FCN segmentation model was trained using a desktop computer (Intel Core i7 3.4GHz ×8 processor, 32 GB memory, 1TB SSD) equipped with GeForce GTX 1050 Ti graphic card and Ubuntu 20.04 operational system. The training time for the OCDC dataset during 400 epochs using 840 images of size 640×640 pixels was about five days. For ORCA dataset the elapsed time to train the model during 100 epochs was about seven days using 4,181 images of size 640×640 pixels. For ORCA 512×512 we train the FCN model during 400 epochs in one day. The elapsed time to train using the CAMELYON16 dataset during 10 epochs using 109,278 images was about 15 days. After training, the elapsed time to process the input images was about 0.8 seconds for the images of 640×640 pixels size and 0.7 seconds for the 512×512 pixels size. To provide a better understanding and make this work as reproducible as possible, the source code is publicly available at <https://github.com/dalifreire/tumor_regions_segmentation>.

4.2.1 Color spaces and image-patch sizes

The influence of different color spaces and different image-patch sizes were evaluated using the OCDC dataset (subsection 3.2.1) to choose the best configuration. The RGB, HSV, and CIE L*a*b* color spaces were chosen for color space evaluation, whereas for image-patch sizes, the following sizes were evaluated: 160×160, 320×320, and 640×640 pixels. The 640×640 pixel size was chosen because this is the largest size that our available

hardware supports. We train our FCN model using these different configurations of color spaces and image-patch sizes to find out the best configuration.

Our proposed FCN model was trained, according to the procedure described in subsection 3.5.1, using the OCDC image-patches training set as presented in subsection 3.2.1 (see Table 1). A separated training was performed during 500 epochs for each different image-patch size (160×160, 320×320, and 640×640 pixels) in combination with different color spaces (RGB, HSV, and CIE L*a*b*). Table 2 summarizes the average quantitative results achieved for each combination of color space and image-patch size.

Table 2 – Average quantitative results for color spaces and image-patch sizes. The image-patch configuration with 640×640 size using L*a*b* color space achieved the best results in the OCDC dataset.

Size	Color	Accuracy	Precision	F1	IOU	Sensitivity	Specificity
160×160	L*a*b*	95.0%	78.8%	84.4%	72.9%	90.5%	95.7%
	HSV	96.9%	89.0%	89.6%	81.2%	90.1%	98.0%
	RGB	96.2%	84.3%	87.8%	78.3%	91.6%	97.0%
320×320	L*a*b*	96.2%	86.9%	87.3%	77.5%	87.8%	97.7%
	HSV	96.5%	86.3%	88.7%	79.7%	91.3%	97.4%
	RGB	96.8%	87.8%	89.4%	80.9%	91.1%	97.8%
640×640	**L*a*b***	**97.6%**	**91.1%**	**92.0%**	**85.2%**	**92.9%**	**98.4%**
	HSV	97.1%	89.2%	90.3%	82.4%	91.5%	98.0%
	RGB	97.2%	91.0%	90.8%	83.1%	90.5%	98.4%

From Table 2 we can see that small image-patches performed worse because of the little contextual information while the image-patches with a larger size contain more contextual information. The color spaces that better separate the lightness information from the color information, like the CIE L*a*b* color space, combined with larger image-patch sizes achieved the best results for segmentation purposes. Only the best configuration (CIE L*a*b* color space with the largest image-patch size) was used in the further experiments.

4.2.2 Evaluation of data augmentation strategies

The influence of different data augmentation transformations on the training of FCN models to segment OSCC tumor regions was evaluated using the three OSCC datasets: OCDC, ORCA, and ORCA 512×512. The following setup was evaluated: no augmentation; color transfer augmentation; image inpainting augmentation; geometric augmentation; distortion augmentation; standard augmentation using all transformations, and the proposed pipeline with random compositions of all transformations on the fly (RCAug, as described in Section 3.3). The FCN segmentation model was separately trained using each of these different configurations. The OCDC and ORCA 512×512 datasets were trained during 400 epochs while the ORCA dataset was trained during 100 epochs. The number of training epochs was chosen based on the empirical performance of the model

Table 3 – Average quantitative results for test images from the OCDC dataset.

Augmentation	Accuracy	Precision	F1	IOU	Sensitivity	Specificity
No augmentation	92.44%	89.83%	68.41%	51.98%	55.24%	**98.91%**
Color transfer	95.50%	**90.70%**	83.65%	71.90%	77.62%	98.61%
Image inpainting	96.36%	90.54%	87.29%	77.44%	84.27%	98.47%
Geometric	96.43%	86.38%	88.21%	78.91%	90.13%	97.53%
Distortion	96.70%	89.49%	88.78%	79.82%	88.08%	98.20%
All (standard)	95.72%	88.11%	85.08%	74.04%	82.26%	98.07%
RCAug	**96.99%**	88.29%	**90.06%**	**81.92%**	**91.91%**	97.88%

Table 4 – Average quantitative results for test images from the ORCA dataset.

Augmentation	Accuracy	Precision	F1	IOU	Sensitivity	Specificity
No augmentation	82.73%	78.75%	79.13%	65.46%	79.50%	84.99%
Color transfer	84.58%	**82.27%**	80.98%	68.04%	79.73%	**87.98%**
Image inpainting	82.87%	74.50%	81.01%	68.08%	**88.77%**	78.74%
Geometric	84.02%	81.10%	80.43%	67.26%	79.77%	86.99%
Distortion	84.44%	79.11%	81.73%	69.10%	84.53%	84.38%
All (standard)	84.42%	79.61%	81.53%	68.82%	83.54%	85.03%
RCAug	**84.60%**	78.58%	**82.14%**	**69.69%**	86.04%	83.59%

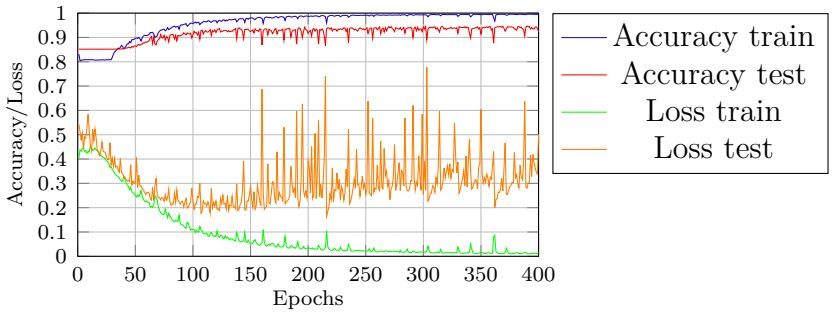
Table 5 – Average quantitative results for test images from the ORCA 512×512 dataset.

Augmentation	Accuracy	Precision	F1	IOU	Sensitivity	Specificity
No augmentation	85.45%	72.22%	75.45%	60.57%	78.97%	88.01%
Color transfer	84.62%	69.59%	74.88%	59.85%	**81.04%**	86.03%
Image inpainting	86.18%	74.41%	76.16%	61.50%	78.00%	89.42%
Geometric	85.71%	**76.63%**	73.83%	58.51%	71.22%	**91.43%**
Distortion	86.23%	73.93%	76.51%	61.96%	79.28%	88.97%
All (standard)	85.92%	76.45%	74.48%	59.34%	72.62%	91.17%
RCAug	**86.46%**	73.81%	**77.16%**	**62.81%**	80.83%	88.68%

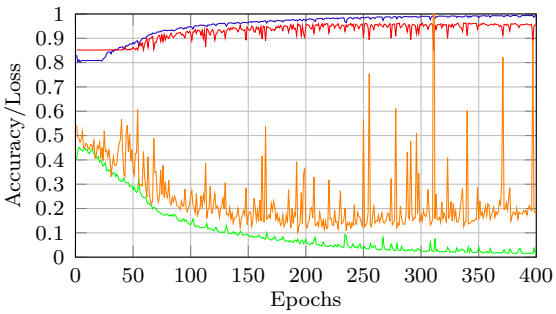
for the 640×640 size images. In the case of the ORCA database, this value was selected based on the experiments of (MARTINO et al., 2020).

Figures 38, 39, and 40 show the model learning performance over training epochs using different data augmentation transformations for each dataset: OCDC, ORCA, and ORCA 512×512 respectively. Tables 3, 4, and 5 summarize the average quantitative results achieved through the discussed data augmentation transformations in OCDC, ORCA, and ORCA 512×512, respectively.

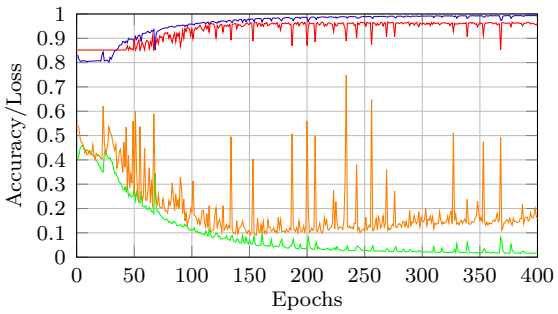
From Tables 3, 4, and 5, we can see that the RCAug strategy is able to improve the segmentation results in all three evaluated datasets, as indicated in bold. The learning curves depicted in Figures 38, 39, and 40 also reveal that our proposed data augmentation strategy, randomly composing different augmentation transformations, turns the training



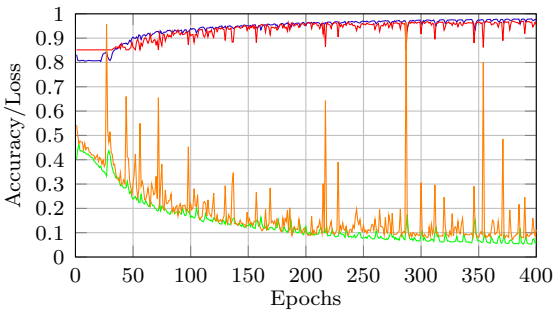
(a) No augmentation.



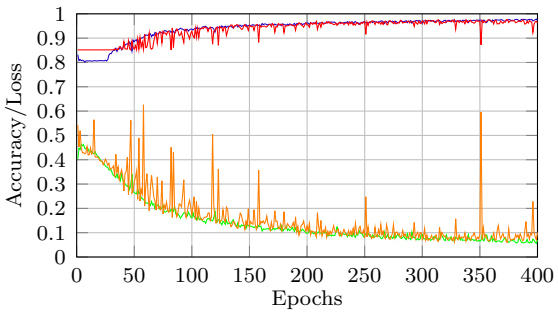
(b) Color transfer augmentation.



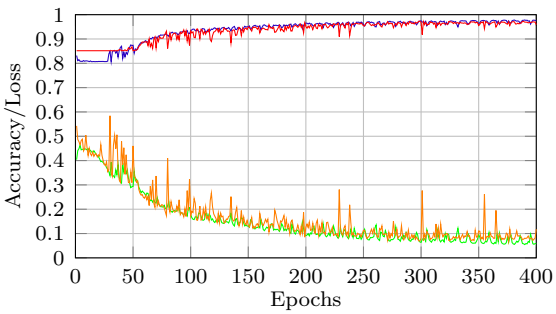
(c) Image inpainting augmentation.



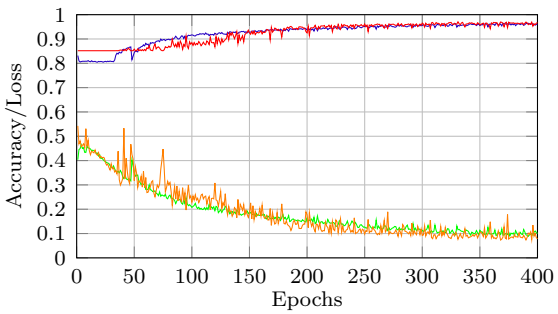
(d) Geometric augmentation.



(e) Distortion augmentation.



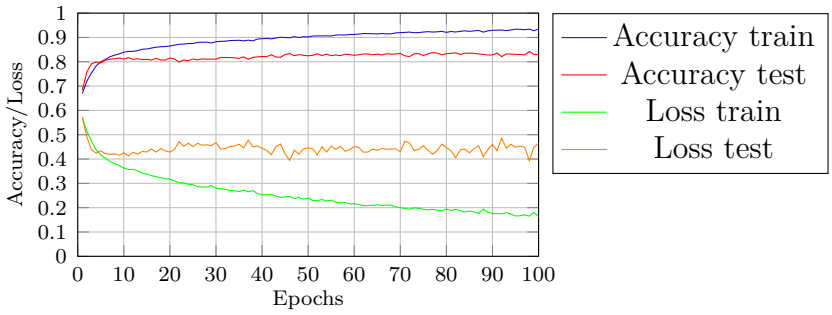
(f) All (standard).



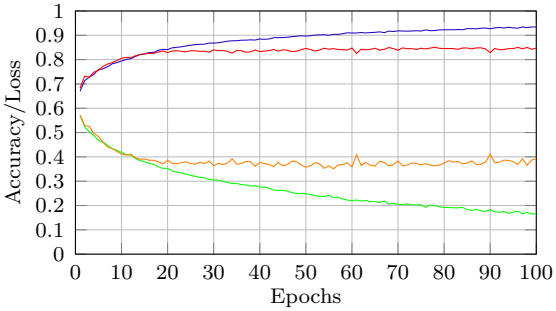
(g) RCAug (proposed).

Figure 38 – Training/testing learning curves for accuracy and loss of the FCN-based segmentation model for the OCDC dataset. Smaller loss values indicate better learning and a value of 0.0 indicates that the dataset was learned perfectly and no mistakes were made.

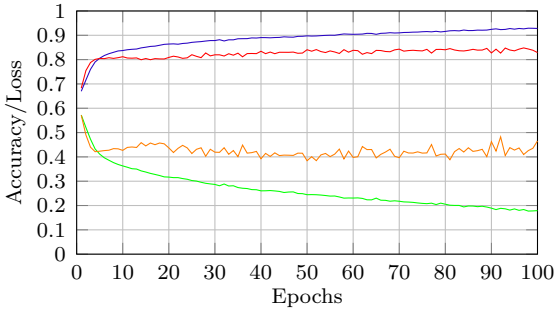
process more stable, reducing the gap between the training and test curves. The learning



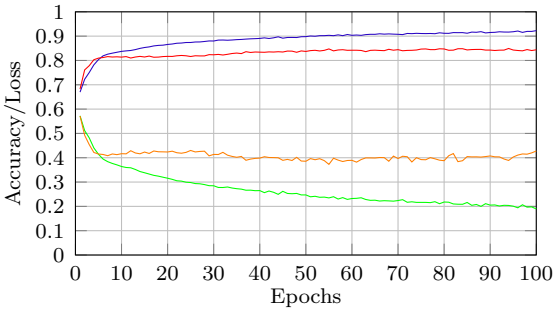
(a) No augmentation.



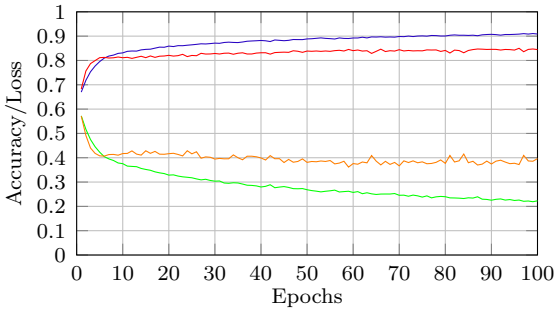
(b) Color transfer augmentation.



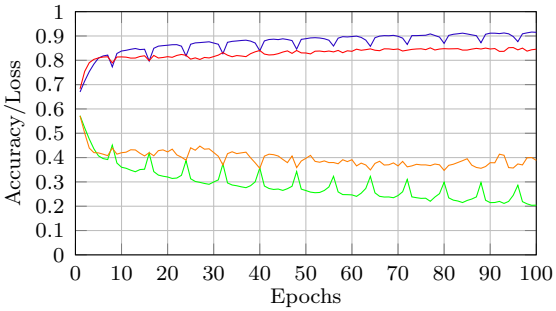
(c) Image inpainting augmentation.



