Universidade Estadual de Campinas
Instituto de Computação

Elvis Rusnel Capia Quispe

Learning non-maximum suppression to improve lung nodule detection in CT images

Aprendendo a suprimir não-máximos para aperfeiçoar a detecção de nódulos pulmonares em imagens de CT

CAMPINAS
2020

Elvis Rusnel Capia Quispe

Learning non-maximum suppression to improve lung nodule detection in CT images

Aprendendo a suprimir não-máximos para aperfeiçoar a detecção de nódulos pulmonares em imagens de CT

Dissertação apresentada ao Instituto de Computação da Universidade Estadual de Campinas como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Dissertation presented to the Institute of Computing of the University of Campinas in partial fulfillment of the requirements for the degree of Master in Computer Science.

Supervisor/Orientador: Prof. Dr. Alexandre Xavier Falcão

Este exemplar corresponde à versão final da Dissertação defendida por Elvis Rusnel Capia Quispe e orientada pelo Prof. Dr. Alexandre Xavier Falcão.

CAMPINAS
2020

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

C172L

Capia Quispe, Elvis Rusnel, 1993-
Learning non-maximum suppression to improve lung nodule detection in CT images / Elvis Rusnel Capia Quispe. – Campinas, SP : [s.n.], 2020.

Orientador: Alexandre Xavier Falcão.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Computação.

1. Tomografia computadorizada por raios X. 2. Processamento de imagens - Técnicas digitais. 3. Pulmões - Câncer. 4. Diagnóstico auxiliado por computador. 5. Redes neurais profundas. I. Falcão, Alexandre Xavier, 1966-. II. Universidade Estadual de Campinas. Instituto de Computação. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Aprendendo a suprimir não-máximos para aperfeiçoar a detecção de nódulos pulmonares em imagens de CT
Palavras-chave em inglês:
X-rays computed tomography
Image processing - Digital techniques
Lungs - Cancer
Computer-aided diagnosis
Deep neural networks
Área de concentração: Ciência da Computação
Titulação: Mestre em Ciência da Computação
Banca examinadora:
Alexandre Xavier Falcão [Orientador]
Fabiano Reis
Sandra Eliza Fontes de Avila
Data de defesa: 26-06-2020
Programa de Pós-Graduação: Ciência da Computação

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0003-1824-0566
- Currículo Lattes do autor: http://lattes.cnpq.br/6014727378542158

Universidade Estadual de Campinas
Instituto de Computação

Elvis Rusnel Capia Quispe

Learning non-maximum suppression to improve lung nodule detection in CT images

Aprendendo a suprimir não-máximos para aperfeiçoar a detecção de nódulos pulmonares em imagens de CT

Banca Examinadora:

- Prof. Dr. Alexandre Xavier Falcão
  Instituto de Computação, UNICAMP
- Prof. Dr. Fabiano Reis
  Faculdade de Ciências Médicas, UNICAMP
- Profa. Dra. Sandra Eliza Fontes de Avila
  Instituto de Computação, UNICAMP

A ata da defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

Campinas, 26 de junho de 2020

Acknowledgements

I would like to express my gratitude to Prof. Dr. Alexandre Xavier Falcão, my advisor. He gave me the opportunity to be part of the Laboratory of Image Data Science (LIDS) and guided me not only in the right direction along the master's degree but also in my personal life. He also took the time to answer all my questions regarding my master's degree project, discuss new ideas to address it, and I have learned a lot from his experience and knowledge. I am very happy and proud of working with him during my master's degree.

Besides him, I would like to thank my parents, Pastor Capia Cama and Alicia Quispe Quea to support me in my education together with my brothers. I would also like to thank my friends Kent Yucra, Walter Erquinigo, Chen Soncco, and Dmytro Soboliev to support and guided me prior and during the master's degree.

The authors would like to thank CNPq (303808/2018-7) and grant 2014/12236-1, São Paulo Research Foundation (FAPESP) for funding our project. Last but not the least, I would like to thank my coworkers from LIDS, I have learned a lot from them and shared unique moments with them during the master's degree.

Resumo

O câncer de pulmão é o tipo mais comum de câncer em homens e o terceiro mais comum em mulheres. Devido ao mau prognóstico, o câncer de pulmão é responsável pela maior taxa de mortalidade, atingindo 1,8 milhão de mortes por ano. O diagnóstico e o tratamento nos estágios iniciais podem aumentar as chances de sobrevivência. A tomografia computadorizada (TC) é a modalidade de imagem preferida para detectar e diagnosticar câncer de pulmão, pois fornece imagens 3D do tórax em alta resolução, facilitando a detecção de pequenos nódulos. No entanto, a natureza 3D das imagens dificulta sua análise visual. Como conseqüência, o número de falsos positivos ainda é alto e, mesmo contando com a opinião de vários especialistas, o diagnóstico é frequentemente sujeito a alguma falta de consenso.

Os sistemas de Diagnóstico Assistida por Computador (CAD) foram desenvolvidos para solucionar o problema, auxiliando especialistas na tarefa de detecção e classificação mais rápidas e precisas de anormalidades. As técnicas usadas nos sistemas CAD podem ser divididas em dois grupos: sistemas CAD que exploram features de imagem baseados em conhecimento e sistemas CAD que aprendem os features de imagens anotadas, principalmente baseadas em aprendizado profundo por meio de redes neurais convolucionais (CNNs).

Na última década, muitos métodos computacionais (sistemas CAD) foram desenvolvidos para auxiliar os médicos na detecção de nódulos pulmonares. Tais métodos são baseados principalmente em CNNs, que alcançaram resultados promissores na detecção precoce de nódulos pulmonares. No entanto, esses métodos geram várias regiões candidatas por nódulo, de modo que um algoritmo de não-máxima supressão (NMS) é necessário para selecionar uma única região por nódulo, eliminando as redundantes. O GossipNet é uma rede neural 1D para NMS, que pode aprender os parâmetros do NMS em vez de confiar nos parâmetros artesanais. No entanto, o GossipNet não tira proveito dos features de imagem para aprender NMS.

Neste trabalho, propomos um sistema CAD automatizado para detecção de nódulos pulmonares, que consiste em quatro módulos: pré-processamento, a definição de uma região de interesse (por exemplo, por segmentação pulmonar), detecção de nódulos e a eliminação de candidatos redundantes. Para a segmentação pulmonar, usamos uma abordagem recente baseada em sequências de transformações florestais de imagem (IFTs) denominada ALTIS, fornecendo uma segmentação mais precisa dos pulmões em comparação com o método usado no desafio LUNA16. Para a detecção de nódulos e a eliminação de candidatos redundantes, usamos o 3D Faster R-CNN com ResNet18 para a detecção de regiões candidatas com nódulos e apresentamos *FeatureNMS* — uma rede neural que fornece features de imagem adicionais à entrada do GossipNet, que resultam de uma transformação sobre as intensidades de voxel de cada região candidata na imagem da TC. Para validação, usamos o conjunto de dados de desafio LUNA16.

Abstract

Lung cancer is the most common type of cancer in men and the third most common one in women. Due to poor prognosis, lung cancer is responsible for the largest mortality rate, reaching 1.8 million deaths per year. Diagnosis and treatment at the early stages can increase the chances of survival. Computerized Tomography (CT) is the imaging modality of preference to detect and diagnose lung cancer since it provides high-resolution 3D images of the thorax, facilitating the detection of small nodules. However, the 3D nature of the images makes their visual analysis difficult. As a consequence, the number of false positives is still high and, even by counting on the opinion of multiple specialists, the diagnosis is often subjected to some lack of consensus.

Computer-Aided Detection (CAD) systems have been developed to address the problem, assisting to specialists in the task of quicker and more accurate detection and classification of abnormalities. The techniques used in CAD systems may be divided into two groups: CAD systems that explore knowledge-based image features and CAD systems that learn the features from annotated images, mostly based on deep learning through Convolutional Neural Networks (CNNs).

In the last decade, many computational methods (CAD systems) have been developed to assist physicians in lung nodule detection. Such methods are mostly based on CNNs, which have achieved promising results in early detection of lung nodules. However, these methods generate several candidate regions per nodule, such that a Non-Maximum Suppression (NMS) algorithm is required to select a single region per nodule while eliminating the redundant ones. GossipNet is a 1D Neural Network (NN) for NMS, which can learn the NMS parameters rather than relying on handcrafted ones. However, GossipNet does not take advantage of image features to learn NMS.

In this work, we propose an automated CAD system for lung nodule detection which consists of four modules: pre-processing, the definition of a region of interest (e.g., by lung segmentation), nodule detection, and the elimination of redundant candidates. For lung segmentation, we use a recent approach based on sequences of Image Foresting Transforms (IFTs) named ALTIS providing a more accurate segmentation of the lungs compared to the method used in the LUNA16 challenge. For nodule detection and the elimination of redundant candidates, we use 3D Faster R-CNN with ResNet18 for the detection of candidate regions with nodules and present *FeatureNMS* — a neural network that provides additional image features to the input of GossipNet, which result from a transformation over the voxel intensities of each candidate region in the CT image. For validation, we use the LUNA16 challenge dataset.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

1.1 Motivation and overview

Lung cancer is the most common type of cancer in men and the third most common one in women. According to the International Agency of Research on Cancer (IARC), lung cancer is the leading type of cancer with new cases — 2.1 million in 2018. Due to poor prognosis, lung cancer is also responsible for the largest mortality rate, reaching 1.8 million deaths per year. Diagnosis and treatment at the early stages can increase the chances of survival. According to the World Health Organization (WHO) and the Forum of International Respiratory Societies (FIRS), the premature detection of lung cancer provides 5-year survival rates within $70\% - 90\%$. However, in the majority of the cases, lung cancer is only detected at an advanced stage, dropping the one-year survival rate from $80\% - 85\%$ to $15\% - 19\%$, when it is compared with the detection at a premature stage.

A lung nodule is a small round growth of tissue within the chest cavity (see Figure 2.11). Generally, nodules are considered less than 30 mm in size, larger sizes are called masses and are presumed to be malignant (cancerous). Nodules between $5 - 30$ mm may be benign or malignant, and the probability that a nodule be malignant increases with its size.

Initially, chest radiography screenings (chest X-ray) were used to detect and diagnose lung cancer, but chest radiography screenings present two main problems: (i) lack of 3D information about the nodules and (ii) occlusion due to the 2D projection. This makes Computerized Tomography (CT) the method of preference since it provides high-resolution 3D images of the thorax, facilitating the detection of small nodules. However, the 3D nature of the images makes their visual analysis difficult. As a consequence, the number of false positives is still high and, even by counting on the opinion of multiple specialists, the diagnosis is often subjected to some lack of consensus.

Computer-Aided Detection (CAD) systems have been developed to address the problem, assisting to specialists in the task of quicker and more accurate detection and classification of abnormalities. A full CAD system to detect and diagnose lung cancer consists of two modules: CADe for the detection of nodules and CADx for the classification of the degree of malignancy of nodules. The techniques used in CAD systems may be divided into two groups: CAD systems that explore knowledge-based image features and CAD

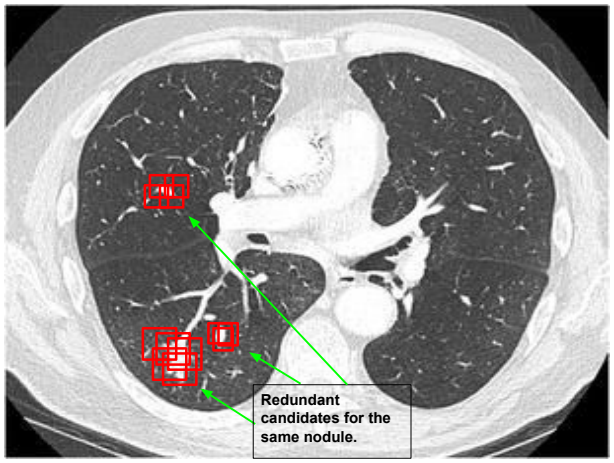


Figure 1.1: Candidate regions derived from a nodule detection algorithm based on CNN [46]. Note that there are redundant candidates for each nodule.

systems that learn the features from annotated images, mostly based on deep learning through Convolutional Neural Networks (CNNs). This work is focused on CADe systems — for the detection of lung nodules.

As it is often the case with medical images, CAD systems that learn features mostly based on deep learning have been limited by the relatively small number of annotated images. However, the publicly available LIDC-IDRI dataset [3] and the LUNA16 challenge [35] have had a big influence on deep learning methods proposed for lung nodule detection. Relevant work has been published in the context of the LUNA16 challenge.

In the last decade, many computational methods (CAD systems) have been developed to assist physicians in lung nodule detection [11, 10, 5, 46]. Such methods are mostly based on CNNs, which have achieved promising results in early detection of lung nodules. However, these methods usually generate redundant candidate regions for the same object (Figure 1.1). Traditional Non-Maximum Suppression (NMS) algorithms [26, 32] have been proposed to eliminate redundant candidates and select one candidate region per object, usually based on handcrafted parameters. In [16], these algorithms are referred to as the GreedyNMS approach since they select the region with the highest object score among each group of overlapping candidate regions. GreedyNMS presents two major difficulties: (i) the overlapping percentage for rejection must be high enough to eliminate regions with high scores containing the same object, while (ii) it must be low enough to avoid eliminating regions of detected objects that are close to each other — a common situation in images with several nodules. Improvements in NMS have also been proposed in the past years. The Soft-NMS [4] algorithm applies a decay function to the object scores of candidate regions to avoid losing neighboring objects. GossipNet [16] uses a 1D NN to analyze geometric features from overlapping candidate regions and modify their scores, such that each object should be represented by a single region with the highest score among the overlapping ones. Different from other NMS approaches, GossipNet can learn the NMS parameters rather than relying on handcrafted ones. However, GossipNet does not take advantage of image features to learn NMS.

