Universidade Estadual de Campinas
Instituto de Computação

Julio César Mendoza Bobadilla

Lung Nodule Classification Based on Deep Convolutional Neural Networks in Chest X-Ray Images

Classificação de Nódulos Pulmonares Baseada em Redes Neurais Convolucionais Profundas em Radiografias

CAMPINAS
2017

Julio César Mendoza Bobadilla

Lung Nodule Classification Based on Deep Convolutional Neural Networks in Chest X-Ray Images

Classificação de Nódulos Pulmonares Baseada em Redes Neurais Convolucionais Profundas em Radiografias

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.

Thesis 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. Hélio Pedrini

Este exemplar corresponde à versão final da Dissertação defendida por Julio César Mendoza Bobadilla e orientada pelo Prof. Dr. Hélio Pedrini.

CAMPINAS
2017

Agência(s) de fomento e nº(s) de processo(s): CAPES
ORCID: https://orcid.org/0000-0001-5820-2615

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

M523L

Mendoza Bobadilla, Julio César, 1990-
Lung nodule classification based on deep convolutional neural networks in chest X-ray images / Julio César Mendoza Bobadilla. – Campinas, SP : [s.n.], 2017.

Orientador: Hélio Pedrini.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Computação.

1. Pulmões - Câncer. 2. Diagnóstico auxiliado por computador. 3. Processamento de imagens. 4. Redes neurais (Computação). I. Predrini, Hélio, 1963-. II. Universidade Estadual de Campinas. Instituto de Computação. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Classificação de nódulos pulmonares baseada em redes neurais convolucionais profundas em radiografias
Palavras-chave em inglês:
Lungs - Cancer
Computer-aided diagnosis
Image processing
Neural networks (Computer science)
Área de concentração: Ciência da Computação
Titulação: Mestre em Ciência da Computação
Banca examinadora:
Hélio Pedrini [Orientador]
Marcelo Zanchetta do Nascimento
Esther Luna Colombini
Data de defesa: 18-12-2017
Programa de Pós-Graduação: Ciência da Computação

Universidade Estadual de Campinas
Instituto de Computação

Julio César Mendoza Bobadilla

Lung Nodule Classification Based on Deep Convolutional Neural Networks in Chest X-Ray Images

Classificação de Nódulos Pulmonares Baseada em Redes Neurais Convolucionais Profundas em Radiografias

Banca Examinadora:

- Prof. Dr. Hélio Pedrini
  IC/UNICAMP
- Prof. Dr. Marcelo Zanchetta do Nascimento
  FACOM/UFU
- Prof. Dra. Esther Luna Colombini
  IC/UNICAMP

A ata da defesa com as respectivas assinaturas dos membros da banca encontra-se no processo de vida acadêmica do aluno.

Campinas, 18 de dezembro de 2017

Resumo

O câncer de pulmão, que se caracteriza pela presença de nódulos, é o tipo mais comum de câncer em todo o mundo, além de ser um dos mais agressivos e fatais, com 20% da mortalidade total por câncer. A triagem do câncer de pulmão pode ser realizada por radiologistas que analisam imagens de raios-X de tórax (CXR). No entanto, a detecção de nódulos pulmonares é uma tarefa difícil devido a sua grande variabilidade, limitações humanas de memória, distração e fadiga, entre outros fatores. Essas dificuldades motivam o desenvolvimento de sistemas de diagnóstico por computador (CAD) para apoiar radiologistas na detecção de nódulos pulmonares. A classificação do nódulo do pulmão é um dos principais tópicos relacionados aos sistemas de CAD. Embora as redes neurais convolucionais (CNN) tenham demonstrado um bom desempenho em muitas tarefas, há poucas explorações de seu uso para classificar nódulos pulmonares em imagens CXR. Neste trabalho, propusemos e analisamos um arcabouço para a detecção de nódulos pulmonares em imagens de CXR que inclui segmentação da área pulmonar, localização de nódulos e classificação de nódulos candidatos. Apresentamos um método para classificação de nódulos candidatos com CNNs treinadas a partir do zero. A eficácia do nosso método baseia-se na seleção de parâmetros de aumento de dados, no projeto de uma arquitetura CNN especializada, no uso da regularização de *dropout* na rede, inclusive em camadas convolucionais, e no tratamento da falta de amostras de nódulos em comparação com amostras de fundo, balanceando mini-lotes em cada iteração da descida do gradiente estocástico. Todas as decisões de seleção do modelo foram tomadas usando-se um subconjunto de imagens CXR da base *Lung Image Database Consortium and Image Database Resource Initiative* (LIDC/IDRI) separadamente. Então, utilizamos todas as imagens com nódulos no conjunto de dados da *Japanese Society of Radiological Technology* (JSRT) para avaliação. Nossos experimentos mostraram que as CNNs foram capazes de alcançar resultados competitivos quando comparados com métodos da literatura. Nossa proposta obteve uma curva de operação (AUC) de 7.51 considerando 10 falsos positivos por imagem (FPPI) e uma sensibilidade de 71.4% e 81.0% com 2 e 5 FPPI, respectivamente.

Abstract

Lung cancer, which is characterized by the presence of nodules, is the most common type of cancer around the world, as well as one of the most aggressive and deadliest cancer, with 20% of total cancer mortality. Lung cancer screening can be performed by radiologists analyzing chest X-ray (CXR) images. However, the detection of lung nodules is a difficult task due to their wide variability, human limitations of memory, distraction and fatigue, among other factors. These difficulties motivate the development of computer-aided diagnosis (CAD) systems for supporting radiologists in detecting lung nodules. Lung nodule classification is one of the main topics related to CAD systems. Although convolutional neural networks (CNN) have been demonstrated to perform well on many tasks, there are few explorations of their use for classifying lung nodules in CXR images. In this work, we proposed and analyzed a pipeline for detecting lung nodules in CXR images that includes lung area segmentation, potential nodule localization, and nodule candidate classification. We presented a method for classifying nodule candidates with a CNN trained from the scratch. The effectiveness of our method relies on the selection of data augmentation parameters, the design of a specialized CNN architecture, the use of dropout regularization on the network, inclusive in convolutional layers, and addressing the lack of nodule samples compared to background samples balancing mini-batches on each stochastic gradient descent iteration. All model selection decisions were taken using a CXR subset of the Lung Image Database Consortium and Image Database Resource Initiative (LIDC/IDRI) dataset separately. Thus, we used all images with nodules in the Japanese Society of Radiological Technology (JSRT) dataset for evaluation. Our experiments showed that CNNs were capable of achieving competitive results when compared to state-of-the-art methods. Our proposal obtained an area under the free-response receiver operating characteristic (AUC) curve of 7.51 considering 10 false positives per image (FPPI), and a sensitivity of 71.4% and 81.0% with 2 and 5 FPPI, respectively.

List of Figures

List of Tables

List of Abbreviations

AAM	Active Appearance Model
ACM	Active Contour Model
ANN	Artificial Neural Network
AR	Adaptive Ring
ARG	Average Radial Gradient
ASM	Active Shape Model
AUC	Area Under the Curve
CAD	Computer-Aided Detection
CI	Convergence Index
CNN	Convolutional Neural Network
ConvNet	Convolutional Neural Network
CT	Computer Tomography
CXR	Chest X-Ray
DoG	Difference of Gaussians
DoH	Determinant of Hessian
FROC	Free-response Receiver Operating Characteristic
GPU	Graphics Processing Unit
GT	Ground Truth
JRST	Japanese Society of Radiological Technology
LCE	Local Contrast Enhancement
LIDC/IDRI	Lung Image Database Consortium and Image Database Resource Initiative
LoG	Laplacian of Gaussian
MNIST	Modified National Institute of Standards and Technology
MRI	Magnetic Resonance Imaging
PA	Procrustes Analysis
PCA	Principal Component Analysis
ReLU	Rectified Linear Unit
ROC	Receiver Operating Characteristic
SB	Sliding Band
SCR	Segmentation on Chest Radiographs
SGD	Stochastic Gradient Descent
SIFT	Scale-Invariant Feature Transform
SVM	Support Vector Machine
WMCI	Weighted Multi-scale Convergence Index

Contents

Chapter 1

Introduction

Lung cancer is the most common type of cancer in men, and the third in women around the world. Moreover, it is one of the most aggressive cancers with a survival rate of $10-15\%$, as well as the deadliest cancer with 20% of total cancer mortality [53]. Figure 1.1 shows the percentages of cancer deaths worldwide. Lung cancer is characterized by the presence of nodules. Notwithstanding this fact, lung nodules can also be caused by scar tissues, inflammations or infections related to benign conditions, such as tuberculosis [9, 38].

Lung cancer screening is performed by radiologists through the analysis of computer tomography (CT) or chest X-ray (CXR) images. Although screening using CT is superior to screening with CXR [88], the acquisition process is more expensive. In consequence, lung cancer screening with CXR are considered by countries that cannot afford a screening program with CT due to its cost and availability [81].

Lung cancer screening can be improved with computational tools that assist radiologists in their image analysis. Computer-aided detection (CAD) systems detect the location of lung nodules automatically with the objective of enhancing the performance of radiologists by considering the output of the system as a second opinion [13, 37]. Thus, CAD systems can potentially detect nodules missed by the radiologists, increasing their detection rate. Furthermore, CAD systems can reduce the reading time allowing the radiologists to focus on the image interpretation tasks, and have the potential to provide additional information such as differentiation among benign, malign and non-nodule structures [2, 13].

CAD systems generally involve three main stages for detecting lung nodules: lung segmentation, candidate detection, and nodule classification. Lung nodules can be detected by considering bright masses of expected size, shape and texture [37]. Then, the methods used for the candidate detection and classification stages should consider these features to recognize potential nodule locations.

1.1 Motivation

One of the main motivations to develop CAD systems for detecting lung nodules is to support cancer screening procedures. Lung cancer screening is performed through the analysis of low-dose CT, CXR or sputum cytology samples [81]. Several studies have been

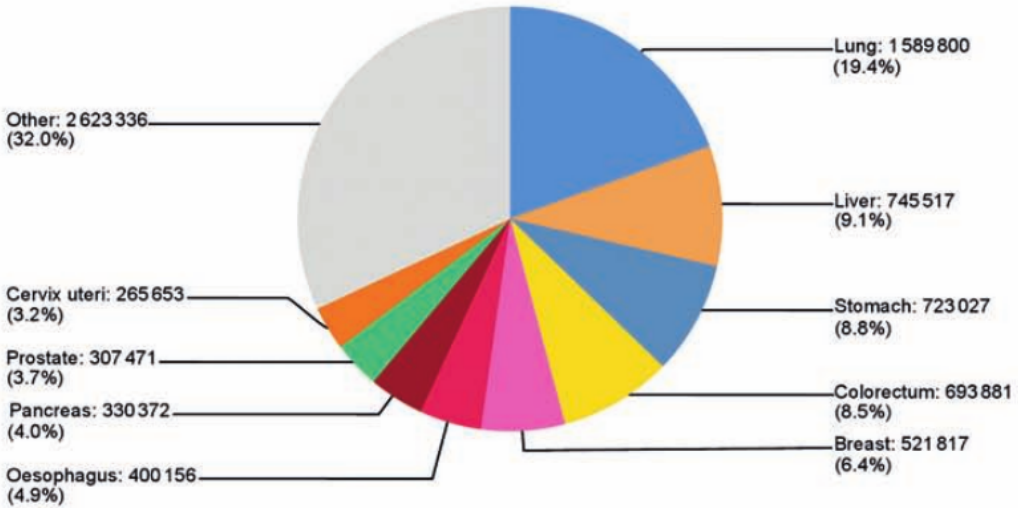


Figure 1.1: Estimated number of cancer deaths on both sexes (Source: World Cancer Report 2014 [53]).

conducted to analyze the effectiveness for lung cancer screening with low-dose CT and CXR images in high risk patients [6, 21, 58, 88]. These studies suggest that screening with CXR increases the detection rate, but not reduce lung cancer mortality. Screening with low-dose CT has better detection rate compared to screening with CXR [88]. Moreover, the usage of low-dose CT reduces lung cancer mortality, as well as mortality caused by other diseases [88].

However, screening low-dose CT has some potential risks. For instance, the exposure to CT ionizing radiation may increase the risk of cancer in patients, and the usage of low-dose CT has higher false positive rate than CXR screening. Therefore, it requires further investigation and also additional costs [81]. On the other hand, lung cancer screening with CXR is cheaper, as previously mentioned. Furthermore, screening with CXR can be improved when radiologists use CAD systems in clinical routine [81]. Thus, the potential role of CXR on lung cancer screening motivates research efforts to improve the effectiveness of CAD systems in CXR, and to measure the impact of these systems on the radiologist performance.

Detecting lung nodules is a challenging task due to their wide variability. In CXR images, lung nodules may have low contrast, can overlap with ribs, large pulmonary vessels and other anatomical structures outside the lung, as well as they usually have irregular size, shape and opacity [37]. Besides the variability of lung nodules, another factor that leads to misdetection of lung nodules on screening is that radiologists are subject to the human limitations of memory, distraction and fatigue. Radiologists usually detect lung nodules by considering brightness and shape of circular objects within the thoracic cavity.

CAD systems for lung nodule detection on CXR images also consider nodule features. Traditional approaches are based on the design of hand-crafted features that characterize lung nodules. This approach presents some drawbacks, such that manual design and tunning may work well on one dataset, but may not work on a new dataset [76]. Moreover, designing hand-crafted features is time-consuming, and typically requires expert knowl-

edge [84]. On the other hand, CAD systems based on deep learning approaches have advantages over traditional methods, such as the automated feature learning based on an objective function that optimizes nodule classification or detection effectiveness [76]. Besides, deep learning based approaches face challenges on lung nodule detection, for instance, the need of large labeled datasets for training, which is also a challenge for several medical imaging tasks.

1.2 Hypotheses

A research question often found in the literature, also considered in this dissertation, is: *Can CNNs be used in medical imaging applications effectively, overcoming the difficulties present in this domain?*.

Several approaches have been proposed to adopt CNNs in medical imaging tasks. Three common approaches explored to use CNNs are: (i) training CNNs from scratch, (ii) use of off-the-shelf CNN features, (iii) use of transfer learning with fine-tuning [73]. From these strategies, transfer learning with fine tuning has been demonstrated to be the best option in diverse tasks with different imaging modalities [73, 78].

In this work, we explore the first approach. The main hypothesis is that *CNNs trained from scratch can perform well on lung nodule classification by setting proper regularization and optimization methods*. Secondary hyphoteses studied in this work include:

- Does the design of a specialized architecture for lung nodule classification improve the performance of the model?
- Are regularization effects of data augmentation, dropout and weight penalties critical for lung nodule classification due to the amount of samples used in training?
- Do regularization and optimization effects of unsupervised objectives in loss functions improve CNN performance in lung nodule classification?

1.3 Contributions

The main contributions of this dissertation are:

- *an analysis and evaluation of lung area segmentation and candidate nodule localization*: we analyzed and compared the performance of relevant approaches to candidate nodule classification, and we evaluated the performance of statistical shape model for lung area segmentation.
- *the proposition of a method for lung nodule classification based a CNN trained from scratch*: our approach investigates the selection of data augmentation parameters, the design of a specialized CNN architecture, the use of dropout regularization on the network, including convolutional layers, and the lack of nodule samples compared to background samples balancing mini-batches in each stochastic gradient descent iteration.

1.4 Text Structure

This dissertation is organized as follows. Chapter 2 describes relevant concepts and work related to the topic under investigation. Chapter 4 presents the methodology proposed in this work. Chapter 5 describes and discusses our experimental results. Chapter 6 concludes this work with final remarks and directions for future work.

Chapter 2

Background

In this chapter, we provide an overview of the technical background for computer-aided detection (CAD) of lung nodules in chest radiography images. First, we introduce relevant concepts related to lung nodule detection. Then, we present three subproblems considered in most of the CAD systems used for detecting lung nodules, and we describe the methods used in this dissertation to address these subproblems. Finally, we present the related work focusing on the methods used in every stage of lung nodule detection.

2.1 Related Concepts

Medical imaging is itself a major subject that involves many areas, from physics to medicine, to build visual representations of the interior of a human body for clinical analysis. Usage of medical images in the diagnosis, detection and treatment of diseases is of increasing importance due to the need for more accurate, cost-effective and less invasive procedures, as well as the advances in medical imaging technologies [85].

Human experts, such as physicians and radiologists, read and interpret medical images. An important component of image interpretation is the identification of anatomical structures present in the images. However, image interpretation is difficult due to the wide variability of anatomical structures, such as human limitation of memory, distraction, fatigue, among others. These difficulties motivate the development of image analysis tools, such as CAD systems, to assist human experts in image analysis.

2.1.1 Computer Aided Detection of Lung Nodules

CAD refers to the procedure that assists radiologists in the analysis and evaluation of medical images. CAD systems can help radiologists in the reading process, for instance, highlighting suspicious locations that may be related to a pathology condition and indicate their degree of confidence.

CAD systems are used for interpretation of images in various modalities such as CXR, CT, magnetic resonance imaging, and ultrasound. Moreover, CAD systems assist radiologists in several tasks, such as breast mass detection, polyp detection and lung nodule detection, which is the focus of this dissertation.

There are two common modes to use CAD systems by radiologists [10]. First, radiologists consider the CAD system as a first reader such that they need only analyze suspicious regions detected by the system. This mode helps radiologists reduce the time spent in locating suspicious regions, prevents fatigue, and allows them to focus on image analysis. However, it also limits its effectiveness to the CAD system performance. Second, radiologists can analyze the CXR first, and consider the CAD system as a second reader, such that missed nodules may be detected by the system, improving the effectiveness of the radiologists.

The impact of CAD systems on the radiologists performance depends on the system effectiveness and radiologists experience. There is evidence that CAD systems used as a second reader enhance the performance of unexperienced radiologists, while not benefiting significantly experienced radiologists [20].

CAD systems for detecting lung nodules in CXR images receive as input a CXR image and give as output the same image highlighting the locations of suspicious regions found by the system. Moreover, CAD systems usually detect nodules solving three subproblems: lung area segmentation, candidate nodule detection, and false positive reduction though candidate classification.

2.1.2 Lung Area Segmentation

Lung area segmentation is a binary segmentation problem in which, given an input CXR image, our objective is to determine a function $f : \Omega \to \{0, 1\}$ such that, for every pixel $P \in \Omega$, where Ω is the lattice that represents the image, $f(P) = 1$ if the pixel belongs to the lung area, and 0 otherwise.

The function f can be represented by a binary mask and, according to the type of segmentation method used, it can be calculated in various ways. For instance, deformable model based methods find a set of representative points along the lung boundary, also called landmarks. Thus, landmarks linked with lines can approximate the lung boundary, and can be used to generate the lung area mask.

