UNIVERSIDAD NACIONAL DE COLOMBIA

A framework for interactive training of automatic image analysis models based on learned image representations, active learning and visualization techniques

Juan Sebastian Otálora Montenegro

Universidad Nacional de Colombia
Engineering School, Systems and Industrial Engineering Department
Bogotá D.C., Colombia
2016

A framework for interactive training of automatic image analysis models based on learned image representations, active learning and visualization techniques

Juan Sebastian Otálora Montenegro

In fulfillment of the requirements for the degree of:
Master of Systems and Computing Engineering

Advisor:
Fabio Augusto González Osorio, Ph.D.

Research Field:
Machine Learning
Research Group:
MindLab

Universidad Nacional de Colombia
Engineering School, Systems and Industrial Engineering Department
Bogotá D.C., Colombia
2016

This work is dedicated to all the people with I've shared.

Acknowledgements

I would like to express special gratitude to my supervisor, Professor Fabio González for his support since 2011 when he gave me an overview of how academia looks like, now in perspective I think that I found the right person to give me advice, not only technical, but also personal. Thanks to all my coleagues in the MindLab laboratory who offered great assistance, and good coffee/beer times when it was necessary. Thanks to all the members of Henning Müller's group MedGIFT for an amazing internship at their lab in Sierre, Switzerland. Infinite gratitude to my family, they provide me food, shelter, and other basic supplies in order to concentrate on my studies and achieve my goals looking forward to a good future. Thanks to the people in my country which stills believes in peace as the first step to overcome with our tragic past. Thank you Paola for making all of this make even more sense.
This work was partially funded by projects Diseño e implementación de un sistema de cómputo sobre recursos heterogéneos para la identificación de estructuras atmosféricas en predicción climatológica number 1225-569-34920 through Colciencias contract number 0213-2013 and Multimodal Image Retrieval to Support Medical CaseBased Scientific Literature Search, ID R1212LAC006 by Microsoft Research/LACCIR. Also thanks to Colciencias grant 645 Jóvenes Investigadores e Innovadores de Colciencias 2014 and the HES-SO for their support while my internship in Switzerland.

Abstract

In this thesis work, the problem of applying active learning for a label efficient training of deep learning models is addressed.
Firstly, in chapter one, the problem is introduced as well as the objectives and results of this thesis work. In the chapter 2, a state of the art of active learning and deep learning models is presented with a particular emphasis in medical scenarios. In chapter three an active learning approach based on the expected gradient length, is introduced for deep convolutional neural networks for applying in medical problems where data is scarse and train deep models could be unfeasible due the the lack of annotated samples.
In chapter four an implemented framework for interactively training of deep learning models based on the previous discused algorithms is presented, where the active learning techniques improves the random selection strategy to classify between healthy eyes patches and patches that contains an early stage of diabetic retinopathy.
Finally in the last chapter, the conclusions of this thesis work are discussed as well as some promising lines of work for further research.

Keywords: Deep Learning, Active Learning, Medical Imaging, Expected Gradient Length, On-line Learning

Resumen

En ésta tesis, se estudia el problema del entrenamiento eficiente de modelos de aprendizaje automático basados en redes neuronales profundas para el caso en el que se cuenta con pocos ejemplos anotados para su entrenamiento. Para esto se presentara una estrategia de aprendizaje activo la cual hace mas eficiente el aprendizaje de una representación profunda utilizando los ejemplos que mas cambios aportan al modelo.

En el primer capítulo, se introduce el problema así como los objetivos y resultados de este trabajo de tesis. Una revisión de los trabajos recientes en el área de aprendizaje activo y modelos de aprendizaje profundo, con énfasis en escenarios médicos se presenta en el capítulo 2.

En el capítulo 3, se presenta el enfoque propuesto de aprendizaje activo para modelos de aprendizaje profundos basado en la longitud esperada del gradiente, el cual resulta útil para la solución de problemas de imágenes médicas donde no se cuenta con la suficiente cantidad de ejemplos anotados.

En el capítulo 4, un marco experimental es implementado para el entrenamiento de modelos basados en redes neuronales profundas, se muestra la aplicación de esta estrategía para clasificar parches de imágenes de fondo de ojo con pacientes sanos y en una etapa temprana de retinopatía diabética. Se muestra que el algoritmo propuesto mejora el desempeño del modelo comparandolo con la estrategía clásica de selección aleatoria de ejemplos.

Finalmente en el último capítulo se discuten las concluciones de este trabajo y también se esbozan algunas lineas de trabajo prometedoras para el futuro.