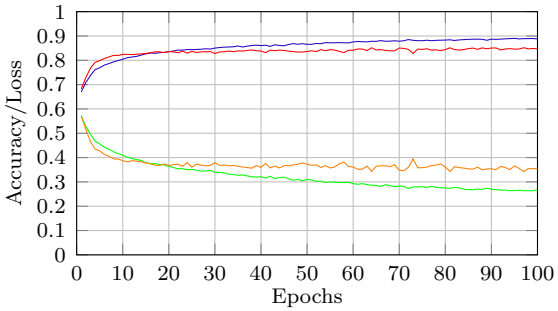
(d) Geometric augmentation.



(e) Distortion augmentation.



(f) All (standard).

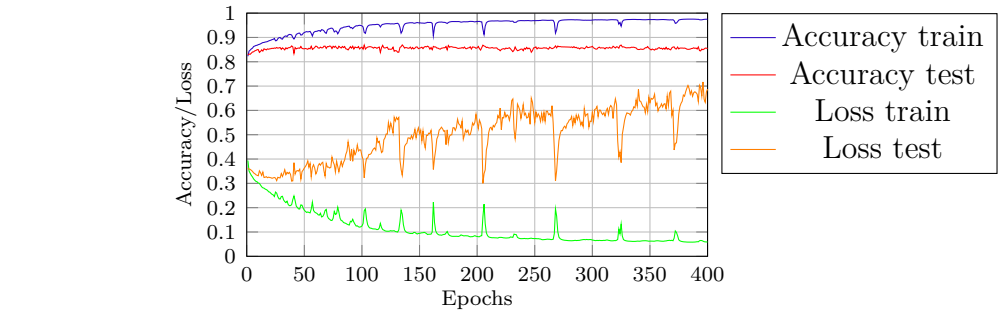


(g) RCAug (proposed).

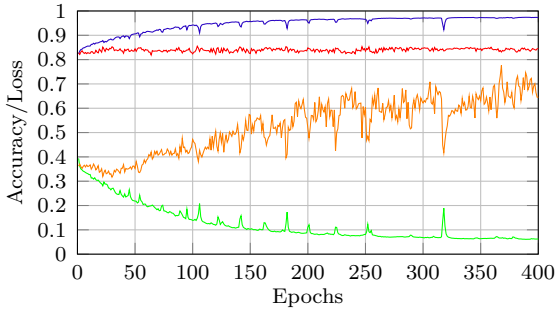
Figure 39 – Training/testing learning curves for accuracy and loss of the FCN-based segmentation model for the ORCA dataset.

curves also show that the test loss becomes higher overtime when no data augmentation is used and, by using RCAug, the process of weight adjustments is made more smoothly.

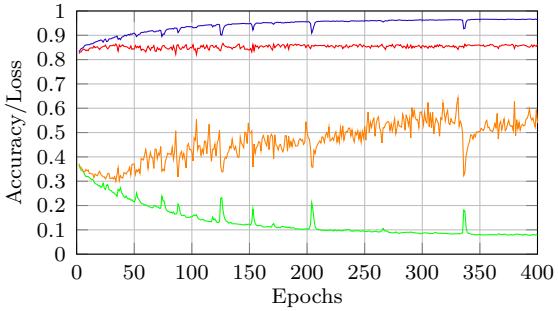
The experimentation results show also that each augmentation transformation (color



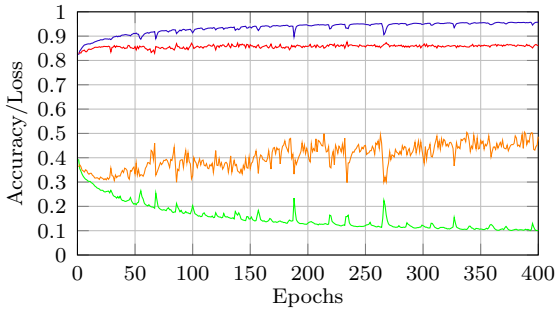
(a) No augmentation.



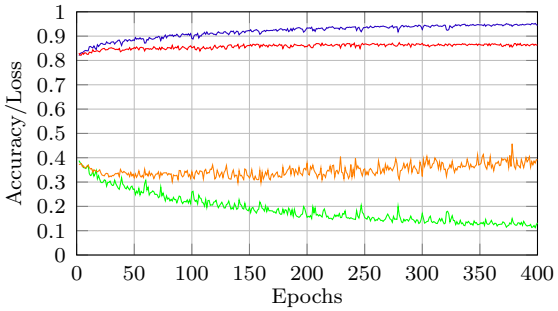
(b) Color transfer augmentation.



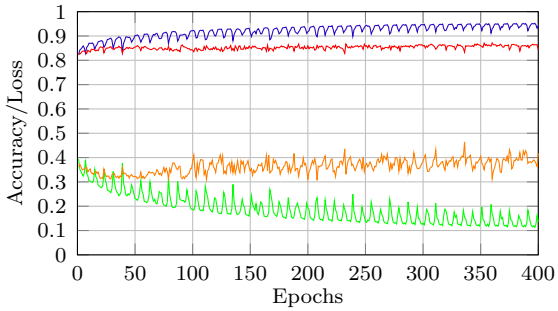
(c) Image inpainting augmentation.



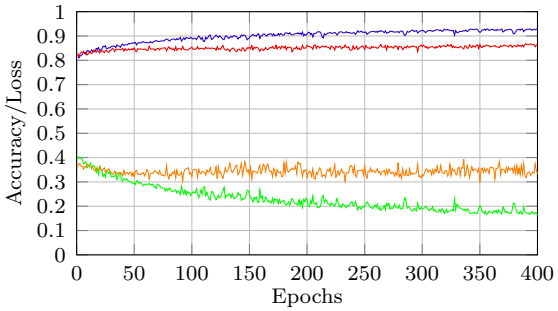
(d) Geometric augmentation.



(e) Distortion augmentation.



(f) All (standard).



(g) RCAug (proposed).

Figure 40 – Training/testing learning curves for accuracy and loss of the FCN-based segmentation model for the ORCA 512×512 dataset.

transformation, GAN-based generative image inpainting transformation, geometric transformations, and distortion transformations) even when used separately improves the segmentation results (compared when no augmentation transformation is applied). However,

as observed in the quantitative results, the color transformation and the GAN-based image inpainting transformation had less influence on the segmentation improvement than geometric and distortion transformations. Furthermore, using a down-sampled version of the original image (ORCA and ORCA 512×512 datasets) negatively impacts the model learning process and consequently the resulting segmentation performance.

4.2.3 External validation

To our knowledge, the only public labeled dataset containing OSCC images available to date is the ORCA dataset (MARTINO et al., 2020). The ORCA dataset (subsection 3.2.3) is not a WSI dataset but is a collection of 200 labeled OSCC tissue pieces (TMA) split into 100 images to train and 100 to test. In addition to the ORCA dataset, for external validation, we tested our proposed segmentation method on the CAMELYON16 dataset (LITJENS et al., 2018). The CAMELYON16 dataset (subsection 3.2.2) is a publicly available dataset with 399 WSIs of tissue sections from another disease: breast cancer-derived lymph node metastases. The 399 WSIs are split into 270 images to train and 129 to test. Unfortunately, the provided annotations for the CAMELYON16 dataset are not pixel-accurate annotations (the tumor annotations may include some small normal areas). Table 1 summarizes information from all the image datasets used in this work.

Figures 41, 42, 43, and 44 show comparative results between gold-standard and our proposed method in the OCDC, ORCA, ORCA 512×512, and CAMELYON16 datasets for qualitative analysis at the full image resolution level. Figure 45 shows some qualitative image-patch segmentation results for the OCDC dataset. Table 6 summarizes the average results achieved for each quantitative measure in all test datasets.

Table 6 – Average quantitative results for test image-patches from our OCDC dataset and from the datasets used as external validation: ORCA, ORCA 512×512, and CAMELYON16.

Dataset	Accuracy	Precision	F1	IOU	Sensitivity	Specificity
OCDC	96.99%	88.29%	90.06%	81.92%	91.91%	97.88%
ORCA	84.60%	78.58%	82.14%	69.69%	86.04%	83.59%
ORCA 512x512	86.46%	73.81%	77.16%	62.81%	80.83%	88.68%
CAMELYON16	92.63%	94.69%	91.43%	84.21%	88.39%	96.02%

The qualitative results on the OCDC dataset demonstrate that our proposed method using FCN, CIE L*a*b* color space, and larger image-patch sizes is able to detect and segment tumor regions in WSIs. As shown in Figures 41 and 45, our method produces refined image-patch segmentation results being able to use contextual information to discriminate between different types of cells and regions such as keratin, lymphocytes, blood, glands, muscle, and tumor cells. Other methods that use CNN-based models need to use

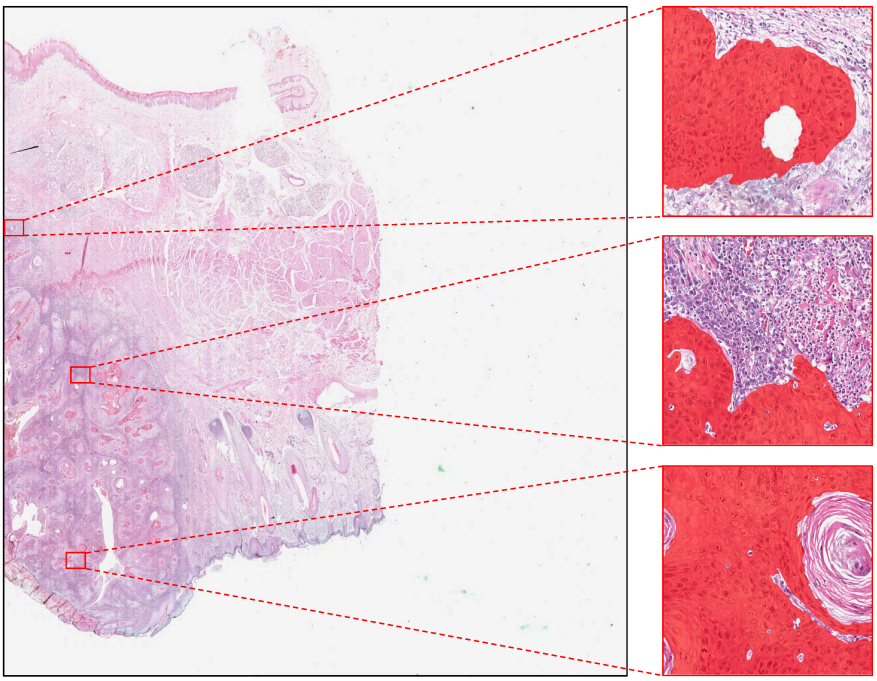
small image-patches, losing valuable context information, to obtain a reasonable resolution of the resulting segmented regions from WSI. Our proposed method achieved for the OCDC dataset quantitative values higher than 90% for almost all measures evaluated, including 97.88% for specificity and 96.99% for image-patch segmentation accuracy metric (see Table 6).

The ORCA dataset is a challenging dataset built using varied tissue samples extracted from different OSCC tumor samples. As shown in Table 6, despite slightly inferior compared to the OCDC dataset, our segmentation method achieved results higher than 80% for almost all evaluated quantitative metrics in the ORCA dataset. As an external comparative result, while (MARTINO et al., 2020) obtained 56% of IOU, in their best configuration, our segmentation method was six percentage points superior, achieving 62.81% of IOU in the experimental evaluation using the same test dataset (as shown in Table 6). (MARTINO et al., 2020) evaluated different color channel compositions using the RGB and HSV color spaces with four different network architectures: SegNet (KENDALL; BADRINARAYANAN; CIPOLLA, 2017), U-Net (RONNEBERGER; FISCHER; BROX, 2015), U-Net with VGG16 (SIMONYAN; ZISSERMAN, 2015) encoder, and U-Net with ResNet50 (HE et al., 2016) encoder. When compared with our patch-based approach, where the original images are used without resizing, our results are even better: our segmentation method was thirteen percentage points superior, achieving 69.69% of IOU (as shown in Table 6). Considering the results achieved by (MARTINO et al., 2020), our results showed that using a down-sampled version of the original image could negatively impact the segmentation performance.

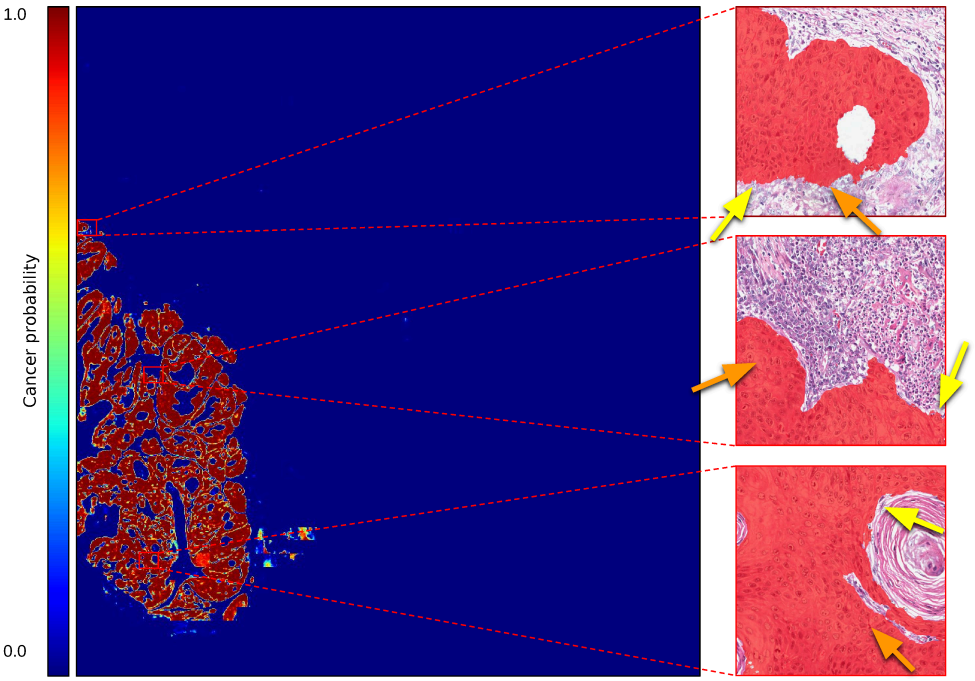
The ground-truth annotations for CAMELYON16 are not precise and may confuse our model, once the provided annotations contain some normal regions annotated as tumor regions as shown in Figure 44. Some false-negative regions (identified by yellow arrows) are actually imprecise ground-truth annotations. Even so, Table 6 shows that our proposed method also achieved good results for the evaluated quantitative segmentation metrics on the CAMELYON16 dataset.

The AUC and FROC metrics used in the CAMELYON16 challenge were calculated, using an evaluation framework provided by the organizers, for comparison (CAMELYON16, 2022). The AUC metric was employed for discriminating between slides containing metastasis and normal slides, while the FROC metric was employed for the tumor localization task inside each slide. Our proposed method achieved an AUC of 0.77 and FROC of 0.55 when our results were evaluated using the evaluation framework provided by the CAMELYON16 organizers. It is important to note that the methods that achieved the best results at the CAMELYON16 challenge were specifically designed for the challenge and all of them execute post-processing stages to extract and classify features from the output segmentation results. Among the twenty-three registered competitor teams until the event closing date, we would be classified among the top five competitors for the

tumor localization task (FROC) and among the top fifteen for the whole-slide image classification task (AUC) in the original CAMELYON16 challenge (BEJNORDI et al., 2017).