In this work, we propose an automated CAD system for lung nodule detection which

consists of four modules: pre-processing, the definition of a region of interest (e.g., by lung segmentation), nodule detection, and the elimination of redundant candidates. For lung segmentation, mostly methods [10, 11, 46] have used the provided segmentation in the LUNA16 challenge which is based on [44]. However, we use a recent approach based on sequences of Image Foresting Transforms (IFTs) named ALTIS providing a more accurate segmentation of the lungs which has a positive impact on the results. For nodule detection and the elimination of redundant candidates, we use 3D Faster R-CNN with ResNet18 [46] for the detection of candidate regions with nodules and propose a transformation over the voxel intensities of those regions in the CT image to include image features in GossipNet [16] for NMS. The new neural network with a modified GossipNet is called *FeatureNMS* (Figure 4.3). For validation, we use the LUNA16 challenge dataset.

1.2 Objectives

The aim of this work is to develop an automated CAD system for lung nodule detection with high sensitivity and low false-positive rate, taking as input CT images. In order to achieve high sensitivity and low false-positive rate in our CAD system, we use the ALTIS algorithm for lung segmentation and propose the FeatureNMS for the elimination of redundant candidates.

1.3 Contributions

In this work, we propose an automated CAD system for lung nodule detection which consists of four modules: pre-processing, the definition of a region of interest (e.g., by lung segmentation), nodule detection, and the elimination of redundant candidates. In order to achieve high sensitivity and low false-positive rate in our CAD system, differently from other methods using the lung segmentation provided in the LUNA16 challenge, we use the ALTIS algorithm for lung segmentation providing a slight improvement when it is tested on LUNA16 dataset. On average, the number of voxels segmented by the ALTIS algorithm is 8% less than the number of voxels segmented by the method used in the LUNA16 challenge. This shows that the ALTIS algorithm is more accurate.

For the elimination of redundant candidates, after the 3D Faster R-CNN with ResNet18 [46] produces the candidate regions with nodules, we propose a transformation over the voxel intensities of those regions in the CT image to include image features in GossipNet [16] for NMS. The new neural network with a modified GossipNet is named *FeatureNMS* (Figure 4.3). When the FeatureNMS is tested on the LUNA16 dataset, results indicate a considerable improvement when it is compared to the original GossipNet and the best GreddyNMS.

To summarize, this work has two main contributions. The first is the incorporation of the ALTIS algorithm for lung segmentation providing a slight improvement when it is compared to the method used in the LUNA16 challenge. The second contribution is the proposed FeatureNMS for the elimination of redundant candidates. Results indicate that FeatureNMS outperforms the original GossipNet and the best GreedyNMS.

The results of the proposed FeatureNMS for the elimination of redundant candidates as part of our CAD system for lung nodule detection were published in [6].

1.4 Organization

The present work is organized as follows: Chapter 2 presents the fundamentals to understand the main concepts used in the remaining chapters. Chapter 3 presents the related works for lung nodule detection and NMS algorithms. The proposed method is presented in Chapter 4. The experiments that were performed to validate the proposed method are presented in Chapter 5. Finally, Chapter 6 presents the conclusions of the work and discusses opportunities for future work.

Chapter 2

Fundamentals

2.1 Anatomy of the lung

The purpose of the lungs is to provide oxygen to the blood. They are a pair of large and spongy organs that are localized in the thorax lateral to the heart and on the upper part of the diaphragm (see Figure 2.1). Each lung is surrounded by a membrane called pleura that provides the lung with space to expand. The left and right lungs are slightly different in size and shape due to the heart which is located near the left lung. Therefore, the left lung is slightly smaller than the right lung and consists of 2 lobes while the right lung has 3 lobes. The interior of the lungs is made of around 30 million sacks which are called the alveoli. Alveoli are lined with thin simple squamous epithelium that allows air entering the alveoli to exchange its gases with the blood passing through the capillaries.

The air, which contains oxygen and other gases, comes into the body through the lungs. In the lungs, the oxygen is moved into the blood-stream and carried through the body. Red blood cells collect the carbon dioxide and transport it back to the lungs, where it leaves the body when we exhale.

2.2 Lung cancer

Lung cancer is commonly due to smoking and it is mainly caused by the uncontrollable irregular growth of cells in lung tissue. A lung nodule is a small round growth of tissue within the chest cavity (see Figure 2.11). Generally, nodules are considered less than 30 mm in size, larger sizes are called masses and are presumed to be malignant (cancerous). Nodules between $5 - 30$ mm may be benign or malignant, and the probability that a nodule be malignant increases with its size. There are three main types of lung cancer [1], which are non-small cell lung cancer (NSCLC), small cell lung cancer, and lung carcinoid tumor. NSCLC is the most common type of lung cancer, which constitutes about 85% of all lung cancers.

Lung cancer is the leading cause of cancer-related death in the world [2], representing 18.4% in 2018. Detection and treatment at an early stage can effectively overcome this burden and increase the chance of survival of patients. According to the World Health Organization (WHO) and the Forum of International Respiratory Societies (FIRS), the

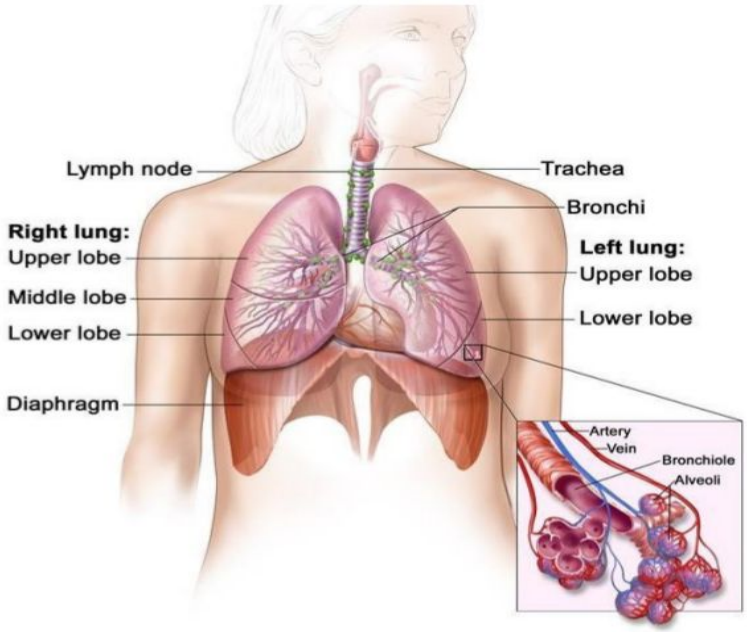


Figure 2.1: Anatomy of the respiratory system [5].

premature detection of lung cancer provides 5-year survival rates within $70\% - 90\%$. However, in the majority of the cases, lung cancer is only detected at an advanced stage, dropping the one-year survival rate from $80\% - 85\%$ to $15\% - 19\%$, when it is compared with the detection at a premature stage.

2.3 Computerized Tomography imaging

Initially, chest radiography screenings (chest X-ray) were used to detect and diagnose lung cancer. However, chest radiography screenings present two main problems: (i) lack of 3D information about the nodules and (ii) occlusion due to the 2D projection. Therefore, a more effective non-invasive imaging technique is required that enables specialists to see inside the body without the risks of exploratory surgery and thus the diagnostic of the presence of nodules.

Computerized Tomography (CT) is the method of preference since it forms 3D images of the thorax, resulting in a higher resolution of nodules and tumor pathology, facilitating the detection of small nodules. CT images are primarily used today to assist specialists in the diagnosis of abnormalities, however, using CT images for nodule detection and classification present three main problems:

- Due to the 3D nature of CT images, it requires analyzing hundreds of images at a time, specialists are overwhelmed with the amount of data to process since it is time-consuming and the fatigue also reduces the effectiveness.

- As a consequence, the number of false positives generated after analyzing CT images remain high.

- As multiple specialists are consulted to analyze the same exam of CT, the lack of consensus on a diagnosis can arise, demanding that more specialists have to analyze that image to obtain a verdict.

2.4 Computer-Aided Detection systems

The Computer-Aided Detection (CAD) system has been developed to assist specialists in the task of quicker and more accurate detection and classification of abnormalities. Existing CAD systems may be divided according to the techniques used into two groups: (i) CAD systems that explore knowledge-based image features and (ii) CAD systems based on deep learning with automatic feature extraction. Approaches in the first group often are called traditional techniques and measure radiological trails such as nodule size, location, shape, texture, and apply a classifier to determine if a nodule is malignant or benign. In the second group, such models based on deep neural networks can automatically learn features for the detection and diagnosis of lung nodules in CT images. In the last years CAD systems based on deep learning have shown more promising results for lung nodule detection and classification [34, 11, 10, 5, 7, 46, 41, 37, 36, 45, 20, 47, 8].

Traditional CAD systems involve manually designed features or descriptors for lung nodule detection [25, 21]. Thus traditional CAD systems use image processing techniques to generate a large number of candidates for the location of nodules [35], followed by a false positives reduction, and finally, classification is used to know the degree of malignancy of nodules. The process of detecting and classifying nodules in CT images is still primarily performed using traditional techniques and trained pulmonary specialists.

A full CAD system to detect and diagnose lung cancer consists of a detection system (often abbreviated as CADe) and a diagnostic system (often abbreviated as CADx). On one hand, the CADe system detects candidates being nodule or non-nodule performing an initial high-sensitive nodule detection. On the other hand, the goal of the CADx system is to classify the degree of malignancy of nodules.

2.5 Deep learning

Deep learning is a new area in the machine learning field, which has been growing very fast in general data analysis in the last years. Deep learning is an improvement of artificial neural networks (ANN), consisting of more layers that allow a higher level of abstraction and improved prediction from data [23]. To date, deep learning is emerging as the leading machine learning tool in the general imaging and computer vision domains. The major prove is that deep learning is surpassing and improving the traditional techniques based on image processing and traditional machine learning algorithms like support vector machine or shallow neural network, in every critical task: image recognition, recognizing speech, characterizing images, generating natural [39].

Since deep learning depends on a lot of data to learn, training a complex neural network from scratch on lung nodule images may not prove very successful. However, transfer learning, or training a network on a large dataset and then using these trained weights for

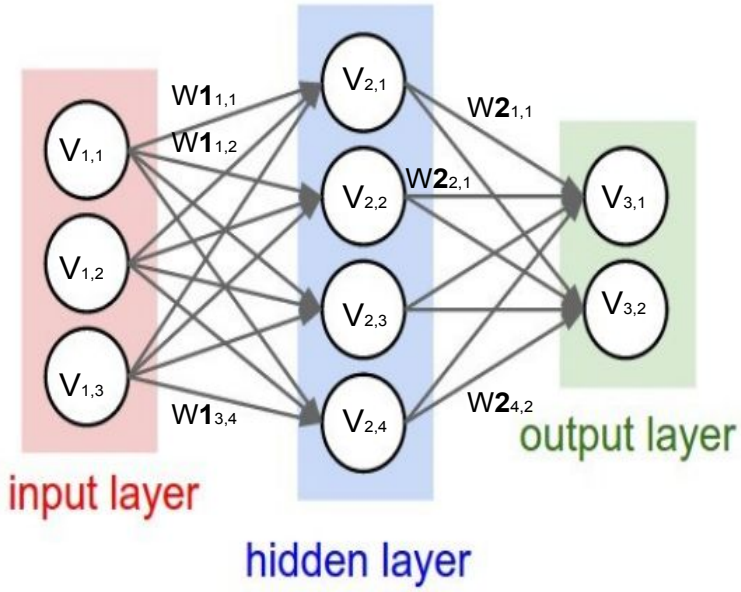


Figure 2.2: **3-Layer neural network architecture**. The input layer has 3 nodes, hidden layer has 4 nodes and output layer has 2 nodes. $V_{i,j}$ is the value of node j in layer i. $Wk_{i,j}$ is the weight associating the node i in layer k with node j in layer $k+1$.

a new task on new datasets has been shown to work well for a wide range of image datasets and tasks [29]. The main advantage is that deep learning algorithms extract the features in the data by themselves. Therefore, there is no need for human intervention during the training process. Besides, this feature extraction mechanism generates features that are hard for a human to think and implement. In this section, we explain the theoretical background of deep learning.

2.5.1 Multilayer Perceptron

Multilayer Perceptron (MLP) consists of at least three layers of nodes and weights (edges) associating nodes in consecutive layers (see Figure 2.2). The first layer is called the input layer, the middle layers are called the hidden layers, and the last layer is called the output layer. The number of nodes in the input layer and the output layer depends on the data and the problem. For example, in order to design a network architecture for handwritten digit recognition where numbers are stored in 28×28 size images, there will be 784 nodes (one input node for one pixel, $28 \times 28 = 784$) in the input layer and 10 nodes (one node for each number) in the output nodes.

The value of a node in an MLP system is obtained by a linear combination of predecessor nodes with their corresponding weights. Therefore, in order to distinguish data that is not linearly separable, non-linear activation functions are applied to node values. The most commons activations functions used are: Sigmoid, Tanh, and Rectified Linear Unit (Relu).