Palabras clave: Aprendizaje de máquina, Redes Neuronales Profundas, Aprendizaje Activo, Aprendizaje de la Representación, Análisis de Imágenes Médicas, Aprendizaje en Linea, Longitud esperada del gradiente.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Nowadays data is a part of our daily lives like never before, we live in the so called information age where everything flows in bits at incredible speeds. Our basic social interactions and high tech products each day relies more and more on data to produce a more personalized and natural interaction. There are academic centexxxrs and industries that have taken the more advantage in the use of algorithms to process the large amount of data that we have available today e.g. social networks in sites like Facebook or Twitter, music and media recommendation are at the order of the day in mobile apps like Spotify and Netflix, autonomous driving is emerging with dedicated research departments at Google, Tesla Motors and Uber, and this would not have been possible by the impressive academic results in the last years thanks to algorithms and the dedicated hardware that allow us to process and learn from quantities and sources of data that exceeds in orders of magnitude of what we had just a decade before. The field of deep learning[43, 3] borns with those premises in mind, showing major advances in computer vision and speech recognition, training high capacity models with hundreds of thousands labeled samples in dedicated graphic processing units.

One of the challenging fields which has partially benefited from such advantages is medical imaging[25], were the patterns to look for reside in big digitalized histology images, functional magnetic resonance images, computer tomography volumes, eye fundus images among others. This benefit is partial because for succesfull training of the deep learning models a large corpora of labeled data is required, which, in general, is a costly and difficult setup to find in the biomedical imaging community

Making more label-efficient computer-aided systems would be of necessary step for (i) reducing costs in building medical image datasets where the experts annotations take much time and are costly and (ii) for succesfull usage of deep learning algorithms in the medical imaging workflow[10, 22].

1.1 Problem definition

The approach in this thesis addresses the problem of training of automatic image analysis models using a reduced number of labeled instances. Specifically, the goal is to devise a method based on learned image representations, active learning and visualization techniques to automatically find visual patterns that compactly explain the visual richness of images and the relationships between visuassl images and their meaning in an interactive setup, but limiting as much as possible the interaction of the user with the system. Working out this problem requires to solve three main subproblems: first, to find an appropriate image representation that takes into account the structure of the image collection and its feasibility of use for an interactive approach; second, to devise appropriate active learning algorithms that, based on the image collection representation, could extract visual, semantic and meaningful relationships between them using only a limited number of annotated samples (or interactions); and third, how to visualize the internal representations and generate an interactive visualization of the patterns found by the algorithm. In order to address this research problem, my approach will be to work out the following research questions:

- How to interactively train and learn a semantic enriched image representation from an image collection?
- How to use the learned representation to find interesting patterns that connect visual content and its meaning?
- Does this kind of representation improve the performance in tasks such as automatic segmentation, image annotation and image captioning?
- How to use the learned representation to make image analysis models more interpretable?

1.2 Objectives

The main objective of this thesis is to develop a method for interactive training of automatic image analysis models based on learned image representations, active label learning and visualization techniques. To cope with the objective, this work has been divided into the following especific objectives:

1. To propose a strategy for generating image representations from image collections suitable to be used in an active learning strategy.

2. To propose or adapt a strategy for active learning in a collection of images based on learned image representations.

3. To develop a method to visualize patterns automatically learned by deep learning models from image collections.

4. To develop and evaluate a prototype for interactive training of image analysis models.

1.3 Thesis Scope

This research thesis will deal mainly with the underlying algorithmic issues of a framework for interactive training of automatic image analysis models based on learned image representations and active label learning. It is limited to the design and implementation of the algorithms in a coherent framework, and the subsequent tasks of programming and divulgation of results of the aforementioned methods. Even though the research will develop general methods that can be applied to different kind of images, the main focus of the project will be in biomedical and eye fundus image collections where the potential for the methods could lead to a big gain in specialized and costly time of training such automatic methods. All the side-activities and research papers product of this thesis will help to contribute in one or more research tasks assigned, even if the problem is not directly related with biomedical and/or eye fundus image collections analysis.

1.4 Results and contributions

The results and contributions of this work can be summarized as follows:

- **Sebastian Otálora** et.al. , "Training Convolutional Neural Networks with Active Learning for exudate classification in eye fundus images"

  In this work, we introduce the expected gradient length algorithm into the training of deep convolutional neural networks for exudate classication in eye fundus images. Our proposed method was able to signicantly reduce training time obtaining a really good performance. My contributions in this work include the code for the algorithms, design and execution of experiments, writing of the draft and the nal submitted paper to the international conference: Information Processing in Medical Imaging - IPMI 2017, the algorithm and the approach are depicted in the chapters 3 and 4 of this document.