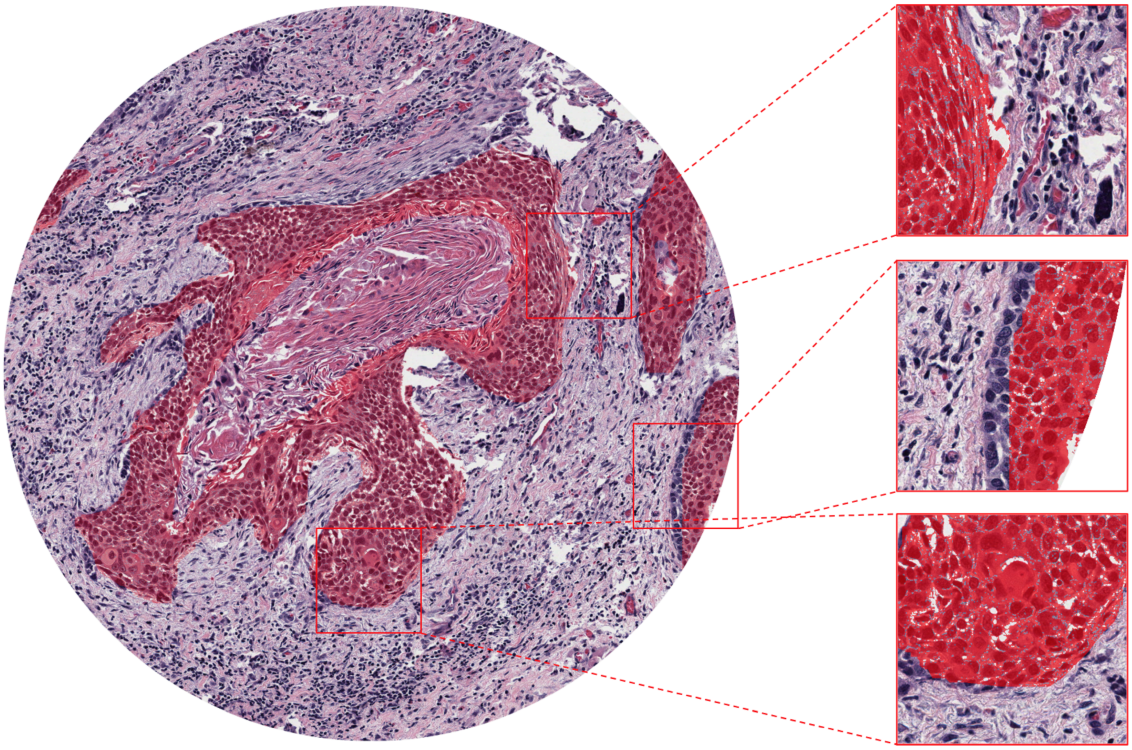


(a) Original WSI and three image-patches with tumor regions manually segmented by the specialist.

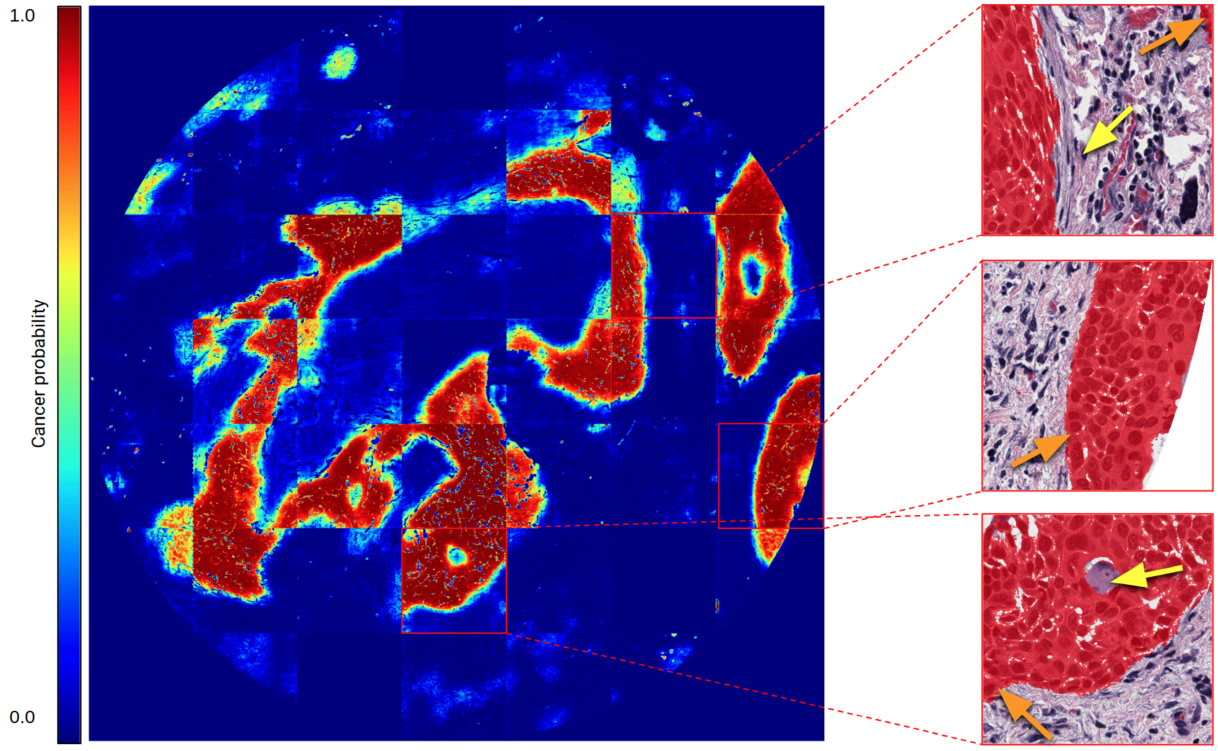


(b) Cancer probability heatmap for the WSI and three image-patch segmentation results.

Figure 41 – One of the WSIs from our OCDC dataset and three corresponding image-patches from the tumor areas. Identified tumor regions are highlighted in red; orange arrows indicate some false positive regions; yellow arrows indicate some false negative regions.



(a) Original image and three selected image-patches (the provided ground-truth regions are overlayed in red).



(b) Heatmap for tumor regions and three image-patch segmentation results for (a).

Figure 42 – One of the images from the ORCA dataset and three corresponding image-patches extracted from the tumor areas. Identified tumor regions are highlighted in red; orange arrows indicate some false positive regions; yellow arrows indicate some false negative regions.

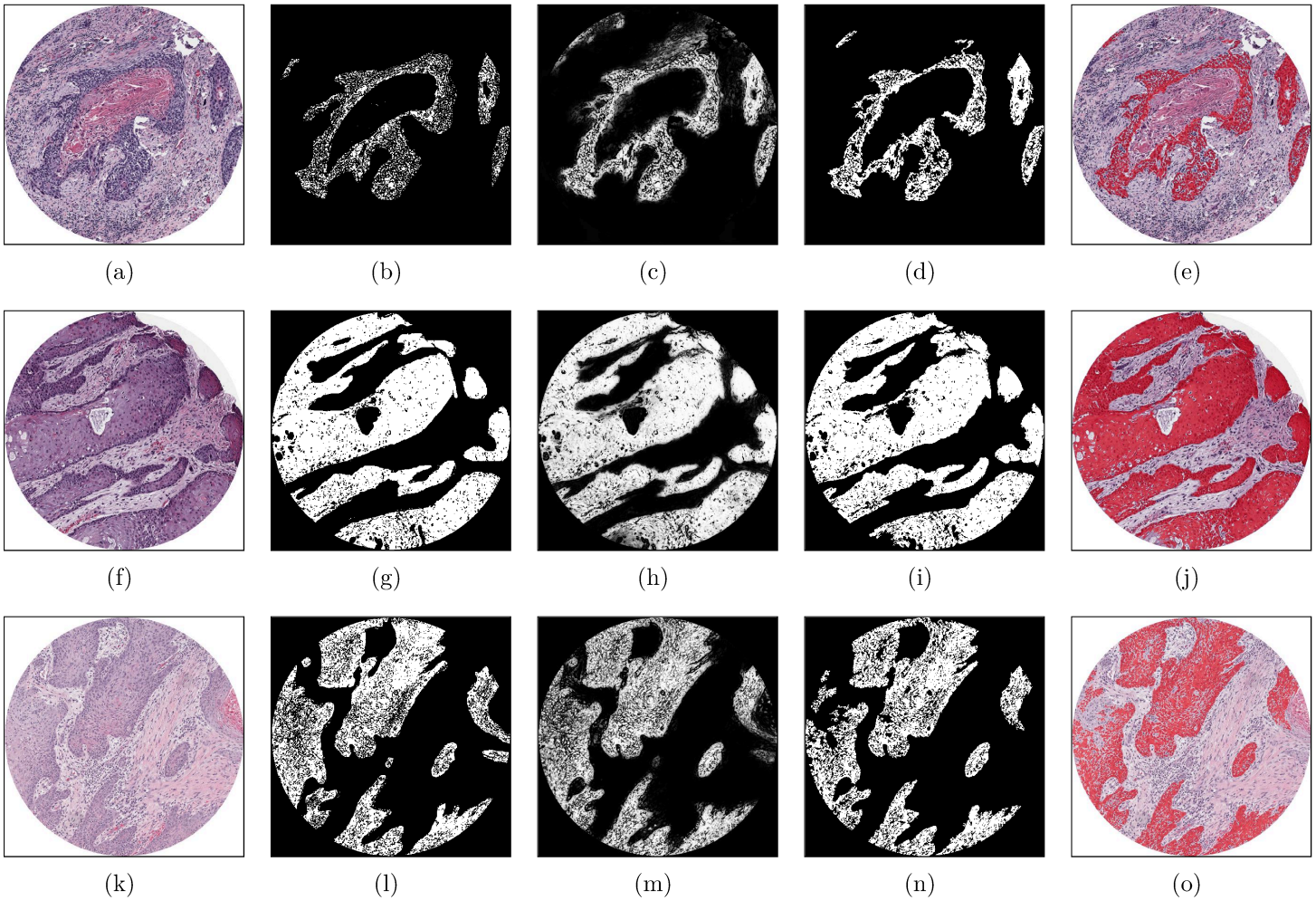
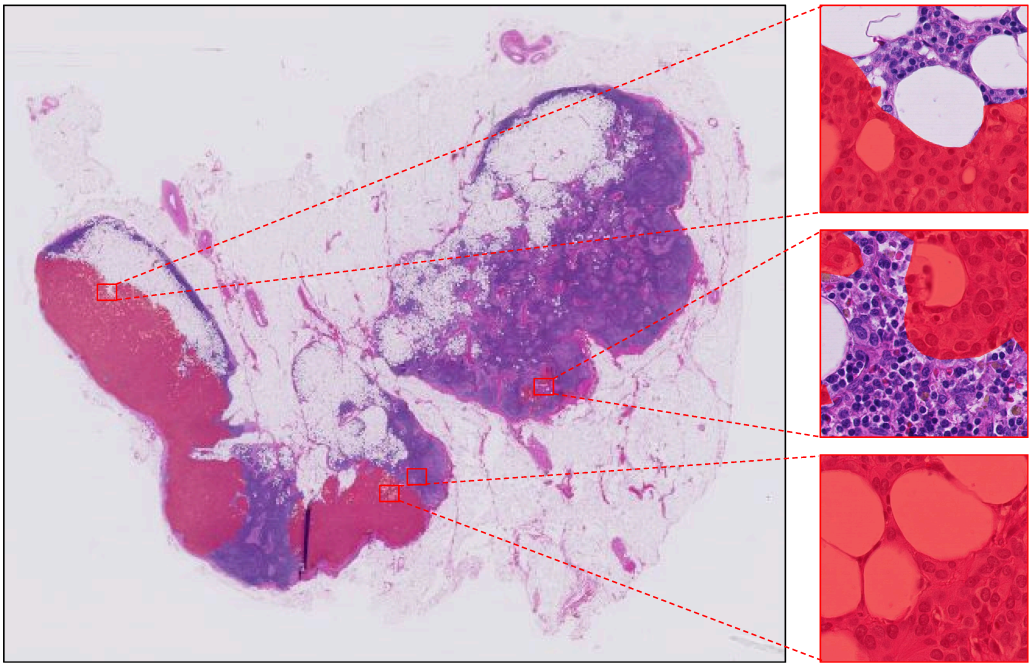
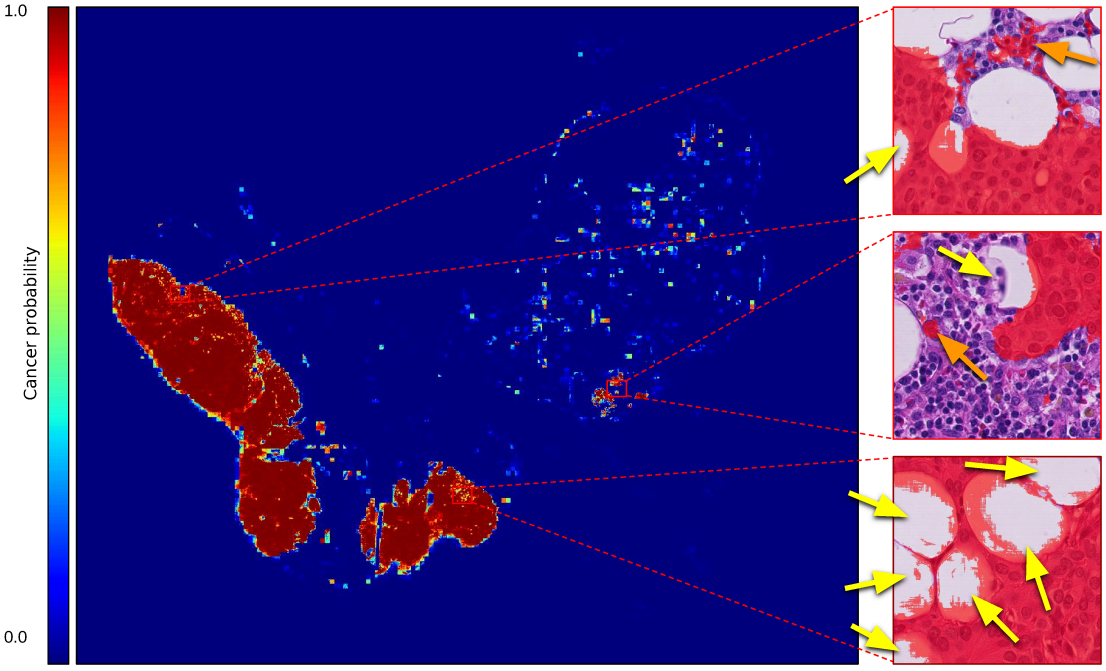


Figure 43 – Qualitative segmentation results for the ORCA 512×512 dataset. First column (a/f/k) shows the input image-patches; second column (b/g/l) shows the gold-standard; third column (c/h/m) shows the output cancer probability for each input image; fourth column (d/i/n) shows the final tumor regions segmentation result; fifth column (e/j/o) shows the input image with the identified tumor regions overlayed in red.



(a) Original WSI with three selected image-patches. Note that for the CAMELYON16 dataset the provided tumor regions annotations include both normal and tumor cells – the white areas are adipose tissues (fat).



(b) Heatmap for tumor regions and three image-patch segmentation results for (a).

Figure 44 – One of the WSIs from the CAMELYON16 dataset and three corresponding image-patches extracted from the tumor areas. Identified tumor regions are highlighted in red; orange arrows indicate some false positive regions; yellow arrows indicate some false negative regions.