The aim of this stage is to separate the lung area in order to prevent the detection of structures similar to nodules located outside the lung. The accuracy of the segmentation method influences directly the lung nodule detection performance. If the method under-segments the lung area, some nodules located on missed regions would not be found in posterior stages of the CAD system. If the method over-segments the lung area, more candidate regions that are not nodules would be detected, increasing the false positives on the last stages.

Deformable Models

Deformable models correspond to a class of segmentation methods that estimate the shape of an object starting from an initial contour, which is improved iteratively by shrinking or expansion operations. Deformable models interpret the segmentation as a global optimization problem. The goal is to minimize a cost function that reaches its optimum when the shape perfectly fits the object boundary. Optimization is performed using numerical optimization methods, such as gradient descent or Gauss-Newton methods.

Deformable models can be grouped into two classes according to the way in which they represent the shapes: parametric and geometric models [54]. Parametric models represent shapes explicitly with their parametric forms. Geometric models represent shapes implicitly as a level set of a higher-dimensional scalar function.

In this work, we focus on a family of parametric methods called statistical shape models [16, 54] conformed by active shape models (ASM) [17], active appearance models (AAM) [18], deformable templates [35], among others. The main characteristic of these models is that they learn statistical information about the object to segment. Statistical shape models are particularly well suited for problems in which the landmarks in the boundary of the object to segment are available as part of the ground truth.

In this work, we focus on AAM for segmentation. We introduce concepts related to AAMs in the next section.

Active Appearance Models

AAMs are generative non-linear parametric models that describe variations of deformable objects in terms of shape and appearance [3]. By estimating the shape and appearance basis of an object class in a set of images, AAM parameters can be adjusted to generate synthetic images as similar as possible to unseen images. Moreover, AAM parameters can be used for segmentation, such as in this work, and for image interpretation [18, 50, 83].

AAMs are constructed from a set of images with a set of v landmark points $\boldsymbol{x_i} = (x_i, y_i)^T \in \mathbb{R}^2$, defined a priori, that represents the shape $\boldsymbol{s} = (x_1, y_1, \cdots, x_v, y_v)^T \in \mathbb{R}^{2v\times 1}$ of the target object on the image [3]. AAMs can approximate instances of an object with a model that explains its shape and appearance, and a function to map the estimated appearance from the AAM reference shape to the estimated shape.

Shape Model. The shape is modeled to allow linear variation such that it could be expressed as the sum of the mean shape $\bar{\boldsymbol{s}}$ and a linear combination of n shape vectors $\boldsymbol{s}_i$:

$$\boldsymbol{s} = \bar{\boldsymbol{s}} + \sum_{i=1}^{n} p_i \boldsymbol{s_i} \tag{2.1}$$

$$= \bar{\boldsymbol{s}} + \boldsymbol{Sp} \tag{2.2}$$

where $\bar{\boldsymbol{s}} \in \mathbb{R}^{2v\times 1}$, and $\boldsymbol{S} \in \mathbb{R}^{2v\times n}$ and $\boldsymbol{p} \in \mathbb{R}^{n\times 1}$ describe the shape bases and parameters, respectively. Object shapes are first normalized using procrustes analysis (PA) to remove variation using a global shape transformations. Then, principal component analysis (PCA) is applied to obtain the shape bases. PCA only considers local shape non-rigid variations due to the application of PA normalization [50].

Appearance Model. The appearance is modeled for all pixels $\boldsymbol{x}$ in a reference frame $\bar{\boldsymbol{s}}$. AAMs are built to allow linear appearance variation. The model can be expressed as the sum of the mean appearance $\bar{A}(\boldsymbol{x})$ and a linear combination of m appearance bases $A_i(\boldsymbol{x})$ with parameters c_i. Let $\boldsymbol{a}$ be the vectorized representation of $A(\boldsymbol{x})$, such that the

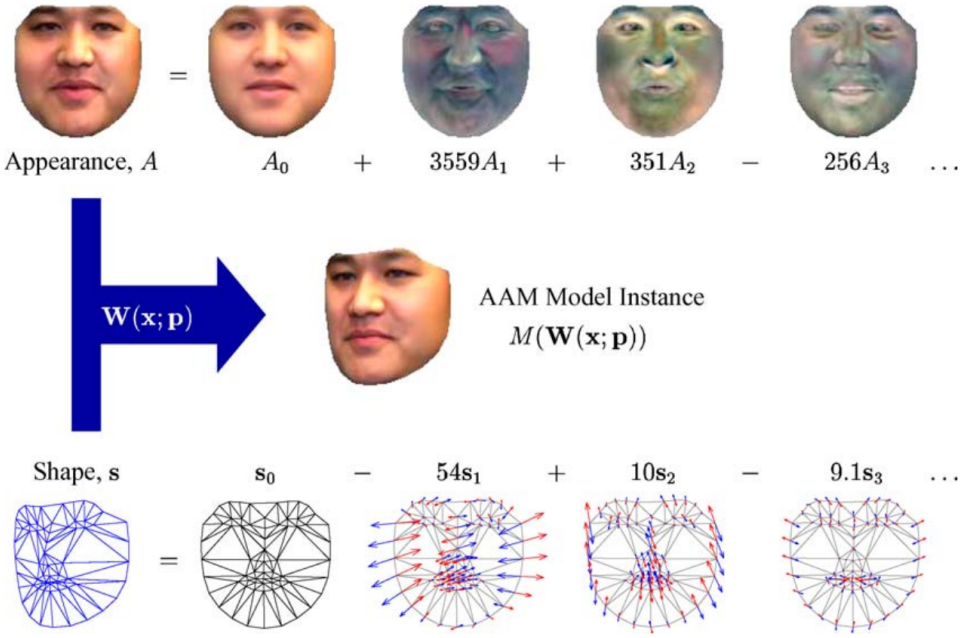


Figure 2.1: An example of AAM instantiation for face modeling (Source: Matthews and Baker [50])

model can be denoted as follows:

$$A(\boldsymbol{x}) = \bar{A}(\boldsymbol{x}) + \sum_{i}^{m} c_i A_i(\boldsymbol{x}) \tag{2.3}$$

$$\boldsymbol{a} = \bar{\boldsymbol{a}} + \boldsymbol{A}\boldsymbol{c} \tag{2.4}$$

where $\bar{\boldsymbol{a}} \in \mathbb{R}^{F\times 1}$ is the mean appearance, $\boldsymbol{A} \in \mathbb{R}^{F\times n}$ and $\boldsymbol{c} \in \mathbb{R}^{n\times 1}$ describe the appearance bases and parameters, respectively.

Model instantiation. The aforementioned models describe shape and appearance variations, but they do not describe how to use the models to approximate a new image I with a target object. Given the shape parameters $\boldsymbol{p}$, we can generate a shape $\boldsymbol{s}$. Given the appearance parameters $\boldsymbol{c}$, we can generate an appearance $\boldsymbol{a}$ defined in a reference shape $\bar{\boldsymbol{s}}$. Then, we generate an approximation to I by using a warp function $\mathcal{W}(\boldsymbol{x}, \boldsymbol{p})$ that maps the pixel positions $\boldsymbol{x}$ from the frame of the shape $\bar{\boldsymbol{s}}$ to the estimated shape $\boldsymbol{s}$ with parameters $\boldsymbol{p}$ [50]. Figure 2.1 shows an example of model instanciation on face modeling.

Fitting AAMs. Given a new image I, a shape model, an appearance model and a warp function, we wish to generate a synthetic image M as similar as possible to the new image. In other words, the difference between I and M must be minimal. We measure this difference as the sum of squared differences. Then, the fitting process could be defined as the following optimization problem:

$$\boldsymbol{p}^*, \boldsymbol{c}^* = \arg\min_{\boldsymbol{p},\boldsymbol{c}} \|\boldsymbol{i}[\boldsymbol{p}] - (\bar{\boldsymbol{a}} + \boldsymbol{A}\boldsymbol{c})\|^2 \tag{2.5}$$

where $\boldsymbol{i}[\boldsymbol{p}]$ is the vectorized representation of $I(\mathcal{W}(\boldsymbol{x}, \boldsymbol{p}))$.

2.1.3 Detection of Nodule-like Regions

Detection of nodule-like regions, also called candidate regions, consists in finding a set of locations $\boldsymbol{c} = \{(x_i, y_i)\} \in \mathbb{R}^{n\times 2}$, given an input image, such that the region of interest around each location (x_i, y_i) is likely to contain a lung nodule.

Methods for candidate detection can provide us with additional information about the candidates besides their location, such as size and the likelihood of being a nodule. For instance, under the assumption that lung nodules are circular objects, scale-space based detectors can give us an approximated radius of the candidates. Moreover, convergence index based detectors calculate the convergence index, that is directly proportional to the nodule likelihood, for all pixels on the image, including candidate locations.

Scale-Space Representation based Detectors

Nodule-like regions appear as blobs of variable size on the image. An intuition to detect a blobs is to describe a property that differentiates the blob from the background. However, the detection methods should consider that such description depends on the scale in which the blob is represented on the image. Since we do not know the scale in which the object is best described, we can represent the image at multiple scales such that one the scales will have the desired description.

An approach for representing the image at multiple scale is the scale-space framework [44]. The central idea is that, given an image, we can create a multi-scale representation by generating a family of derived images where fine-scale details are removed sequentially by a scale-space smoothing mechanism.

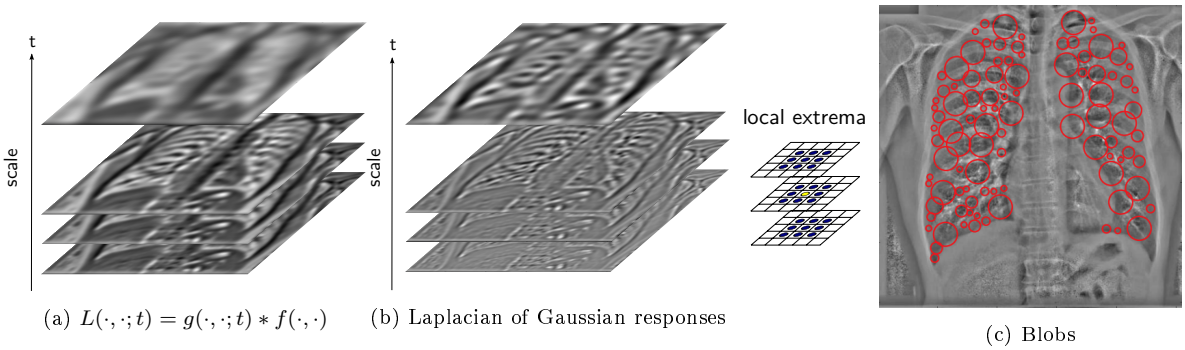
Formally, let a 2D image be denoted as $f : \mathbb{R}^2 \to \mathbb{R}$. The scale-space representation $L : \mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}$ is defined as the family of images that result from the convolution of the image with various Gaussian kernels $g : \mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}$ with a variance $t = \sigma^2$, which is referred to as the scale parameter [46]:

$$L(\cdot, \cdot, t) = g(\cdot, \cdot; t) * f(\cdot, \cdot) \tag{2.6}$$

We can find blobs using differential operators over L at a fixed scale. For instance, detectors such as Laplacian of Gaussian (LoG) [44] derive from second order partial derivatives of L. The LoG corresponds to the trace of the Hessian matrix, and its local extrema indicate bright or dark blobs on the image. The Difference of Gaussians (DoG) [48] is an approximation of the LoG, therefore, the local extrema of its response also correspond to blobs in the image. The Determinant of Hessian (DoH) [46] is another detector in which the extrema responses correspond to blobs and saddle points on the image.

We can find blobs at multiple scales calculating local extrema over multiple scales. However, we cannot compare responses of different scales directly since numerical values of the derivatives of L decrease when scale t increases due to the non-enhancement property of spatial smoothing. Instead, multi-scale detectors localize blob-like structures by finding

local extrema on the response of differential expressions calculated using normalized-derivatives of L [45]. Figure 2.2 depicts multi-scale LoG detector.



(a) $L(\cdot,\cdot;t) = g(\cdot,\cdot;t) * f(\cdot,\cdot)$ (b) Laplacian of Gaussian responses (c) Blobs

Figure 2.2: Scale-space blob detection with the LoG operator.

Convergence Index based Detectors

The convergence index (CI) measures the degree in which gradients around a point of interest are oriented to it. We can define convergence degree of the gradient at the point Q in the neighborhood or support region R of a point of interest P as $\cos\theta(k,l)$, which is the projection of the unitary vector with same orientation as the gradient, with respect to the line $\overline{PQ}$ as we show in Figure 2.3. Then, the convergence index point of interest P is the average of the convergence degrees of the gradients of all pixels in its support region R [39]. Formally, the convergence index at the point P with the location (i,j) can be calculated as follows:

$$CI(i,j) = \frac{1}{M} \sum_{(k,l)\in R} \cos\theta(k,l) \tag{2.7}$$

where (k,l) is the position of the point Q, and M is the number of pixels in R. Several CI-based filters such as coin, iris, adaptive filters were proposed in [39, 86] for detecting lung nodules.

2.1.4 Classification for False Positive Reduction

Candidate classification can be seen as a binary classification problem in which we aim to estimate a function $f : \mathbb{R}^n \to \mathbb{R}$ such that, for a candidate region $r \in \mathbb{R}^{wh}$ centered at the location (x,y), we have $y = f(r)$ where y is the estimated probability of the region r may contain a truth nodule, and w and h are the dimensions of the candidate region.

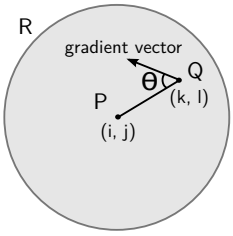


Figure 2.3: CI filter computation at the pixel (i,j)

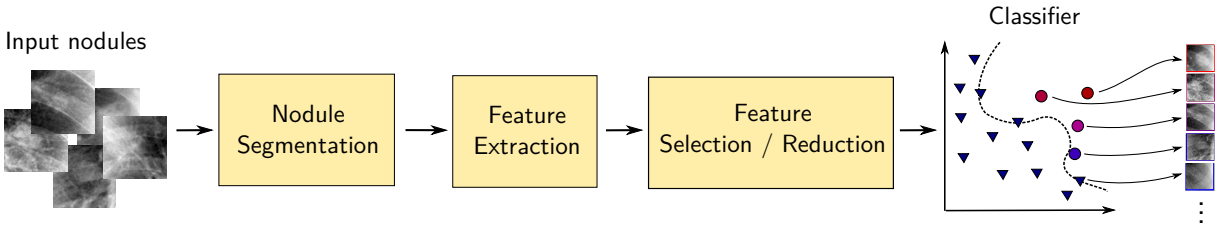


Figure 2.4: Classical pattern recognition pipeline for lung nodule classification

In the end, we have a list of candidate locations sorted by probabilities estimated by the classification method. False positive reduction is performed by considering only the first elements of the sorted list of candidates that have higher probability of being true positives.

Most of the available methods follow a classical pattern recognition pipeline that we show in Figure 2.4 . Typically, the first stage is preprocessing. Preprocessing may include image resizing, noise removal, contrast enhancement and nodule feature enhancement, nodule segmentation, among others. The second stage is feature extraction, which involves representing characteristics of the nodule that will help to differentiate them from background tissue. The third stage is feature selection, which aims to reduce the number of dimensions of the feature vector to save computational time and reduce irrelevant or noisy features. Finally, the classification stage consists in using the features selected as representations of the candidates, and using them for building a classification model.

On the other hand, deep learning approaches have been used in several tasks related to medical imaging, including false positive reduction on CAD systems [29, 76]. However, they have been rarely used for lung nodule classification on CXR images [84]. In the next sections, we introduce some concepts of deep learning and their use for classifying lung nodules.

2.1.5 Deep Learning

Deep learning is a subarea of machine learning that studies methods for solving tasks from the experience by learning a hierarchy of concepts, in which abstract concepts are defined as compositions of concepts. In this section, we provide an overview of the history of deep learning. Then, we introduce some concepts related to architecture and learning procedure of CNNs.

Deep Learning History

Several approaches have contributed to the development of deep learning. These contributions occurred in three waves [26, 70]: the first wave began with the development of biologically inspired linear models, the second wave began with the connectionism movement, and the third wave began with the use of greedy layerwise pre-training and the popularization of deep learning.

The predecessors of deep learning models were simple linear models inspired by the brain function. The McCulloch-Pitts Neuron [52] was an earlier linear model able to classify two categories of inputs, but the weights of the models should be set manually.

Emergence of learning algorithms [32] inspired the development of supervised and unsupervised linear models. The Perceptron [65] was the first model that learned weights given inputs and its categories. The adaptive linear element [90] was a linear model trained with an algorithm that is a special case of stochastic gradient descent algorithm, which is nowadays used to train deep learning models. Later, limitations of linear models reduced the interest of researchers on brain inspired models in general [55].

The biologically inspired idea that computational units can achieve complex behaviors via their interactions motivated the development of models that introduced concepts that are being used in modern deep networks. The Cognitron [23] introduced a more complicated version of the rectifier linear unit [25, 56], and a simplified version is widely used today. The Neocognitron [24] introduced a model architecture that combines convolution, weight or parameter sharing and subsampling mechanisms. This architecture was the basis of modern CNNs [41]. The Cresceptron introduced Max-Pooling as subsampling layer to gain tolerance to local location deformations [89].

The second wave began in part with the connectionism movement [51, 66], which in the context of cognitive science has as central idea that simple computational units can accomplish intelligent behaviors when interconnected together. The connectionism movement introduced the concept of distributed representation [34]. Another advance was the successful application of backpropagation algorithm to deep networks with internal representations in hidden layers [67], followed by its application to earlier CNNs [41].

Interest in neural networks diminished due to the good results of methods from other machine learning areas in several tasks. The algorithms proposed in this wave were quite well to train deep networks, however, limited amount of data and hardware available at that time did not allowed much experimentation.

The third wave began with the use of layerwise unsupervised pre-training to deep belief networks efficiently [33]. This pre-training scheme was used to train other kinds of deep networks [7, 64]. Then, the interest in the deep learning research increased, first motivated by the success of unsupervised learning algorithms for allowing deep networks to generalize well from small datasets, and followed by the success of supervised learning algorithms for leveraging large labeled datasets. Deep networks improved the state-of-the-art on several tasks, such as object recognition [40], speech recognition [33], pedestrian detection [71], image segmentation [22]. Moreover, deep networks are used in top technology companies and have supported research in other sciences [42].

Convolutional Neural Networks

CNNs are a category of neural networks for processing data with grid-like topology such as time-series which can be seen as 1D grid, images which can be seen as 2D grid, and volumes and videos which can be represented as 3D grids [26]. CNNs introduce two concepts that other types of neural networks do not consider: convolution and pooling.