- Oscar Perdomo, **Sebastian Otálora**, Fabio A.González, "A novel machine learning model based on exudate localization to detect diabetic macular edema". MICCAI 2016, the 19th International Conference on Medical Image Computing and Computer Assisted Intervention. Athens, Greece. October 17th to 21st, in proceedings.

  In this work, we introduce a novel convolutional neural network architecture to detect diabetic macular edema. Our propossed method was able to surpass in performance baseline CNN models. My contributions in this work included the efficient generation of predictions and design of experiments, corrections and the final submission of the paper to a workshop event in the MICCAI 2016 conference.

- **Sebastian Otálora**, Angel Cruz-Roa, John Arevalo, Manfredo Atzori, Anant Madabhushi, Alexander Judkins, Fabio A.González, Henning Müller and Adrien Depeursinge. "Anaplastic Medulloblastoma tumor differentiation by combining Unsupervised Feature Learning and Riesz wavelets for histopathology image representation". MICCAI 2015, the 18th International Conference on Medical Image Computing and Computer Assisted Intervention. Munich, Germany. October 5th to 9th, in proceedings.

  In this work, we show that the combination of two complementary approaches for feature learning (unsupervised and supervised) improves the classification performance for medulloblastoma tumor differentiation. Our approach outperforms the best methods in literature by 2.5% achieving 99% accuracy over region-based data comprising 7,500 square regions from 10 cases diagnosed with medulloblastoma (5 anaplastic and 5 non-anaplastic). My contributions in this work includes the development of the code, design and execution of experiments, writing of the draft, corrections for the final submission of the paper and elaboration of the poster presented at the main conference event in Munich, Germany by one of the co-authors.

- John Arevalo, **Sebastian Otálora**, Julien Wist and Fabio A. González. "Automatic Infrared spectroscopy signal analysis with unsupervised feature learning and neural networks". 9th Colombian Computing Congress. Pereira, Colombia, September 3-5, 2014 9ccc proceedings.

  In this paper we presented a novel method for the prediction of molecular fragments from infrared spectra based on unsupervised feature learning. We evaluated this method on a set of 6373 infrared signals obtaining an improvement in the prediction stage with an automatically learned representation using an unsupervised learning method. Our model improved the interpretation of the results by allowing us to compute the best signal from a given structure label. One key advantage in the proposed approach is that it is required only one model that used a shared representation to predict 512 labels. My contributions in this work include participations in the development of the code, design and execution of experiments, writing of the paper and corrections for the final submission for later presentation (oral presentation) of the work at the 9th Colombian Computing Congress conference event. This paper is annexed to this document.

From February to June of 2015 I had the great oportunity to do an internship in the Swiss research group MedGIFT in head of of professor Henning Müller's, in those months I was able to work in my thesis project in a really nice enviroment with great collaborators to write the third listed paper above and more important to strengthen the research link between our MindLAB group at Bogotá. Thanks HES-SO for providing me the francs necessarly to cover most of the internship costs and also for preparing such nice activities with foreign students.
Besides my main thesis work I also collaborated on several projects of my research group and in some cases those collaborations lead to an academic product, notably:

- Sanandres C. Eliana , **Sebastián Otálora**. "Una aplicación de topic modeling para el estudio del trauma: El caso de Chevron-Texaco en Ecuador." Investigación & Desarrollo 23.2 (2015).

  In this work, we introduce topic modeling techniques using Latent Dirichlet Allocation for working qualitatively with large amounts of data addressing the emergence of the trauma process resulting from the Chevron-Texaco case in Ecuador. My contributions in this work include participations in the development of the code, design and execution of experiments, writing of the method section and corrections for the final submission of the paper for the submission to the journal. This paper is annexed to this document.

- Jorge A. Vanegas, John Arevalo, **Sebastian Otálora**, Fabián Páez, Santiago A. Pérez-Rubiano, and Fabio A. González. MindLab at ImageCLEF 2014: Scalable Concept Image Annotation. CLEF (Working Notes) 2014: 404-410

  This paper describes the participation of the MindLab research group of Universidad Nacional de Colombia at the ImageCLEF 2014 Scalable Concept Image Annotation challenge. Our strategy mainly relies in finding a good visual representation based on deep convolutional neural networks. Despite the simplicity of the proposed classifier which allows to deal with the large-scale nature of this task, we can achieve good performance (our proposed approach achieved the best MAP) thanks to the rich visual representation based on learned features. My contributions in this work include programming the features extraction process and the evaluation using the proposed experimental setup. This paper is annexed to this document.

- de Herrera, Alba Garcıa Seco, Henning Müller, and Stefano Bromuri. "Overview of the ImageCLEF 2015 medical classification task." Working Notes of CLEF 2015 (2015).