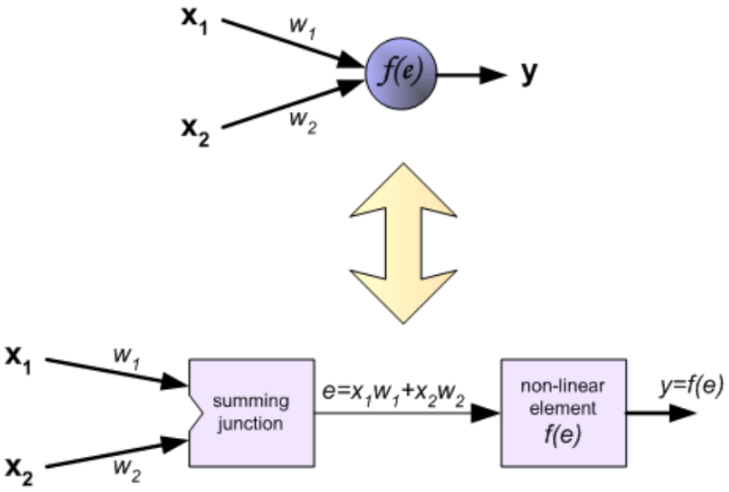


Figure 2.3: Feedforward process [27]

2.5.2 Training algorithm

When a MLP is trained, there are two main steps, the feedforward and backward (backpropagation).

- **Feedforward step:** In the feedforward step (see Figure 2.3), starting from input nodes, they are multiplied by their corresponding weights, and results are summed. The resulting summation is processed with a non-linear activation function. The output of the activation function becomes the value of that node. The process ends when the node values of the output layer are calculated.
- **Backward step:** Once the feedforward step is completed, there is an error that can be easily computed between the output layer and the real output values according to the trained data. In order to measure how good an MLP system is, we define a cost function, being the objective to minimize this cost function. The backpropagation algorithm computes the partial derivates of the cost function with respect to each weight and updates its value using the chain rule, in order to minimize the cost function (Figure 2.4).

The training algorithm can be summarized as follows:

1. Initialize the weights.
2. Apply the feedforward procedure for each sample.
3. Use backpropagation to update weights.
4. Repeat steps 2 and 3 until there is a convergence.

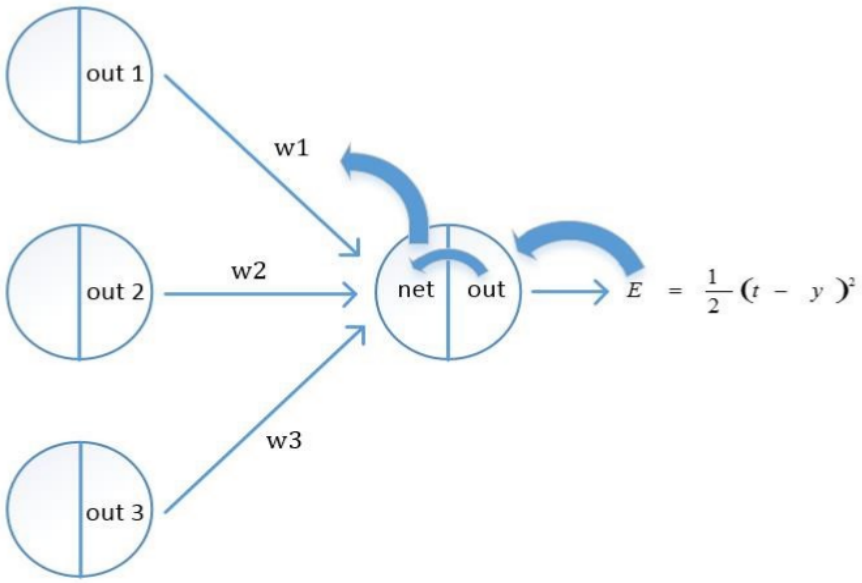


Figure 2.4: Update procedure of an output weight. Arrows show the derivative steps [27]

2.5.3 Convolutional Neural Networks

One of the problems using MLP is discarding the spatial relationship between adjacent pixels, which in images is very important. For example, in the classification problem of hand-written digits (28×28 pixels), MLP gets each pixel individually and forms a vector whose size is 784. However, by vectorizing the input, the spatial relationship between adjacent pixels is lost. The same situation happens for natural language processing, each word or syllable depends on its previous or next word or syllable.

In the Convolutional Neural Networks (CNNs), this spatial information between adjacent pixels is taken into account by a convolution step. CNNs are usually composed of convolutional layers, pooling layers, and fully connected layers. CNNs extract the specific patterns by using the filters, then pooling layers help the model ignore redundant data. By applying convolutions and pooling respectively, only certain patterns remain and as a final step, the resulting data is vectorized and MLP is used in the last step.

- **Convolutional layer:** Convolution is a mathematical operation of two functions. The main function of a convolutional layer is to extract features from the input data. In the CNNs, the convolutional operation slides a kernel function, which is also called filter, over the main data by performing an inner product among the elements. For each sliding window, the sum-up of the elements is the output. The whole output of the convolutional operation is called a feature map. In the convolution layer, many kernels are used on the original data. Each of these kernels learns different patterns and features of the input data. For example, they can learn to detect the edges, curves, blobs, and smooth areas. In Figure 2.5, a matrix of size 5×5 is convolved with a kernel of size 3×3. When considering the convolution operation in CNNs, there are three design issues to think about: kernel size, number of kernels, and stride. These hyper-parameters have an effect on the shape of the output data and memory usage.

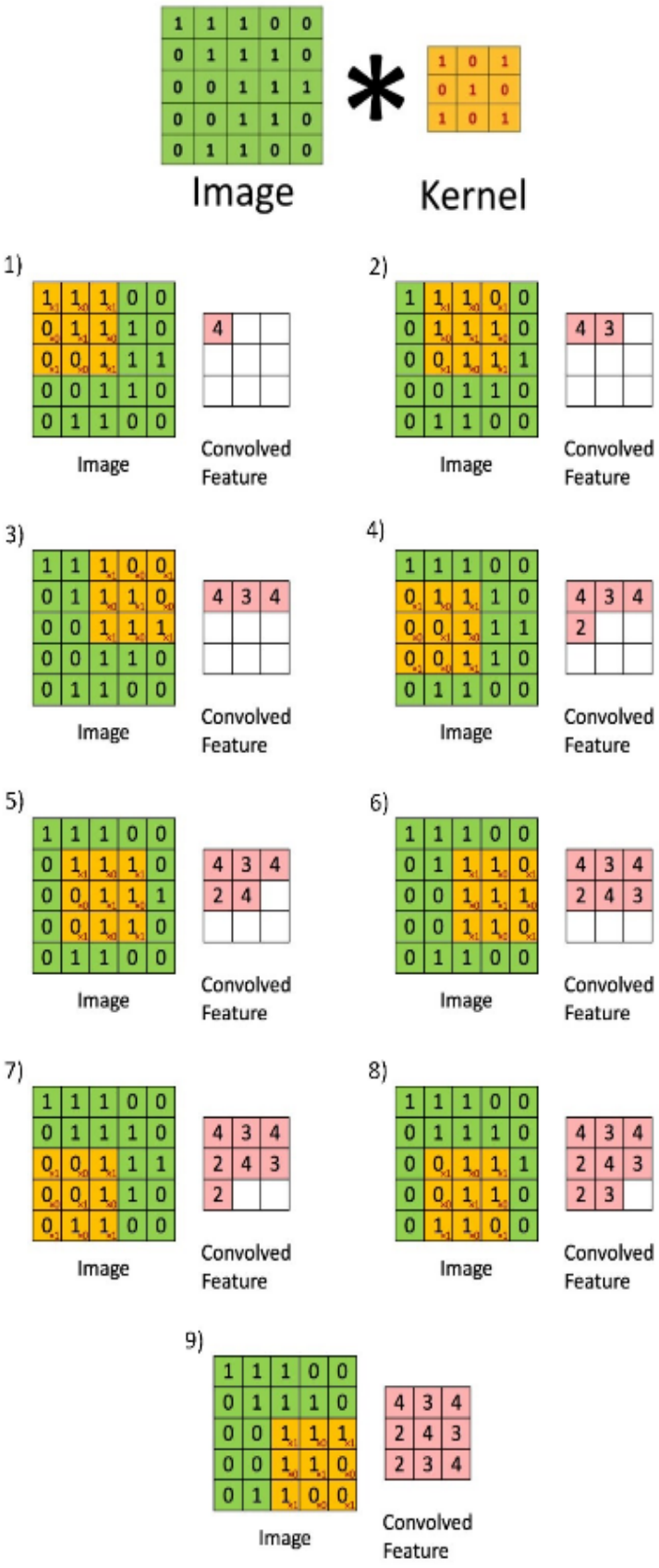


Figure 2.5: Convolution operation [27]

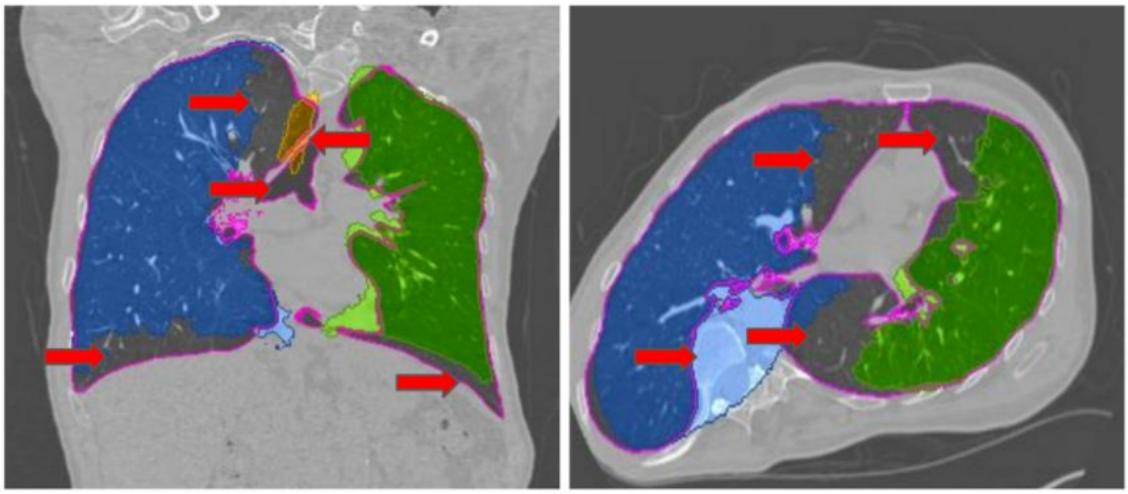


Figure 2.6: Due to the curved spine of a patient with scoliosis, image segmentation based on shape models may fail [40]. The right and left lungs, and trachea are represented by the colors blue, green, and yellow, respectively. Magenta lines describe the expected boundaries of the objects.

- **Pooling layer:** The pooling layer reduces the spatial size of the output of the convolutional layer in order to reduce the number of parameters and computation in the network, also it is used for controlling overfitting. Max pooling and average pooling are two of the most used pooling methods. For a given window, max-pooling takes the maximum value in that window and average pooling takes the average value of the values in the window. For pooling operation, there are two important hyper-parameters which are window size and stride value.

- **Fully Connected Layer:** Convolution and pooling layer generate rectangular-shaped outputs. These outputs are flattened so that they can be multiplied by their weights. For example, if there are 64 feature map layers each of which has $5 \times 5 \times 3$ voxels, in the fully connected layer these volumes are flattened to 4800×1 ($5 \times 5 \times 3 \times 64 = 4800$). The layer before the fully connected layer represents high-level features. With the help of a fully connected layer, these high-level features can be multiplied by the weights of the hidden layers. The rest of the system works as MLP does.

2.6 Automatic Lung and Trachea Image Segmentation

Automatic Lung and Trachea Image Segmentation (ALTIS) [40] is a fast method for lungs and trachea segmentation. It relies on the Image Foresting Transform (IFT) framework [12] to design a fast sequence of image processing operations based on relative-shape and intensity-based features, and image properties that are robust to account for most appearance variations of abnormal lungs, separating the trachea, left lung, and right lung. One of the main advantages of ALTIS against other methods is its robustness when lungs are deformed by disease or abnormal shape of the thoracic cage (Figure 2.6).

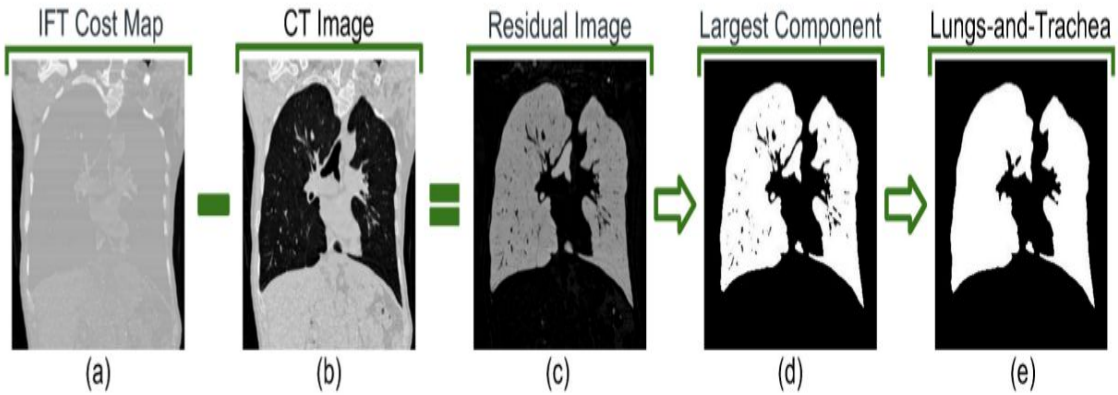


Figure 2.7: Lungs-and-trachea extraction pipeline [40]. (a) Minimum cost map after the first axial slice-by-slice IFT. (b) Original image. (c) Residual image obtained by subtracting (b) from (a). The lungs-and-trachea are enhanced. (d) Thresholding and largest component selection from (c). (e) Morphological closing from (d).

ALTIS consists of a sequence of Image Foresting Transforms (IFTs) organized in three main steps: (i) lungs-and-trachea extraction, (ii) seed estimation inside background, trachea, left lung, and right lung, and (iii) their delineation such that each object is defined by an optimum-path forest rooted at its internal seeds.