Convolution operates on two functions: the input and the kernel, and produces an output known as feature map. In its discrete representation the input, kernel and the feature map are represented as multi-dimensional arrays known as tensors. Convolution operator adds relevant properties to CNNs [26]: (i) sparse interactions, which refers to reducing

the interaction of output units with a few local input units (receptive field), allowing the network to detect meaningful features efficiently, (ii) parameter sharing, which refers to learning the same set of parameters for every output unit, and (iii) equivariance to translation, which indicates that if a feature is translated on the input, its representation is translated on the output.

Pooling allows for approximating invariance to small local translations by replacing the content of the convolutional output at specific locations with a statistic calculated over its neighborhood, and by reducing the dimensionality of the output via subsampling [91]. Invariance to local translations is useful when knowing the presence of an object in an image is more important than the exact location of the object. Moreover, pooling improves the efficiency due to the size reduction of feature maps. Modern CNN usually perform pooling selecting the maximum value over the neighborhood, however other functions can be used such as the average, weighted average and L2-norm [26].

Building CNNs, as other kinds of neural networks, involves making decisions about the proper hidden unit, regularization, and learning mechanisms. Neural networks use nonlinear activation functions in their hidden units to increase their learning capacity. CNNs, similarly to other kinds of deep neural networks, use a rectified linear unit (ReLU) as nonlinearity due to its effectiveness in large models [36]. Moreover, several generalizations of ReLU have been proposed and successfully applied to CNNs such as leaky ReLU [49], parametric ReLU [31] and Maxout units [27].

The score function maps the input of the network to probabilities for the target classes. A score function is more appropriate according to the task and its configuration. For instance, the softmax function is a common choice for classification problems in which the output is categorical distribution. The cost function measures the cost of configuration of the network according to the agreed between its estimation of the network over the training set and the ground truth (GT) labels. In the same way as the score function, the choice of the cost function depends on the task, however, most of the CNNs are trained using the cross-entropy cost [26].

CNNs can be trained using gradient-based learning algorithms such as stochastic gradient descent (SGD). Besides, several variants of SGD have been proposed. SGD with momentum has been proposed to accelerate learning [63]. Another recent variant is SGD with Nesterov momentum [77].

Training CNNs with Imbalanced Datasets

Classification with imbalanced data is an important problem due to the existence of several application with skewed class distribution. For instance, applications in which the class of interest has the minority of samples, such as fraud detection in banking operations, abnormal cell detection in histological images, and lung nodule detection in medical images since the amount of background samples obtained in candidate detection stage is much greater than true nodule samples.

There are two main approaches to addressing the classification problem with imbalanced datasets: cost-sensitive learning and sampling-based methods. The central idea of cost-sensitive learning is to assign a higher cost to misclassification of minority samples

in the learning stage. Sampling based approaches include oversampling and subsampling. In oversampling, the samples of the minority class are replicated to make the dataset balanced. Oversampling may lead to overfitting [93]. In subsampling, the majority samples are reduced to make the dataset balanced. However, subsampling involves ignoring many samples of the majority class, producing loss of information.

2.2 Related Work

In this section, we briefly describe some relevant approaches used in CAD systems to detect lung nodules. The first step usually is to downsample the image to reduce computational time [11, 69]. Moreover, downsampling is commonly followed by a noise reduction procedure. The next step is to segment the lung region to reduce the search space of lung nodules. Then, methods aim to localize candidate nodules through methods for enhancing nodule like-regions and for selecting local maxima. The last step consists in reducing false positives found in the candidate detection stage using a classification method that scores candidates allowing for the user to have a high sensitivity to select the most suspicious candidates.

Lung field segmentation methods can be grouped into rule-based methods, pixel classification methods, deformable model-based methods, and hybrid methods [14, 83]. Deformable model-based methods, such as ASM, are broadly used for lung field segmentation. Schilham et al. [69] used ASM to segment the lung fields. Hardie et al. [30] segmented the lung with an ASM. In order to produce a better delineation, the ASM was applied first in the images with low resolution, and later refined with the original image. The low resolution image was generated by resampling the input image to reduce noise and processing time using a finite impulse response low-pass anti-aliasing filter followed by bilinear interpolation.

Chen and Suzuki [13] argued that ASM cannot cover variations in the boundaries of the lungs accurately because they model shape with equidistant landmarks along the boundary, and they cannot fit segments with high curvature. They used multi-segment active shape model (M-ASM) in which a proper number of landmarks is fixed according to the structural properties of the segments of lung boundaries. Then, boundary segments with high curvature have more landmarks than flat segments. Wang et al. [84] segmented the lung region using ASM with an objective function based on distance and edge constraints.

Shiraishi et al. [75] and Campadelli et al. [11] segmented the lung fields on the derivatives of the image. Shiraishi et al. [75] proposed a method that segments the lung field using ribcage edge detection method that works on the derivative of the image. Campadelli et al. [11] segmented the lung area using a multi-scale edge-tracking method that finds the lung boundaries on the edge image which they found through the first derivatives of Gaussian filters taken at different orientations. Campadelli et al. [11] proposed a method that segments the hidden lung area such as the regions of the diaphragm and heart. Segmentation masks produced with the method described by Shiraishi et al. [75] also segmented hidden regions of the lung.

The next step is the candidate nodule detection. However, before extracting candidates, preprocessing techniques are usually applied to enhance characteristics of nodule-like regions. Rib suppression methods were used to improve the visibility of candidates overlapped with ribs. A method proposed by Chen and Suzuki [13] uses Dual-Energy CXRs and a regression model based on artificial neural networks (ANNs) to suppress rib-like structures by learning to generate virtual ribs enhanced CXR supervisedly, and subtracting these images from the conventional CXRs. In contrast, unsupervised methods do not require Dual-Energy CXR to suppress ribs. Wang et al. [84] removed ribs in conventional CXR with an unsupervised method based on PCA.

Most of the methods for detecting candidate nodules can be grouped into scale-space representation based methods [11, 69, 84], radial gradient filters [13, 14, 75], and convergence index filters [30, 62, 87].

Campadelli et al. [11] detected candidate nodules in two stages: first, they used a multi-scale enhancement procedure that highlighted high-frequency regions in the image. Then, they used a multi-scale candidate detection approach based on selecting regions with high contrast between their center and the surrounding area. Moreover, they reduced the amount of candidates using a set of rules derived from observations of nodule characteristics.

Schilham et al. [69] searched for candidates in the scale-space using LoG filters. Wang et al. [84] localized candidate regions using generalized LoG filters at multiple scales and orientations. In contrast with symmetric LoG filters, generalized LoG can capture blobs with elliptical shapes. Shiraishi et al. [75] used an average radial gradient (ARG) filter to enhance nodule-like regions and multi-level thresholding to localize an initial set of candidates. Moreover, they reduced the amount of candidates with a set of rules based on nodule characteristics. Chen and Suzuki [13] used a directional gradient magnitude filter on the enhanced image to produce a likelihood map in which local maxima were taken as candidate nodules.

Wei et al. [87] used an adaptive ring (AR) filter, a CI-based filter that enhances positions with high convergence degree with a ring-like shape support regions, and selected local peaks as candidate locations. Hardie et al. [30] showed that a CI-based detector outperform LoG blob detector and ARG detector. They introduced the Weighted Multi-scale Convergence Index (WMCI) filter. WMCI filter is a set of weighted CI-filters at different scales in which each filter increases the contribution of pixels closer to the center of the structure. Finally, the local maxima values were taken as candidate nodules.

Pereira et al. [62] proposed the sliding band (SB) filter to enhance the image. SB-filter is a CI-based filter with a support region with a non-uniform band shape. After image enhancement, Watershed segmentation is used to divide the image in non-overlapping regions, and the maxima values in regions are taken as candidate nodules.

The candidate nodule detection step is followed by the classification stage. Several approaches available in the literature have followed a common pattern recognition pipeline that consists in two main steps: feature extraction and classification [11, 13, 30, 69, 75, 87]. On the other hand, some methods classify candidates directly from the pixels using neural networks [19, 43, 47], and other use an hybrid approach [84].

Several methods perform feature selection to find a better set of features before classifi-

cation. Chen and Suzuki [13] used stepwise feature selection method on the morphological and gray-level-based features extracted from candidate nodules in the original and rib-suppressed images. Then, a SVM classifier with Gaussian kernel is used for classification. Campadelli et al. [11] extracted a set of 160 features from candidates. Then, they perform univariate feature selection using the Golub statistic and classification using a SVM with polynomial kernel.

Moreover, some methods transform the candidate regions to highlight different features and improve the representation capability of their descriptors. For instance, Shiraishi et al. [75] described candidate nodules using geometric, intensity, background and edge gradient features from the original and density corrected, nodule enhanced, and contralateral subtracted images of each suspicious nodule. Classification is performed by applying ANNs in three stages: first, they reduce the amount of false positives using simple ANNs trained on feature pairs; then, they trained ANNs for four sets of features taken from transformations of the original images; finally, they used an ANN trained with the outputs of latter ANNs.

In addition, candidate representation can also be improved by performing nodule segmentation to describe nodule area and its context separately. Wei et al. [87] proposed an optimal feature set extracting features from candidates in the original and transformed images. Transformed images were obtained with Adaptive Ring filter, Iris filter, and Sobel filter. Wei et al. [87] segmented the candidate nodules using an active contour model (ACM). Then, geometric features, contrast features, first and second order statistics were extracted from the inner and outer regions on the original and transformed candidates at multiple scales. Next, Wei et al. [87] selected features using the forward stepwise selection, and classified candidates using a statistical method based on the Mahalanobis distance measure.

Schilham et al. [69] segmented candidate nodules with a multi-scale approach based on ray-casting to find nodule boundaries. They described the candidates using statistics extracted from their inner and band regions in the output of the convolution of the image with Gaussian filters. They also used the candidate position and the Hessian matrix for nodule description. The authors proposed a two-stage classification approach using an approximate k-nearest neighbor algorithm, first to reduce the number of candidates before nodule segmentation and, finally, to estimate the probability of being a nodule for each candidate. The authors found that the two-stage classification boosted the method performance, whereas candidate segmentation did not improve their results.

Hardie et al. [30] proposed an adaptive distance-based thresholding approach to nodule segmentation. Then, a set of geometric, gradient and intensity features were extracted from the local contrast enhanced, normalized, nodule segmentation and WCMI versions of each candidate region. Finally, they used a sequential forward selection method to find a proper subset of features and Fisher Linear Discriminant for classification.

Lo et al. [47] used a three-layer CNN to classify candidates, whereas Lin et al. [43] used a two-layer CNN. CNNs proposed in [43, 47] are shallow when compared with CNNs used nowadays due to hardware limitations in terms of computational time in which these works were conducted. Coppini et al. [19] used a multi-scale ANN architecture. At each scale, Gabor filters at various orientation and a LoG filter were applied to the candidate

region. The filters outputs were used to train an ANN to learn the subtlety degree. Then, an optimal linear combiner merged the outputs of the ANNs at each scales to determine the nodule likelihood.

Recently, Wang et al. [84] showed that fusing handcrafted features with features extracted from a deep model can obtain good results for lung nodule classification. They extracted a set of geometric, intensity, contrast and first and second order filter derived features, then extracted deep features using the AlexNet model [40]. They used PCA to reduce the dimensionality of the fused descriptor, and then they used a cost-sensitive random forest for classification. The authors used data augmentation in the positive samples to balance the amount of positive and negative samples.

Chapter 3

Materials

In this chapter, we present the datasets used in the experiments to validate our results, as well as the hardware and software resources employed in the implementation and execution of the codes.

3.1 Datasets

We used three datasets to evaluate the effectiveness of our approach. Two datasets that contain chest X-rays with/without nodules were used to train and evaluate the detection and classification stages. One dataset with landmarks on the border of lung fields was used to train and evaluate the segmentation stage.

In the following subsections, we briefly describe each dataset, including details about its purpose, data, and annotation process.

3.1.1 Japanese Society of Radiological Technology (JSRT) Dataset

This publicly available dataset [74], used in the evaluation of several approaches in the literature, was created by the Japanese Society of Radiological Technology (JSRT) for research on digital medical imaging, training, education, and quality assurance. The parameters used in the imaging acquisition were defined aiming to be appropriate for lung nodule detection. The dataset is composed of 247 posterior-anterior chest images collected from 14 medical centers. Each image is represented by a 2048×2048 matrix, with a pixel spacing of 0.175 mm, and 12-bit gray levels.

The dataset was annotated by three experienced chest radiologists. There are 154 images with single nodules that may be benign or malignant, and 93 images without nodules. Malignant nodules can be associated to certain conditions such as lung and metastatic cancers, whereas benign nodules can be associated to conditions such as tuberculosis and hematoma.

The annotations include information about patients: age and sex; and nodules: size, anatomical location, x and y coordinates, degree of subtlety, malignancy condition, and final diagnosis [74]. Occurrence of nodules were confirmed by CT examination. Malig-

nant condition was confirmed by histologic and cytologic examination. Benign condition was confirmed by histology examination, definitive isolation of a pathogenic organism, shrinkage and disappearance with the use of antibiotics, or no change observed during a follow-up period of 2 years [74].

The dataset contains a wide range of nodules. Table 3.1 reports the distribution of nodules according to their size and subtlety level. In particular, the degree of subtlety is suitable for showing the variability of the nodules on the dataset, and how difficult is to recognize them.

Table 3.1: Number of nodules by subtlety level and effective diameter.

	Effective diameter						
Degree of subtlety	0–10 mm	11–15 mm	16–20 mm	21–25 mm	26–30 mm	31–60 mm	Total
Extremely Subtle	10	8	2	2	3	0	25
Very Subtle	5	11	7	4	2	0	29
Subtle	10	18	11	6	4	1	50
Relatively Obvious	5	12	14	2	4	1	38
Obvious	1	3	2	1	3	2	12
Total	31	52	36	15	16	4	154

Figure 3.1 illustrates samples selected randomly by each subtlety level. We excluded nodules located on opaque areas, that overlap with tissues outside of the lung such as the hearth, the mediastum, and the diaphragmatic region.

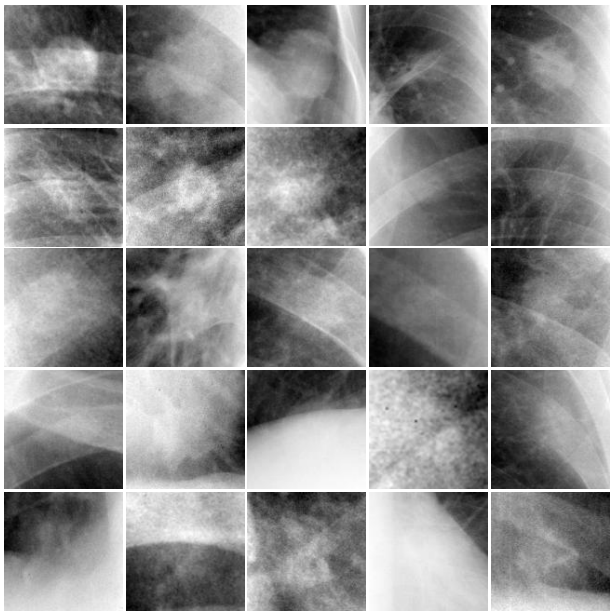


Figure 3.1: Patches extracted from the JSRT dataset that correspond to nodues arranged by subtlety level. Each row shows the patches that correspond to very obvious, obvious, subtle, very subtle and extremely subtle nodules.

3.1.2 Segmentation on Chest Radiographs (SCR) Dataset

This publicly available dataset [83] was created by the Image Sciences Institute of the University Medical Center Utrecht, whose main purpose is to evaluate methods for segmentation of lungs, heart and clavicles in chest X-rays. The dataset is composed of a set of landmarks extracted from the listed anatomical structures of the images on JSRT dataset.

The dataset was annotated by two observers with the supervision of an experienced radiologist. First, observers marked points on the boundaries of the anatomical structures. Several segmentation methods usually require a fixed number of landmarks. To obtain a fixed number of landmarks, observers annotated additional distinguishable points that indicate anatomical or other characteristics. These additional points are considered as landmarks. Then, a fixed number of points is equidistantly sampled using initial boundary points. These points are also considered as landmarks. Figure 3.2 shows the initial points, as well as the final landmarks.

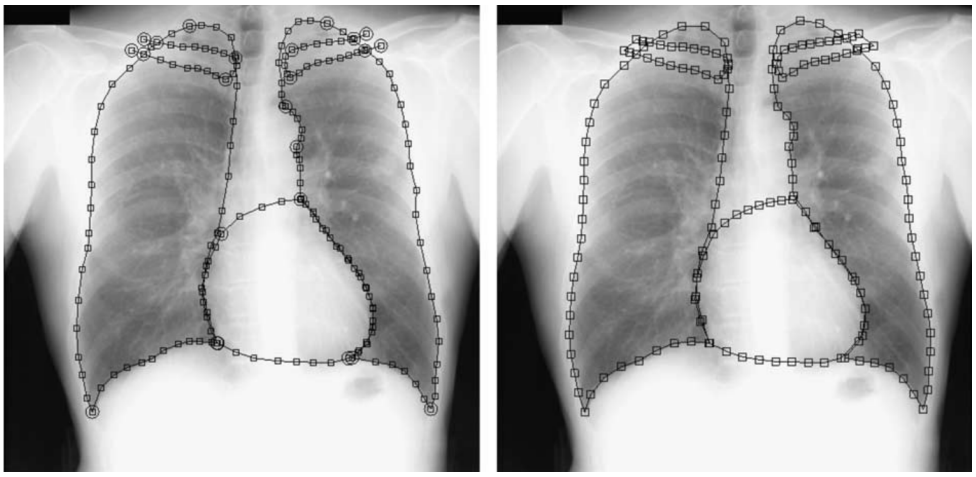


Figure 3.2: Boundary annotations on a sample of the JSRT dataset. On the left side: the initial points marked by an observer to delineate the anatomical structures on the image. Distinguishable points are marked with circles. On the right side: the interpolated landmarks along the boundary for use of segmentation methods [83].

The first observer was a medical student, whereas the second one was a computer science student specializing in medical image analysis. Both observers were supervised by the experienced radiologist until their segmentation outputs were considered reliable. The landmarks annotated by the first observed were used as ground truth, and the landmarks of the second observer were used to compare the human performance with several automated segmentation methods [83].

3.1.3 Lung Image Database Consortium and Image Database Resource Initiative (LIDC/IDRI) Dataset

This publicly available dataset [5] was created by the Lung Image Database Consortium and Image Database Resource Initiative (LIDC/IDRI), a group of public and private

medical institutions, with the purpose of development of CAD methods for detection, classification and quantitative assessment of lung nodules.