  This paper describes the participation of the MindLab research group of Universidad Nacional de Colombia at the ImageCLEF 2015 Multi–label Classification subtask of the ImageCLEF 2015 medical classification task. Our strategy was based on building a visual representation by means of deep convolutional neural networks, by relying on the theory of transfer learning which is based in the ability of a system to recognize and apply knowledge learned in previous domains to novel domains, we obtain the best Hamming Loss value, demonstrating again the suitability of aplying deep learning representations to biomedical domains. My contributions in this work were programming of the experiments, evaluation and a small paragraph explaining our strategy. This paper is annexed to this document.

Most of the code is available in personal web repositories:

- The code for the articles "Training Convolutional Neural Networks with Active Learning for exudate classification in eye fundus images" and "A novel machine learning model based on exudate localization to detect diabetic macular edema" is available at

  `https://bitbucket.org/sebastianffx/paper_labels_miccai`

- The code for the articles "MindLab at ImageCLEF 2014: Scalable Concept Image Annotation" and "Overview of the ImageCLEF 2015 medical classification task" is available at

  `https://bitbucket.org/sebastianffx/imageclef-2014-scia`

1.5 Document structure

This thesis is divided in 5 chapters. Chapter 1 presents the thesis problem statement, scope, objectives, results and contributions, and this document structure. Chapter 2 presents a brief review of the state of the art in automatical image analysis in the biomedical domain using active learning algorithms and deep learning representations. In chapter 3, the expected gradient length algorithm is presented as well as its adaptation to use as a sample and image selector for sample-eficient training of deep convolutional neural networks. Chapter 4 presents the application of the previously presented strategy to detect exudates in eye fundus images with minimal number of labeled samples used, in this chapter a visualization strategy to identify interesting regions in the image is presented as well. Finally, in chapter 5 concluding remarks and future work in this line of research are discussed.

Chapter 2

State of the art

In medical imaging workflows the use of the interdisciplinary computer vision techniques to analysis the wide spectra of diseases present in digitalized images has been an usefull and important aplication of computer science since the birth of X-rays in XIX century, with several milestones that allowed medical content obtained by different modalities and sources to be analysed with aid of a computer [40], with the evolution of the computer vision techniques along with the clear definition of the challenges and application areas for specific diseases more computer aided systems and tools are being used into the medical practice but stills there is a limited reach of this techniques, a concrete modern example is digital pathology, which is an image-based information workflow that is enabled by computer technologies and machine learning methods that allows for the management of information generated from a light microscopy-digital slide. With the advent of Whole-Slide Imaging, the field of digital pathology has exploded and is currently regarded as one of the most promising avenues of diagnostic medicine in order to achieve even better, faster and cheaper diagnosis, prognosis and prediction of cancer and other important diseases In the digital pathology workflow there are main challenges to be faced: In first place, the semantic gap between image descriptors and the complex histopathology patterns involved in the domain, secondly the noise in histopathology images and its subsequent feature extraction process[32], and in addition to these, there are several challenges in making the process of annotation and visualization of the features extracted more efficient and useful for the pathologist, particularly the problem of minimize the labeling effort from the pathologist has not been fully addressed [17], for this reasons there are few successful attempts of introducing computer-aided decision support systems into the medical practice [23].

This thesis can be classified in the research areas of active learning [44] algorithms, deep learning [3] and visualization. Its applications are in the emerging research areas of digital and computational pathology [32, 10, 22, 23] and also in more classical applications such as automatic biomedical image analysis and understanding, computer-aided diagnosis, natural image analysis.

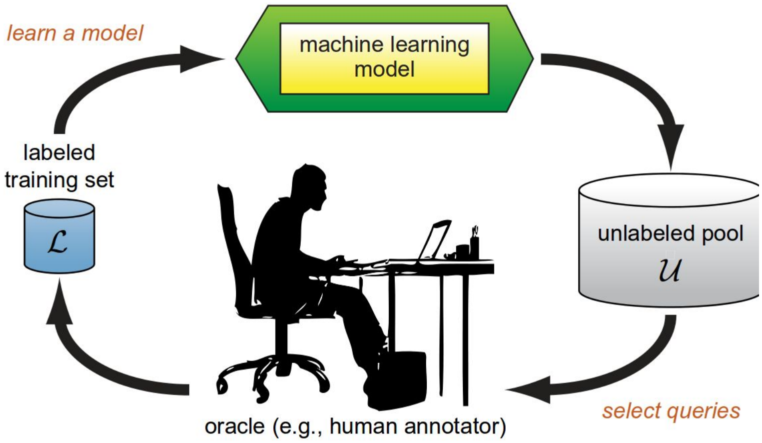

Figure **2-1**: The active learning framework, in which the label oracle is represented by a human annotator that dynamically annotates unlabeled samples to include them in the training set, image taken from [44]

2.1 Active Learning and applications to medical imaging

In a classical supervised (passive) machine learning model, there is a whole annotated data set from which the model learns the patterns given the pairs of samples and labels, in active learning (AL) [44]the main assumption is that one can learn such model with a reduced number of labels if one is allowed to choose from which samples to learn. Figure 1. shows a visual depiction of an active learning scenario.