2.6.1 Lungs-and-trachea extraction

The strategy for lungs-and-trachea extraction starts by enhancing the majority of voxels inside the lungs and trachea, which appears darker than the surrounding tissues. For that, a minimum cost map slice-by-slice is computed using IFT. Then a residual image is obtained subtracting the original image from the minimum cost map image. A thresholding and largest component selection is applied, and finally a morphological closing to get the volume of interest for the lungs-and-trachea extraction. Figure 2.7 shows the lungs-and-trachea extraction pipeline.

2.6.2 Seed estimation

For the given volume of interest from lungs-and-trachea extraction, it is estimated the markers (seed set) outside the lungs-and-trachea object, markers inside each of the lungs, and markers inside the trachea for the subsequent object delineation by optimum seed competition.

The lungs and background seeds are obtained by simultaneous dilation and erosion from the volume of interest. In order to get the trachea seed, a Geodesic distance map from the lung seeds is computed, and then thresholding and highest component are applied. Figure 2.8 shows the seed estimation, seed labeling, and object delineation pipeline.

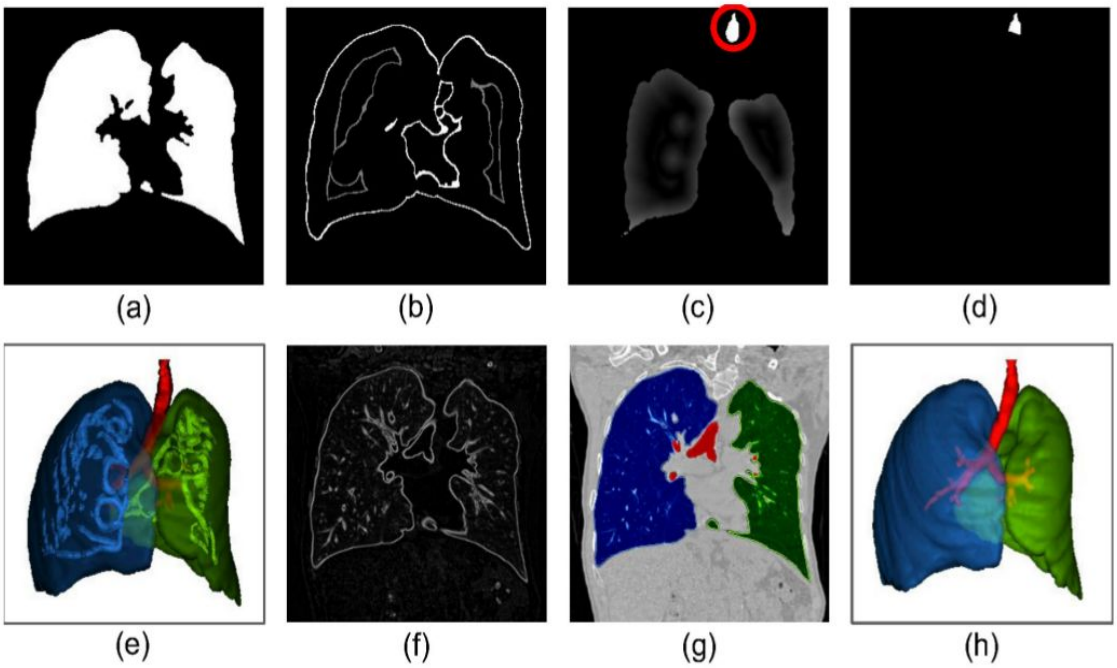


Figure 2.8: Seed estimation, seed labeling, and object delineation pipeline [40]. (a) Volume of interest obtained from the lungs-and-trachea extraction. (b) Lungs and background seeds obtained by simultaneous dilation and erosion from (a). (c) Geodesic distance map from the lung seeds. (d) Trachea seeds obtained by thresholding and highest component selection from (c). (e) Labeled seeds seen from the 3D rendition: green for the right lung seeds, blue for the left lung seeds, red for the trachea seeds, and white for the background seeds. (f) Gradient image of I 2 (Fig. 6c). (g) Resulting object delineation and (h) its 3D rendition.

2.6.3 Seed labeling and object delineation

The objects of interest are delineated by optimum seed competition. Once the lungs, trachea, and background seeds are estimated, they compete with each other to propagate the corresponding labels to their most closely connected voxels. The process is implemented by a sequence of two IFTs.

The first IFT allows propagating along optimum paths to create a label map. This first object delineation can correctly segment most parts of the lungs and trachea, but the high gradient values at voxels in narrow parts of these objects may leave those parts conquered by background seeds 2.9. The second IFT is used to solve this problem.

2.7 Non-Maximum Suppression algorithm

Object detectors in the last years have moved to the end-to-end learning paradigm: proposals, features, and the classifier becoming one neural network improving results on general object detection. The task of object detection can be interpreted as mapping an image to a set of candidate regions: one candidate per object of interest in the image and each candidate enclosing as much as possible an object. In this way, detectors should return exactly one candidate per object.

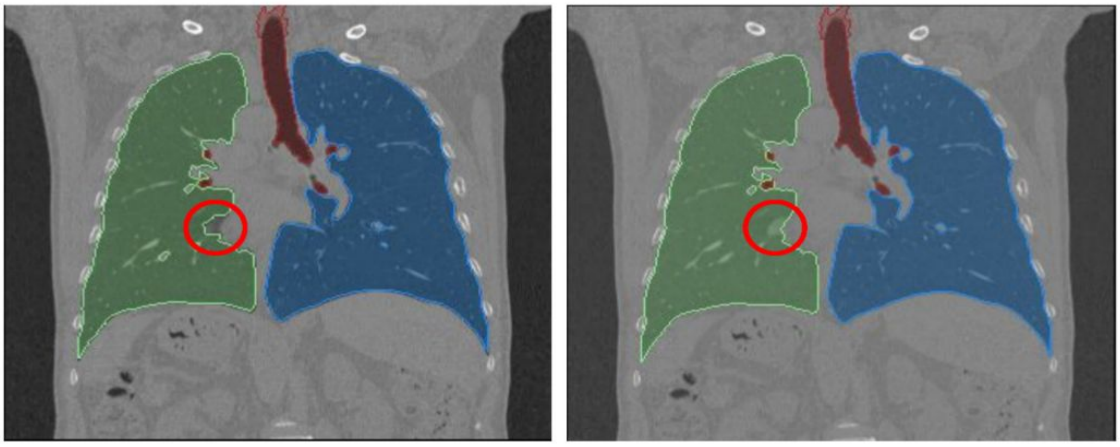


Figure 2.9: Example of a lung region left unsegmented (red circle) by the IFT with the first seed set (left). By increasing the internal seed set, the second IFT corrects segmentation (right) [40].

Uncertainty is an inherent part of the detection process causing each candidate to be associated with a confidence score. The detection process estimates the probabilities of object classes being present for every candidate in an image. Initially, these candidates are proposed from a search space (e.g. sliding window, proposals), and the estimation of class probabilities for each candidate is computed independently of any other candidate. Therefore, current detectors generate several candidate regions for the same object with different scores, being necessary a post-processing step to eliminate redundant candidates for the same object.

The Non-Maximum Suppression (NMS) algorithm has been proposed to eliminate redundant candidates and select one candidate region per object. This NMS algorithm relies on handcrafted parameters and greedy clustering with a fixed distance threshold which forces a trade-off between recall and precision.

The input for the NMS algorithm is a list of candidates C, corresponding confidence scores S and overlap threshold N and the output is a list of filtered candidates D. The following steps summarized the algorithm:

1. Select the candidate with the highest confidence score from C and remove it from C, then add it to the final candidate list D (initially D is empty).

2. Now compare this candidate with all remaining candidates from C — calculate the IOU (Intersection over Union) of this candidate with every remaining candidate from C. If the IOU is greater than the threshold N, remove that candidate from C.

3. This process is repeated until there are no more candidates left in C.

When we analyze the NMS algorithm, this presents two major difficulties: (i) the overlapping percentage for rejection must be high enough to eliminate regions with high scores containing the same object, while (ii) it must be low enough to avoid eliminating regions of detected objects that are close to each other — a common situation in images with several nodules.

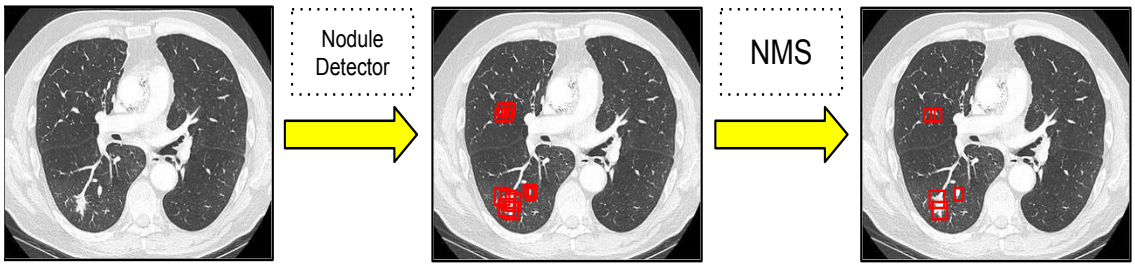


Figure 2.10: The pipeline of an automated CAD system for lung nodule detection using the NMS algorithm to eliminate redundant candidates.

In figure 2.10 we have the pipeline of an automated CAD system for lung nodule detection. We can see that the nodule detector generates several candidate regions for the same nodule with different scores, being necessary a post-processing step for the elimination of redundant candidates. In this case, the NMS algorithm was used as the post-processing step, however, we still have some redundant candidates for the same nodule.

2.8 LUNA16 challenge

The LUng Nodule Analysis 2016 (LUNA16) challenge invites participants to develop a CAD system that automatically detects pulmonary nodules in CT scans. The challenge provides the dataset and the reference annotations from the publicly available LIDC-IDRI database [3]. This dataset can be used for training of the systems and the evaluation of the algorithms is performed on the same dataset. To prevent biased results as a result of training and testing on the same dataset, participants are instructed to perform cross-validation.

2.8.1 Data

In the LUNA16 challenge, the dataset was collected from the publicly available LIDC-IDRI database [3]. The LIDC-IDRI database contains a total of 1018 CT scans. CT images come with associated XML files with annotations from four experienced radiologists. Scans with a slice thickness greater than 3 mm were excluded. On top of that, scans with inconsistent slice spacing or missing slices were also excluded. This led to the final list of 888 scans considered for the LUNA16 challenge. These scans were provided as MetaImage (.mhd) images that can be accessed and downloaded from the LUNA16 website.

Each LIDC-IDRI scan was annotated by experienced thoracic radiologists in a two-phase reading process [3]. In the first phase, four radiologists annotated the scans independently. All lesions were marked as nodule $\geq$ 3 mm; nodule $<$ 3 mm; non-nodule (any other pulmonary abnormality). For lesions annotated as nodule $\geq$ 3 mm, diameter measurements were provided. In the second phase, the anonymized blinded results of all other radiologists were revealed to each radiologist, who then independently reviewed all marks. No consensus was forced.

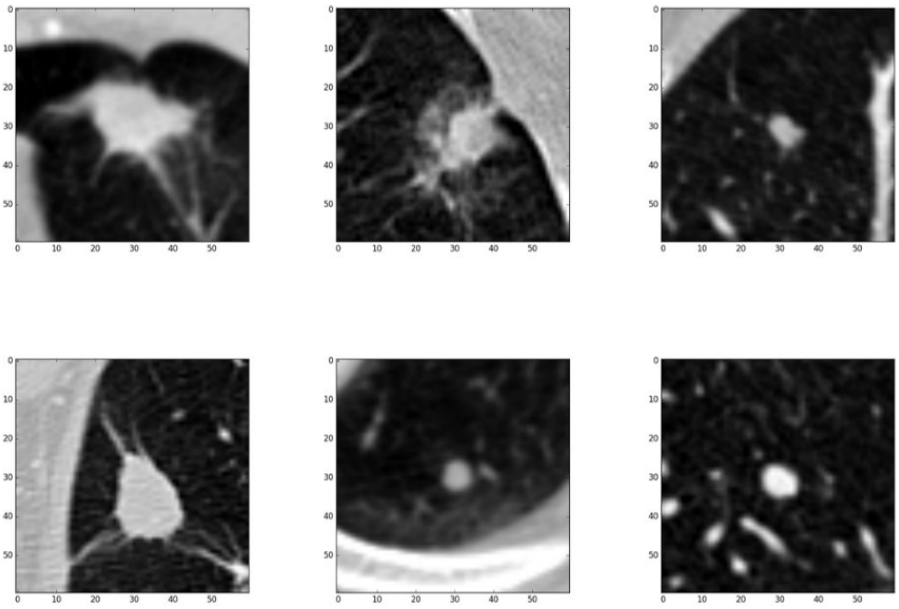


Figure 2.11: Nodule annotations from LUNA16 challenge [27].

In the 888 scans, a total of 36,378 annotations were made by the radiologists. The LUNA16 challenge only considered annotations categorized as nodules 3 mm as relevant lesions, as nodules $<$ 3 mm, and non-nodule lesions are not considered relevant for lung cancer screening protocols. This resulted in a set of 2290, 1602, 1186, and 777 nodules annotated by at least 1, 2, 3, or 4 radiologists, respectively. The LUNA16 challenge considered the 1186 nodules annotated by the majority of the radiologists (at least 3 out of 4 radiologists) as the positive examples. Figure 2.11 shows some nodules (true positives). These are the lesions that the algorithms should detect.