The dataset contains a set of 1018 cases with series of CT scans, and a subset of 290 cases with X-rays associated to these CT scans. This dataset was collected from eight medical institutions. Each case of dataset is associated to an XML file with the annotations performed by four experienced radiologists. The annotations contain the coordinates of suspicious lesions categorized as non-nodules, nodules with diameter $<$ 3mm, and nodules with diameter $>=$ 3 mm. In addition, each radiologist independently assessed characteristics of the nodule, such as subtlety and level of malignancy [5].

The confidence on the annotated regions depends on the consensus of radiologists. Nodules annotated by few radiologists have lower confidence than those annotated by all radiologists. Nodules size also determines if a nodule is important for certain condition. For instance, nodules with diameter lower than 3 mm are not considered relevant for cancer screening. Based on a competition hosted in an international conference that uses a subset of CT scans of this dataset [72], we filtered the annotation by selecting nodules with diameter greater or equal than 3 mm accepted by at least 3 radiologists. A single nodule could have several center coordinates annotated by different radiologists. We set the center of a nodule as the medoid of its corresponding annotations.

We evaluated heuristics to select more significant nodules. First, we include nodules annotated by two radiologists when the sum of the confidence level is greater or equal than 7. Then, we excluded nodules that have an average subtlety level lower than 2. We decided to use these heuristics because their positive impact on the performance on preliminary experiments. In addition we excluded X-rays that are not on in posterior-anterior orientation. The final dataset is composed of 265 chest images, where 138 images do not have nodules, 98 images have a single nodule, and 29 images have multiple nodules. Table 3.2 shows the distribution of nodules according the approximated size and average subtlety level.

Table 3.2: Number of nodules by subtlety level and effective diameter.

	Effective diameter						
Degree of subtlety	0–10 mm	11–15 mm	16–20 mm	21–25 mm	26–30 mm	31–60 mm	Total
Very Subtle	27	18	5	3	0	0	53
Subtle	14	11	11	3	1	0	40
Relatively Obvious	13	14	16	10	4	0	57
Obvious	0	1	3	6	2	2	14
Total	54	44	35	22	7	2	164

Figure 3.3 illustrates samples selected randomly by each subtlety level. As well as in JSRT dataset, We excluded nodules located on opaque areas, that overlap with tissues outside of the lung.

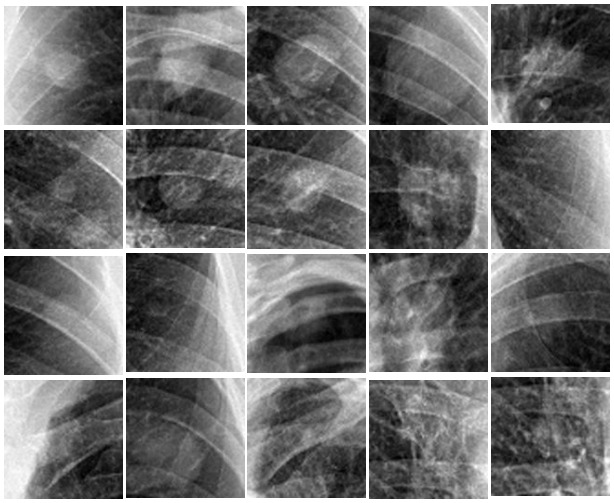


Figure 3.3: Patches extracted from CXRs of the LIDC-IDRI dataset arranged by subtlety level. Each row shows the patches that correspond to very obvious, obvious, subtle, and very subtle nodules.

3.2 Hardware and Software Platform

We used convolutional neural networks (CNNs) in several experiments conducted in this dissertation. The major computational burden on working with deep learning models, such as CNNs, is associated with tensor operations. Graphics processing units (GPUs) are suitable for efficiently performing these operations, contributing significantly to the advance of deep learning [70].

GPUs were crucial for this research by speeding up the execution of the codes. We conducted all experiments on Linux machines with Intel Core-i7 CPUs, 32 GB of RAM, and GTX TITAN Black and GeForce GTX 660 GPUs.

Our approaches were implemented in Python programming language. The most important libraries used were OpenCV [1], Scikit-Image [82] and Menpo [4] for preprocessing, detection and segmentation methods; and Theano [79] , Keras [15] and Scikit-Learn [61] for the classification and evaluation.

Chapter 4

Proposed Method

The success of convolutional neural networks (CNNs) on several fields has motivated their use on medical imaging applications. Recent works [8] have explored the application of CNNs to CAD systems. Recent research on CAD systems for detecting pulmonary nodules on CXR and CT images have used CNNs on lung segmentation, candidate nodule detection, and classification.

Related work has reported high sensitivity values, however, with too many false positives per image. Although several approaches have been explored in classification stage for false positive reduction (Section 2.1.4), there is a performance gap that should be reduced to make the CAD systems more useful for clinical usage. We focus on the use of CNNs in the classification task to reduce such performance gap.

Given a set of CXR images, our method returns an list of locations ordered by their probability of being a lung nodule. The goal of our method is to achieve a high sensitivity on detecting lung nodules by selecting a few suspicious locations. Figure 4.1 illustrates an overview of our method, which is divided into three stages: segmentation (Section 4.1), candidate detection (Section 4.2), and classification (Section 4.3).

We segment the input images through Active Appearance Models (AAMs). This method is trained with a set of CXR images and a set of landmarks located in the border of the lungs, in order to model their shape and appearance. Given a new CXR, AAMs are capable to estimate the landmark points corresponding to the lung border. We use the landmarks to generate a mask of the lung area.

We then detect the candidate locations by applying a Sliding Band (SB) filter to the input images. This filter assigns a convergence index (CI) to all pixels. For SB filters, the convergence index measures the degree in which the gradients on a band region are oriented to the pixel of interest. Then, we select the local maxima of the convergence index image as candidate locations. Next, we exclude candidates outside the lungs using the mask produced in the segmentation stage.

We use a CNN to estimate the probability of a nodule being in suspicious locations found on the detection stage. Then, radiologists can set the amount of candidates (more likely of being nodules) to be shown by the system. We evaluated empirically several CNN architectures, regularization schemes, and transformations used for data augmentation to find the best configuration for classifying lung nodules in CXR images.

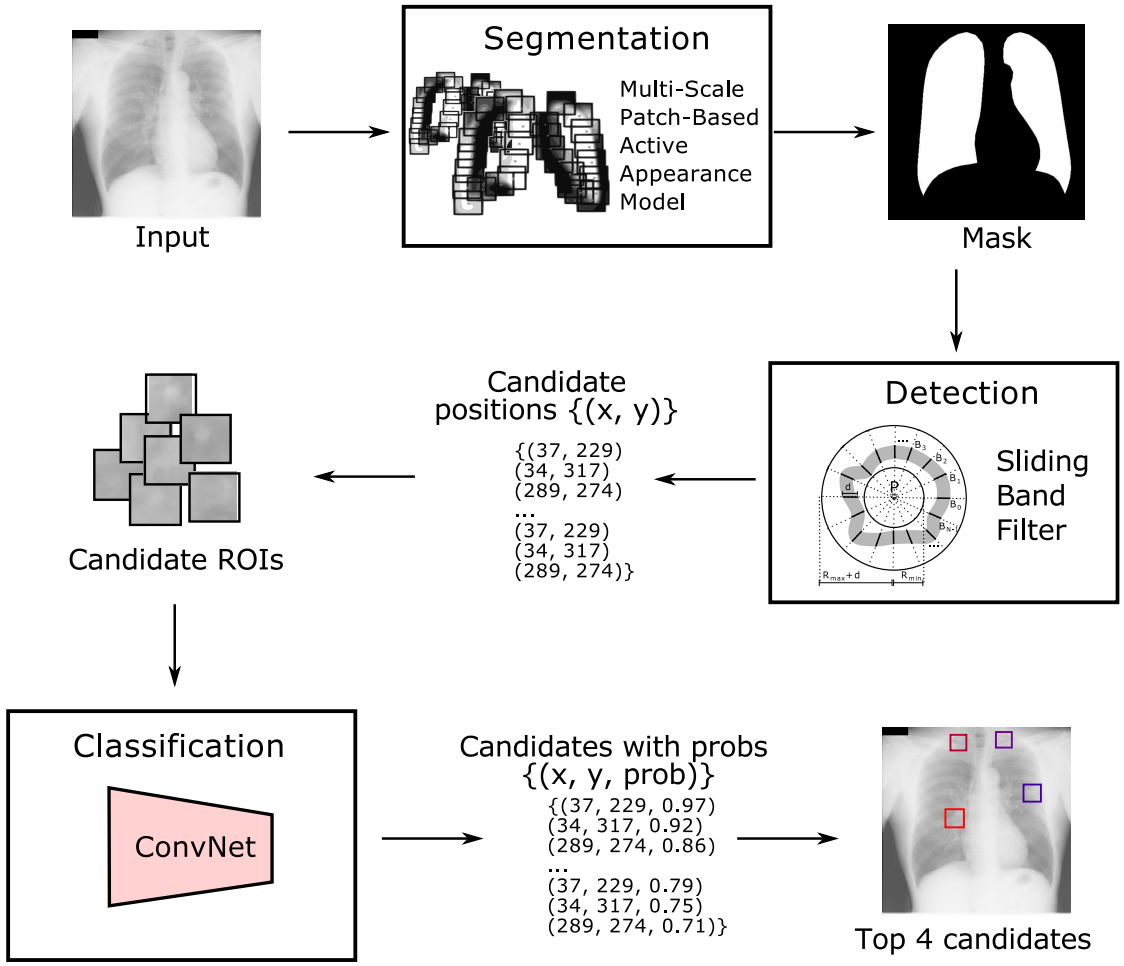


Figure 4.1: Overview of the main steps of the proposed methodology. First, (a) we segment the lung area using patch-based AAM. Then, (b) a SB detector finds candidate locations. Candidates outside the lungs are excluded using the masks generated in the segmentation stage. Finally, (c) we estimate the probability of nodules being candidates using a CNN.

4.1 Segmentation

In this section, we describe the method proposed for the lung segmentation. Figure 4.2 depicts an overview of the segmentation stage. We downsampled the images to 512×512 pixels before the segmentation process to reduce the computational time.

Given a training set with n CXR images and a set of landmarks $\boldsymbol{L} = \{\boldsymbol{s}_i\} \in \mathbb{R}^{2v \times n}$ with $\boldsymbol{s}_i = (s_1, s_2, \ldots, s_v) \in \mathbb{R}^{2v}$ denoting the boundaries of the lung region for the image I_i. We follow the method explained in Section 2.1.2 for normalizing $\boldsymbol{L}$ using Procrustes Analysis (PA).

Then, we build the shape and appearance models using Patch-based AAM. Given an unseen image I, we initialize the landmark points of I to the mean landmarks. Using the parameters of the appearance and shape models, we fit the landmark points to lung boundaries in I. Finally, we generate the segmentation mask S by joining the landmarks with straight lines.

Patch-based AAM works by sampling the patches centered around the landmarks

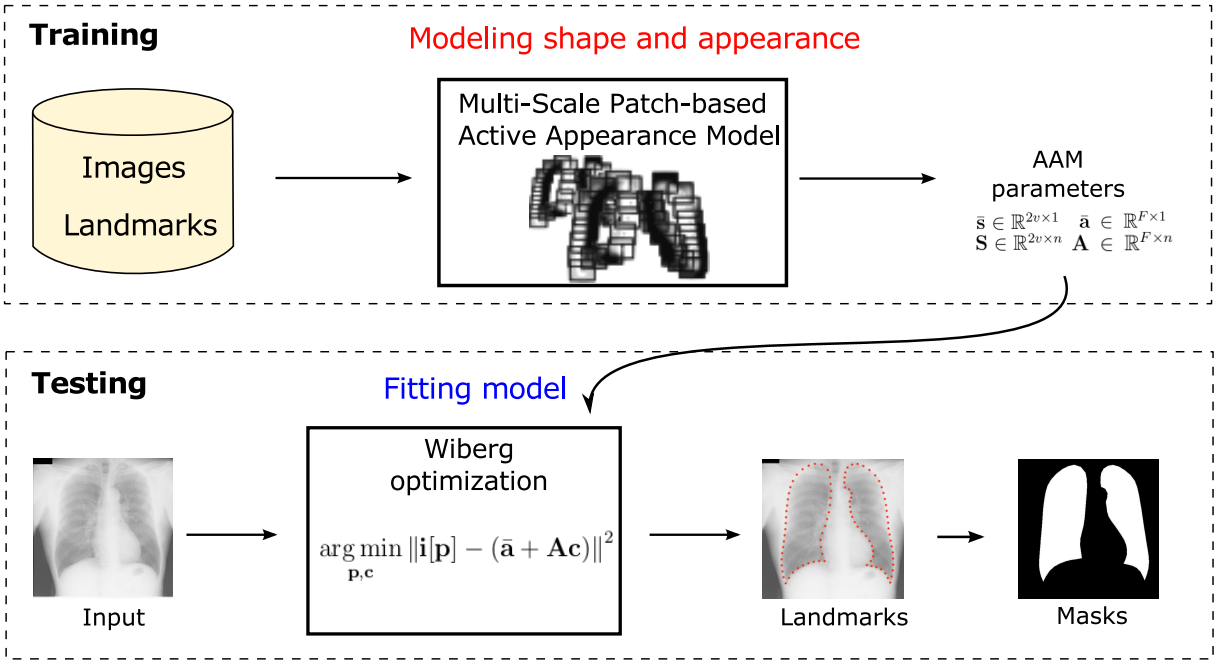


Figure 4.2: Overview of the segmentation stage.

points. The warp function $\mathcal{W}(\boldsymbol{x}, \boldsymbol{p})$ maps the pixel positions $\boldsymbol{x}$ from the patches on the reference shape $\bar{\mathbf{s}}$ to the patches on the estimated shape $\mathbf{s}$.

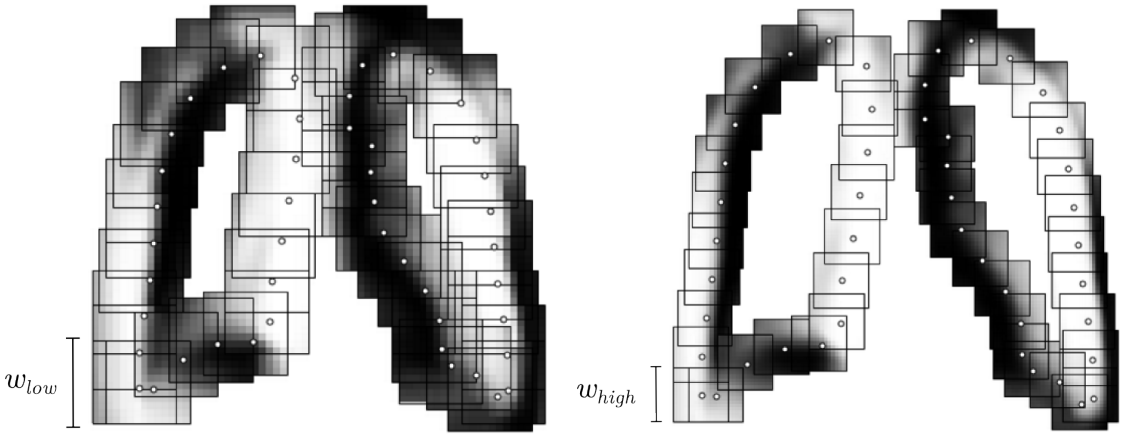
There are two main differences between the standard AAM, described in Section 2.1.2, and the Patch-based AAM we used for lung segmentation. First, we model the appearance on a representation space. We applied the dense Scale-Invariant Feature Transform (SIFT) [48] descriptor to represent the patches more distinctively. Then, we modeled the appearance by selecting the most discriminative components $pc_{appearance}$ obtained from the representations.

Second, we model the lung boundary variation in two resolutions. Thus, we use the low resolution model to provide a robust initial segmentation. Then, we refine the initial segmentation with the high resolution model. For the low resolution model, we downscale the input image by a factor of s_{low}, and extract patches of width w_{low}. For the high resolution model, we maintain the resolution of the images and extract patches of width w_{high}. Figure 4.3 shows a representation of the Patch-based AAM.

We find the shape and appearance parameters using the Wiberg Inverse-Composition optimization method [80]. Figure 4.4 shows how the shape changes while searching the best shape and appearance parameters. Once we find the best shape parameters $\boldsymbol{p}^*$, we project them onto the image space to generate the final landmarks for the image I.

4.2 Detection

In this section, we describe the methods used for candidate nodule detection. We evaluate several detectors and group them into two sets: detectors based on scale-space representation and detectors based on convergence index. Before using a candidate detector, we processed the images from JSRT and LIDC-IDRI datasets. First, we smoothed the images



(a) Low resolution model (b) High resolution model

Figure 4.3: Visualization of Patch-based AAM (for low and high resolution models) fitted on the sample `JPCLN001.IMG` from the JSRT dataset.

(a) Mean shape (b) Iteration 4 (c) Iteration 9 (d) Iteration 13 (e) Iteration 18

Figure 4.4: Shape parameters projected onto the image space throughout iterations of the Wiberg method. (a) mean shape initialization, (b-e) shape on the 4, 9, 13 and 18 iterations.

using a Gaussian filter with an 11×11 kernel and resampled the images to 512×512 to decrease noise and reduce computational time. We describe the evaluated detectors in the following subsections.

4.2.1 Detectors based on Scale-Space Representation

Nodules are difficult to recognize due not only to low contrast between the lesion and background tissue, but also due to internal structures such as ribs and vessels. Low contrast makes difficult to find nodules with methods based on spatial differences. Inspired by the works [30, 69], we deal with the low contrast problem by normalizing the contrast using local contrast enhancement (LCE).

We perform LCE to normalize the contrast of candidates located in regions with different levels of contrast to improve the performance of the candidate detector. Normalization is performed by scaling the high-frequency component of I with a factor inversely proportional to the local standard deviation. LCE can be formally defined as follows:

$$I_{LCE} = (I - G * I)\frac{1}{[G * (I^2) - (G * I)^2]^{\frac{1}{2}}} \tag{4.1}$$

where I_{LCE} is the output of the local contrast enhancement, G is a Gaussian filter with

zero mean and σ_{LCE} standard deviation, and $*$ is the convolution operator.

We evaluated Laplacian of Gaussian (LoG), Difference of Gaussians (DoG) and Determinant of Hessian (DoH) blob-like detectors. We follow the same steps of all detectors. Given an image I, we preprocess the image using LCE. Then, we find candidate nodules using a detector. Methods based on scale-space representation detect candidate nodule by enhancing blob-like features at different levels of scale of each image I, and selecting the local maxima greater than a threshold t over all scales in the enhanced image [46]. Finally, we prune the candidates smaller than an adjacent candidate that has an overlapping area greater than 50%.