An active learner may ask label queries for a given unlabeled sample and then, with an *informativeness* measure decide if it is included in training dataset along with its label, in this way the expensive time of the human annotator is reduced. Active learning has been a topic of significant research over the past decades with a growing attention for both, the theoretical and practical considerations of leveraging knowledge with few data samples[44]. The active learner aims to achieve high accuracy using as few labeled instances as possible, thereby minimizing the cost of obtaining labeled data, e.g. in the medical imaging domain, where the time for a medical specialist to exhaustively examine the images where the patterns reside is expensive. Probably the most cited and and relevant survey of the AL field is [44] where the author compiles the relevant literature of the field and present it in a coherent way up to 2008, nevertheless there are other comprehensive references for active learning that includes [13][4][7].

There have been several attempts to apply the concepts and algorithms of active learning in the medical domain. In [21] the author explores active learning algorithms as a way to reduce the requirements for large training sets in medical text classification tasks obtaining statistically significant better performance that the passive algorithms. In [27] Hoi et. al.

present a framework for batch active learning that applies the Fisher information matrix to select a number of informative examples simultaneously. They tackle the computational challenge of how to efficiently identify the subset of unlabeled examples that can result in the largest reduction in the Fisher information. To resolve this challenge, they propose an efficient greedy algorithm that is based on the property of submodular functions, their results with five UCI datasets and one real world medical image classification show that the proposed batch mode active learning algorithm is more effective than the state-of-the-art algorithms for active learning.
There is a particular interest in the community of active learning on the task of automatic segmentation, for instance In [50] the authors formulate interactive 3D image segmentation in an AL framework. Specifically, they evaluate a given segmentation by constructing an uncertainty field over the image domain based on boundary, regional, smoothness and entropy terms. Their contributions is to being the first work to formulate interactive 3D image segmentation as a formal AL process. They validate their method against random plane selection showing an average Dice score improvement of 10% in the first five plane suggestions (batch queries). Furthermore, their experiments shows that the method saves 64% of user's time, on average.
Wang et. al. [53] focus on how to actively recommend crucial regions to reduce user inputs. The main contribution lies in two aspects: first, they propose an approach which can successively recommend informative regions based on random walks; second, they propose a novel criterion, maximal Expected Cost of Change, which aims to select regions that will change most on the expected confidence over all unlabeled ones. Experiments on the GrabCut database demonstrate that, compared with conventional interactive segmentation methods, their approach can significantly reduce user efforts and help more quickly achieve satisfactory results.
In [31] The authors introduced an approach to exploiting the geometric priors inherent to images to increase the effectiveness of Active Learning for segmentation purposes. For 2D images, it relies on an approach to Uncertainty Sampling that accounts not only for the uncertainty of the prediction at a specific location but also in its neighborhood. For 3D image stacks, it adds to this the ability to automatically select a planar patch in which manual annotation is easy to do.
In [52] the authors presented an approach for interactive segmentation that combines active learning with the GrowCut interactive segmentation. Using a two-way interaction approach their algorithm suggests locations for drawing gestures to the user, who in turn can label the pixels as suggested or pick where to draw. They showed that using active learning guided gesture suggestions reduces the variability of the segmentation and reduces the user interactions by almost (50%) compared to segmenting the novel images with no learning. Additionally, the learning is completely transparent to the user and does not require the user to explicitly provide a lot of labelled data for learning.
In [19] Ertekin et. al. deals with the class imbalance problem which has a negative impact

on the classifiers performance. The paper show how it is possible to use an active learning method that selects informative instances from a randomly picked small pool of examples rather than making a full search in the entire training set. By focusing the learning on the instances around the classification boundary, more balanced class distributions can be provided to the learner in the earlier steps of the learning. Their empirical results on a variety of real-world datasets shows that active learning is comparable or even better than other popular re-sampling methods in dealing with imbalanced data classification, nevertheless, the authors don't mention the sample bias, which is the problem of the learner to be biased by the samples that are close to the boundary.

In the PhD. thesis of Monteleoni [38] she analyzes and designs algorithms for learning under the following online constraints: i) the algorithm has only sequential, or one-at-a-time access to data; ii) the time and space complexity of the algorithm must not scale with the number of observations. This was an important advance in understanding more theoretical aspects of active learning.