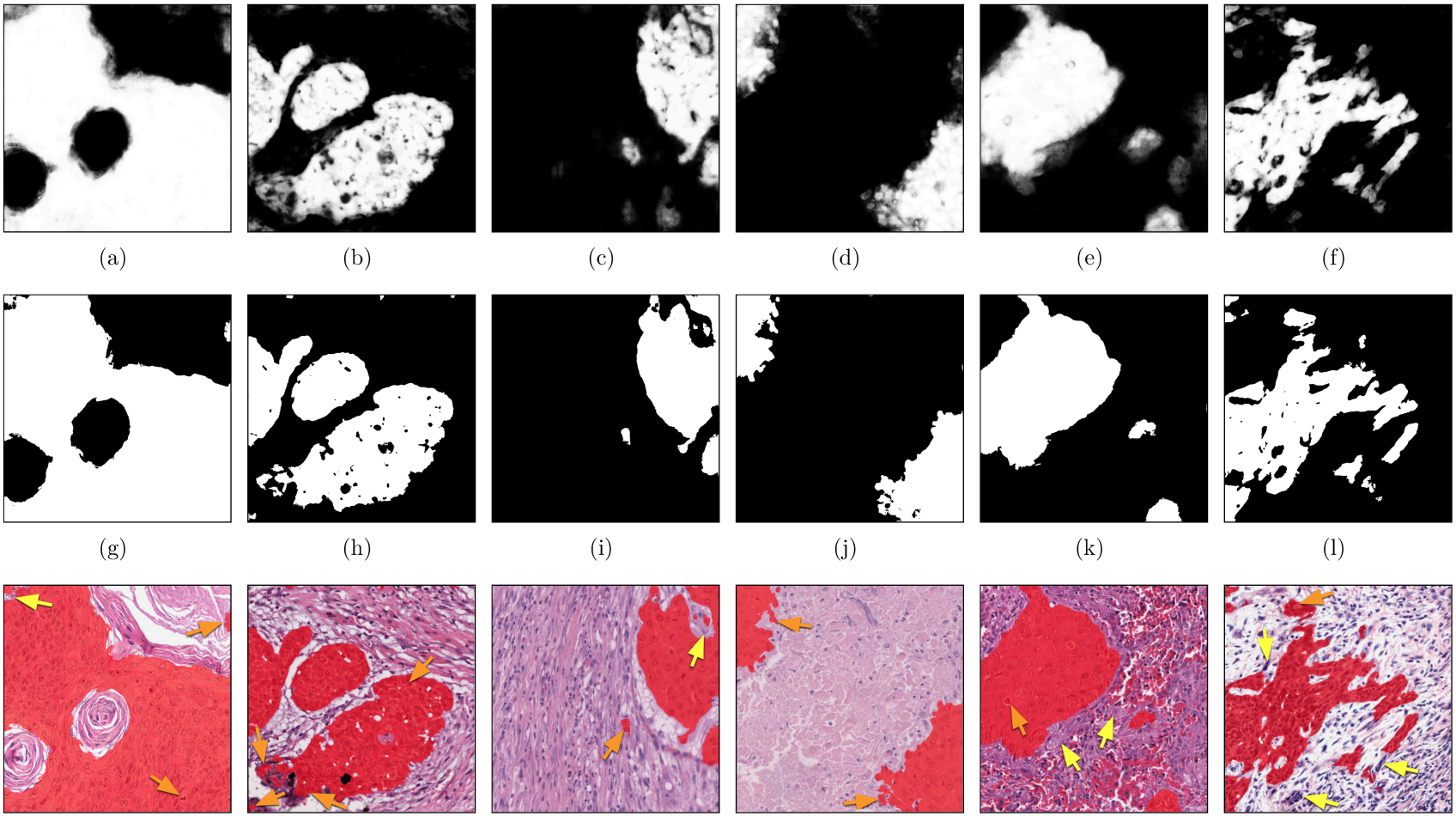


Figure 45 – Qualitative image-patch segmentation results for the input image-patches showed in Fig. 24 (g-l). First row (a-f) shows the output cancer probability for each input image; second row (g-l) shows the final tumor regions segmentation result; third row (m-r) shows the input image with the identified tumor regions overlayed in red – orange arrows indicate some false positive regions and yellow arrows indicate some false negative regions.

4.3 Experiments for Automated Nuclei Segmentation on Oral Histological Images

Nuclei detection and segmentation are critical tasks for cancer diagnosis and grading in CAD systems. As a subsequent step in the detection of tumor regions, a dataset of regions of interest was used to evaluate the nuclei segmentation task. This phase can be improved as further work in order to analyze the nuclei of the identified tumor regions. The effectiveness of our proposed model on automated nuclei segmentation of previously selected ROI images was evaluated using the 24 ROI images from the ONHI dataset test subset (described in subsection 3.2.4). The color and contrast differences caused by the staining and/or digitization process, which is one of the most prevalent artifacts that can lower the quality of the digitized histological images, also were evaluated using this dataset.

The performance of segmentation methods was obtained as described in Section 4.1. First, we compared quantitatively and visually the results of our proposed method with the nuclei manually segmented by the specialist; second, we compared our results with other four algorithms for nuclei segmentation and, finally, we investigated the influence of color normalization algorithms on our proposed method and on the other four evaluated nuclei segmentation algorithms and the impact in different color space transformations.

The quantitative metrics were calculated for each of the 24 ROI images from the ONHI test subset. Figure 46 shows some comparative results between the gold-standard and our proposed method for qualitative analysis. Figure 47 depicts the visual analysis results among the segmentation methods employed in this experimental evaluation. Table 7 summarizes the average results for each quantitative measure.

It is important to note that despite visually presenting good results and contours close to the gold-standard, it is possible to observe by the red and yellow arrows that our method presents some false negative and positive regions. From Table 7, we can see that our proposed method generally outperformed all the other segmentation methods, as indicated in bold. For specificity metric, the method proposed by (WIENERT et al., 2012) achieved the best result. Specificity is used to quantify the percentage of correctly defined true negatives, so, although the specificity metric was higher for the (WIENERT et al., 2012) method, as shown in Figure 47(f), the segmentation process degrades part of the nuclei structures present in the image.

Color is one of the most distinguishable low-level features and it is one of the features that can be highly influenced by biological variability and by the image acquisition process (KHAN et al., 2014). As color variations have a high impact on the automated analysis of histological images, color normalization methods have been used to minimize the influence of color variations, improving the performance of automated segmentation methods (TOSTA et al., 2019). The main objective of these techniques is to promote the

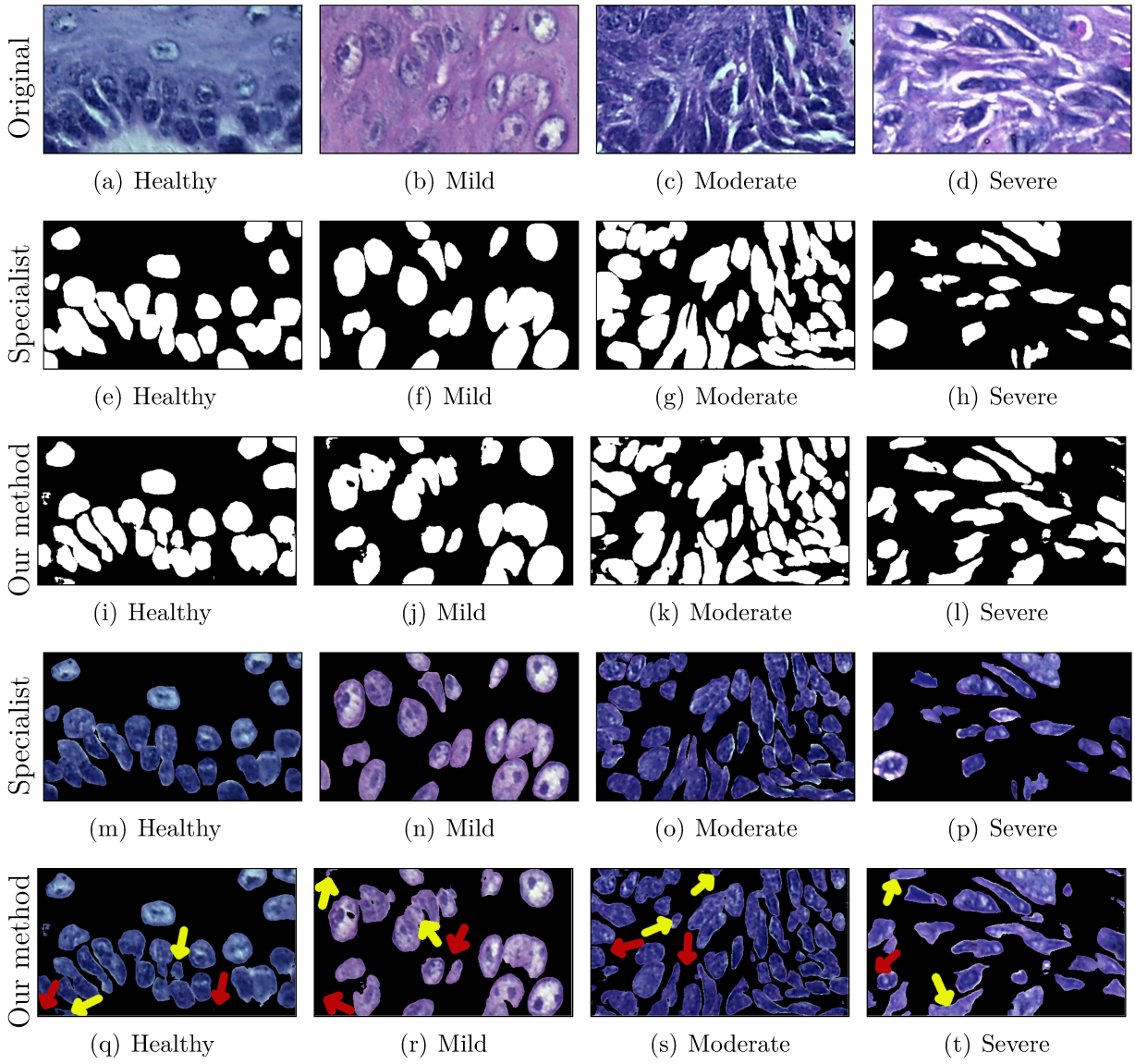
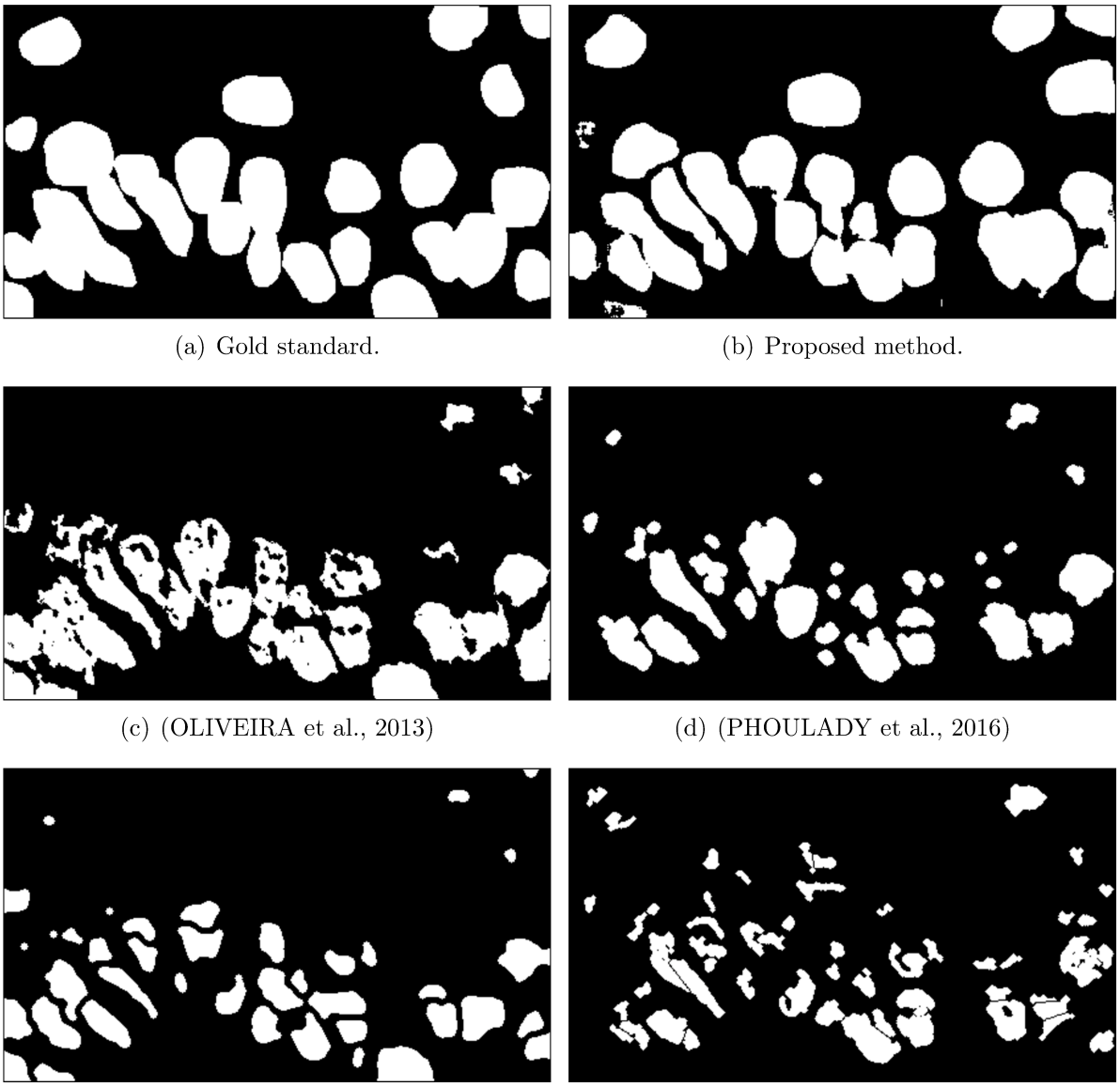


Figure 46 – Qualitative analysis: 1st row (a-d) shows original ROI images for each class; 2nd row (e-h) shows nuclei masks manually delimited by the specialist; 3rd row (i-l) shows nuclei masks obtained by our method; 4th row (m-p) shows segmentation results using the masks manually delimited by the specialist; 5th row (q-t) shows segmentation results by our method – red arrows indicate some false negative areas; yellow arrows indicate some false positive areas.

color adjustment obtained by different stains allowing to eliminate color variations that affect the biological data. This improvement may contribute to increase the sensitivity of the segmentation process, feature extraction, and classification of CAD systems. To investigate the influence of color normalization algorithms on nuclei segmentation methods, we applied the 120 ROI images from the original training and test sets into four different color normalization methods to obtain four new training/test datasets. Figure 48 shows some images from these new training and test datasets.

Our proposed model was retrained using the new training sets separately (four sets



(a) Gold standard. (b) Proposed method.

(c) (OLIVEIRA et al., 2013) (d) (PHOULADY et al., 2016)

(e) (VAHADANE; SETHI, 2013) (f) (WIENERT et al., 2012)

Figure 47 – Qualitative analysis indicating the nuclei masks obtained by different segmentation methods. The original ROI image can be viewed in Figure 46(a).

of 96 images) and a new evaluation was performed using the new test sets (four sets of 24 images). The same ROI images employed to evaluate our method were used to run the evaluated segmentation methods. Table 7 summarizes the average results for the color normalization and nuclei segmentation methods. Figure 49 shows the evolution of accuracy results by each color normalization method.

From Table 7 and Figure 49 we can see that, overall, the color normalization methods were able to improve the nuclei segmentation task results in the traditional methods – except the method proposed by (MACENKO et al., 2009). The same was not true for our proposed method, where the color normalization operations maintained the overall performance of our method (slight decrease/increase).