In Figure 2.12, three different views of a CT scan from the LUNA16 challenge dataset can be seen. The upper left view is called the axial plane, the upper-right view is called the sagittal plane, and the lower right view is called the coronal plane. When 2D images are used in the algorithms, generally, the axial plane is used.

2.8.2 Challenge tracks

The participants develop their algorithms and upload their predictions in one of two separate tracks: (1) the complete nodule detection track or (2) the false-positive reduction track.

In the complete nodule detection track, participants require to develop an entire CAD system to predict nodules locations, taking as input only the CT scans. In order to train the algorithms, the 1188 nodules locations are provided with their diameters. For the false-positive reduction track, participants require to classify a number of locations in each scan as being a nodule or not. A list of candidates are provided to the participants, in total, there are 551,066 candidate points.

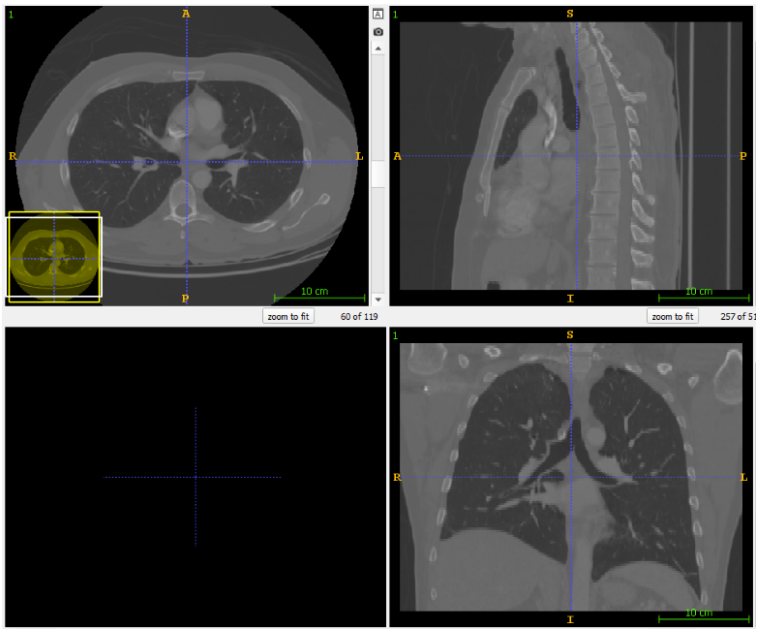


Figure 2.12: Views of a CT scan from LUNA16 challenge [27].

2.8.3 Cross-validation

Participants have to perform the 10-fold cross-validation when they use the provided dataset in LUNA16 challenge both as training and test data. The dataset has been randomly split into ten subsets of equal size on a patient level. The following steps describe how to perform the 10-fold cross-validation for the fold n ($n = 1, 2, ..., 10$):

1. Use the subset n as the test set and the remaining 9 sets as the training set.
2. Train the algorithm with the training set.
3. Test the training algorithm on the test set and generate the result file.
4. After iterating this process for all folds, merge the result files to get the result for all cases.

2.8.4 Evaluation

For the complete nodule detection track, participants are required to submit their results in the form of a comma-separated value (CSV) file. The CSV file contains all the candidates produced by the CAD system. For each CAD mark, the position (x, y, and z coordinates) and the probability to be a nodule should be provided.

A CAD mark is considered a true positive if it is located within a distance r from the center of any nodule included in the reference standard (true positives), where r is set to the radius of the reference nodule. When a nodule is detected by multiple CAD marks, the CAD mark with the highest score is selected. CAD marks that detect irrelevant findings are discarded from the analysis and are not considered as either false positive or true positive. CAD marks not falling into previous categories are marked as false positives.

Results are evaluated using the Free-Response Receiver Operating Characteristic (FROC) analysis (International Commission on Radiation Units and Measurements, 2008). The sensitivity is defined as the fraction of detected true positives (TPs) divided by the number of nodules in the reference standard. In the FROC curve, sensitivity is plotted as a function of the average number of false positives per scan (FPs/scan). For each scan, it is taken a maximum of 100 CAD marks that were given the highest scores. The 95% confidence interval of the FROC curve is computed using bootstrapping with 1000 bootstraps, as detailed in Efron and Tibshirani (1994). In order to evaluate and compare different systems easily, it is defined one overall output score. The overall score is defined as the average of the sensitivity at seven predefined false positive rates: 1/8, 1/4, 1/2, 1, 2, 4, and 8 FPs per scan. The performance metric was introduced in the ANODE09 challenge and is referred to as the Competition Performance Metric (CPM) in Niemeijer et al. (2011) .

The evaluation script is publicly available on the LUNA16 website and can thus be viewed and used by all participants.

Chapter 3

Related works

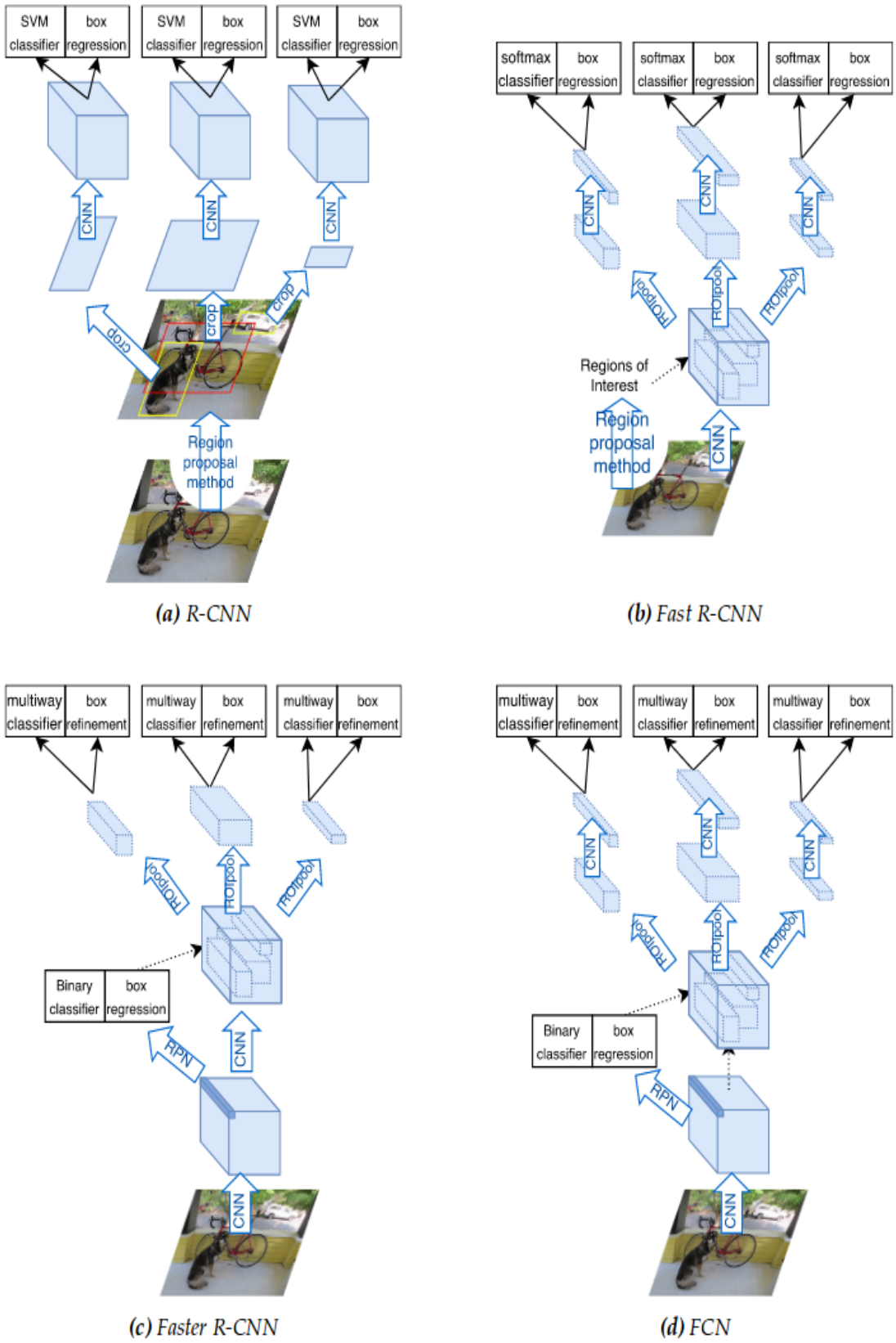
3.1 Detection framework

One of the first attempts for the detection problem using a Convolutional Neural Network (CNN) was made in 2014 by LeCun et al., proposing the OverFeat [33] — a network with a small receptive field applied to the input image at different scales and positions in a sliding window fashion. In the same year, a region-based CNN (R-CNN) [15] (see Figure 3.1a) was published, which consists of four modules: (i) find regions of interest, called region proposal, (ii) feature extraction using a CNN for each proposal, (iii) the feature vector is fed into support vector machine (SVM) for classification and (iv) linear regression for the localization.

Later, due to the high amount of computation required by R-CNN, the Fast R-CNN [14] fixed this issue. The main improvement in the Fast R-CNN comes from the feature extraction directly performed on the entire image using a CNN, the generated feature map is then cropped based on the proposed regions. Then a region of interest pooling layer (RoIpool) is applied with the assumption that the feature map is still embedding some spatial information. The sub-patches of the feature map go into the final part of the network, which performs the classification and regression of the bounding box (see Figure 3.1b). However, the Fast R-CNN model still relies on an external region proposal algorithm which is the bottleneck of the computations.

The first end-to-end deep learning detection algorithm is the Faster R-CNN [31]. Shaoqing et al. introduced a Region Proposal Network (RPN), generating region proposals with a CNN. The main improvement aiming to reduce computations is sharing parameters between RPN and the detection network (Fast R-CNN) by using the same first convolutional layers (see Figure 3.1c). Also, they use a key concept used later by all detection frameworks called anchors [13].

Region-based Fully Convolutional Network (R-FCN) [9] is a modified version of the Faster R-CNN. Compared with Faster R-CNN, the feature map is cropped later on, on the very last convolution before the classification and regression of the bounding box (see Figure 3.1d). Other systems remove the need for RPN as they perform the object classification and class-specific bounding box regression directly (see Single Shot MultiBox Detector (SSD) [24] and You Only Look Once (YOLO) [30]). At the end of each of these algorithms, a post-processing operation is conducted, the NMS algorithm [26].



(a) R-CNN *(b) Fast R-CNN*

(c) Faster R-CNN *(d) FCN*

Figure 3.1: Pipeline for different detection frameworks [5]

3.2 Lung nodule detection and NMS algorithms

As it is often the case with medical images, deep learning methods have been limited by the relatively small number of annotated images and several researchers have used different datasets being difficult to compare techniques with the same dataset. However, the LUNA16 challenge [35] has had a big influence on deep learning methods proposed for lung nodule detection. Relevant work has been published in the context of the LUNA16 challenge.

The common pipeline in the detection consists of nodule detection and false positive reduction. The nodule detection works on the images and outputs often a high number of candidate regions, producing high sensitivity but at the same time high false-positive rate. This motivates to use the false-positive reduction stage.

In 2017, due to the successful results obtained by the Faster R-CNN in the general object detection, Ding *et al.* [10] proposed a 2D Faster R-CNN [31] (RPN + Fast R-CNN) with a feature extractor based on VGG-16 [38] for lung nodule detection, and a 3D CNN with 6 convolutional layers and 3 fully connected layers for false-positive reduction stage. The principal problem of original 2D Faster R-CNN lies in the fact that lung nodules are much smaller than natural objects, thus the original 2D Faster R-CNN, which utilizes five group convolutional layers of VGG-16 for feature extraction, cannot explicitly capture the features of nodules and result in a limited performance for lung nodule detection. To address this problem a deconvolutional layer was added, after the last layer of the original feature extractor. They use 6 anchors of sizes ranging from 4 mm to 32 mm leading to high computational cost. As the network only sees three consecutive slices at a time, screening a full CT scan is expensive and time-consuming.

Dou *et al.* [11] propose a method that employs a 3D fully convolutional network (FCN) with online sample filtering to detect candidate nodules followed by a residual convolutional network for false positive reduction. The 3D FCN can not only leverage rich volumetric spatial information to extract high-level features for accurate nodule detection but also rapidly produce the probability prediction in a volume-to-volume manner. The 3D FCN consists of 5 convolutional layers and 1 max-pooling layer. The model is trained with small 3D patches of nodules and non-nodules, and tested on the entire CT scan in a fully convolutional manner. A residual convolutional network with Hybrid-Loss 3D Residual Learning is used for false positive reduction. However, in order to reduce the workload of the residual network, a GreedyNMS algorithm is applied over the candidates in order to obtain one detected region per object.

In 2018, Qin *et al.* [28] explore the use of a full 3D CNN model that employs 3D U-Net architecture as the backbone of a region proposal network (RPN) to generate nodule candidates. A multi-task residual learning and online hard negative example mining strategy are adopted to accelerate the training process and improve the accuracy of nodule detection. Then, a 3D DenseNet-based model is presented to reduce false-positive nodules. Similarly, Tang *et al.* [42] propose the use of a U-Net-inspired 3D Faster R-CNN trained using online hard negative mining for candidate extraction and a 3D CNN classifier for false-positive reduction.

Antoine *et al.* [5] use an only stage for lung nodule detection based on the RPN of the

Faster R-CNN [31] framework with RestNet for feature extraction. The network uses 3D convolutions and requires a volume as input. The final network relies on fewer convolutions and fewer trainable parameters compared to previous methods [10, 11, 28, 42]. As a post-processing filtering operation a GreedyNMS algorithm is applied after the candidate selection to obtain one candidate region per object.