4.2.2 Detectors based on Convergence Index

On the other hand, convergence index (CI) based filters address the low contrast problem considering only the orientation of the gradients. We evaluated two CI-based filters: first, the Weighted Multi-Scale Convergence Index (WMCI) filter which is a multi-scale convergence index filter that weights the gradient contributions according to its distance of the center [30]; and the Sliding Band (SB) filter which measures the convergence degree of the gradients considering a support region with non-uniform shape.

Weighted Multi-Scale Convergence Index Filter

The WMCI filter differs from the formulation of CI filter that we described in Section 2.1.3 because it considers two observations: (i) CI filters averages the convergence degree of the gradients on the support regions uniformly. Hardie et al. [30] proposed to weight the convergence degree contributions according to the application. For candidate nodule detection, the authors proposed a weight function that decrease the contribution step-wise while the gradient moves away from the interest point; (ii) since nodule sizes are variable, it is not possible to use only 1 set of weights. Then, the authors proposed to use various weighted CI filters to cover nodule size variability. We refer the reader to reference [30] for a detailed explanation.

Sliding Band Filter

WCMI defines a support region as a circular region around a point of interest. However, the definition of support region can be simplified as a union of N half-lines radiating from the point of interest. For SB filter, we select the band with a fixed width d with the maximum convergence index for each half-line. Then, the support regions is defined as the union of these selected bands.

Figure 4.5 shows the support region of the SB filter in the pixel P at the position (x, y), where B_i represents the band with maximum convergence degree in the half-line i, $R_{\min}$ and $R_{\max}$ represents the lower and upper bound of the range in which we search the bands.

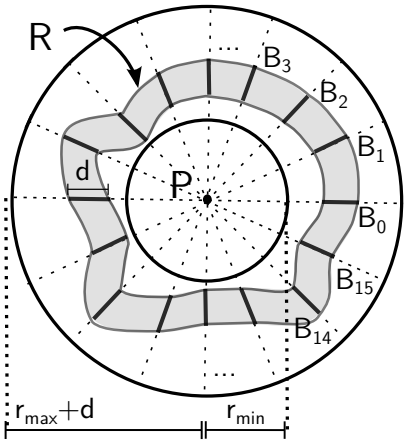


Figure 4.5: Visualization of the region of support of a SB filter with $N = 16$ bands on a nodule sample.

Given an image I, the output of the SB filter for a pixel $P \in I$ is defined formally as:

$$SB(x,y) = \frac{1}{N}\sum_{i=0}^{N-1}\left(\max_{R_{min}\leq n\leq R_{max}}\left(\frac{1}{d}\sum_{m=n}^{n+d}\cos\theta_{i,m}\right)\right) \tag{4.2}$$

where N is the number of half-lines radiating from P, and d is the width of the band.

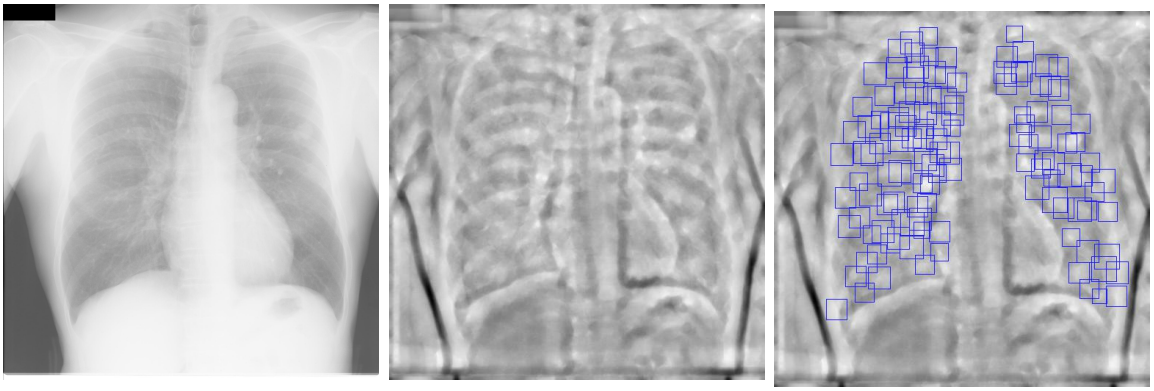
We choose SB filter because it has features that are suited for nodule detection: (i) the distance between bands and the point of interest can be vary in different half-lines that allows detecting nodules with irregular shapes; (ii) the position of the bands depends only on where the gradients have more convergence index that allows detecting nodules regardless of its size; (iii) SB filter calculates the convergence index for pixels in the region around the boundaries, but not the center of the nodules that may have a random degree of convergence [62].

Finally, we set as candidates the regions with a center pixel that is local maximum in its circular neighborhood with radius r_{sbf}. We obtain better results with SB filter 5.2 and use it as candidate detector in our methodology. The output of the detection stage is a set of candidate locations $\boldsymbol{C} = \{\boldsymbol{c_i}\}$ where $\boldsymbol{c_i} = \{(x_j, y_j)\} \in \mathbb{R}^{m_i \times 2}$ is the set of m_i candidates found on the image I_i. Figure 4.6 shows the candidate detection stage.

4.3 Classification

Our method performs candidate analysis directly on the pixel of images. Figure 4.7 shows an overview of the classification stage. For training phase, we extract the squared regions centered on the candidate locations found on detection stage to construct the training dataset. Then, we preprocessed and augmented the samples.

We search the parameters used for the transformations on data augmentation. Then, we search an optimal CNN for our problem by evaluating several architectures on the training set. For a new image, after performing segmentation and candidate detection, we use the best CNN we found on training phase to estimate the probabilities of being a nodule for the candidate locations in the new image. Finally, we remove candidates that



(a) Input X-ray (`JPCLN001.IMG`) (b) SB filter output (c) Non-maxima suppression with $r_{sbf} = 7$px

Figure 4.6: Visualization of the candidate detection stage. (a) Input x-ray (`JPCLN001.IMG`); (b) SB filter output; (c) candidates detected selecting the local maxima with $r_{sbf} = 7$px on the neighborhood region.

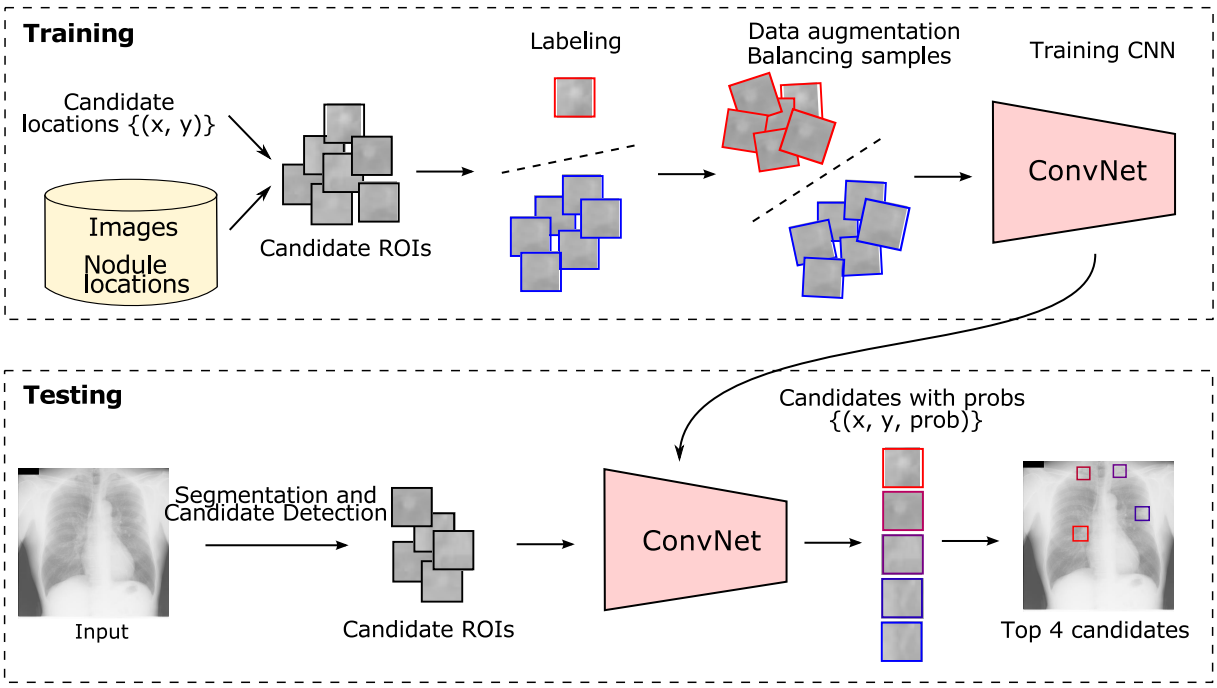


Figure 4.7: Overview of the classification stage

may belong to the same true nodule through an adjacent rejection rule.

4.3.1 Data Preparation and Augmentation

Given a set of candidate locations $\boldsymbol{C}$, we build a dataset $\boldsymbol{R} = \{\boldsymbol{r_i}\}$, where $\boldsymbol{r_i} \in \mathbb{R}^{m_i \times w \times w}$ is a set of m_i squared patches with side w around each location on the set of candidates $\boldsymbol{c_i}$ of the image I_i. Although there is no unique labeling scheme for candidate nodules, we choose the scheme proposed by Hardie et al. [30], which consists in labeling every candidate located within 25 mm a truth nodule as true.

We applied a z-score normalization to the pixel intensities over the dataset $\boldsymbol{R}$ to have zero mean and unit variance.

The candidate detection stage produces approximately several negative samples for each positive one. To balance the dataset and prevent overfitting, we augment the data available by increasing the amount of positive samples and applying the following transformations to all candidates, including negative samples: (i) horizontal/vertical random shift with a factor of each dimension of the region of interest, selected over a uniform distribution between 0 and t_{max}; (ii) random zoom with a scale factor s selected over a logarithmic distribution of 1.0 and s_{max}; (iii) horizontal flip: yes or no with a probability of 0.5; (iv) random rotation with an angle θ between θ_l and θ_r degrees, and (v) intensity shift v selected over a uniform distribution between $i * (-\sigma)$ and $i * \sigma$ where i is a scale factor and σ is the standard deviation of all intensities on $\boldsymbol{R}$.

We search heuristically the optimal parameters t_{max}, s_{max}, θ_l, θ_r and i for the transformations in data augmentation. First, we search the optimal values for each parameter empirically and independently. Then, we combine transformation including iteratively the transformations with greater impact at first.

4.3.2 Architecture Design

Motived by the work [28], we consider a family of CNNs with alternating convolutional and max-pooling, followed by fully connected layers. We can describe each instance of this family as *ConvNet(c, k, f)*, where c is the number of convolutional layers separated by c max-pooling layers. The number of filters in each convolutional layer increases linearly with a factor of k. For instance, the i-th convolutional layer is supposed to have ik filters. Finally, there are f fully connected layers followed by the output layer. Figure 4.8 depicts an arbitrary convolutional neural network.

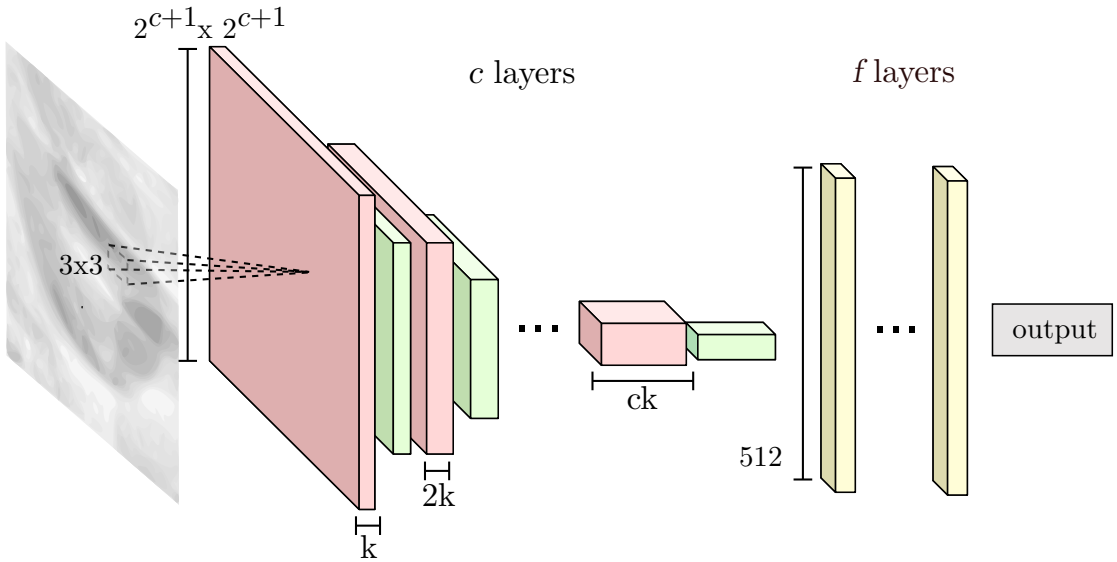


Figure 4.8: Representation of the architecture *ConvNet(c, k, f)*. The network is fed with images with size $2^{c+1} \times 2^{c+1}$. Composed by c alternating convolutional (in red), and max-pooling (in green) layers, followed by f fully connected layers (in yellow)

Given a network *ConvNet(c, k, f)*, we rescale the inputs to $w_{convnet} = 2^{(c+1)}$. Convolutional layers have a receptive field of 3 with stride 1. Max-pooling is performed over

windows of 2×2, with stride 2. We use leaky Rectifier Linear Units (ReLU) with alpha of 0.333 for all convolutional layers. We apply dropout with a probability of d_i for the i-th convolutional layer, and 0.5 for fully connected layers. The output of the last layer is connected to a softmax layer, which produces a distribution over the two classes.

We search the optimal network heuristically. We use the network *ConvNet(5, 64, 1)*, similar to the network used in the work [28] for classifying digits of the MNIST database, as a start-point. Then, for every parameter we search the optimal value keeping the rest of parameters fixed. For instance: we search the optimal c^* keeping fixed $k = 64$ and $f = 1$. Then, we search the optimal k^* keeping fixed $c = c^*$ and $f = 1$ and so on.

Dropout on convolutional layer has proven to have a positive impact of CNN performance [28, 92]. We explore two schemes of data augmentation: We set a fixed dropout probability p_c for all convolutional layers. Moreover, we can set a linearly increasing dropout probability of $d_i = i * m$, where d_i is the dropout probability for the i-th convolutional layer, and m is the reason in which dropout increases. Then, the dropout probability is small for lower convolutional layers which are fairly immune to overfitting [28]. We set a linear increasing dropout on the last 4 convolutional layers such that $d_i = \max(0, 4 - c + i * m)$ where c is the number convolutional layers.

4.3.3 Learning

For the learning stage, we use the Stochastic Gradient Descent (SGD) with Nesterov momentum. We initialized the weights on each layer with the orthogonal random matrix initialization procedure proposed by Saxe et al. [68]. We train until the area under the free-response receiver operating characteristic (FROC) curve[1] stops improving. We experimented with weight decay regularization, however, it did not improve our results. Therefore, the final model does not use weight decay regularization.

Motivated by the work proposed by Yan et al. [93] we propose the Algorithm 1, we balance the amount of positives and negatives samples for each batch used to feed the CNN in the training stage. For positives, we select half of the batch size samples randomly. For negatives, we select half of the batch size samples iteratively. Positive and negative samples are perturbed with the transformations described in Section 4.3.1 before they feed the CNN on each SGD iteration.

[1] In this dissertation, we will refer to the area under the free-response receiver operating characteristic (FROC) curve as AUC.

Algorithm 1 Stochastic Gradient Descent with balanced samples.

procedure SGD-BalancedSamples
require: Network parameters θ
require: Positive samples P
require: Negative samples N
    while AUC does not stop improving **do**
        Select the first half of the minibatch from N iteratively.
        Select the second half of the minibatch from P randomly.
        Apply the data augmentation transformations to all samples of the minibatch.
        Compute gradient estimate with the minibatch.
        Update network parameters θ.
    end while
end procedure

4.3.4 Reducing Adjacent Candidates

We select the candidate location as the pixels with maximum convergence index on its radial neighborhood of 7 pixels as we described in Section 4.2.2. Therefore, several candidates may be related to the same true nodule when the nodule radius is greater than the neighborhood radius.

We use the adjacency rejection rule proposed by Hardie et al. [30]. This procedure reduces the number of final detections by preserving the candidates that have the maximum detection probability on their radial neighborhood of 22 mm.

Chapter 5

Experimental Results

In this chapter, we present the empirical results that we obtained by evaluating the methods developed for lung field segmentation, candidate nodule detection and candidate nodule classification stages. First, we report our results with path-based AAM, and we compare them to methods available in the literature. Then, we present the results achieved with our approaches to candidate detection.

Next, we present the results for candidate classification. We evaluate several transformations individually and composed for data augmentation. Then, we compare two dropout regularization schemes on convolutional layers. Moreover, we show the impact of the parameters considered to design our CNN architecture. Finally, we present a visual analysis and results obtained with our best architecture and compare its performance against competing approaches.

5.1 Segmentation

We aim to demonstrate that patch-based AMM is capable of achieving competitive results compared to state-of-the art methods and, therefore, it does not have a negative effect on the later stages of our pipeline.

5.1.1 Evaluation

We evaluate our segmentation method following the protocol developed by Van Ginneken et al. [83] in order to allow the comparison of our results to the state-of-the-art. We divide the images from the JSRT dataset and their corresponding masks from the SCR dataset into two folds. One fold contains 124 images corresponding to odd numbered cases, whereas the other fold contains 123 images corresponding to even numbered cases. We use the images in one fold to train a model that is used to generate the segmentation of the images in the other fold, and vice-versa. Once we have the segmentation results for all images in the dataset, we use a metric to assess the quality of our segmentation output.

We compare our segmentation method to approaches of the literature using the Jaccard similarity coefficient, denoted Ω. This metric quantifies the agreement between the pixels

in the estimated segmentation mask (S) and the ground truth (GT).

$$\Omega = \frac{|S \cap GT|}{|S \cup GT|} = \frac{|TP|}{|FP| + |TP| + |FN|} \tag{5.1}$$

where TP denotes the pixels inside the lung that are correctly classified, FP denotes the pixels in the background incorrectly classified as lung, FN denotes the pixels in the lung incorrectly classified as background. Similarly to [83], we quantify the performance of our method by averaging the Jaccard coefficients calculated on the left and right lung segmentation results.