In [56] the same line of online active learning research is explored where an active learning from data streams algorithms is devised and also they develop a minimal variance measure to derive weight updating rule for the optimization problem, in a similar way to expected change cost, similarly in [12] where the authors propose a maximum classification optimization method for actively selecting unlabeled images to acquire labels.

In [42] the authors propose a method called rule induced active learning query for constructing generic active learning queries based on rule induction. Their method is able to construct shorter and more intuitive queries that are easier for a human oracle to answer, allowing to better utilize human resources.

The authors of [20] propose an interesting statistical framework for AL called model retraining improvement. This approach is both theoretical and practical, giving new insights into AL, and competitive AL algorithms for applications which inspires the experimental setup for various experiments made for this thesis.

In [16] a generalization of the concept of active learning is introduced, called proactive learning, is designed to relax unrealistic assumptions and thereby reach practical applications, whereas in traditional AL the oracle is assumed to be infallible (never wrong), indefatigable (always answers), individual (only one oracle), and insensitive to costs (always free or always charges the same). Proactive learning relaxes all four of these assumptions, relying on a decision-theoretic approach to jointly select the optimal oracle and instance, by casting the problem as a utility optimization problem subject to a budget constraint. Results on multi-oracle optimization over several datasets demonstrating the competitivity their approach over the single-imperfect-oracle baselines in most cases.

Another important issue in the AL framework is the scalability of the algorithms, in [29] the authors partially tackle this by considering the binary feedback scenario in a multi-class classification problem and proposing an algorithm based on information theoretic measures.

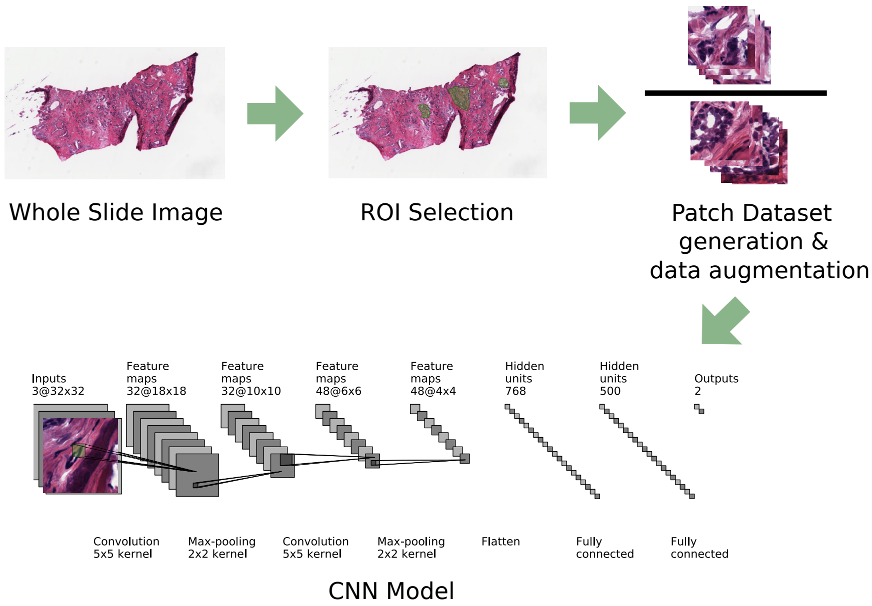

Figure **2-2**: A typical framework for applying supervised deep learning approaches like CNN to patch classication in histopathology images

2.2 Deep Learning and Unsupervised Representations

Deep learning (DL) techniques are the most studied and successful kinds of machine learning nowadays. In the last decade, DL based techniques have been shown to outperform classical machine learning algorithms with hand-crafted features in diverse fields such as computer vision [33], speech recognition[15], natural language processing [24] and also recently in biological elds like functional genomics [39].

Since early 2010's the neural networks based representations have regained interest in the machine learning and particular in computer vision and speech recognition communities because they have been systematically surpassing state of the art hand-crafted representations in computer vision challenges and standard benchmark datasets [43], this is in part due to the following factors: feasibility of training models with a great learning capacity with a large number of hidden layers, improved accuracy and relatively fast training times with the aid of GPU computing[35, 37, 3, 18, 34], a large number of practical considerations that have been studied in the last decade for monitoring the training process and selection of hyperpameters of deep models [2]. In figure 2.3, a comparison between the trends of search for this field of study is compared with the machine learning one, an interesting shared elbow in the graph that could be related with the increasing of attention by the media and several companies that have made use of such technologies is evident. The convolutional neural network (CNN) is one of the most studied deep supervised models nowadays, with this model is possible to learn hierarchical visual representations which are of particular interest for biomedical applications where one is interested in finding