Wentao *et al.* [46] introduced a fully automated lung computed tomography cancer diagnosis system called DeepLung. Deeplung consists of two components: nodule detection (identifying the locations of candidate nodules) and classification (classifying candidate nodules into benign or malignant). A 3D Faster R-CNN with 3D dual path blocks is designed for nodule detection and a U-net-like encoder-decoder scheme is employed to effectively learn the features. For nodule classification, a gradient boosting machine (GBM) with 3D dual path network features is proposed. Since the detector produces several redundant candidates for the same nodule, a GreedyNMS algorithm was applied to obtain one candidate region per nodule.

The majority of the previous lung detectors generate several candidate regions for the same nodule even by considering a false-positive reduction stage. Therefore, a post-processing step is required for the elimination of redundant candidates. The GreedyNMS algorithm was employed to address this problem. However, the GreedyNMS present difficulties to eliminate effectively the redundant candidates, so several works have been proposed in the last years.

In 2014, Rothe, Guillaumin, and Gool [32] propose the use of Affinity Propagation Clustering for NMS by treating the object score as a similarity factor to map the detected regions into a similarity space. Thereby, Affinity Propagation is applied to select the most likely candidate to contain the object.

In 2017, to avoid the greedy factor of traditional NMS, Bodla *et al.* [4] propose Soft-NMS — a decay function to decrease the object score of detected regions instead of rejecting the regions. However, this method generates false positives when a new object score of a false positive region is above the score of a real object detection.

Hosang, Benenson, and Schiele [16] propose a 1D NN to learn NMS, namely GossipNet. This network analyzes all pairs of detected regions with some degree of overlapping. Its output is a new object score for each detected region and, ideally, each nodule should be represented by a single region with the highest object score among the overlapping ones. Different from other NMS approaches, GossipNet can learn the NMS parameters rather than relying on handcrafted ones. However, GossipNet does not take advantage of image features to learn NMS.

Chapter 4

Proposed method

The proposed method for lung nodule detection consists of four modules: pre-processing, the definition of a region of interest (e.g., by lung segmentation), nodule detection, and the elimination of redundant candidates (see Figure 4.1).

For lung segmentation, mostly methods [10, 11, 46] have used the provided segmentation in the LUNA16 challenge which is based on [44]. However, we use a recent approach based on sequences of Image Foresting Transforms (IFTs) named ALTIS [40] providing a more accurate segmentation of the lungs which has a positive impact on the results. For nodule detection and the elimination of redundant candidates, inspired by recent and promising works [16] and [17], we use 3D Faster R-CNN with ResNet18 for the detection of candidate regions with nodules and propose a transformation over the voxel intensities of those candidates in the CT image, such that the resulting image features are presented as an additional input to GossipNet, rather than using the all-zero input vectors of the original GossipNet. The new network for NMS with this modified GossipNet is named *FeatureNMS* (Figure 4.3). Note that our proposed FeatureNMS aims to reduce the number of redundant candidates without losing true positives. For validation, we use the LUNA16 challenge dataset.

In the following sections are described each module composing the entire proposed method.

4.1 Pre-processing

The pre-processing is performed before the data is processed. The main objective of pre-processing is to remove noise and irrelevant information during the data acquisition. CT scans come in different scales and resolutions. Therefore, we resample the CT scans to an isotropic resolution of 1 mm between the center of two consecutive voxels in the axial, coronal and sagittal views. We use interpolation to determine the final voxel values.

The pixels' values of a CT scan in the LUNA16 challenge are given in Hounsfield units (HU). Values in HU lower than -1000 do not have any semantic meaning and they are used for padding. On the other hand, values in HU higher than 400 do not bring any information for lung nodule detection — they represent bones or foreign bodies like pacemakers. Therefore, we have clipped the pixels' values between -1000 and 400 HU.

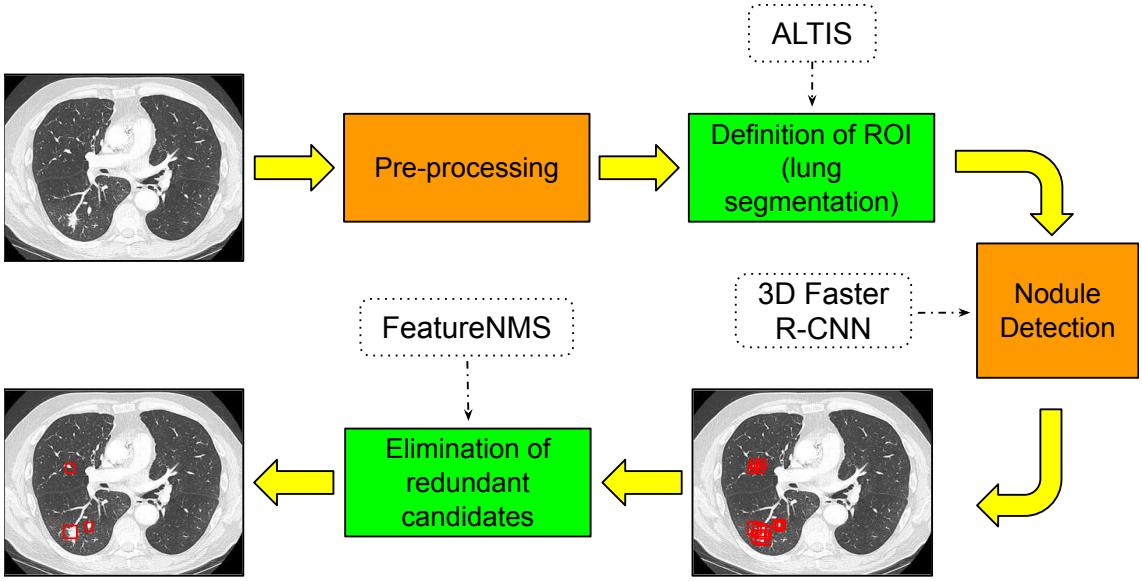


Figure 4.1: The pipeline of the proposed method. Each module is represented by a color rectangle (orange or green), while white rectangles are the techniques used for each module. In the output CT image, red rectangles are the detected nodules.

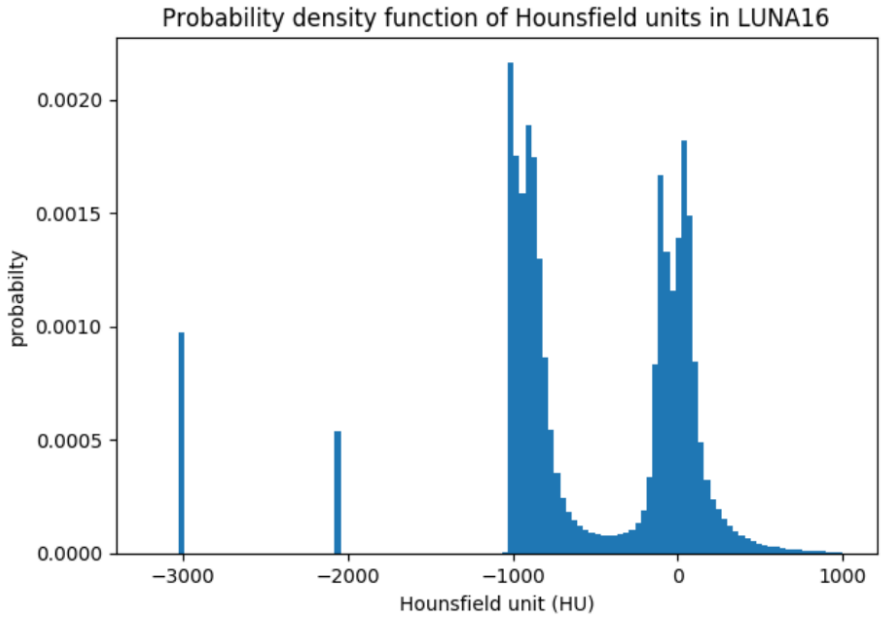


Figure 4.2: HU values distribution on LUNA16 [5].

Figure 4.2 represents the HU distribution in the LUNA16 challenge dataset. As a final step, pixels' values are standardized such that they have a zero mean and unit variance.

4.2 Definition of a region of interest

We use the segmentation of the lungs as a crucial step since the nodules are only present inside the lungs. Mostly methods [10, 11, 46] have used the provided segmentation in the LUNA16 challenge which is based on [44] and other methods [34] have not even considered

any lung segmentation module, producing candidate regions outside the lungs, which only increase the number of false-positive candidates and the computational cost.

In the present work, the segmentation of lungs refers to the process of partitioning the pre-processed CT image into multiple regions to separate the pixels belonging to lung tissue from the surrounding anatomy. Differently from other methods, we use the ALTIS [40] algorithm for lung segmentation providing a more accurate segmentation of the lungs which has a positive impact on the results.

The ALTIS algorithm segments a CT image into three regions: the left lung, the right lung, and the trachea. For the purpose of our method, only the left lung and the right lung are considered.

4.3 Nodule detection

Inspired by the successful results obtained by the Faster R-CNN in object detection, we use 3D Faster R-CNN with ResNet18 for the detection of candidate regions with nodules. The Faster R-CNN consists of two modules. The first module is called Region Proposal Network (RPN) which is a fully convolutional network that proposes candidate regions, and the second module is the Fast R-CNN detector that uses those candidate regions. The complete system is a single and unified network for object detection. The Region Proposal Network tells to the Fast R-CNN where to look.

The Region Proposal Network (RPN) takes as input an image and outputs a set of rectangular object proposals, each of them with an object score. The RPN and the Fast R-CNN share common convolutional layers for feature extraction, in our case, we use the first 17 layers from ResNet18. The RPN slides a small network (sliding window) over the feature map of the last shared convolution layer and it predicts several bounding boxes (k) at each sliding window. These k reference boxes are actually the anchors [31] which are centered at the sliding window and have different scales and aspect ratio. Finally, the features for each anchor is fed into two sibling fully-connected layers: (i) a box-regression layer and an (ii) box-classification layer. The multi-scale and aspect ratios issues in object detection are addressed using multiple anchors.

The Fast R-CNN takes as input the candidate regions proposed by the RPN (with an object score), deciding whether a region is a nodule or not. The same convolutional layers used for feature extraction in the RPN are also used in the Fast R-CNN. In the Fast R-CNN a Region of Interest (RoI) Pooling Layer is slide to map each candidate region to a small feature map with a fixed spatial size of $W \times H$. The RoI Pooling Layer works dividing the candidate region in $W \times H$ grid of sub-windows and then max-pooling the values in each sub-window. After RoI Pooling Layer is applied, a fully connected network is added, which is composed of two 4096-way fully connected layers. Finally, a classifier and a regression produce the probability for the candidate region to be a nodule or not, and the bounding box respectively.

4.4 Elimination of redundant candidates

For the elimination of redundant candidates, after 3D Faster R-CNN with ResNet18 produces the candidate regions with nodules, we propose a transformation over the voxel intensities of those candidates in the CT image, such that the resulting image features are presented as an additional input to GossipNet, rather than using the all-zero input vectors of the original GossipNet. The new network for NMS with this modified GossipNet is named *FeatureNMS* (Figure 4.3). Note that our proposed FeatureNMS aims to reduce the number of redundant candidates without losing true positives.

Additionally to the all-zero input vector, GossipNet receives geometric features from each pair of detected regions with some degree of overlapping — in our case, pairs of regions with more than 15% of Intersection over Union (IoU).

FeatureNMS essentially adapts GossipNet to receive image features from the detected regions. This modification aims not only to improve nodule detection but also to speed up the overall convergence process.

4.4.1 FeatureNMS Network

Architecture

FeatureNMS adds a key ingredient to the architecture of GossipNet: a 3D RoI pooling layer that extracts voxel intensities from each detected region in the CT image and transforms it into suitable image features. Since the diameter of the nodules can largely vary, we use a 3D RoI pooling on each region with output size of $4 \times 4 \times 4$ voxels for standardization, flatten the ROI into a feature vector, and increase its size by a Fully Connected (FC) layer into 128 dimensions — the same size of the all-zero input vector of GossipNet.

When analyzing geometric information from all pairs of overlapping candidate regions with more than 15% of IoU, GossipNet processes all of them in the same mini-batch through three FC layers before feeding the result into each of its blocks. In the same spirit, the image features from the corresponding candidate regions are also processed in the same mini-batch and provided as input to the first block of GossipNet (see Figure 4.3).

The geometric features of each pair of detected regions under consideration are twelve: (1) the intersection over union (IoU); (2-4) their normalized distances along x, y, and z; (5-7) the normalized L2 distances between their centers in the xy, xz, and yz planes; (8) the normalized L2 distance between their centers; (9-10) the normalized ratio between their diameters; and (11-12) their detection scores by the 3D Faster R-CNN.

Loss function

Lung nodule detectors can be judged based on a benchmark evaluation criterion, which defines a matching strategy to decide whether or not a nodule has been correctly detected.

The matching strategy in the LUNA16 Challenge [35] considers as true positive the region with the highest score among the candidate ones whose location is within a distance R from the center of the real nodule, whereas all other overlapping candidates for that nodule are ignored. In this work, $R = 0.5d$, where d is the diameter of the real nodule.