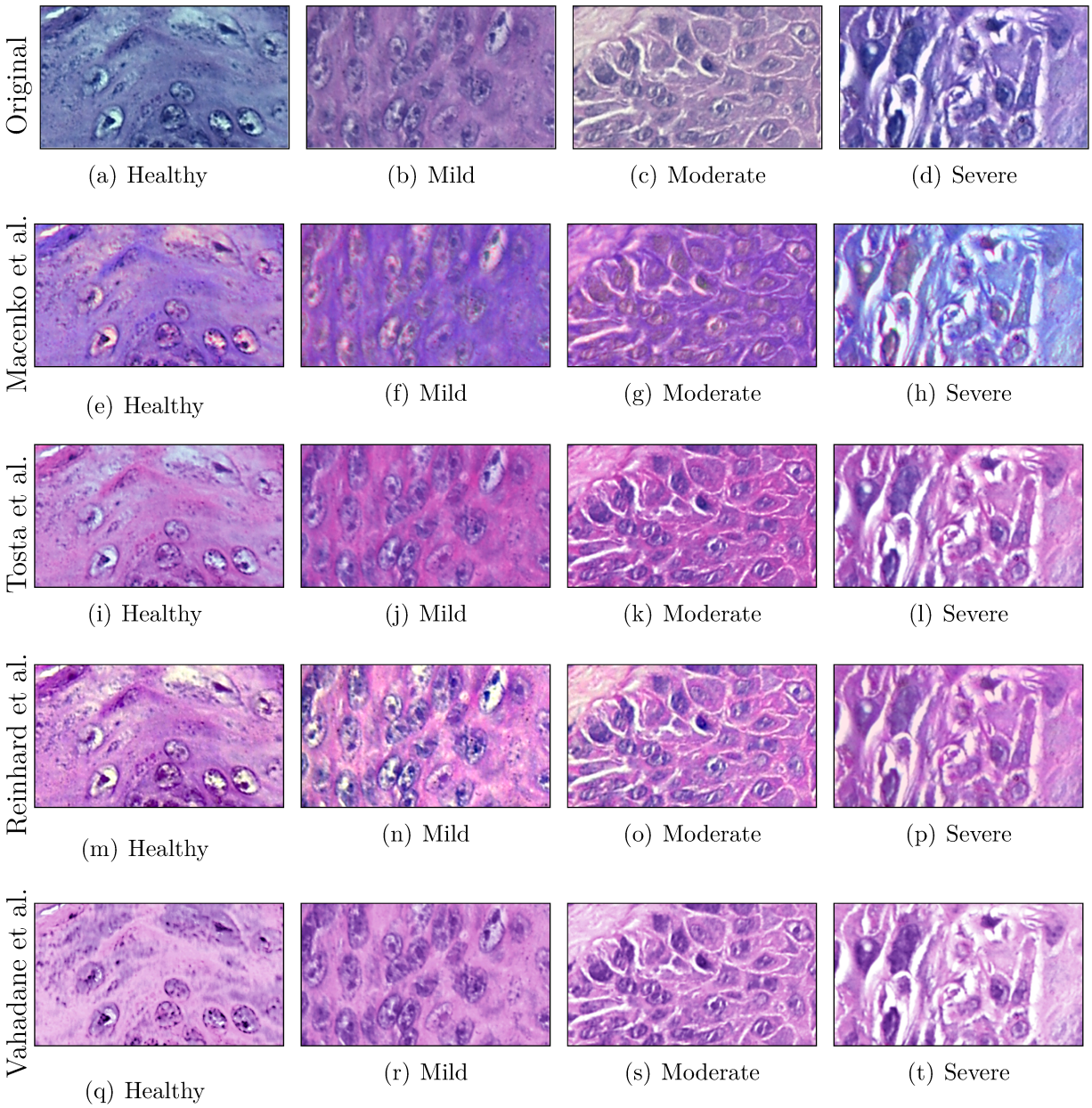


Figure 48 – ROI images applied to different color normalization methods: 1st column (a, e, i, m, q) shows samples of healthy mucosa class; 2nd column (b, f, j, n, r) shows mild dysplasia class; 3rd column (c, g, k, o, s) shows moderate class and 4th column (d, h, l, p, t) shows severe class.

Table 7 – The average quantitative results by color normalization and nuclei segmentation methods applied to test set images.

Color Normalization	Nuclei Segmentation Method	Accuracy	Precision	F1/Dice	Jaccard	Sensitivity/Recall	Specificity
No color normalization	Proposed method	**0.879**	**0.793**	**0.820**	**0.699**	**0.860**	0.891
	(OLIVEIRA et al., 2013)	0.748	0.671	0.629	0.468	0.658	0.816
	(PHOULADY et al., 2016)	0.744	0.753	0.495	0.332	0.388	0.937
	(VAHADANE; SETHI, 2013)	0.725	0.727	0.468	0.312	0.372	0.918
	(WIENERT et al., 2012)	0.722	0.731	0.405	0.257	0.291	**0.950**
(MACENKO et al., 2009)	Proposed method	**0.874**	**0.843**	**0.800**	**0.668**	**0.773**	**0.931**
	(OLIVEIRA et al., 2013)	0.590	0.492	0.472	0.318	0.552	0.625
	(PHOULADY et al., 2016)	0.684	0.566	0.378	0.238	0.295	0.886
	(VAHADANE; SETHI, 2013)	0.623	0.503	0.346	0.214	0.314	0.795
	(WIENERT et al., 2012)	0.674	0.557	0.301	0.180	0.214	0.914
(TOSTA et al., 2019)	Proposed method	**0.854**	0.761	**0.789**	**0.658**	**0.846**	0.869
	(OLIVEIRA et al., 2013)	0.785	0.789	0.617	0.450	0.547	0.923
	(PHOULADY et al., 2016)	0.759	0.810	0.522	0.356	0.404	0.953
	(VAHADANE; SETHI, 2013)	0.738	**0.824**	0.446	0.292	0.324	0.961
	(WIENERT et al., 2012)	0.730	**0.824**	0.391	0.247	0.266	**0.973**
(REINHARD et al., 2001)	Proposed method	**0.874**	**0.788**	**0.813**	**0.688**	**0.853**	0.887
	(OLIVEIRA et al., 2013)	0.791	0.738	0.652	0.490	0.620	0.892
	(PHOULADY et al., 2016)	0.754	0.761	0.523	0.356	0.412	0.938
	(VAHADANE; SETHI, 2013)	0.745	0.778	0.483	0.325	0.368	0.946
	(WIENERT et al., 2012)	0.729	0.776	0.406	0.258	0.284	**0.963**
(VAHADANE et al., 2016)	Proposed method	**0.871**	0.799	**0.805**	**0.678**	**0.827**	0.900
	(OLIVEIRA et al., 2013)	0.778	0.819	0.580	0.419	0.481	0.947
	(PHOULADY et al., 2016)	0.747	0.800	0.484	0.323	0.360	0.955
	(VAHADANE; SETHI, 2013)	0.742	**0.855**	0.436	0.286	0.306	**0.976**
	(WIENERT et al., 2012)	0.729	0.819	0.401	0.260	0.288	0.967

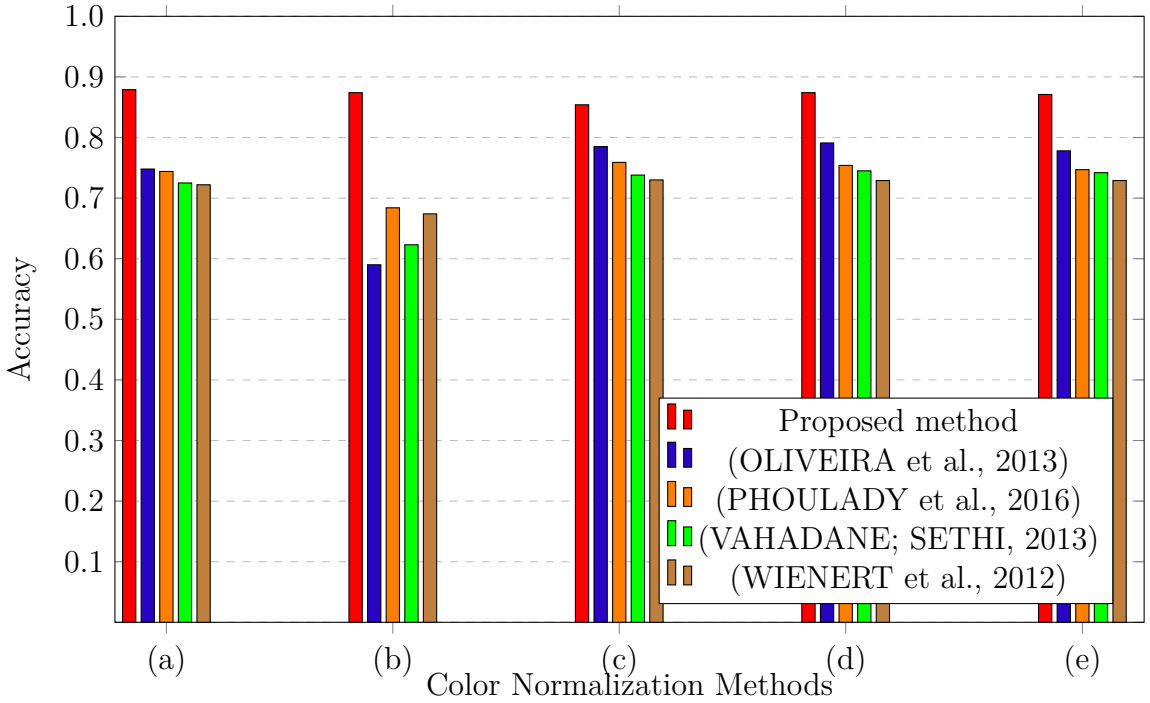
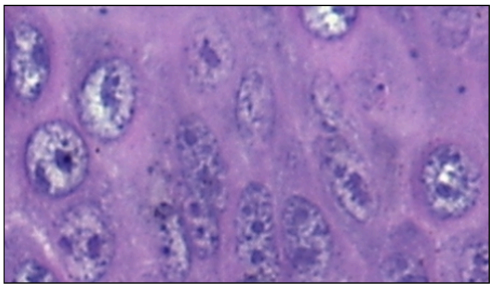


Figure 49 – Accuracy results by nuclei segmentation and color normalization methods: (a) No color normalization; (b) (MACENKO et al., 2009); (c) (TOSTA et al., 2019); (d) (REINHARD et al., 2001); (e) (VAHADANE et al., 2016).

In this stage, we also investigate the influence of color space transformations on the proposed nuclei segmentation method, the 120 ROI images from the original training and test sets were transformed into three different color spaces (RGB, HSV, and CIE L*a*b*). Our nuclei segmentation model was retrained using the grayscale image and each color space and their respective individual channels separately, resulting in 13 new training datasets (thirteen sets of 96 images). The experimental evaluations were performed using the 13 configurations for the 24 ROI images from the test subset: Grayscale image; RGB, only the R channel from RGB, only the G channel from RGB, only the B channel from RGB; HSV, only the H channel from HSV, only the S channel from HSV, only the V channel from HSV; CIE L*a*b*, only the L* channel from CIE L*a*b*, only the a* channel from CIE L*a*b* and only the b* channel from CIE L*a*b*.

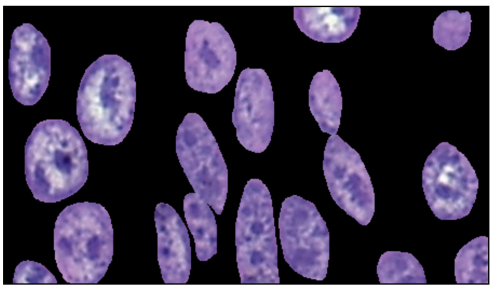
Figure 50 shows one selected image from the dataset as a reference for the qualitative analysis. The original image is shown in Figure 50(a), its corresponding nuclei mask manually delimited by the specialist is shown in Figure 50(b), and the segmentation result using the masks manually delimited by the specialist is shown in Figure 50(c). Figure 51 reveals the results obtained by our automated nuclei segmentation method for each configuration on RGB color space and the grayscale image. Figure 52 depicts the results for HSV color space transformation and its respective individual channels. Figure 53 shows the results obtained by CIE L*a*b* color space transformation and its respective individual channels.



(a) Original input ROI image.



(b) Manually segmented nuclei.



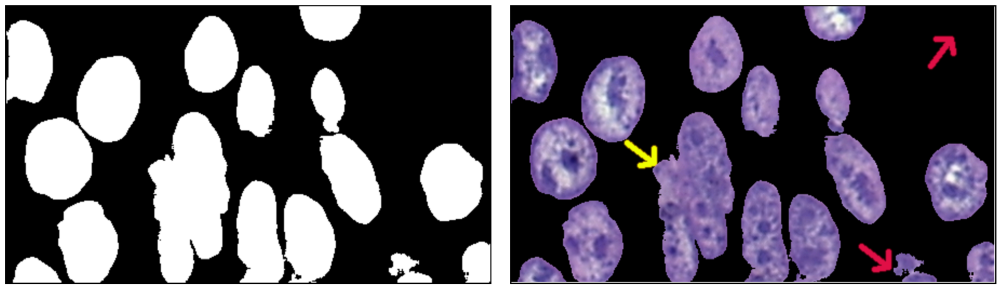
(c) Nuclei segmentation result.

Figure 50 – Gold-standard nuclei segmentation.

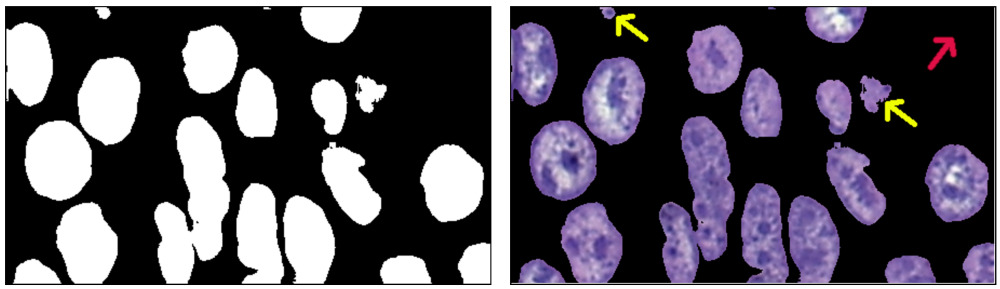
As we can see in Figures 51, 52, and 53 and Table 8, the original RGB and the CIE L*a*b* color spaces present the best results, outperforming all the configurations. As indicated in the images by the red and yellow arrows, all configurations present some false negative and false positive regions. The worst results were presented by the CIE a* and HSV H color space configurations. It is worth noting that these results indicate that the chrominance information alone, or the color values, does not play the most significant role for nuclei segmentation purposes on oral epithelial image datasets. The color space channels that indicate lightness information (Grayscale, CIE L*, HSV S/V) performed very close to the best configurations, revealing that lightness information plays a pivotal role for nuclei segmentation purposes (note also that grayscale and L* channel are computed from a very similar equation from the RGB color space).

Regarding quantitative analysis, Table 8 summarizes the average results for each quantitative measure in the test subset. The same 24 ROI images (test subset transformed thirteen times for each corresponding color space configuration) were employed to evaluate all configurations.

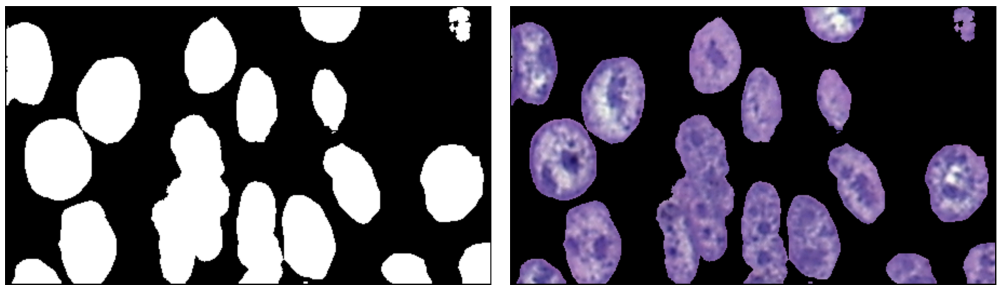
The FCN-based nuclei segmentation model was implemented using the PyTorch framework (PASZKE et al., 2017). The model was trained using a Microsoft Azure Data Science Virtual Machine with a size 'Standard NC12_Promo GPU Family' (12 vcpus, 112 GiB memory) and Ubuntu 16.04 operational system. The elapsed time to train the model with 500 epochs was about 350 minutes. After training, the elapsed time to process an input image was about 0.3 seconds. The source code is publicly available at <https://github.com/dalifreire/dysplastic_oral_tissues_segmentation>.