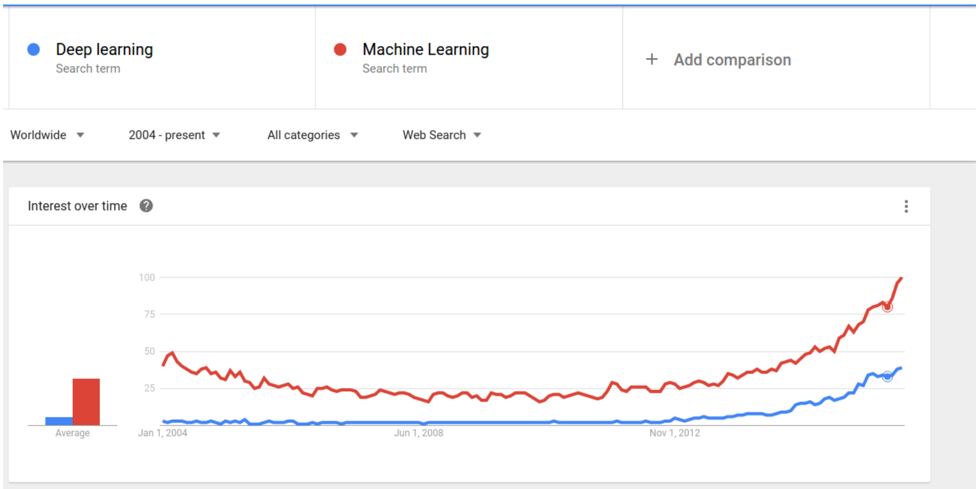

Figure **2-3**: Interest over time of the of active learning (above) and machine learning-deep learning concepts , image taken from google trends

the building blocks of the high level features extracted with the aim of havin additional interpretability, an example CNN for histopathology patch-based image classification is depicted in figure 2.2.
In Cruz-Roa's PhD. thesis the suitability of these methods for the histopathology image analysis domain is studied with promising results[54, 1, 11, 9, 8] and providing some insights about the future research work on this field. One of the research focus of this work, will be to study and devise the use of these deep learning techniques in the context of active learning.

Chapter 3

Training Convolutional Neural Networks with Active Learning

3.1 Introduction

Deep convolutional Neural Networks (CNN) are a particular kind of a supervised multi-layer perceptrons inspired by the visual cortex[6]. The CNN are able to detect visual patterns with minimal pre-processing, trained with the robustness to respond to the distortion, variability and invariances to the exact position of the pattern, and being benefited of data augmentation that subtle transform the inputs for learning more invariances[33]. We will call an *architecture* an arrangement of the parameters that are learned by an optimization algorithm. The architecture of the CNN is usually composed by stacking convolutional, pooling, normalization and fully connected layers, a typical CNN architecture is depicted in Figure3.1. The convolutional layer is a set of learnable windows or filters moving through a stride with a kernel size that represent the receptive field. Each window is convolved computing the dot product between the filter and the input generating an activation map for that filter. Pooling layer is a non-linear function to reduce the size of the convolutional layer by extracting the most representative value in a window defined by a kernel with a given stride. Max Pooling is a particular pooling function that is commonly used in architectures in computer vision and it works by choosing the maximum activation of the filter in a particular neighbourhood. Usually the last layer is a fully-connected that is a layer where all the neurons have full connections among all the neurons in the previous layer, and its non-linear function is a soft-max activation function, that outputs the probability for each of the output classes.
The deep learning algorithms have the drawback of being really data intensive algorithms, because for a successful application of this models, thousands or hundreds of thousands, sometimes even millions of labelled data samples are required for the model to converge, this is in part due to the high capacity that has to be fit and codified in a really big parameter matrix where all the weights that represents the internal configuration of the

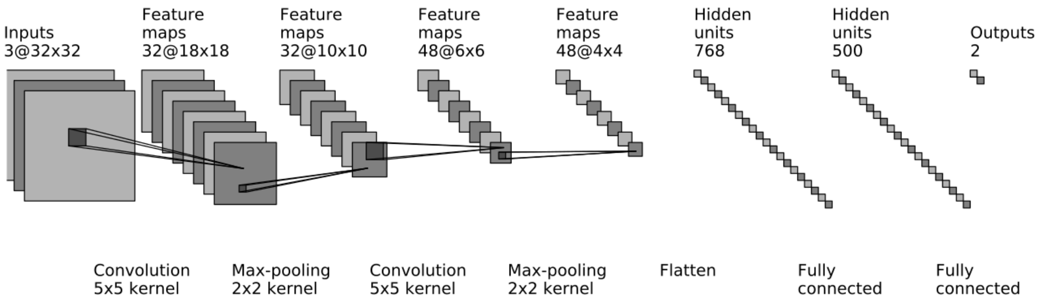

Figure **3-1**: A typical deep convolutional neural network architecture, with 7 layers and a final layer with two outputs that represents the probabilities for the sample to belong to the output classes.

network have to be set with an optimization algorithm.