For the lung nodule detection pipeline, we divide the JSRT dataset into two subsets: a subset that includes images with a nodule, and the other subset with images without any nodule. We use the first subset to train the segmentation model used to generate masks for the images from the LIDC-IDRI dataset and for the images that contain nodules from the JSRT dataset. These masks are used to exclude findings outside the lung in candidate detection stage.

5.1.2 Results

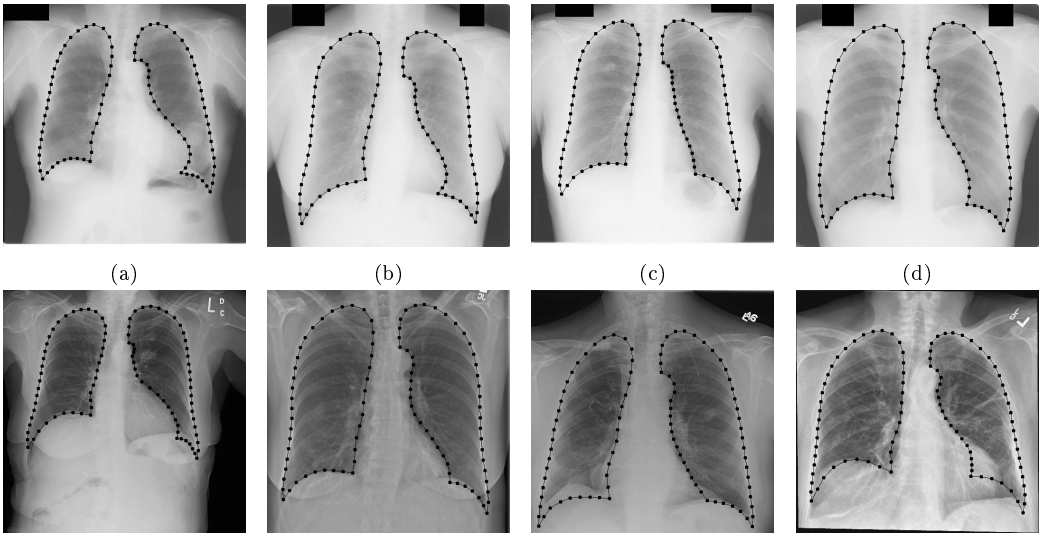
We used a patch-based AAM with $pc_{shape} = 20$ and $pc_{appearance} = 150$ principal components to model shape and appearance, respectively. Our patch-based AAM models the lung boundaries in two resolutions. We downsample the image with a scaling factor of $s_{low} = 0.5$ for the low resolution model and maintain the dimensions for the high resolution model. We extract patches with sides $w_{low} = 20$ pixels and $w_{high} = 12$ pixels for low and high resolution models, respectively.

Figure 5.1 shows estimated segmentation results of our model trained with the samples without nodules of JSRT. The first and second rows show the segmentation results estimated for samples of the subset with nodules on the JSRT dataset and for samples of the LIDC-IDRI dataset, respectively.

Segmentation results on JSRT samples are more accurate than on LIDC-IDRI samples because the model was trained with images of the same distribution. In contrast, segmentation results estimated on samples of the LIDC-IDRI are less accurate because lung boundaries have a different contrast and more shape variability than lung boundaries on JSRT dataset. For instance, the segmentation estimated for the sample (Figure 5.1(g)) fails including pixels of the diaphragm, and the segmentation estimated for the sample (Figure 5.1(h)) classifies pixels on the heart as part of the lung.

Table 5.1 shows that the patch-based AAM obtains results comparable to methods available in the literature. Ngo and Carneiro [57] proposed an iterative method that obtains impressive results, but it depends on the precision of the initial segmentation. Their method obtained an mean similarity coefficient of $\Omega = 0.985$, but it requires initializations with a mean similarity coefficient of $\Omega = 0.955$.

An advantage of the graph-cut based method proposed by Candemir et al. [12] and pixel classification method proposed by Van Ginneken et al. [83] is that they perform segmentation at pixel level. In contrast, the accuracy of patch-based AAM relies on the distribution of the landmarks on the lung boundary. Moreover, a drawback of patch-



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

(e) (f) (g) (h)

Figure 5.1: Segmentation results using AAM. The first and second rows show the landmarks and boundaries found on samples of the positive subset of JSRT dataset and LIDC-IDRI datasets, respectively.

Table 5.1: Segmentation results in terms of the Jaccard coefficient Ω. Methods are ranked according to their mean.

	Ω		
	$\mu \pm \sigma$	Min	Max
Level Set + Deep Belief Network [57]	0.985 ± 0.003	0.972	0.991
Non-Rigid Registration using Atlas [12]	0.954 ± 0.015	-	-
Human Observer [83]	0.946 ± 0.018	0.822	0.972
Pixel Classification Post-Processed [83]	0.945 ± 0.022	0.823	0.972
Proposed method (Patch-based AAM with DSIFT)	0.935 ± 0.019	0.786	0.964
ASM tunned [83]	0.927 ± 0.032	0.745	0.964
Mean Shape	0.714 ± 0.075	0.461	0.889

based AAM is that the lung mask is created by joining landmarks with straight lines, therefore, ignoring the information contained in pixel intensities about the shape of the lung boundaries. Thus, a post-processing method that considers pixel intensities between landmarks may improve the performance of the method considerably. The inclusion of appearance information through dense SIFT descriptor improves the results of our model compared to the ASM approach [83] that describes landmark position using only gradient profiles.

5.2 Detection

We evaluate the detectors on scale-space representation and based on the convergence-index to determine the method that produces better set of candidates for the classification

stage.

5.2.1 Evaluation

We use the LIDC-IDRI dataset for evaluation. We selected the best method based on the area under the FROC curve (AUC) between the operating points with 0 and 100 false positives per image. The criterion that we use to consider a candidate as true positive verify if exists a ground truth nodule in its neighborhood, such that the distance between the center of the candidate and the center of a ground truth nodule is less than or equal to true nodule radius.

5.2.2 Results

Candidate detectors have a threshold parameter t that controls the sensitivity of the detectors. Smaller t values produce less false positive per image than larger t values. We choose the smallest t that does not reduce the AUC score of the candidate detector on the LIDC-IDRI dataset.

Figure 5.2 shows the FROC curves of detectors on the LIDC-IDRI dataset. Sliding band (SB) filter detector with $t = 0.5$ obtains better AUC score than DoG with $t = 1.0$, LoG with $t = 0.5$, DoH with $t = 0.05$, and WMCI with $t = 0.0005$. We choose the SB method for the detection stage. Finally, we used SB detector to generate candidates on the JSRT dataset to evaluate our system on the classification stage.

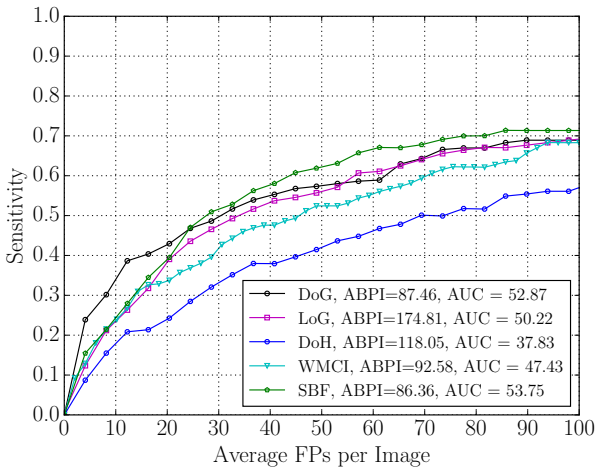


Figure 5.2: FROC curves for detectors evaluated using LIDC-IDRI dataset.

5.3 Classification

The main objective of the experiments on classification is to search an optimal configuration of our classification model for the lung nodule detection problem.

5.3.1 Evaluation

There are no standard protocols for evaluating nodule detection methods using either the JSRT dataset or the CXR set of the LIDC-IDRI dataset. Evaluation protocols used in the methods that report results using JSRT have many sources of variation that we will discuss later (Subsection 5.3.2). To the best of our knowledge, there are a few methods [60] that report results using the CXR subset of LIDC-IDRI. Moreover, LIDC-IDRI dataset is more used on the evaluation of lung nodule detection methods in low-dose CT images.

We search for parameters for learning method, data augmentation, dropout, and architecture design completely on LIDC-IDRI dataset using 5-fold cross-validation. For evaluation, we perform 5-fold cross-validation on the entire JSRT dataset using the best parameters. We excluded opaque cases that correspond to the nodules that overlap with anatomical structures outside the lung.

Several competing methods [19, 30, 59, 69, 84] report results with a cross-validation strategy using only JSRT dataset. Therefore, we evaluate our CNN, configured with the best parameters found on the LIDC-IDRI dataset, with 5-fold cross-validation.

5.3.2 Results

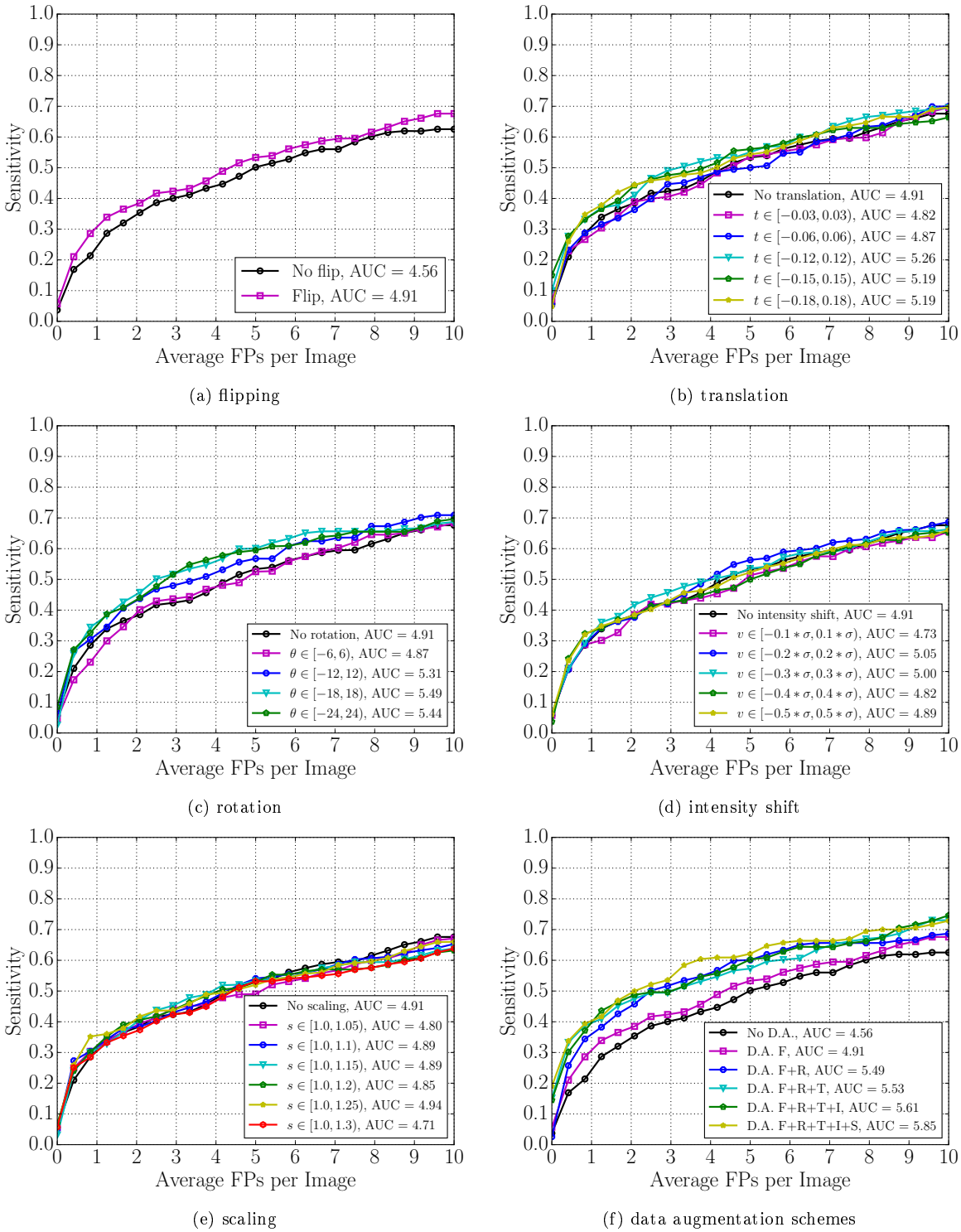
We use a *ConvNet(5, 60, 1)*, similar to the used by [28] for digit classification MNIST dataset, as a baseline architecture. We create the training dataset $\boldsymbol{R}$ following the procedure that we described in Section 4.3.1. We determined the best patch side $w = 32$ for candidate regions empirically on the LIDC-IDRI dataset. Then, we evaluate the impact of data augmentation, dropout regularization on convolution layers, and architecture parameters.

Data Augmentation

We evaluated several transformations to increase the amount of images used for training our model following the procedure described in Section 4.3.1. First, we find the best parameters for each transformation independently. Then, we combine the transformation including first the transformation that has better impact.

Figure 5.3 shows the FROC curves of *ConvNet(5, 60, 1)* with transformations. Figure 5.3(a) shows that a horizontal flip improves the results significantly. We search for the parameters for the rest of transformations combined with horizontal flip. Figure 5.3(b) shows that the performance of our baseline CNN increases when we augment candidates by performing translation with small $t_{\max}$ to the patches, but decreases when translation can move the location of the candidate too far from the center. The baseline CNN performs better with $t_{\max} = 0.12$ when AUC stops improving. Figure 5.3(c) shows that our baseline model improves by augmenting samples with rotation until $\theta_{\max} = 18$. Then, increasing $\theta_{\max}$ does not vary the AUC substantially. Figure 5.3(d) shows that performance increases when we augment the samples with intensity variations. We set $v_{\min} = -0.2 * \sigma$ and $v_{\max} = 0.2 * \sigma$.

Figure 5.3 shows that the transformations are complementary. The baseline CNN performs better by combining all transformations than including each transformation



(a) flipping (b) translation

(c) rotation (d) intensity shift

(e) scaling (f) data augmentation schemes

Figure 5.3: We evaluate the impact of several transformations on the performance. We show in (a) that flip transform increases the AUC significantly. In (b)-(e), we search for the best parameters for translation, rotation, intensity shift and scaling transformations. In (f), we compare the performance of various data augmentation schemes based on including the transformation incrementally.

individually. We obtain and improvement of 1.29 in the AUC. Moreover, we see that translation and rotation increase the AUC more than all transformations individually.

However, Figure 5.3(f) shows that translation does not increase the AUC considerably when combined with flip and rotation, whereas scaling, which does not improve the model performance individually, but increases the AUC when combined with all transformations. Further experiments are performed to consider all transformations.

Dropout on Convolutions

In this section, we evaluate the impact on using dropout regularization after the convolutional blocks in the baseline CNN. We evaluate the performance of the baseline CNN using the dropout with constant inclusion probability p_c. Figure 5.4(a) shows FROC curves by varying p_c. We observe that independently on the p_c values that we evaluated, dropout improves the performance of the CNN. We have better results with $p_c = 0.25$, obtaining a gain of 0.14 on AUC compared with the same architecture without dropout after convolutional blocks.

Figure 5.4(b) shows the FROC curves using dropout with an increasing dropout probability in the last 4 convolutional blocks. We evaluated the baseline CNN varying the increasing reason m. This approach performs better with all m values that we evaluated compared to the same CNN without dropout after convolutional blocks. We obtain better results with $m = 0.125$, which assigns 0.0, 0.125, 0.25, 0.375 and 0.5 dropout probabilities to the five convolutional layers of the baseline CNN.

In contrast with assigning constant dropout probabilities, the latter approach assigns small dropout values to the first convolutional blocks. This observation suggests that first convolutional layers are less prone to overfitting for our problem setting. We observed that convergence is faster and the AUC values have lower variance in training when we set increasing dropout probabilities to convolutional layers. We use the latter approach in further experiments.

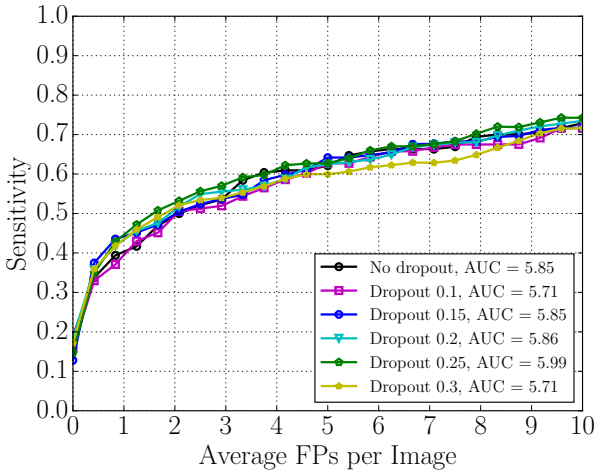
Architecture Design

In this section, we describe the experiments we conducted to find the proper architecture for our problem setting. We evaluate several architectures varying the number of convolution layers c, filters by convolution layers k, and fully connected layers f.

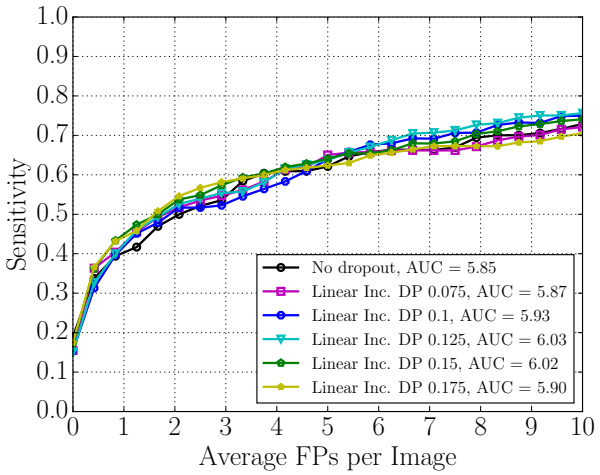
First, we evaluate architectures varying c with $k = 64$ and $f = 1$. Figure 5.5(a) shows that deeper CNN performs better until $c = 6$. We did not evaluate CNN with more than six convolutional blocks due to the GPU memory limitation of the hardware we used to conduct our experiments. However, the small performance gain we obtained with a CNN with 6 convolution layers compared to the CNN with 5 convolutional layers suggests that adding more convolutional layers will not improve the results drastically. We use $c^* = 6$ convolutional layers in further experiments.

In Figure 5.5(b), we report the FROC curves of the architectures that we explored varying k with $c = 6$ and $f = 1$. In contrast with the other architecture parameters, f impacts considerably on the amount of parameters of the network. *ConvNet(6,96,1)* has $\times 28$ parameters than *ConvNet(6,16,1)*. We obtain better results with $k = 32$ filters.