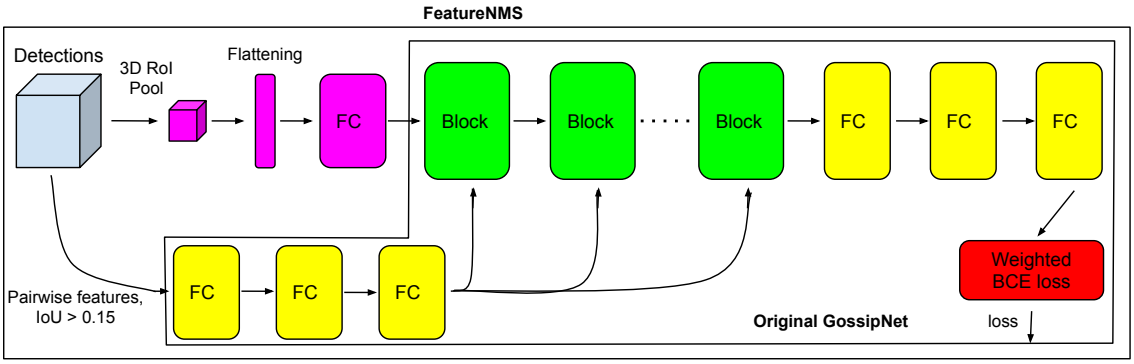


Figure 4.3: Architecture for the FeatureNMS Network.

Nodule candidates falling outside the distance R are considered false positives decreasing the precision of the detector. We use this matching strategy to define labels at training time for our FeatureNMS network.

After candidate region detection by the 3D Faster R-CNN, successfully matched candidates receive positive labels, while unsuccessfully matched candidates are assigned to negative labels for training the FeatureNMS network. Due to the severe class imbalance, we adopt the weighted binary cross-entropy loss function during the training phase.

Chapter 5

Experiments and results

5.1 Dataset and evaluation metric

For validation, we use the LUNA16 challenge dataset [35] which contains 888 CT scans with annotations from the publicly available LIDC-IDRI dataset [3] for pulmonary nodules. We applied the 10-fold cross validation, as described in the LUNA16 challenge.

In the LUNA16 challenge [35], the performance for nodule detection is evaluated by the Free-Response Receiver Operating Characteristic (FROC) [22] analysis, which quantifies a trade-off between sensitivity and specificity. We use the same evaluation metric, by measuring the detection sensitivity and false positives per scan (FPs/scan). The Final Evaluation Metric (FEM) is the average of sensitivity at seven different FPs/scan rates: 0.125, 0.25, 0.5, 1, 2, 4, 8.

5.2 ALTIS algorithm

For the definition of a region of interest by lung segmentation, we use the ALTIS algorithm which assumes that the patient orientation in the CT scan is from inferior to superior along the axial slices (z-axis), from right to left along the sagittal slices (x-axis), and from anterior to posterior along the coronal slices (y-axis). Therefore, in a coronal slice, the lungs and trachea appear in the upright position, being the right lung on the left side of the slice.

We linearly interpolated the CT scans from the LUNA16 challenge dataset to the same voxel size, $1.25 \times 1.25 \times 1.25 mm^3$ as it is described in [40]. Then, we apply the ALTIS algorithm to obtain the lung segmentation taking only the left lung and right lung. Finally, we interpolate again the CT scans to an isotropic resolution of $1 \times 1 \times 1 mm^3$ in order to train and test the nodule detector and the elimination of redundant candidates.

5.3 Training

To train the 3D Faster R-CNN responsible for nodule candidate selection, we use the 10-fold cross validation of LUNA16, as shown in [35]. The network is trained with the stochastic gradient descent (SGD) optimizer for 100 epochs, with momentum of 0.9, weight

decay of 0.0001, and initial learning rate of 0.01, which is decreased after the 50^{th} and 80^{th} epochs.

For a given left-out fold, once the Faster R-CNN training is complete, the corresponding model is used to generate nodule candidates for the entire dataset. These candidates are used to train, validate, and test the GossipNet and FeatureNMS, maintaining the same training-validation split used in the Faster R-CNN. Both training processes (Faster R-CNN and GossipNet/FeatureNMS) are performed in sequence for each fold, but both the GossipNet and FeatureNMS are trained on samples preprocessed by the Faster R-CNN. The number of candidates produced is approximately 70k for training, 10k for validation, and another 10k for testing.

The optimizer used for GossipNet and FeatureNMS was ADAM with 200k iterations, starting with a learning rate of 0.001 and weight decay of 0.1 after 10k and 80k iterations.

5.4 Baseline

For the lung segmentation, we compare the ALTIS algorithm with the provided segmentation in the LUNA16 challenge which is based on [44]. Furthermore, we report results without using any lung segmentation in order to show the importance of this module for lung nodule detection.

Since the 3D Faster R-CNN is common to all methods, we compare FeatureNMS with the original GossipNet [16] and a GreedyNMS algorithm (used in [46]) using several overlapping thresholds.

5.5 Results and discussion

In order to perform the comparison among the different techniques for each module of our proposed method, we experiment for three scenarios: (i) using the lung segmentation provided in the LUNA16 challenge which is based on [44], (ii) using the ALTIS algorithm for lung segmentation and (iii) without using any lung segmentation at all. For these three scenarios, we employ 3D Faster R-CNN with RestNet18 for candidate region detection and compare our proposed FeatureNMS with the original GossipNet [16] and a GreedyNMS algorithm (used in [46]) using several overlapping thresholds.

5.5.1 Using the lung segmentation from LUNA16 challenge

For all these experiments, the lung segmentation provided in the LUNA16 challenge is used. Table 5.1 shows the FEM results for GreedyNMS at various overlapping thresholds. Starting with a threshold of $\theta = 0$, the nodule candidate with the highest score suppresses all touching neighbors, including real nearby nodules which reduces the number of true positives detected by the method (low sensitivity). As θ increases, the numbers of both true positives and false positives increase as well. This shows a trade-off between sensitivity and specificity based on the selected threshold. The best threshold for GreedyNMS occurs when $\theta = 0.20$, which results in 83.12% of FEM on the LUNA16 challenge dataset.

Table 5.1: FEM results for GreedyNMS at several overlapping thresholds (θ). In this case, the lung segmentation provided in the LUNA16 Challenge is used.

θ	> 0.0	> 0.15	> 0.2	> 0.25	> 0.3
FEM	69.54%	82.18%	83.12%	81.45%	80.24%

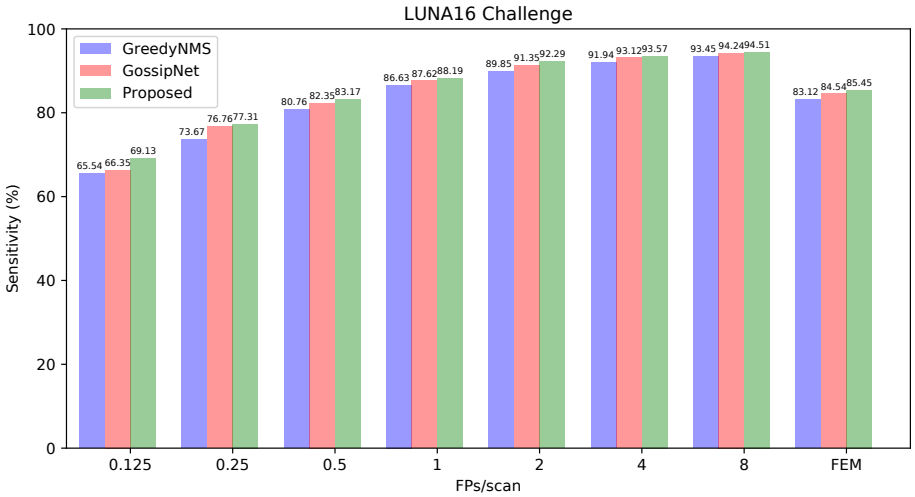


Figure 5.1: Performance comparison among GreedyNMS, GossipNet, and FeatureNMS (proposed). FeatureNMS achieves superior results as compared to the other methods. In all cases, the lung segmentation provided in the LUNA16 challenge is used.

The performance comparison among the best GreedyNMS, GossipNet, and our FeatureNMS is summarized in Figure 5.1. Notice that GossipNet already improves nodule detection by 1.42%, when compared to the best GreedyNMS. On its turn, FeatureNMS achieves an improvement of 2.33% as compared to the best GreedyNMS, and a further boost of 0.91%, when compared to GossipNet. These results indicate the effectiveness of feeding image features into the first block of GossipNet and that FeatureNMS is a competitive method against state-of-the-art approaches. Notice that no false-positive reduction was applied to the detected nodules.

Figure 5.2 shows the performance of FeatureNMS when compared to the traditional GreedyNMS and GossipNet on a randomly selected CT image from LUNA16. Note that the FeatureNMS suppresses more redundant regions per nodule than the GreedyNMS and GossipNet.

5.5.2 Using the lung segmentation performed by the ALTIS algorithm

For all these experiments, we use the ALTIS algorithm for lung segmentation. Table 5.2 shows the FEM results for GreedyNMS at various overlapping thresholds. The best threshold for GreedyNMS occurs when $\theta = 0.20$, which results in 84.08% of FEM on the LUNA16 challenge dataset.

The figure 5.3 summarize the performance comparison among the best GreedyNMS,

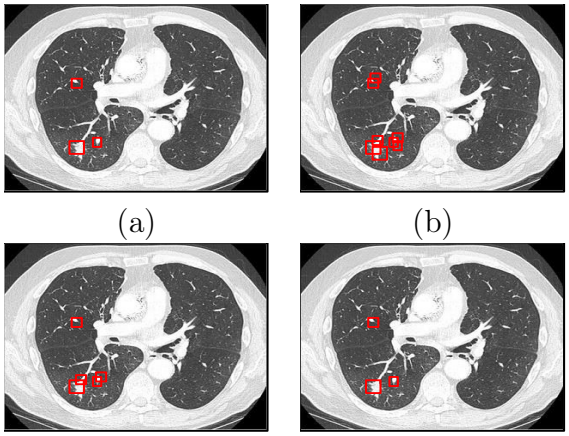


Figure 5.2: A qualitative comparison among the methods. (a) Real nodules. Results of (b) GreedyNMS, (c) GossipNet, and (d) FeatureNMS. Note that FeatureNMS can eliminate all redundant regions, differently from the others. In all cases, the lung segmentation provided in the LUNA16 challenge is used.

Table 5.2: FEM results for GreedyNMS at several overlapping thresholds (θ). In this case, the ALTIS algorithm is used for lung segmentation.

θ	> 0.0	> 0.15	> 0.2	> 0.25	> 0.3
FEM	70.23%	83.12%	84.08%	82.97%	81.15%

GossipNet, and our FeatureNMS using the ALTIS algorithm for lung segmentation. The results show that GossipNet already improves nodule detection by 1.44%, when it is compared to the best GreedyNMS. On the other hand, our FeatureNMS achieves an improvement of 2.40% as compared to the best GreedyNMS, and a further boost of 0.96%, when it is compared to GossipNet. These results confirm again that FeatureNMS is a competitive method against state-of-the-art approaches. Notice that no false-positive reduction was applied to the detected nodules.

The performance comparison among methods using the ALTIS algorithm for lung segmentation and the lung segmentation provided in the LUNA16 challenge is summarized in the table 5.3. The GreedyNMS improves nodule detection by 0.96% using the ALTIS algorithm for lung segmentation rather than the lung segmentation provided in the LUNA16 challenge. On their turn, GossipNet and FeatureNMS improve nodule detection by 0.98% and 1.03% respectively by using the ALTIS algorithm for lung segmentation. These results indicate the effectiveness of using the ALTIS algorithm for lung segmentation rather than the lung segmentation provided in the LUNA16 challenge.

Table 5.3: FEM results among methods using the ALTIS algorithm for lung segmentation, the lung segmentation provided in the LUNA16 challenge, and without using any lung segmentation.

Lung segmentation	From LUNA16	From ALTIS	Without any segmentation
GreedyNMS	83.12%	84.08%	67.56%
GossipNet	84.54%	85.52%	67.72%
FeatureNMS	85.45%	86.48%	68.02%

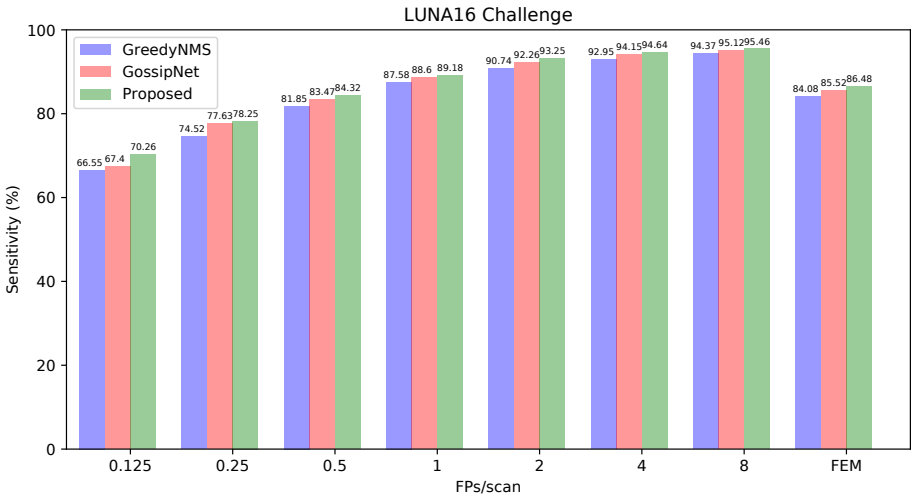


Figure 5.3: Performance comparison among GreedyNMS, GossipNet, and FeatureNMS (proposed). FeatureNMS achieves superior results as compared to the other methods. In all cases, the ALTIS algorithm is used for lung segmentation.

5.5.3 Without using any lung segmentation at all

For the proposed method, the definition of Region of Interest through lung segmentation is a crucial module since it constraint the searching inside the lungs avoiding candidate regions with nodules outside of them. In order to show the crucial role that plays the lung segmentation for nodule detection, we experiment the proposed method without using any lung segmentation, that is, processing the entire CT image for lung nodule detection.