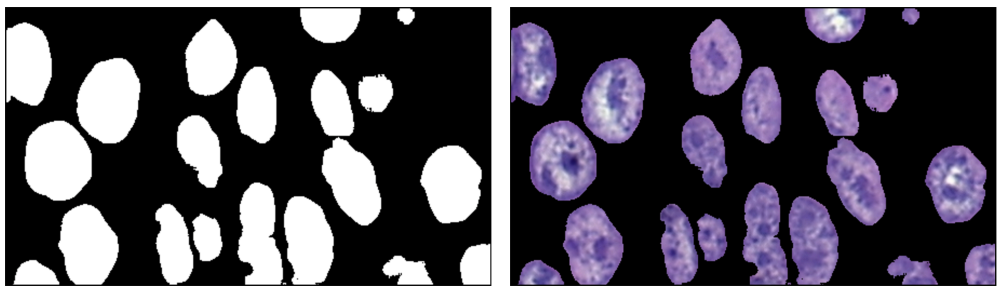
(a) Grayscale mask. (b) Grayscale segmentation.



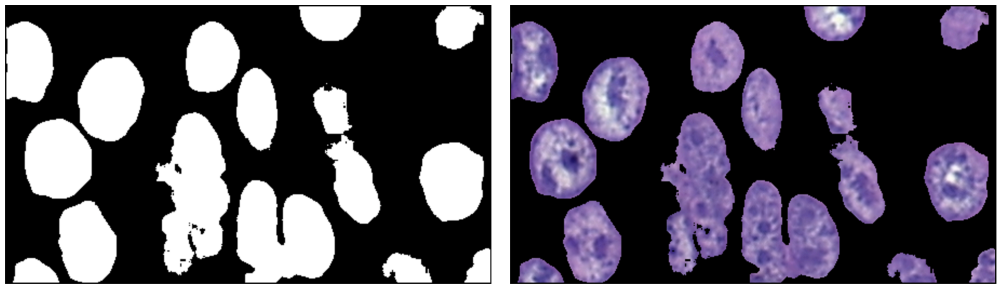
(c) RGB mask. (d) RGB segmentation.



(e) RGB R mask. (f) RGB R segmentation.

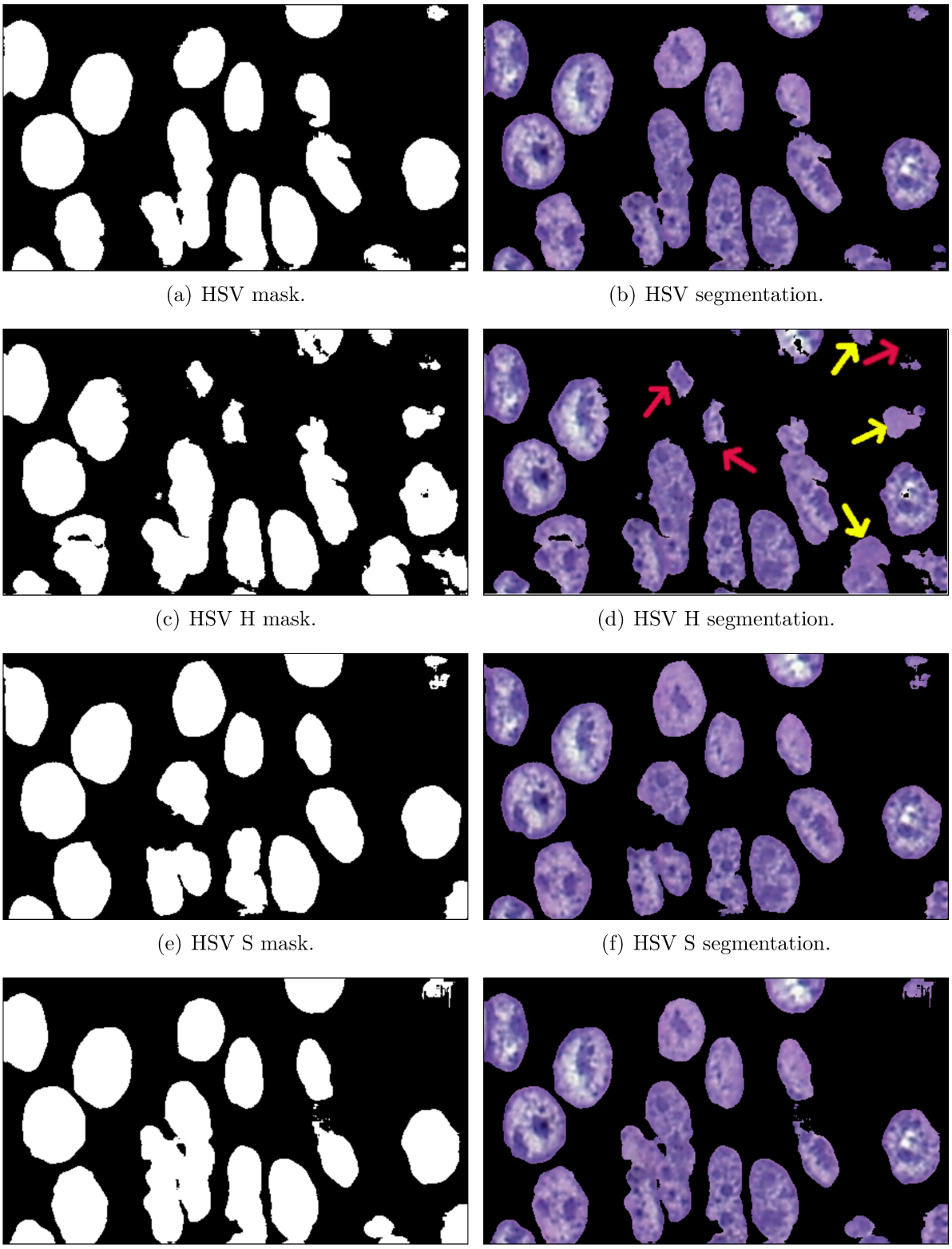


(g) RGB G mask. (h) RGB G segmentation.



(i) RGB B mask. (j) RGB B segmentation.

Figure 51 – Qualitative analysis for RGB color space: 1st column shows nuclei masks obtained; 2nd column shows the final segmentation results – red arrows indicate some false negative areas; yellow arrows indicate some false positive areas.



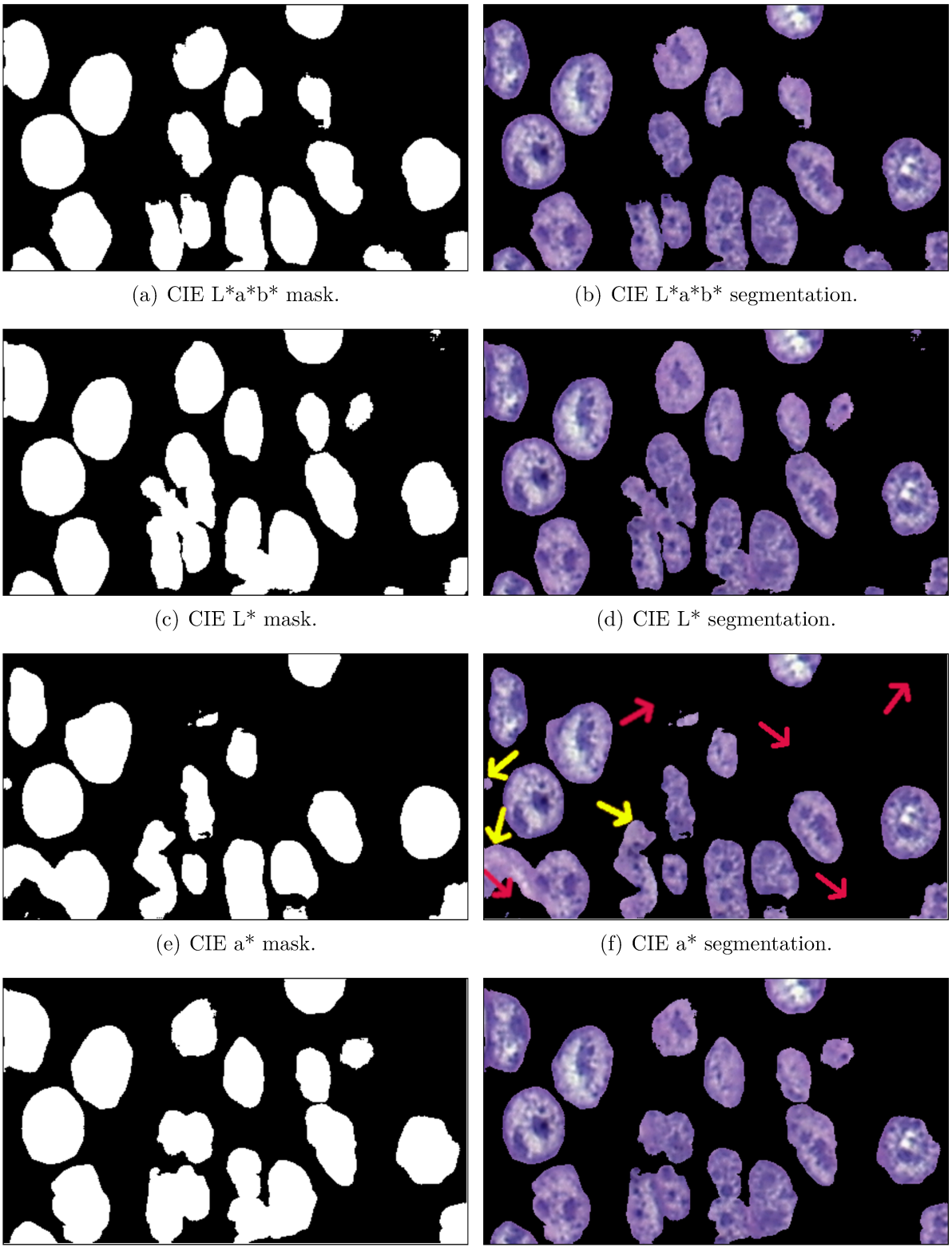
(a) HSV mask. (b) HSV segmentation.

(c) HSV H mask. (d) HSV H segmentation.

(e) HSV S mask. (f) HSV S segmentation.

(g) HSV V mask. (h) HSV V segmentation.

Figure 52 – Qualitative analysis for HSV color space segmentation: 1st column (a, c, e, g) shows nuclei masks obtained; 2nd column (b, d, f, h) shows the final segmentation results – red arrows indicate some false negative areas; yellow arrows indicate some false positive areas.



(a) CIE L*a*b* mask. (b) CIE L*a*b* segmentation.

(c) CIE L* mask. (d) CIE L* segmentation.

(e) CIE a* mask. (f) CIE a* segmentation.

(g) CIE b* mask. (h) CIE b* segmentation.

Figure 53 – Qualitative analysis for CIE L*a*b* color space segmentation: 1st column (a, c, e, g) shows nuclei masks obtained; 2nd column (b, d, f, h) shows the final segmentation results – red arrows indicate some false negative areas; yellow arrows indicate some false positive areas.

Table 8 – The average quantitative results by each color space transformation applied on the test set.

Color Space	Accuracy	Precision	F1/Dice	Jaccard	Sensitivity/Recall	Specificity
Grayscale	86.2%	76.9%	79.5%	66.7%	84.2%	87.7%
RGB	87.9%	79.3%	**82.0%**	**69.9%**	**86.0%**	89.1%
R channel from RGB	87.4%	78.8%	81.1%	68.9%	84.8%	89.3%
G channel from RGB	86.0%	79.4%	78.0%	64.7%	78.2%	90.8%
B channel from RGB	86.2%	77.5%	79.4%	66.6%	83.2%	88.5%
HSV	88.1%	80.7%	81.9%	69.8%	84.4%	90.4%
H channel from HSV	78.4%	69.1%	65.9%	50.3%	64.8%	85.8%
S channel from HSV	84.2%	75.2%	76.5%	62.8%	80.1%	87.1%
V channel from HSV	86.3%	78.3%	79.2%	66.3%	81.7%	89.5%
Lab	**88.2%**	**85.3%**	81.4%	68.9%	78.%8	**93.7%**
L channel from Lab	86.5%	77.5%	79.7%	67.0%	83.5%	88.6%
a channel from Lab	75.8%	63.7%	64.8%	49.0%	67.9%	80.3%
b channel from Lab	81.9%	71.1%	73.8%	59.0%	79.0%	84.1%

4.4 Final Considerations

This chapter presented the experimental evaluations for the proposed FCN-based model for the detection and segmentation of tumor regions in OSCC-derived-H&E-stained-WSIs. The proposed method uses a color-based tissue detector in a preprocessing step to remove background and scanning artifacts and an FCN to localize and segment 640×640 image-patch based tumor regions on WSIs. To train and validate the method, a new WSI dataset of H&E-stained tissues from OSCC regions was created. As no other similar public dataset is available, for comparison we tested our method on a publicly available dataset of WSI from sentinel lymph node with breast cancer metastases – CAMELYON16. The proposed method also was evaluated on the ORCA dataset, a public collection of labeled OSCC tissue piece images in TMA pattern.

Experimental results revealed that our proposed method is robust and capable to detect and perform refined segmentation of tumor regions in H&E-stained histological WSIs. The proposed method is generalizable, achieving good results in datasets with images of different diseases: OSCC and breast cancer-derived lymph node metastases. By using RCAug, we improved the FCN-based segmentation method from 0.51 to 0.81 of IOU in a whole slide image dataset and from 0.65 to 0.69 of IOU in a tissue microarray images dataset. The experiments using different augmentations also revealed that geometric and distortion transformations performed better than the others, such as color transformations, in the segmentation task of OSCC histological images. Altogether were performed experimental evaluations using 90,102 image-patches of size 640×640 pixels and our method achieved an averaged segmentation accuracy of 91.4%.

This chapter also presented our FCN-based method to address the automated nuclei segmentation task on oral histological image datasets. The impacts of different color space transformations and color normalization techniques in automated nuclei segmentation methods using H&E-stained histological images were evaluated. To train and validate the proposed method, a dataset with H&E-stained histological images of tongue tissues from 30 dysplasia-harboring mice was used.

Experimental evaluations, using H&E-stained images applied to different color normalization methods, revealed that color normalization plays an important role in traditional nuclei segmentation methods. However, for our FCN model, it was observed that the color normalization preprocessing step does not bring significant improvements, therefore it is not necessary. Furthermore, our proposed method presented nuclei segmentation results very close to the gold-standard one. Experimental results, using thirteen different color space configurations, also revealed that the chrominance information does not play the most significant role for nuclei segmentation purposes in H&E-stained histological images. Furthermore, the results indicates that most significant role for nuclei segmentation purposes may be played by the lightness information contained in the color spaces.

Chapter 5

Conclusion

This chapter presents the conclusions of this research including its contributions to the state of the art. The publications derived from this research, as well as co-authoring collaborations, are enumerated. Future works directions have also been described that aim to improve the proposed methods.

5.1 Conclusions

The diagnosis of different types of cancer, including oral cavity-derived cancer, is made by a pathologist through complex and time-consuming microscopic analysis of tissue samples. Despite being a gold-standard method, the microscopic investigation is plagued by inter-and intra-observer variability and considered a time-consuming task so that it requires well-trained and experienced pathologists to perform such a task.