Figure 5.5(c) shows the results of the architectures obtained varying f with the best $c = 6$ and $k = 32$. We obtain better results with the initial $f = 1$. The heuristic we



(a) constant dropout



(b) linear increasing dropout

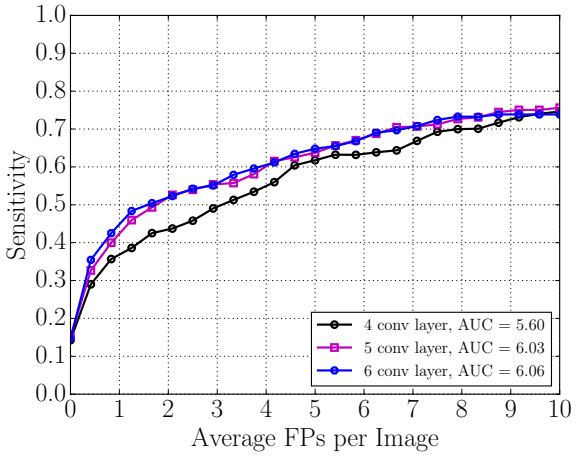
Figure 5.4: Comparison of constant dropout and linear increasing dropout on convolution layers.

use to search for a proper architecture requires less evaluations and it is simpler when compared to grid search, random search or optimization-based methods. However, the search is biased to the initial configuration.

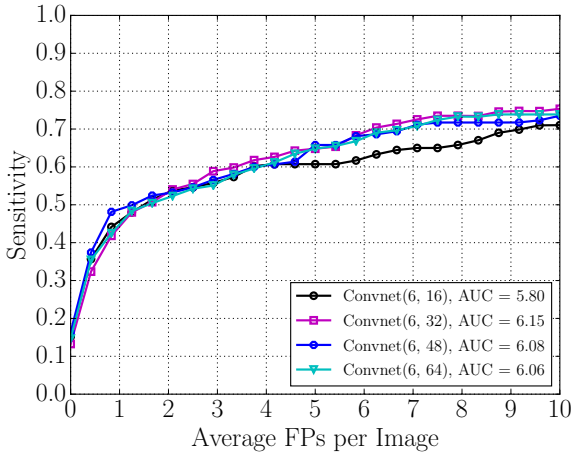
Result Comparison

Several CAD systems report their results in the public JSRT dataset. However, making a proper comparison is still difficult [13, 59] due to the variability of evaluation protocols used in related work. The main sources of variation are the usage of entire JSRT dataset or a subset by excluding opaque cases, different labeling criterion, different validation procedures, parameter optimization considering the test set such that generalization capacity is not properly measured, and results reported on different operating points. Table 5.2 summarizes the main sources of variations on competing methods.

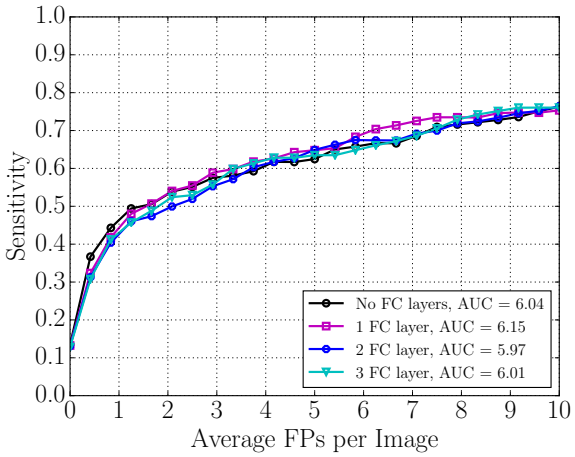
We address the dataset variability considering that the evaluated methods excluded the 14 opaque cases. Thus, we adjust the sensitivity obtained with the JSRT subset by 0.9091 for comparison with the methods that report results on the entire JSRT dataset.



(a) varying c with $k = 64$ and $f = 1$



(b) varying k with $c = 6$ and $f = 1$



(c) varying f with $c = 6$ and $k = 32$

Figure 5.5: FROC curves varying the parameters c, k and f.

Most of the state-of-the-art methods report their results using a cross-validation strategy. Thus, we use a 5-fold cross-validation. There are three common criteria to consider a candidate nodule as true positive: if the distance between the candidate and a ground-truth location is less than a *distance* threshold, if *centroid* of the candidate is inside a ground-truth region, and if the *overlap* between the candidate region and the ground-truth regions is greater than a threshold. We consider that comparison with a individual method should be conducted through the same labeling criterion. However, due to computational time limitation, we report our results using only the distance criterion. We calculated the effectiveness of our method on the operating points reported by competing methods.

Figure 5.6 shows the FROC curve of our CNN and the main operating points reported in state-of-the-art approaches. Table 5.3 compares the sensitivity of our CNN with the evaluated methods at the same FPPI values.

Wang et al. [84] obtained proper results by fusing handcrafted and deep features for classification, but using a lighter data augmentation scheme in contrast with our scheme. Chen and Suzuki [13] reported high sensitivity values. However, their system uses dual-energy radiographs for rib-suppression on training, which is not commonly available as

CAD System	Datasets	Protocol	Labeling Criteria	Parameter Optimization
Chen and Suzuki [13]	Private dataset, JSRT excluding 14 opaque cases	Train on private dataset, test on JSRT subset	Distance (25 mm)	Considers test results
Hardie et al. [30]	Private dataset, JSRT excluding 14 opaque cases	Train on a private dataset, test on JSRT, 10-fold cross-validation on JSRT subset	Distance (25 mm.)	Considers only training set
Schilham et al. [69]	JSRT dataset	5-fold cross-validation on JSRT	Overlap (> 0%)	Considers test results
Shiraishi et al. [75]	Private dataset augmented with JSRT images	Train and test on the merged dataset	Distance (22mm and 24mm)	Considers only training set
Coppini et al. [19]	Private dataset, JSRT subset of 140 samples	5 partitions of the JSRT subset	Centroid	Considers test results
Wei et al. [87]	JSRT dataset	Leave-one-out cross-validation	Not specified	Considers test results
Wang et al. [84]	JSRT excluding opaque cases	10-fold cross-validation	Distance (Not specified)	Not specified

Table 5.2: Main sources of variation of the evaluation protocol used in competing methods.

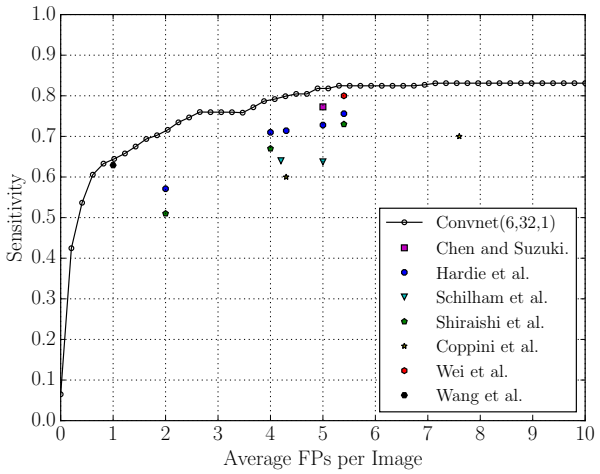


Figure 5.6: Comparison of the method performance on the JRST database. Sensitivity values are adjusted by considering opaque cases as missed.

conventional CXR. Wei et al. [87] obtained the best results among the methods that use hand-crafted features. Our proposal was evaluated on the 154 images with nodules of the JSRT dataset. We obtained an area under the-free-response receiver operating characteristic (FROC) curve of 7.51 considering 10 false positives per image (FPPI), and a sensitivity of 71.4% and 81.0% with 2 and 5 FPPI, respectively.

Table 5.3: CAD system performance comparison.

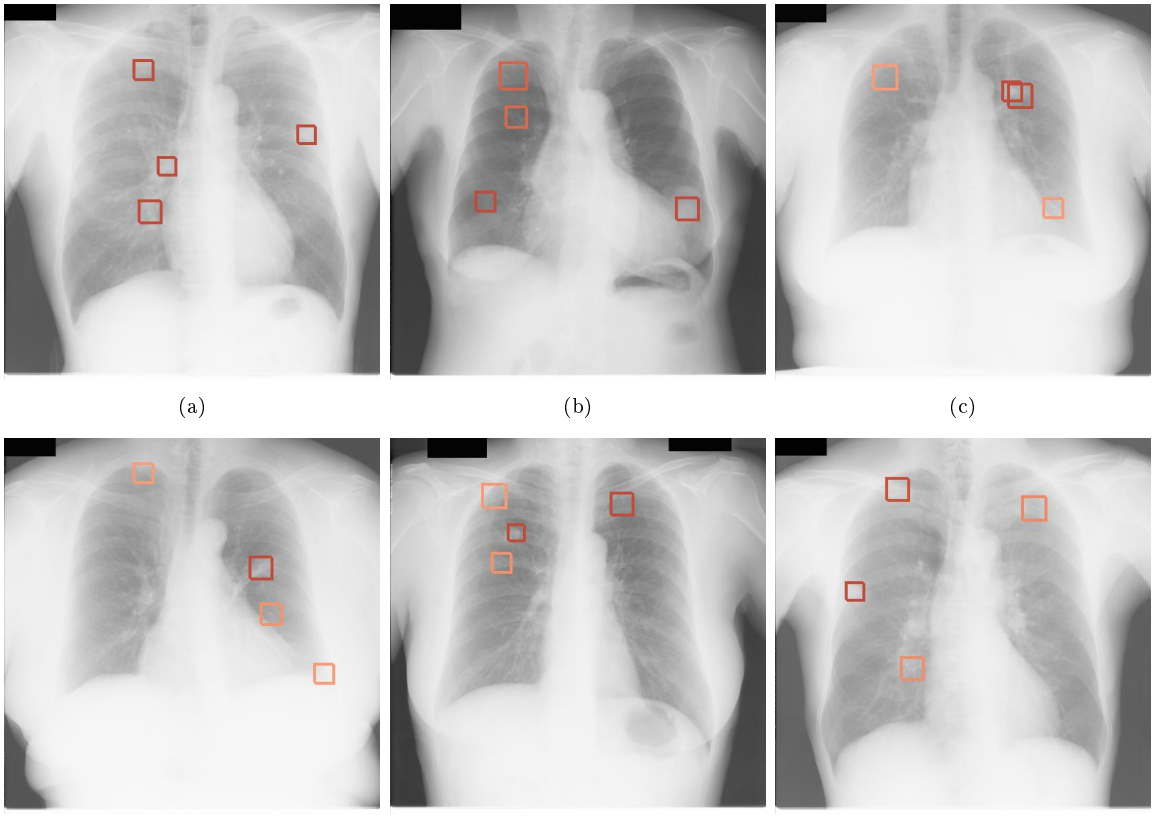
Average FPPI	Method	Reported Sensitivity (%)	CNN Sensitivity (%)
5.0	Chen and Suzuki [13] [a,c]	85.0	**90.0**
5.0	Hardie et al. [30] [a]	80.1	**90.0**
2.0	Schilham et al. [69]	51.0	**71.4**
4.0	Schilham et al. [69]	71.0	**79.2**
5.0	Shiraishi et al. [75] [a,b]	70.1	**90.0**
4.3	Coppini et al. [19]	60.0	**79.9**
5.4	Wei et al. [87]	80.0	**82.5**
1.19	Wang et al. [84] [a]	69.3	**72.1**

[a]Results reported in this row exclude opaque cases.
[b]Results based on 924 chest radiographs that include the JRST cases.
[c]A private database was used for training and the JSRT for testing.

Visual Analysis

Figure 5.7 shows the output of the system. We highlight the top 4 locations with high probability of being nodules. Suspicious locations are connected with a red bounding box with a saturation directly proportional to the nodule probability. Figure 5.7(c) shows that our model proposes two locations related to the same region that were not filtered by the adjacent candidate rejection rule after classification.

Figure 5.8 shows top findings on the JSRT dataset. In the first row, we show the top true positives detected through our network. These samples are well centered nodules. The sample shown in Figure 5.8(a) shows that there are cases with low contrast even on the samples with high probability. The second row shows the top false positives. An interesting observation here is that the top six false positives samples we selected contain non centered nodules. This suggests that the CNN recognizes nodules even if they are no well centered. We marked the regions that contain these nodules as non-nodule because the distance between the center of the ROI and the ground truth annotation is larger than the threshold distance of our labeling criterion. The third row shows six randomly selected false positives found by our CNN. These samples contain ribs and large vessels that are miss-classified as nodules by the network.



(a) (b) (c)

(d) (e) (f)

Figure 5.7: Visual results of our system. We highlight the 4 top positive candidates detected by each image. The saturation of each bounding box is proportional to the probability of the candidate of being a true nodule.

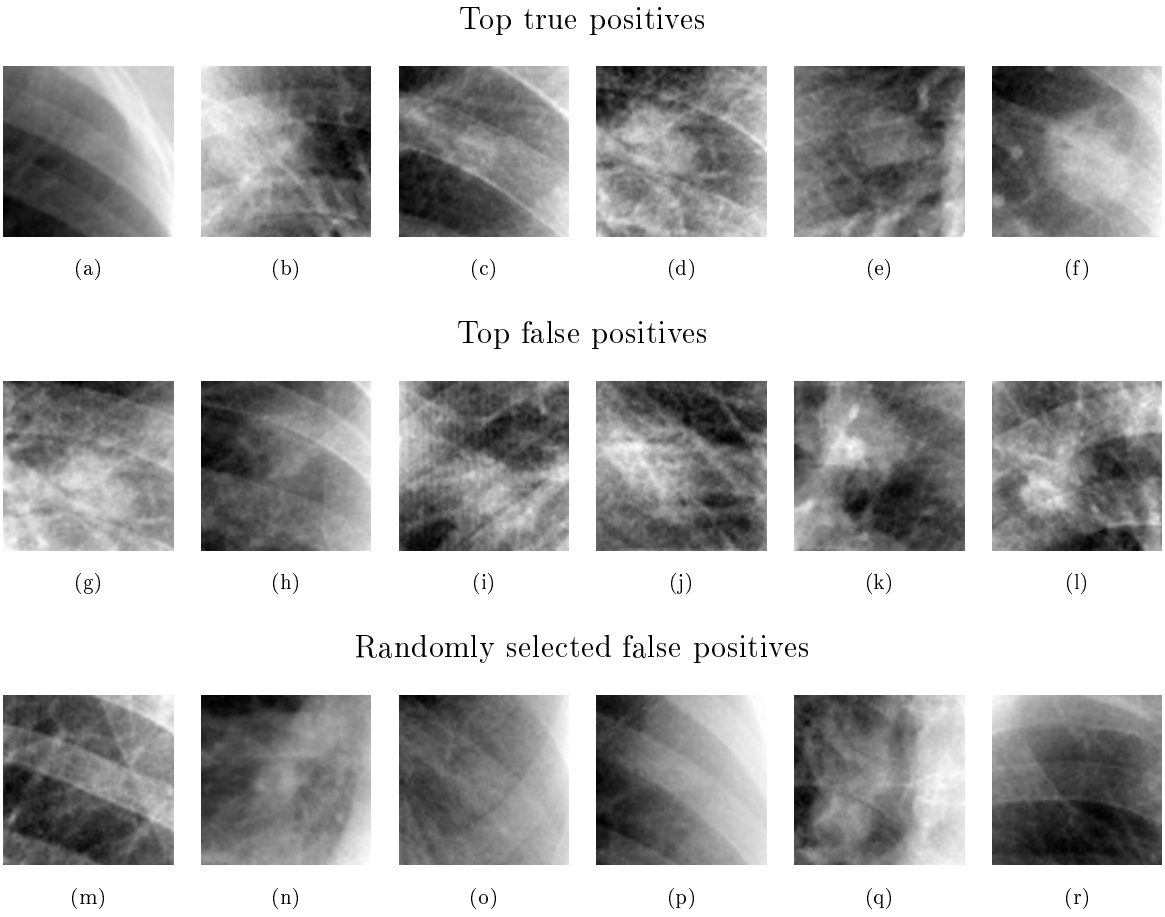


Figure 5.8: Detection results using *ConNet(6, 32, 1)*. The first and second rows show the landmarks and boundaries found on samples of the positive subset of JSRT dataset and LIDC-IDRI datasets, respectively.

Chapter 6

Conclusions and Future Work

In this dissertation, we proposed and analyzed a pipeline for lung nodule detection on CXR images that consists in three stages: (i) lung area segmentation, (ii) candidate nodule detection, and (iii) candidate nodule classification. An important goal of our method is to detect potential location of nodules with high sensitivity through a few false positives per image. To achieve this aim, we explore the effectiveness of CNNs for reducing false positives on the candidate classification stage.

To compare our results with competing approaches, we used a subset of CXR images on the LIDC-IDRI dataset and the JSRT dataset. Both datasets contain annotations with the locations and other characteristics of the nodules. Moreover, we used the SCR dataset, which allowed the lung are segmentation in the images of the JSRT dataset. We trained our segmentation model with the images of the JSRT dataset, which do not contain nodules and their annotations on the SCR dataset. The segmentation model was used to generate the lung masks for the LIDC-IDRI dataset and the images with nodules on the JSRT dataset. We used the LIDC-IDRI dataset and the images with nodules of the JSRT dataset for the evaluation of candidate lung detection and classification methods.

In lung area segmentation, we evaluated the performance of a multi-scale patch-based AAM and compared its performance with various segmentation methods. We found that our method obtained competitive results when compared to other approaches of the literature. We observed that the methods that surpassed the AAM effectiveness perform segmentation at pixel level. We concluded that a post-processing method for finding accurate boundaries between landmarks may improve the results considerably. However, we considered that the gain on other stages of the pipeline will produce greater improvements on the overall performance of the system.

We analyzed the effectiveness of the Laplacian of Gaussian, the difference of Gaussian, and the determinant of Hessian, weighted-multiscale-convergence-index filter and the sliding band filter detection methods in terms of sensitivity by the amount of FPPI. In this experiment, we found that the sliding band filter obtained better results than the other detectors.

In the classification stage, we proposed a method based on a CNN trained from the scratch. We analyzed the impact of various data augmentation transformations individually and combined. We found that the transformations individually can improve the results substantially. Moreover, we also found that the transformations are complemen-

tary. Furthermore, we compared two ways to use dropout on convolutional layers. We found that both approaches improved the effectiveness of the model and that assigning increasing dropout probabilities is slightly superior. We also evaluated the impact of CNN architecture parameters on the performance of the system.