The performance comparison among methods whether lung segmentation is used or not is summarized in the table 5.3. The GreedyNMS without using any lung segmentation decreases nodule detection by 16.52% and 15.56% when it is compared to such a method using the ALTIS algorithm for lung segmentation and the lung segmentation provided in the LUNA16 challenge respectively. On their turn, GossipNet decreases nodule detection by 17.80% and 16.82%, and FeatureNMS decreases nodule detection by 18.46% and 17.43%, when they are compared to such methods using the ALTIS algorithm for lung segmentation and the lung segmentation provided in the LUNA16 challenge respectively. These results indicate the crucial role that plays the lung segmentation for lung nodule detection.

Chapter 6

Conclusions and future work

We have presented a method for lung nodule detection consisting of four modules: preprocessing, the definition of a region of interest (e.g., by lung segmentation), nodule detection, and the elimination of redundant candidates. For lung segmentation, we use a recent approach based on sequences of Image Foresting Transforms (IFTs) named ALTIS providing a more accurate segmentation of the lungs which has a positive impact on the results. For nodule detection and the elimination of redundant candidates, we rely on a 3D Faster R-CNN with ResNet18 for the detection of candidate regions followed by the proposed FeatureNMS for NMS. FeatureNMS essentially adapts GossipNet to receive image features from the detected regions. It uses an RoI pooling layer, followed by flattening, and one FC layer to feed image features to GossipNet and obtain a more suitable object score per detected region.

A CAD system for lung nodule detection without using any lung segmentation decrease considerably nodule detection by 16.52%, 17.80%, and 18.46% for the methods GreedyNMS, GossipNet, and FeatureNMS respectively, when they are compared to such methods using the ALTIS algorithm for lung segmentation. These results show the crucial role that plays the lung segmentation for lung nodule detection.

The effectiveness of the ALTIS algorithm has been shown when it is tested in different samples of the LUNA16 dataset. The experiments indicate that when the ALTIS algorithm is used for lung segmentation, an improvement of 0.96%, 0.98%, and 1.03% is reached for the methods GreedyNMS, GossipNet, and FeatureNMS respectively, when they are compared to such methods using the lung segmentation provided in the LUNA16 challenge.

FeatureNMS has proven to be very effective for lung nodule detection when trained and tested in different samples of the LUNA16 dataset. The experiments indicate that FeatureNMS can perform slightly better (0.96%) than the original GossipNet using the ALTIS algorithm for lung segmentation, and an improvement of 0.91% than the original GossipNet using the lung segmentation provided in the LUNA16 challenge. Furthermore, FeatureNMS outperforms the GreedyNMS by 2.40% and 2.33% using the ALTIS algorithm and the adopted LUNA16 method for lung segmentation respectively. Note that we have not used any false-positive reduction step which might further improve our results.

6.1 Future work

In our proposed method we have considered a 3D Faster R-CNN with RestNet18 for the detection of candidate regions with nodules. However, it can be explored other alternatives for the nodule detector and the pre-trained model. For the nodule detector, there are other alternatives such as the Single Shot MultiBox Detector (SSD) [24], You Only Look Once (YOLO) [30], Region-based Fully Convolutional Network (R-FCN) [9] and some recent approaches based only on RPN [5]. Similarly, for the pre-trained model, it can be considered the VGGNet [38], DenseNet [19], or SENet[18]. On top of these alternatives, our proposed FeatureNMS can be used for the elimination of redundant candidates and check its effectiveness for nodule detection.

For the elimination of redundant candidates, we proposed FeatureNMS — a 1D neural network that learns the parameters for NMS instead of relying on handcrafted parameters. It proposes a transformation over the voxel intensities on the candidate regions with nodules in the CT image. However, we did not explore the possibility to consider for these candidate regions the image features from RestNet18 rather than the voxel intensities. We believe that it cannot only lead to better results for the elimination of redundant candidates since FeatureNMS would learn more complex patterns to learn the parameters for NMS, but also to faster convergence. Additionally, it can be considered other image features for the candidate regions rather than the voxel intensities.

As we have mentioned in the present work, we have not used any false-positive reduction stage in our proposed method for lung nodule detection, which might further improve our results. As it was shown in several works [10, 11, 28], adding a false-positive reduction stage can have a positive impact on nodule detection. After the nodule detector generates the candidate regions with nodules, the false-positive reduction stage can be used. Also, it can be explored to use an end-to-end framework for nodule detection, integrating nodule candidate screening, and false-positive reduction into one model, trained jointly as it is presented in [43]. At the end or after generating the candidate regions with nodules, our proposed FeatureNMS can be incorporated for the elimination of redundant candidates.

Most of the current detectors process each candidate region individually, without exploiting their relations during learning. This is the main reason behind detectors generating several candidate regions per object. With the same spirit of GossipNet and our proposed FeatureNMS, it can be used relation networks for object detection [17]. They process a set of objects simultaneously through interaction between their appearance feature and geometry, thus allowing modeling of their relations. For these relation networks, a post-processing step for the elimination of redundant candidates could not be necessary since it exploits implicitly the relations among candidate regions as GossipNet and FeatureNMS do.

Bibliography

[1] American cancer society, lung cancer, [online]. `https://www.cancer.org/cancer/lung-cancer.html`. [Accessed 20 August 2018].

[2] International agency for research on cancer, cancer fact sheets, [online]. `http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx` [Accessed 10 August 2018].

[3] S. G. Armato, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. R. Henschke, and E. A. Hoffman. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on ct scans. *Medcal Physics*, 38, 2011.

[4] N. Bodla, B. Singh, R. Chellappa, and L. S. Davis. Soft-NMS–improving object detection with one line of code. In *Proceedings of the IEEE international conference on computer vision*, pages 5561–5569, 2017.

[5] ANTOINE BROYELLE. *Automated Pulmonary Nodule Detection on Computed Tomography Images with 3D Deep Convolutional Neural Network*. PhD thesis, KTH Royal Institute of Technology, 2018.

[6] E. R. Capia, A. M. Sousa, and A. X. Falcão. Improving lung nodule detection with learnable non-maximum suppression. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1861–1865, 2020.

[7] Yunpeng Chen, Jianan Li, Huaxin Xiao, Xiaojie Jin, Shuicheng Yan, and Jiashi Feng. Dual path networks. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4467–4475. Curran Associates, Inc., 2017.

[8] Yunpeng Chen, Jianan Li, Huaxin Xiao, Xiaojie Jin, Shuicheng Yan, and Jiashi Feng. Dual path networks. *CoRR*, abs/1707.01629, 2017.

[9] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. R-FCN: object detection via region-based fully convolutional networks. *CoRR*, abs/1605.06409, 2016.

[10] J. Ding, A. Li, Z. Hu, and L. Wang. Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 559–567. Springer, 2017.

[11] Q. Dou, H. Chen, Y. Jin, H. Lin, J. Qin, and P.-A. Heng. Automated pulmonary nodule detection via 3D convnets with online sample filtering and hybrid-loss residual learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 630–638. Springer, 2017.

[12] A. X. Falcao, J. Stolfi, and R. de Alencar Lotufo. The image foresting transform: theory, algorithms, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):19–29, Jan 2004.

[13] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1627–1645, Sept 2010.

[14] Ross B. Girshick. Fast R-CNN. *CoRR*, abs/1504.08083, 2015.

[15] Ross B. Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *CoRR*, abs/1311.2524, 2013.

[16] J. Hosang, R. Benenson, and B. Schiele. Learning non-maximum suppression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4507–4515, 2017.

[17] H. Hu, J. Gu, Z. Zhang, J. Dai, and Y. Wei. Relation networks for object detection. *CoRR*, 2017.

[18] J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7132–7141, 2018.

[19] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, 2017.

[20] Sarfaraz Hussein, Kunlin Cao, Qi Song, and Ulas Bagci. Risk stratification of lung nodules using 3d cnn-based multi-task learning. *CoRR*, abs/1704.08797, 2017.

[21] Colin Jacobs, Eva M. van Rikxoort, Thorsten Twellmann, Ernst Th. Scholten, Pim A. de Jong, Jan-Martin Kuhnigk, Matthijs Oudkerk, Harry J. de Koning, Mathias Prokop, Cornelia Schaefer-Prokop, and Bram van Ginneken. Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images. *Medical Image Analysis*, 18(2):374 – 384, 2014.

[22] H. L. Kundel, K. Berbaum, and D. D. Dorfman. Receiver operating characteristic analysis in medical imaging. *J. ICRU*, 8:1–62, 2008.

[23] Yann LeCun, Y Bengio, and Geoffrey Hinton. Deep learning. 521:436–44, 05 2015.

[24] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott E. Reed, Cheng-Yang Fu, and Alexander C. Berg. SSD: single shot multibox detector. *CoRR*, abs/1512.02325, 2015.

[25] Keelin Murphy, B van Ginneken, Arnold Schilham, B J de Hoop, Hester Gietema, and Mathias Prokop. A large-scale evaluation of automatic pulmonary nodule detection in chest ct using local image features and k-nearest-neighbour classification. 13:757–70, 08 2009.

[26] A. Neubeck and L. Van Gool. Efficient non-maximum suppression. In *18th International Conference on Pattern Recognition (ICPR'06)*, 2006.

[27] GÖRKEM POLAT. Classification of lung nodules in ct images using convolutional neural networks. Master's thesis, Middle East Technical University, 2018.

[28] Y. Qin, H. Zheng, Y.-M. Zhu, and J. Yang. Simultaneous accurate detection of pulmonary nodules and false positive reduction using 3d cnns. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018.

[29] Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. CNN features off-the-shelf: an astounding baseline for recognition. *CoRR*, abs/1403.6382, 2014.

[30] Joseph Redmon and Ali Farhadi. YOLO9000: better, faster, stronger. *CoRR*, abs/1612.08242, 2016.

[31] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster R-CNN: towards real-time object detection with region proposal networks. *CoRR*, abs/1506.01497, 2015.

[32] R. Rothe, M. Guillaumin, and L. Van Gool. Non-maximum suppression for object detection by passing messages between windows. In *Asian conference on computer vision*, pages 290–306. Springer, 2014.

[33] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. *CoRR*, abs/1312.6229, 2013.

[34] A. A. A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. J. van Riel, M. M. W. Wille, M. Naqibullah, C. I. Sánchez, and B. van Ginneken. Pulmonary nodule detection in ct images: False positive reduction using multi-view convolutional networks. *IEEE Transactions on Medical Imaging*, 35(5):1160–1169, May 2016.

[35] A. A. A. Setio, A. Traverso, T. de Bel, M. S. N. Berens, C. van den Bogaard, P. Cerello, H. Chen, Q. Dou, M. E. Fantacci, B. Geurts, R. van der Gugten, P.-A. Heng, B. Jansen, M. M. J. de Kaste, V. Kotov, J. Y.-H. Lin, J. T. M. C. Manders, A. S.-M. Mengana, J. C. G.-N. Naranjo, M. Prokop, M. Saletta, C. S. Prokop, Scholten. E. T., L. Scholten, M. M. Snoeren, E. L. Torres, J. Vandemeulebroucke, N. Walasek, G. C. A. Zuidhof, B. van Ginneken, and C. Jacobs. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. *CoRR*, abs/1612.08012, 2016.

[36] Wei Shen, Mu Zhou, Feng Yang, Di Dong, Caiyun Yang, Yali Zang, and Jie Tian. Learning from experts: Developing transferable deep features for patient-level lung cancer prediction. In Sebastien Ourselin, Leo Joskowicz, Mert R. Sabuncu, Gozde Unal, and William Wells, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pages 124–131, Cham, 2016. Springer International Publishing.

[37] Wei Shen, Mu Zhou, Feng Yang, Caiyun Yang, and Jie Tian. Multi-scale convolutional neural networks for lung nodule classification. *Information processing in medical imaging : proceedings of the ... conference*, 24:588–99, 2015.

[38] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

[39] QingZeng Song, Lei Zhao, XingKe Luo, and XueChen Dou. Using deep learning for classification of lung nodules on computed tomography images. 2017:1–7, 08 2017.

[40] Azael Sousa, Samuel Martins, Alexandre Falcão, Fabiano Reis, Ericson Bagatin, and Klaus Irion. Altis: A fast and automatic lung and trachea ct-image segmentation method. *Medical Physics*, 46, 08 2019.

[41] Kenji Suzuki, Feng Li, Shusuke Sone, and Kunio Doi. Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose ct by use of massive training artificial neural network. *IEEE Transactions on Medical Imaging*, 24:1138–1150, 2005.

[42] H. Tang, D. R. Kim, and X. Xie. Automated pulmonary nodule detection using 3d deep convolutional neural networks. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2018.

[43] H. Tang, X. Liu, and X. Xie. An end-to-end framework for integrated pulmonary nodule detection and false positive reduction. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 859–862, 2019.

[44] Eva van Rikxoort, Bartjan Hoop, Max Viergever, Mathias Prokop, and Bram Ginneken. Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection. *Medical physics*, 36:2934–47, 08 2009.

[45] Xingjian Yan, Jianing Pang, Hang Qi, Yixin Zhu, Chunxue Bai, Xin Geng, Mina Liu, Demetri Terzopoulos, and Xiaowei Ding. Classification of lung nodule malignancy risk on computed tomography images using convolutional neural network: A comparison between 2d and 3d strategies. In Chu-Song Chen, Jiwen Lu, and Kai-Kuang Ma, editors, *Computer Vision – ACCV 2016 Workshops*, pages 91–101, Cham, 2017. Springer International Publishing.

[46] W. Zhu, C. Liu, W. Fan, and X. Xie. Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 673–681, 2018.

[47] Wentao Zhu, Qi Lou, Yeeleng Scott Vang, and Xiaohui Xie. Deep multi-instance networks with sparse label assignment for whole mammogram classification. *CoRR*, abs/1612.05968, 2016.