The main goal of this research aims to develop computational techniques for histological image segmentation to support pathologists in decision-making in cases of oral cavity-derived cancer. To achieve this goal, this thesis proposes an FCN-based segmentation method to automatically segment regions of interest from H&E-stained WSIs. The proposed method was applied to two different tasks: tumor regions segmentation and nuclei segmentation. A dataset composed of mice tongue histopathological images was used to evaluate the nuclei segmentation task. The tumor regions segmentation task was performed using WSIs, which are large multi-gigapixel images that digitize the entire tissue section into a single histological image. As none similar OSCC labeled WSI dataset is publicly available to date, a new labeled WSI dataset for studying real OSCC cases was created with the assistance of a pathologist. The influence of different color spaces, stain color normalization methods, and different image-patch sizes were also explored.

Histological image segmentation is often constrained by the availability of labeled training data. Thus, data augmentation strategies become essential to train convolutional models to overcome the overfitting problem when only a few training samples are available. This research also goals to introduce a data augmentation strategy to segment

H&E-stained histological images using FCN-based methods. To address this goal, a data augmentation strategy, named RCAug, was proposed. Given the input image and their corresponding label, a pipeline with a random composition of geometric, distortion, color transfer, and generative image transformations is executed on the fly.

All the proposed methods and performed experiments were made available as open-source and were published at a GitHub repository: <https://github.com/dalifreire/tumor_regions_segmentation>

5.2 List of Publications

Co-authoring collaborations:

- A. B. SILVA et al. Segmentation of Oral Epithelial Dysplasias Employing Mask R-CNN and Color Normalization. In: 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, Korea (South), 2020, pp. 2818-2824. doi: 10.1109/BIBM49941.2020.9313101

Reviewed conference publications:

- D. F. D. dos Santos et al. Automated Nuclei Segmentation on Dysplastic Oral Tissues Using CNN. In: 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), Niterói, Brazil, 2020, pp. 45-50, doi: 10.1109/IWSSIP48289.2020.9145157

- D. F. D. dos Santos et al. Impacts of Color Space Transformations on Dysplastic Nuclei Segmentation Using CNN. In: Anais do XVI Workshop de Visão Computacional. Porto Alegre, RS, Brasil, SBC, 2020, pp. 6–11, doi: 10.5753/wvc.2020.13475

Journal publications:

- D. F. D. dos Santos et al. Automated detection of tumor regions from oral histological whole slide images using fully convolutional neural networks. In: Biomedical Signal Processing and Control, Volume 69, 2021, 102921, ISSN 1746-8094. doi: https://doi.org/10.1016/j.bspc.2021.102921

Journal submissions (under review):

- D. F. D. dos Santos et al. Influence of data augmentation strategies on the segmentation of oral histological images using fully convolutional neural networks. In: Journal of Digital Imaging.

5.3 Future Works

Histological images can suffer from different artifacts that can lower the quality of the digitized images (see Section 2.2 and Figure 6). Pathologists naturally avoid tissue regions with artifacts when analyzing WSIs. Thus, it is expected that CAD tools, including those based on deep neural networks, that are able to detect and avoid these artifacts regions could perform better (SALVI et al., 2021). For future works, we propose optimizing the pre-processing step to remove non-tissue and scanning artifacts regions (such as out-of-focus regions and tissue folding). Further investigations could be evaluated by modifying our FCN-based segmentation method to use overlapping image-patches.

Deep learning techniques traditionally need large amounts of labeled data to be trained. Unfortunately, histological image segmentation is often constrained by the availability of labeled training data once labeling histological images is a highly-skilled, complex, and time-consuming task. One way to surpass this drawback is data augmentation. In future directions, it will also be necessary to incorporate more data augmentation transformations to evaluate which group of augmentation transformations is able to produce better segmentation results in OSCC histological images. Further investigations, using cross-validation in the experiments, to verify whether the best group of augmentation transformations for segmentation tasks are also the best for classification tasks could be evaluated. We also propose to invite more specialists to include ground-truth labels in our OCDC dataset.

As a subsequent step in detecting and segmentation tumor regions, extracting and quantifying features from those tumor regions is also proposed as future work. In this way, a post-processing step based on nuclei features could be performed to quantify and grad the detected tumor regions from OSCC whole slide images. We also propose using transfer learning between the different datasets considered in this study (OCDC, CAMELYON16, and ORCA) to improve the quantitative results achieved.

Bibliography

ALOYSIUS, N.; GEETHA, M. G. A review on deep convolutional neural networks. In: . [S.l.: s.n.], 2017. p. 0588–0592.

ALTAF, F.; ISLAM, S. M. S.; AKHTAR, N.; JANJUA, N. K. Going deep in medical image analysis: Concepts, methods, challenges, and future directions. **IEEE Access**, v. 7, p. 99540–99572, 2019.

ANDREA, C. E. de; BLEGGI-TORRES, L. F.; ALVES, M. T. de S. Anáise da morfometria nuclear: descrição da metodologia e o papel dos softwares de edição de imagem. **Jornal Brasileiro de Patologia e Medicina Laboratorial**, scielo, v. 44, p. 51 – 57, 02 2008. ISSN 1676-2444. Available: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1676-24442008000100010&nrm=iso>.

BADDOUR, H. M.; MAGLIOCCA, K. R.; CHEN, A. Y. The importance of margins in head and neck cancer. **Journal of Surgical Oncology**, v. 113, p. n/a–n/a, 12 2015.

BAIK, J.; YE, Q.; ZHANG, L.; POH, C.; ROSIN, M.; MACAULAY, C.; GUILLAUD, M. Automated classification of oral premalignant lesions using image cytometry and random forests-based algorithms. **Cellular Oncology**, Springer Netherlands, v. 37, p. 193 – 202, 2014. Available: <https://doi.org/10.1007/s13402-014-0172-x>.

BEJNORDI, B. E.; VETA, M.; DIEST, P. J. van; GINNEKEN, B. van; KARSSEMEIJER, N.; LITJENS, G.; LAAK, J. A. W. M. van der; CONSORTIUM the C. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. **JAMA**, v. 318, n. 22, p. 2199–2210, 12 2017. ISSN 0098-7484. Available: <https://doi.org/10.1001/jama.2017.14585>.

BELL, E. S.; LAMMERDING, J. Causes and consequences of nuclear envelope alterations in tumour progression. **European journal of cell biology**, v. 95, p. 449–464, 2016.

BOFFETTA, P.; HECHT, S.; GRAY, N.; GUPTA, P.; STRAIF, K. Smokeless tobacco and cancer. **The lancet oncology**, v. 9, p. 667–75, 08 2008.

BOXBERG, M.; BOLLWEIN, C.; JÖHRENS, K.; KUHN, P.-H.; HALLER, B.; STEIGER, K.; WOLFF, K.-D.; KOLK, A.; JESINGHAUS, M.; WEICHERT, W. Novel prognostic histopathological grading system in oral squamous cell carcinoma based on tumour budding and cell nest size shows high interobserver and intraobserver

concordance. **Journal of Clinical Pathology**, BMJ Publishing Group, v. 72, n. 4, p. 285–294, 2019. ISSN 0021-9746. Available: <https://jcp.bmj.com/content/72/4/285>.

BRAY, F.; FERLAY, J.; SOERJOMATARAM, I.; SIEGEL, R.; TORRE, L.; JEMAL, A. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries: Global cancer statistics 2018. **CA: A Cancer Journal for Clinicians**, v. 68, 09 2018.

BUSLAEV, A.; IGLOVIKOV, V. I.; KHVEDCHENYA, E.; PARINOV, A.; DRUZHININ, M.; KALININ, A. A. Albumentations: Fast and flexible image augmentations. **Information**, v. 11, n. 2, 2020. ISSN 2078-2489. Available: <https://www.mdpi.com/2078-2489/11/2/125>.

CAMELYON16. **The website of The CAMELYON16 challenge**. 2022. Accessed: 2022-03-06. Available: <https://camelyon16.grand-challenge.org/Evaluation/>.

CHATURVEDI, A. K.; ENGELS, E. A.; PFEIFFER, R. M.; HERNANDEZ, B. Y.; XIAO, W.; KIM, E.; JIANG, B.; GOODMAN, M. T.; SIBUG-SABER, M.; COZEN, W.; LIU, L.; LYNCH, C. F.; WENTZENSEN, N.; JORDAN, R. C.; ALTEKRUSE, S.; ANDERSON, W. F.; ROSENBERG, P. S.; GILLISON, M. L. Human papillomavirus and rising oropharyngeal cancer incidence in the united states. **Journal of Clinical Oncology**, v. 29, n. 32, p. 4294–4301, 2011. Available: <https://doi.org/10.1200/JCO.2011.36.4596>.

DAKUA, S. P.; ABI-NAHED, J. Patient oriented graph-based image segmentation. **Biomedical Signal Processing and Control**, v. 8, n. 3, p. 325–332, 2013. ISSN 1746-8094. Available: <https://www.sciencedirect.com/science/article/pii/S1746809412001292>.

DAKUA, S. P.; ABI-NAHED, J.; AL-ANSARI, A. A. Pathological liver segmentation using stochastic resonance and cellular automata. **Journal of Visual Communication and Image Representation**, v. 34, p. 89–102, 2016. ISSN 1047-3203. Available: <https://www.sciencedirect.com/science/article/pii/S1047320315002138>.

DAKUA, S. P.; ABI-NAHED, J.; AL-ANSARI, A. A. A PCA-based approach for brain aneurysm segmentation. **Multidimensional Systems and Signal Processing**, v. 29, p. 257–277, 2018.

DAKUA, S. P.; SAHAMBI, J. S. Detection of Left Ventricular Myocardial Contours from Ischemic Cardiac MR Images. **IETE Journal of Research**, Taylor Francis, v. 57, n. 4, p. 372–384, 2011. Available: <https://www.tandfonline.com/doi/abs/10.4103/0377-2063.86338>.

DAKUA, S. P.; SAHAMBI, J. S. Modified active contour model and Random Walk approach for left ventricular cardiac MR image segmentation. **International Journal for Numerical Methods in Biomedical Engineering**, v. 27, p. 1350 – 1361, 09 2011.

DAS, D. K.; BOSE, S.; MAITI, A.; MITRA, B.; MUKHERJEE, G.; DUTTA, P. Automatic identification of clinically relevant regions from oral tissue histological images for oral squamous cell carcinoma diagnosis. **Tissue and Cell**, v. 53, 06 2018.

DAS, D. K.; CHAKRABORTY, C.; SAWAIMOON, S.; MAITI, A.; CHATTERJEE, S. Automated identification of keratinisation and keratin pearl area from in situ oral histological images. **Tissue and Cell**, v. 47, 06 2015.

DHANACHANDRA, N.; MANGLEM, K.; CHANU, Y. J. Image segmentation using k -means clustering algorithm and subtractive clustering algorithm. **Procedia Computer Science**, v. 54, p. 764–771, 2015. ISSN 1877-0509. Available: <https://www.sciencedirect.com/science/article/pii/S1877050915014143>.

DUMOULIN, V.; VISIN, F. **A guide to convolution arithmetic for deep learning**. 2018. Available: <http://arxiv.org/abs/1603.07285>.

ETTINGER, K. S.; GANRY, L.; FERNANDES, R. P. Oral cavity cancer. **Oral and Maxillofacial Surgery Clinics of North America**, v. 31, n. 1, p. 13–29, 2019. ISSN 1042-3699. Available: <https://www.sciencedirect.com/science/article/pii/S1042369918300670>.

FERNáNDEZ-CARROBLES, M. M.; BUENO, G.; DéNIZ, O.; SALIDO, J.; GARCíA-ROJO, M.; GONZáLEZ-LóPEZ, L. Influence of texture and colour in breast tma classification. **PloS one**, v. 10, 2015.

FILIPCZUK, P.; KOWAL, M.; OBUCHOWICZ, A. Automatic breast cancer diagnosis based on k-means clustering and adaptive thresholding hybrid segmentation. In: **IPC**. [S.l.: s.n.], 2011.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing (3rd Edition)**. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006. ISBN 013168728X.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016. Available: <http://www.deeplearningbook.org>.

GOODFELLOW, I.; POUGET-ABADIE, J.; MIRZA, M.; XU, B.; WARDE-FARLEY, D.; OZAIR, S.; COURVILLE, A.; BENGIO, Y. Generative adversarial nets. **Advances in neural information processing systems**, v. 27, 2014.

GUO, Z.; LIU, H.; NI, H.; WANG, X.; SU, M.; GUO, W.; WANG, K.; JIANG, T. A fast and refined cancer regions segmentation framework in whole-slide breast pathological images. **Scientific Reports**, v. 9, p. 882, 01 2019.

GURCAN, M. N.; BOUCHERON, L. E.; CAN, A.; MADABHUSHI, A.; RAJPOOT, N. M.; YENER, B. Histopathological image analysis: A review. **IEEE Reviews in Biomedical Engineering**, v. 2, p. 147–171, 2009.

HALICEK, M.; SHAHEDI, M.; LITTLE, J.; CHEN, A.; MYERS, L.; SUMER, B.; FEI, B. Head and neck cancer detection in digitized whole-slide histology using convolutional neural networks. **Scientific Reports**, v. 9, 12 2019.

HAO, S.; ZHOU, Y.; GUO, Y. A brief survey on semantic segmentation with deep learning. **Neurocomputing**, v. 406, p. 302–321, 2020. ISSN 0925-2312. Available: <https://www.sciencedirect.com/science/article/pii/S0925231220305476>.

HASHIBE, M.; BRENNAN, P.; BENHAMOU, S.; CASTELLSAGUE, X.; CHEN, C.; CURADO, M. P.; MASO, L. D.; DAUDT, A. W.; FABIANOVA, E.; WüNSCH-FILHO, V.; FRANCESCHI, S.; HAYES, R. B.; HERRERO, R.; KOIFMAN, S.; VECCHIA, C. L.; LAZARUS, P.; LEVI, F.; MATES, D.; MATOS, E.; MENEZES, A.; MUSCAT, J.; ELUF-NETO, J.; OLSHAN, A. F.; RUDNAI, P.; SCHWARTZ, S. M.; SMITH, E.; STURGIS, E. M.; SZESZENIA-DABROWSKA, N.; TALAMINI, R.; WEI, Q.; WINN, D. M.; ZARIDZE, D.; ZATONSKI, W.; ZHANG, Z.-F.; BERTHILLER, J.; BOFFETTA, P. Alcohol Drinking in Never Users of Tobacco, Cigarette Smoking in Never Drinkers, and the Risk of Head and Neck Cancer: Pooled Analysis in the International Head and Neck Cancer Epidemiology Consortium. **JNCI: Journal of the National Cancer Institute**, v. 99, n. 10, p. 777–789, 05 2007. ISSN 0027-8874. Available: <https://doi.org/10.1093/jnci/djk179>.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: **2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2016. p. 770–778.

HEIJMANS, H. J. A. M. Mathematical morphology: Basic principles. In: **Proceedings of Summer School on Morphological Image and Signal Processing**. [S.l.: s.n.], 1995. p. 228–231.

HU, Z.; TANG, J.; WANG, Z.; ZHANG, K.; ZHANG, L.; SUN, Q. Deep learning for image-based cancer detection and diagnosis a survey. **Pattern Recognition**, v. 83, p. 134–149, 2018. ISSN 0031-3203. Available: <https://www.sciencedirect.com/science/article/pii/S0031320318301845>.

INCA. **The website of the Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA)**. 2022. Accessed: 2022-02-19. Available: <https://www.inca.gov.br/tipos-de-cancer/cancer-de-boca>.

IRSHAD, H.; VEILLARD, A.; ROUX, L.; RACOCEANU, D. Methods for nuclei detection, segmentation and classification in digital histopathology: A review current status and future potential. **IEEE Rev. Biomed. Eng.**, v. 7, p. 97–114, 05 2014.