The main goal of the conducted experiments was to find the best configuration for lung nodule classification in CXR images. Then, we compared the results of the best CNN configuration, $ConvNet(6, 32, 1)$, to competing approaches. Due to factors related to variation of the methods, it is difficult to guarantee that our method outperformed competing approaches. However, we showed that, under our considerations with respect to variations, our method achieved good results when compared to the state-of-the-art.

As directions for future work, we conjecture that the exploration of other regularization mechanisms, besides dropout and data augmentation, is an important aspect. Thus, we conducted some preliminary experiments on augmenting a CNN trained from the scratch, with a Stacked-What-and-Where-AutoEncoder [94] such that the learning is performed by considering supervised and unsupervised objectives at the same time. This idea has improved our results and suggests that further investigation in this direction should be conducted. Moreover, we could leverage the insights found in this work to develop a method for lung nodule classification in CT images.

Bibliography

[1] Open Source Computer Vision Library. `https://github.com/opencv/opencv`, 2017.

[2] B. Al Mohammad, P. Brennan, and C. Mello-Thoms. A Review of Lung Cancer Screening and the Role of Computer-Aided Detection. *Clinical Radiology*, 72(6): 433–442, 2017.

[3] J. Alabort-i Medina and S. Zafeiriou. A Unified Framework for Compositional Fitting of Active Appearance Models. *International Journal of Computer Vision*, 121(1):26–64, 2017.

[4] J. Alabort-i-Medina, E. Antonakos, J. Booth, P. Snape, and S. Zafeiriou. Menpo: A Comprehensive Platform for Parametric Image Alignment and Visual Deformable Models. In *ACM International Conference on Multimedia*, pages 679–682, Orlando, FL, USA, 2014. ACM.

[5] S. G. Armato, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. 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. *Medical Physics*, 38 (2):915–931, 2011.

[6] N. Becker, E. Motsch, M.-L. Gross, A. Eigentopf, C. Heussel, H. Dienemann, P. Schnabel, M. Eichinger, D.-E. Optazaite, and M. Puderbach. Randomized Study on Early Detection of Lung Cancer with MSCT in Germany: Results of the First 3 Years of Follow-up After Randomization. *Journal of Thoracic Oncology*, 10(6):890–896, 2015.

[7] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle. Greedy Layer-wise Training of Deep Networks. *Advances in Neural Information Processing Systems*, 19:153, 2007.

[8] J. C. M. Bobadilla and H. Pedrini. Lung Nodule Classification Based on Deep Convolutional Neural Networks. In *XXI Iberoamerican Congress on Pattern Recognition*, pages 117–124, Lima, Peru, Nov. 2016. Springer.

[9] W. E. Brant and C. A. Helms. *Fundamentals of Diagnostic Radiology*. Lippincott Williams & Wilkins, 2012.

[10] N. Camarlinghi. *Pattern Recognition Methods Applied to Medical Imaging: Lung Nodule Detection in Computed Tomography Images*. PhD thesis, Universita degli Studi di Pisa, 2012.

[11] P. Campadelli, E. Casiraghi, and D. Artioli. A Fully Automated Method for Lung Nodule Detection from Postero-anterior Chest Radiographs. *IEEE Transactions on Medical Imaging*, 25(12):1588–1603, 2006.

[12] S. Candemir, S. Jaeger, K. Palaniappan, J. P. Musco, R. K. Singh, Z. Xue, A. Karargyris, S. Antani, G. Thoma, and C. J. McDonald. Lung Segmentation in Chest Radiographs using Anatomical Atlases with Nonrigid Registration. *IEEE Transactions on Medical Imaging*, 33(2):577–590, 2014.

[13] S. Chen and K. Suzuki. Computerized Detection of Lung Nodules by Means of Virtual Dual-Energy Radiography. *IEEE Transactions on Biomedical Engineering*, 60(2):369–378, 2013.

[14] S. Chen, K. Suzuki, and H. MacMahon. Development and Evaluation of a Computer-aided Diagnostic Scheme for Lung Nodule Detection in Chest Radiographs by means of Two-Stage Nodule Enhancement with Support Vector Classification. *Medical physics*, 38(4):1844–1858, 2011.

[15] F. Chollet. Keras. `https://github.com/fchollet/keras`, 2015.

[16] T. F. Cootes and C. J. Taylor. Statistical Models of Appearance for Computer Vision. Technical report, University of Manchester, England, United Kingdom, 2004.

[17] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Training Models of Shape from Sets of Examples. In *British Machine Vision Conference*, pages 9–18. Springer, 1992.

[18] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active Appearance Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[19] G. Coppini, S. Diciotti, M. Falchini, N. Villari, and G. Valli. Neural Networks for Computer-Aided Diagnosis: Detection of Lung Nodules in Chest Radiograms. *IEEE Transactions on Information Technology in Biomedicine*, 7(4):344–357, 2003.

[20] D. W. De Boo, M. Uffmann, M. Weber, S. Bipat, E. F. Boorsma, M. J. Scheerder, N. J. Freling, and C. M. Schaefer-Prokop. Computer-aided Detection of Small Pulmonary Nodules in Chest Radiographs: An Observer Study. *Academic Radiology*, 18 (12):1507–1514, 2011.

[21] L. Dominioni, A. Poli, W. Mantovani, S. Pisani, N. Rotolo, M. Paolucci, F. Sessa, V. Conti, V. D'Ambrosio, and A. Paddeu. Assessment of Lung Cancer Mortality Reduction after Chest X-ray Screening in Smokers: A Population-based Cohort Study in Varese, Italy. *Lung Cancer*, 80(1):50–54, 2013.

[22] C. Farabet, C. Couprie, L. Najman, and Y. LeCun. Learning Hierarchical Features for Scene Labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1915–1929, 2013.

[23] K. Fukushima. Cognitron: A Self-Organizing Multilayered Neural Network. *Biological Cybernetics*, 20(3-4):121–136, 1975.

[24] K. Fukushima. Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. *Biological Cybernetics*, 36(4):193–202, 1980.

[25] X. Glorot, A. Bordes, and Y. Bengio. Deep Sparse Rectifier Neural Networks. In *Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 315–323, 2011.

[26] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[27] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio. Maxout Networks. *arXiv preprint arXiv:1302.4389*, 2013.

[28] B. Graham. Spatially-Sparse Convolutional Neural Networks. *arXiv preprint arXiv:1409.6070*, 2014.

[29] H. Greenspan, B. van Ginneken, and R. M. Summers. Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting new Technique. *IEEE Transactions on Medical Imaging*, 35(5):1153–1159, 2016.

[30] R. C. Hardie, S. K. Rogers, T. Wilson, and A. Rogers. Performance Analysis of a new Computer Aided Detection System for Identifying Lung Nodules on Chest Radiographs. *Medical Image Analysis*, 12(3):240–258, 2008.

[31] K. He, X. Zhang, S. Ren, and J. Sun. Delving Deep into Rectifiers: Surpassing Human-level Performance on Imagenet Classification. In *IEEE International Conference on Computer Vision*, pages 1026–1034, 2015.

[32] D. O. Hebb. *The Organization of Behavior: A Neuropsychological Approach*. John Wiley & Sons, 1949.

[33] G. Hinton, S. Osindero, and Y.-W. Teh. A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, 18(7):1527–1554, 2006.

[34] G. E. Hinton. Learning Distributed Representations of Concepts. In *Eighth Annual Conference of the Cognitive Science Society*, volume 1, page 12. Amherst, MA, USA, 1986.

[35] A. K. Jain, Y. Zhong, and M.-P. Dubuisson-Jolly. Deformable Template Models: A Review. *Signal Processing*, 71(2):109–129, 1998.

[36] K. Jarrett, K. Kavukcuoglu, and Y. LeCun. What is the best Multi-Stage Architecture for Object Recognition? In *IEEE 12th International Conference on Computer Vision*, pages 2146–2153. IEEE, 2009.

[37] S. Katsuragawa and K. Doi. Computer-Aided Diagnosis in Chest Radiography. *Computerized Medical Imaging and Graphics*, 31(4):212–223, 2007.

[38] A. N. Khan, H. H. Al-Jahdali, C. M. Allen, K. L. Irion, S. Al Ghanem, and S. S. Koteyar. The Calcified Lung Nodule: What does it Mean? *Annals of Thoracic Medicine*, 5(2):67, 2010.

[39] H. Kobatake and S. Hashimoto. Convergence Index Filter for Vector Fields. *IEEE Transactions on Image Processing*, 8(8):1029–1038, 1999.

[40] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, pages 1106–1114, 2012.

[41] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 1(4):541–551, 1989.

[42] Y. LeCun, Y. Bengio, and G. Hinton. Deep Learning. *Nature*, 521(7553):436–444, 2015.

[43] J.-S. Lin, S.-C. Lo, A. Hasegawa, M. T. Freedman, and S. K. Mun. Reduction of False Positives in Lung Nodule Detection using a Two-level Neural Classification. *IEEE Transactions on Medical Imaging*, 15(2):206–217, 1996.

[44] T. Lindeberg. Scale-Space Theory in Computer Vision. *The International Series in Engineering and Computer Science*, 1993.

[45] T. Lindeberg. Feature Detection with Automatic Scale Selection. *International Journal of Computer Vision*, 30(2):79–116, 1998.

[46] T. Lindeberg. Image Matching using Generalized Scale-space Interest Points. *Journal of Mathematical Imaging and Vision*, 52(1):3–36, 2015.

[47] S.-C. Lo, S.-L. Lou, J.-S. Lin, M. T. Freedman, M. V. Chien, and S. K. Mun. Artificial Convolution Neural Network Techniques and Applications for Lung Nodule Detection. *IEEE Transactions on Medical Imaging*, 14(4):711–718, 1995.

[48] D. G. Lowe. Distinctive Image Features from Scale-invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[49] A. L. Maas, A. Y. Hannun, and A. Y. Ng. Rectifier Nonlinearities Improve Neural Network Acoustic Models. In *International Conference on Machine Learning*, volume 30, 2013.

[50] I. Matthews and S. Baker. Active Appearance Models Revisited. *International Journal of Computer Vision*, 60(2):135–164, 2004.

[51] J. McClelland, D. Rumelhart, and G. Hinton. The Appeal of Parallel Distributed Processing. *MIT Press*, pages 3–44, 1986.

[52] W. S. McCulloch and W. Pitts. A Logical Calculus of the Ideas Immanent in Nervous Activity. *The Bulletin of Mathematical Biophysics*, 5(4):115–133, 1943.

[53] S. McGuire. World Cancer Report 2014. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press. *Advances in Nutrition: An International Review Journal*, 7(2):418–419, 2016.

[54] P. Mesejo, O. Ibanez, O. Cordon, and S. Cagnoni. A Survey on Image Segmentation using Metaheuristic-based Deformable Models: State of the Art and Critical Analysis. *Applied Soft Computing*, 44:1–29, 2016.

[55] M. Minsky and S. Papert. *Perceptrons*. MIT Press, 1969.

[56] V. Nair and G. E. Hinton. Rectified Linear Units Improve Restricted Boltzmann Machines. In *27th International Conference on Machine Learning*, pages 807–814, 2010.

[57] T. A. Ngo and G. Carneiro. Fully Automated Segmentation Using Distance Regularised Level Set and Deep-Structured Learning and Inference. In *Deep Learning and Convolutional Neural Networks for Medical Image Computing*, pages 197–224. Springer, 2017.

[58] M. M. Oken, W. G. Hocking, P. A. Kvale, G. L. Andriole, S. S. Buys, T. R. Church, E. D. Crawford, M. N. Fouad, C. Isaacs, and D. J. Reding. Screening by Chest Radiograph and Lung Cancer Mortality: the Prostate, Lung, Colorectal, and Ovarian (PLCO) Randomized Trial. *Jama*, 306(17):1865–1873, 2011.

[59] G. Orbán and G. Horváth. Algorithm Fusion to Improve Detection of Lung Cancer on Chest Radiographs. *International Journal of Intelligent Computing and Cybernetics*, 5(1):111–144, 2012.

[60] G. G. Orbán and G. Horváth. Kernel Descriptors for Chest X-ray Analysis. In *SPIE Medical Imaging*, pages 101343O–101343O. International Society for Optics and Photonics, 2017.

[61] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[62] C. S. Pereira, H. Fernandes, A. M. Mendonça, and A. Campilho. Detection of Lung Nodule Candidates in Chest Radiographs. In *Iberian Conference on Pattern Recognition and Image Analysis*, pages 170–177. Springer, 2007.

[63] B. T. Polyak. Some Methods of Speeding up the Convergence of Iteration Methods. *USSR Computational Mathematics and Mathematical Physics*, 4(5):1–17, 1964.

[64] M. Ranzato, C. Poultney, S. Chopra, and Y. LeCun. Efficient Learning of Sparse Representations with an Energy-based Model. In *19th International Conference on Neural Information Processing Systems*, pages 1137–1144, Vancouver, Canada, 2006. MIT Press.

[65] F. Rosenblatt. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, 65(6):386, 1958.

[66] D. E. Rumelhart and J. L. Mcclelland. *Parallel Distributed Processing*. MIT Press, Aug. 1986.

[67] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning Representations by Back-propagating Errors. *Cognitive Modeling*, 5(3):1, 1988.

[68] A. M. Saxe, J. L. McClelland, and S. Ganguli. Exact Solutions to the Nonlinear Dynamics of Learning in Deep Linear Neural Networks. *arXiv preprint arXiv:1312.6120*, 2013.

[69] A. M. Schilham, B. van Ginneken, and M. Loog. A Computer-aided Diagnosis System for Detection of Lung Nodules in Chest Radiographs with an Evaluation on a Public Database. *Medical Image Analysis*, 10(2):247–258, 2006.

[70] J. Schmidhuber. Deep Learning in Neural Networks: An Overview. *Neural Networks*, 61:85–117, 2015.

[71] P. Sermanet, K. Kavukcuoglu, S. Chintala, and Y. LeCun. Pedestrian Detection with Unsupervised Multi-Stage Feature Learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3626–3633. IEEE, 2013.

[72] A. A. A. Setio, A. Traverso, T. de Bel, M. S. Berens, C. v. d. Bogaard, P. Cerello, H. Chen, Q. Dou, M. E. Fantacci, and B. Geurts. Validation, Comparison, and Combination of Algorithms for Automatic Detection of Pulmonary Nodules in Computed Tomography Images: The LUNA16 Challenge. *arXiv preprint arXiv:1612.08012*, 2016.

[73] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*, 35(5):1285–1298, 2016.

[74] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K.-i. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi. Development of a Digital Image Database for Chest Radiographs with and without a Lung Nodule: Receiver Operating Characteristic Analysis of Radiologists' Detection of Pulmonary Nodules. *American Journal of Roentgenology*, 174(1):71–74, 2000.

[75] J. Shiraishi, Q. Li, K. Suzuki, R. Engelmann, and K. Doi. Computer-aided Diagnostic Scheme for the Detection of Lung Nodules on Chest Radiographs: Localized Search Method based on Anatomical Classification. *Medical Physics*, 33(7):2642–2653, 2006.

[76] R. M. Summers. Deep Learning and Computer-Aided Diagnosis for Medical Image Processing: A Personal Perspective. In *Deep Learning and Convolutional Neural Networks for Medical Image Computing*, pages 3–10. Springer, 2017.

[77] I. Sutskever, J. Martens, G. Dahl, and G. Hinton. On the Importance of Initialization and Momentum in Deep Learning. In *International Conference on Machine Learning*, pages 1139–1147, 2013.

[78] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang. Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? *IEEE Transactions on Medical Imaging*, 35(5):1299–1312, 2016.

[79] T. D. Team. Theano: A Python Framework for Fast Computation of Mathematical Expressions. *arXiv e-prints*, abs/1605.02688, May 2016.

[80] G. Tzimiropoulos and M. Pantic. Gauss-Newton Deformable Part Models for Face Alignment In-the-Wild. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1851–1858, 2014.

[81] E. J. van Beek, S. Mirsadraee, and J. T. Murchison. Lung Cancer Screening: Computed Tomography or Chest Radiographs? *World Journal of Radiology*, 7(8):189, 2015.

[82] S. Van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu. Scikit-image: Image Processing in Python. *PeerJ*, 2:e453, 2014.

[83] B. Van Ginneken, M. B. Stegmann, and M. Loog. Segmentation of Anatomical Structures in Chest Radiographs using Supervised Methods: A Comparative Study on a Public Database. *Medical Image Analysis*, 10(1):19–40, 2006.

[84] C. Wang, A. Elazab, J. Wu, and Q. Hu. Lung Nodule Classification using Deep Feature Fusion in Chest Radiography. *Computerized Medical Imaging and Graphics*, 57:10–18, 2017.

[85] S. Wang and R. M. Summers. Machine Learning and Radiology. *Medical Image Analysis*, 16(5):933–951, 2012.

[86] J. Wei, Y. Hagihara, and H. Kobatake. Detection of Cancerous Tumors on Chest X-ray Images-candidate Detection Filter and its Evaluation. In *International Conference on image Processing*, volume 3, pages 397–401. IEEE, 1999.

[87] J. Wei, Y. Hagihara, A. Shimizu, and H. Kobatake. Optimal Image Feature Set for Detecting Lung Nodules on Chest X-Ray Images. In *Computer Assisted Radiology and Surgery*, pages 706–711. Springer, 2002.

[88] R. Wender, E. T. Fontham, E. Barrera, G. A. Colditz, T. R. Church, D. S. Ettinger, R. Etzioni, C. R. Flowers, G. Scott Gazelle, and D. K. Kelsey. American Cancer Society Lung Cancer Screening Guidelines. *CA: A Cancer Journal for Clinicians*, 63 (2):106–117, 2013.

[89] J. Weng, N. Ahuja, and T. S. Huang. Cresceptron: A Self-Organizing Neural Network which Grows Adaptively. In *International Joint Conference on Neural Networks*, volume 1, pages 576–581. IEEE, 1992.

[90] B. Widrow and M. E. Hoff. Adaptive Switching Circuits. Technical report, Stanford University, Stanford, CA, USA, 1960.

[91] D. J. Wu. *End-to-end Text Recognition with Convolutional Neural Networks.* PhD thesis, Stanford University, Stanford, CA, USA, 2012.

[92] H. Wu and X. Gu. Towards Dropout Training for Convolutional Neural Networks. *Neural Networks*, 71:1–10, 2015.

[93] Y. Yan, M. Chen, M. L. Shyu, and S. C. Chen. Deep Learning for Imbalanced Multimedia Data Classification. In *IEEE International Symposium on Multimedia*, pages 483–488, Dec. 2015.

[94] Y. Zhang, K. Lee, and H. Lee. Augmenting Supervised Neural Networks with Unsupervised Objectives for Large-Scale Image Classification. In *International Conference on Machine Learning*, pages 612–621, 2016.