JUNQUEIRA, L. C. U.; CARNEIRO, J. **Histologia básica**. 12. ed. Rio de Janeiro: Editora Guanabara Koogan, 2013. ISBN 9788527723114.

KENDALL, A.; BADRINARAYANAN, V.; CIPOLLA, R. Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. In: . [S.l.: s.n.], 2017.

KHAN, A. M.; RAJPOOT, N.; TREANOR, D.; MAGEE, D. A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. **IEEE Trans. Biomed. Eng.**, v. 61, n. 6, p. 1729–1738, June 2014.

KINGMA, D. P.; BA, J. **Adam: A Method for Stochastic Optimization**. 2014.

KLECZEK, P.; JAWOREK-KORJAKOWSKA, J.; GORGON, M. A novel method for tissue segmentation in high-resolution he-stained histopathological whole-slide images. **Computerized Medical Imaging and Graphics**, v. 79, p. 101686, 2020. ISSN 0895-6111. Available: <https://www.sciencedirect.com/science/article/pii/S0895611119301016>.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: PEREIRA, F.; BURGES, C. J. C.; BOTTOU, L.; WEINBERGER, K. Q. (Ed.). **Advances in Neural Information Processing Systems 25**. [S.l.]: Curran Associates, Inc., 2012. p. 1097–1105. Available: <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. **Commun. ACM**, Association for Computing Machinery, New York, NY, USA, v. 60, n. 6, p. 84–90, maio 2017. ISSN 0001-0782. Available: <https://doi.org/10.1145/3065386>.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 2015.

LECUN, Y.; BOSER, B.; DENKER, J. S.; HOWARD, R. E.; HABBARD, W.; JACKEL, L. D.; HENDERSON, D. Handwritten digit recognition with a back-propagation network. In: _____. **Advances in Neural Information Processing Systems 2**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1990. p. 396–404. ISBN 1558601007.

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. **Proc. IEEE**, v. 86, n. 11, p. 2278–2324, November 1998.

LIAO, L.-J.; HSU, W.-L.; LO, W.-C.; CHENG, P.-W.; SHUENG, P.-W.; HSIEH, C.-H. Health-related quality of life and utility in head and neck cancer survivors. **BMC Cancer**, v. 19, 12 2019.

LITJENS, G.; BANDI, P.; BEJNORDI, B. E.; GEESSINK, O.; BALKENHOL, M.; BULT, P.; HALILOVIć, A.; HERMSEN, M.; LOO, R.; VOGELS, R.; MANSON, Q.; STATHONIKOS, N.; BAIDOSHVILI, A.; DIEST, P.; WAUTERS, C.; DIJK, M. van; LAAK, J. van der. 1399 he-stained sentinel lymph node sections of breast cancer patients: The camelyon dataset. **GigaScience**, v. 7, 05 2018.

LITJENS, G.; KOOI, T.; BEJNORDI, B. E.; SETIO, A. A. A.; CIOMPI, F.; GHAFOORIAN, M.; LAAK, J. A. van der; GINNEKEN, B.; SÁNCHEZ, C. I. A survey on deep learning in medical image analysis. **Medical image analysis**, v. 42, p. 60–88, 2017.

LONG, J.; SHELHAMER, E.; DARRELL, T. J. Fully convolutional networks for semantic segmentation. In: **IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2015. p. 3431–3440.

LUMERMAN, H.; FREEDMAN, P.; KERPEL, S. Oral epithelial dysplasia and the development of invasive squamous cell carcinoma. **Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology**, v. 79, n. 3, p. 321 – 329, 1995. ISSN 1079-2104. Available: <http://www.sciencedirect.com/science/article/pii/S1079210405802264>.

MACENKO, M.; NIETHAMMER, M.; MARRON, J. S.; BORLAND, D.; WOOSLEY, J. T.; GUAN, X.; SCHMITT, C.; THOMAS, N. E. A method for normalizing histology slides for quantitative analysis. In: **IEEE International Symposium on Biomedical Imaging: From Nano to Macro**. [S.l.: s.n.], 2009. p. 1107–1110.

MAHMOOD, H.; SHABAN, M.; INDAVE, B.; SANTOS-SILVA, A.; RAJPOOT, N.; KHURRAM, S. A. Use of artificial intelligence in diagnosis of head and neck precancerous and cancerous lesions: A systematic review. **Oral Oncology**, v. 110, p. 104885, 11 2020.

MARTINO, F.; BLOISI, D. D.; PENNISI, A.; FAWAKHERJI, M.; ILARDI, G.; RUSSO, D.; NARDI, D.; STAIBANO, S.; MEROLLA, F. Deep learning-based pixel-wise lesion segmentation on oral squamous cell carcinoma images. **Applied Sciences**, v. 10, n. 22, 2020. ISSN 2076-3417. Available: <https://www.mdpi.com/2076-3417/10/22/8285>.

MCCANN, M. T.; OZOLEK, J. A.; CASTRO, C. A.; PARVIN, B.; KOVACEVIC, J. Automated histology analysis: Opportunities for signal processing. **IEEE Signal Process. Mag.**, v. 32, n. 1, p. 78–87, 2015. Available: <https://doi.org/10.1109/MSP.2014.2346443>.

MOLINARO, E. M.; CAPUTO, L. F. G.; AMENDOEIRA, M. R. R. **Conceitos e métodos para a formação de profissionais em laboratórios de saúde.** Rio de Janeiro: EPSJV, 2010. v. 2. Available: <https://www.arca.fiocruz.br/handle/icict/13410>. ISBN 9788598768410.

NAJMAN, L.; SCHMITT, M. Watershed of a continuous function. **Signal Processing**, v. 38, n. 1, p. 99–112, 1994. ISSN 0165-1684. Available: <https://www.sciencedirect.com/science/article/pii/0165168494900590>.

NCI. **The website of the US National Cancer Institute (NCI)**. 2022. Accessed: 2022-02-19. Accessed: 2022-02-19. Available: <https://www.cancer.gov/about-cancer/understanding/what-is-cancer>.

NOCK, R.; NIELSEN, F. Statistical region merging. **IEEE transactions on pattern analysis and machine intelligence**, v. 26, p. 1452–8, 12 2004.

OLIVEIRA, D.; NASCIMENTO, M. Z. do; NEVES, L.; GODOY, M.; ARRUDA, P.; NETO, D. Unsupervised segmentation method for cuboidal cell nuclei in histological prostate images based on minimum cross entropy. **Expert Syst. Appl.**, v. 40, p. 7331–7340, 12 2013.

OPENSLIDE. **The website of the OpenSlide Project**. 2022. Accessed: 2022-05-14. Available: <https://openslide.org/formats/aperio/>.

OTSU, N. A threshold selection method from gray-level histograms. **IEEE Trans. Syst., Man, Cybern.**, v. 9, n. 1, p. 62–66, Jan 1979.

PASZKE, A.; GROSS, S.; CHINTALA, S.; CHANAN, G.; YANG, E.; DEVITO, Z.; LIN, Z.; DESMAISON, A.; ANTIGA, L.; LERER, A. Automatic differentiation in PyTorch. In: **NIPS Autodiff Workshop**. [S.l.: s.n.], 2017.

PHOULADY, H. A.; GOLDGOF, D. B.; HALL, L. O.; MOUTON, P. R. Nucleus segmentation in histology images with hierarchical multilevel thresholding. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. **Medical Imaging 2016: Digital Pathology**. [S.l.]: SPIE, 2016. v. 9791, p. 280 – 285.

POLONI, K. M.; FERRARI, R. J. A deep ensemble hippocampal cnn model for brain age estimation applied to alzheimer's diagnosis. **Expert Systems with Applications**, v. 195, p. 116622, 2022. ISSN 0957-4174.

PULTE, D.; BRENNER, H. Changes in survival in head and neck cancers in the late 20th and early 21st century: A period analysis. **The oncologist**, v. 15, p. 994–1001, 09 2010.

REINHARD, E.; ADHIKHMIN, M.; GOOCH, B.; SHIRLEY, P. Color transfer between images. **IEEE Comput. Graph. Appl.**, v. 21, p. 34–41, 2001.

ROBERTSON, A. R. The CIE 1976 color-difference formulae. In: **Color Research & Application**. [S.l.: s.n.], 1977. v. 2, p. 7–11.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: **Medical Image Computing and Computer-Assisted Intervention (MICCAI)**. [S.l.]: Springer, 2015. (LNCS, v. 9351), p. 234–241. (available on arXiv:1505.04597 [cs.CV]).

RUSSAKOVSKY, O.; DENG, J.; SU, H.; KRAUSE, J.; SATHEESH, S.; MA, S.; HUANG, Z.; KARPATHY, A.; KHOSLA, A.; BERNSTEIN, M.; BERG, A. C.; FEI-FEI, L. ImageNet Large Scale Visual Recognition Challenge. **International Journal of Computer Vision (IJCV)**, v. 115, n. 3, p. 211–252, 2015.

SALVI, M.; ACHARYA, U. R.; MOLINARI, F.; MEIBURGER, K. M. The impact of pre- and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis. **Computers in Biology and Medicine**, v. 128, p. 104129, 2021. ISSN 0010-4825. Available: <https://www.sciencedirect.com/science/article/pii/S0010482520304601>.

SHABAN, M.; KHURRAM, S. A.; FRAZ, M. M.; ALSUBAIE, N.; MASOOD, I.; MUSHTAQ, S.; HASSAN, M.; LOYA, A.; RAJPOOT, N. M. A novel digital score for abundance of tumour infiltrating lymphocytes predicts disease free survival in oral squamous cell carcinoma. **Scientific Reports**, Nature Research (part of Springer Nature), v. 9, n. 1, September 2019. Available: <http://eprints.whiterose.ac.uk/152371/>.

SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. **Journal of Big Data**, v. 6, n. 1, p. 60, 2019. ISSN 2196-1115.

SILVA, A. B.; MARTINS, A. S.; NEVES, L. A.; FARIA, P. R. de; TOSTA, T. A. A.; NASCIMENTO, M. Z. do. Automated nuclei segmentation in dysplastic histopathological oral tissues using deep neural networks. In: NYSTRÖM, I.; HEREDIA, Y. H.; NÚÑEZ, V. M. (Ed.). **Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications**. Cham: Springer International Publishing, 2019. p. 365–374. ISBN 978-3-030-33904-3.

SILVA, A. B.; SANTOS, D. F. D. dos; TOSTA, T. A. A.; MARTINS, A. S.; NEVES, L. A.; TRAVENçOLO, B. A. N.; FARIA, P. R. de; NASCIMENTO, M. Z. do. Segmentation of oral epithelial dysplasias employing mask r-cnn and color normalization. In: **2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)**. Los Alamitos, CA, USA: IEEE Computer Society, 2020. p. 2818–2824. Available: <https://doi.ieeecomputersociety.org/10.1109/BIBM49941.2020.9313101>.

SIMONYAN, K.; ZISSERMAN, A. **Very Deep Convolutional Networks for Large-Scale Image Recognition**. 2015.

SOKOLOVA, M.; LAPALME, G. A systematic analysis of performance measures for classification tasks. **Information Processing Management**, v. 45, n. 4, p. 427 – 437, 2009. ISSN 0306-4573. Available: <http://www.sciencedirect.com/science/article/pii/S0306457309000259>.

SRINIDHI, C. L.; CIGA, O.; MARTEL, A. L. Deep neural network models for computational histopathology: A survey. **Medical Image Analysis**, v. 67, p. 101813, 2021. ISSN 1361-8415. Available: <https://www.sciencedirect.com/science/article/pii/S1361841520301778>.

SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCKE, V.; RABINOVICH, A. Going deeper with convolutions. In: **2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2015. p. 1–9.

TCGA. **The website of The Cancer Genome Atlas Program (TCGA)**. 2022. Accessed: 2022-02-19. Available: <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>.

TOSTA, T. A. A.; FARIA, P. R. de; SERVATO, J. P. S.; NEVES, L. A.; ROBERTO, G. F.; MARTINS, A. S.; NASCIMENTO, M. Z. do. Unsupervised method for normalization of hematoxylin-eosin stain in histological images. **Comput. Med. Imaging Graph.**, v. 77, p. 101646, 2019. ISSN 0895-6111.

VAHADANE, A.; PENG, T.; SETHI, A.; ALBARQOUNI, S.; WANG, L.; BAUST, M.; STEIGER, K.; SCHLITTER, A. M.; ESPOSITO, I.; NAVAB, N. Structure-preserving color normalization and sparse stain separation for histological images. **IEEE Trans. Med. Imag.**, v. 35, n. 8, p. 1962–1971, 2016.

VAHADANE, A.; SETHI, A. Towards generalized nuclear segmentation in histological images. In: **13th IEEE International Conference on BioInformatics and BioEngineering**. [S.l.: s.n.], 2013. p. 1–4.

WANG, Z.; WANG, E.; ZHU, Y. Image segmentation evaluation: a survey of methods. **Artif. Intell. Rev.**, v. 53, n. 8, p. 5637–5674, 2020. Available: <https://doi.org/10.1007/s10462-020-09830-9>.

WARNAKULASURIYA, S.; REIBEL, J.; BOUQUOT, J.; DABELSTEEN, E. Oral epithelial dysplasia classification systems: predictive value, utility, weaknesses and scope for improvement. **Journal of Oral Pathology & Medicine**, Wiley-Blackwell, v. 37, n. 3, p. 127–33, 3 2008. ISSN 0904-2512.

WHO. **The website of the World Health Organization (WHO)**. 2022. Accessed: 2022-02-19. Available: <https://www.who.int/health-topics/cancer#tab=tab_1>.

WIENERT, S.; HEIM, D.; SAEGER, K.; STENZINGER, A.; BEIL, M.; HUFNAGL, P.; DIETEL, M.; DENKERT, C.; KLAUSCHEN, F. Detection and segmentation of cell nuclei in virtual microscopy images: A minimum-model approach. **Sci. Rep.**, v. 2, p. 503, 07 2012.

WILS, L. J.; POELL, J. B.; EVREN, I.; KOOPMAN, M. S.; BROUNS, E. R. E. A.; VISSCHER, J. G. A. M. de; BRAKENHOFF, R. H.; BLOEMENA, E. Incorporation

of differentiated dysplasia improves prediction of oral leukoplakia at increased risk of malignant progression. **Modern Pathology**, v. 33, p. 1033–1040, July 2020.

XIAO, Y.; DECENCIèRE, E.; VELASCO-FORERO, S.; BURDIN, H.; BORNSCHLöGL, T.; BERNERD, F.; WARRICK, E.; BALDEWECK, T. A new color augmentation method for deep learning segmentation of histological images. In: **2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)**. [S.l.: s.n.], 2019. p. 886–890.

XING, F.; XIE, Y.; SU, H.; LIU, F.; YANG, L. Deep learning in microscopy image analysis: A survey. **IEEE Transactions on Neural Networks and Learning Systems**, v. 29, n. 10, p. 4550–4568, 2018.

Xing, F.; Xie, Y.; Su, H.; Liu, F.; Yang, L. Deep learning in microscopy image analysis: A survey. **IEEE Trans. Neural Netw. Learn. Syst.**, v. 29, n. 10, p. 4550–4568, Oct 2018.

XUN, S.; LI, D.; ZHU, H.; CHEN, M.; WANG, J.; LI, J.; CHEN, M.; WU, B.; ZHANG, H.; CHAI, X.; JIANG, Z.; ZHANG, Y.; HUANG, P. Generative adversarial networks in medical image segmentation: A review. **Computers in Biology and Medicine**, v. 140, p. 105063, 11 2021.

YU, J.; LIN, Z.; YANG, J.; SHEN, X.; LU, X.; HUANG, T. S. **Generative Image Inpainting with Contextual Attention**. 2018.

ZHONG, Z.; ZHENG, L.; KANG, G.; LI, S.; YANG, Y. Random erasing data augmentation. **Proceedings of the AAAI Conference on Artificial Intelligence**, v. 34, 08 2017.