The stochastic gradient descent (SGD) algorithm [5]has been used as the *de facto* algorithm to optimize loss functions in deep architectures, this optimization algorithm works well with this this kind of models because is able to achieve a good local optima processing iteratively hundreds of thousand of samples packaged in small *batches*. The main hypothesis when using this algorithm is that one have as many labelled samples available as needed for the algorithm to converge, for this reason, in SGD the batch samples are fed randomly to the model up to convergence or over-fitting. Active learning models helps to alleviate the problem of having that many labelled samples by selecting only a few selected annotated samples to be used in the training of such deep models, in the following subsection we explain how this can be accomplished when training deep convolutional neural networks.

3.2 Active Learning Model For CNN

Traditional supervised learning algorithms use whatever labelled data is provided to induce a model. By contrast, active learning gives the learner a degree of control by allowing it to select which instances are labelled and added to the training set. A typical active learner begins learning with a small labelled set $\mathcal{L}$, selects one or more informative query instances from a large unlabelled pool $\mathcal{U}$, learns from these labelled queries (which are then added to $\mathcal{L}$), and repeats[46]. The principle behind active learning is that a machine learning algorithm can achieve similar or even greater accuracy when trained with fewer training labels than the fully supervised one if the algorithm is allowed to choose the data from which it learns from [45]. An active learner may pose queries, usually in the form of unlabelled data instances to be labelled by an oracle (e.g., a human annotator). Active learning is well-motivated in many modern machine learning problems, where unlabelled data may be abundant or easily obtained but labels are not, this is an interesting direction for the so-called *deep learning in the small data regime*[1], where the objective is to train the time consuming and high sample complexity algorithms, with less resources, as in the case of medical imaging.

3.2.1 Expected Gradient Length

SGD works by stochastically optimizing an objective function J with respect to the model parameters θ, this is, finding the model parameters by optimizing with only one random sample or random sample batches instead of the full training dataset:

$$\theta_{t+1} = \theta_t - \eta \nabla J_i(\theta) \tag{3-1}$$

Where $J_i(\theta_t)$ is the objective function evaluated at the i-th sample tuple (x^i, y^i) at iteration t, η is the learning rate and ∇ is the gradient operator.

Since for computing $\nabla J_i(\theta)$ we need the i-th sample representation and its corresponding label, if we measure the norm of this term, i.e. the gradient length term$\|\nabla J_i(\theta)\|$, this will quantifies how much the i-th sample and its label contributes to each component of the gradient vector. A natural choice for selecting the most informative patches for each batch iteration of SGD is to select the instances that gives the highest values for the gradient length weighted by the probability of that sample having the y^i label. In other words, to select that instances that would impact the greatest change to the current model as if we knew their labels:

$$\Phi(x^i) = \sum_{j=1}^{c} p(y^i = j|x^i) \|\nabla J_i(\theta)\| \tag{3-2}$$

Where c is the total number of labels or classes, the Expected Gradient Length algorithm (EGL) works by sorting the Φ values from an unlabelled pool of samples and then adding them to the training dataset by asking an oracle to give us the ground truth label of those samples. The EGL algorithm was firstly mentioned by Settles et. al. [47] in the setting of multiple-instance active learning, nevertheless to the best of our knowledge this is the first time that is used in the selection of samples in CNN.

[1]In 2016 there was a workshop in the International Conference on Machine Learning dealing with this topic https://sites.google.com/site/dlworkshop16/, also in the 2016 International Conference on Medical Image Computing and Computer Assisted Intervention, there was a workshop dedicated to study and evaluate particular solutions to this topic in medical imaging:http://campar.in.tum.de/LABELS2016/WebHome .

3.3 Expected Gradient Length for Convolutional Neural Networks

3.3.1 EGL for Patch Selection in Convolutional Neural Networks

For being able to select the most informative samples in a CNN architecture we have to compute the two terms involved in equation (3.2.2), first, for the probability of a sample having the j-th label we can perform a forward propagation through the network and obtain the corresponding probabilities from the soft-max layer of the network, secondly, to measure the gradient length we can perform a backward propagation through the network to measure the Frobenius norm of the gradient parameters, in a CNN architecture we have the flexibility to compute the backward/forward phases up to a certain layer, in our experiments in next section we made the backward down to first fully connected layer as this values showed no significative difference for in-between layers. This process must be done over all the possible labels for each sample. Once we have computed the Φ values for all the samples, we sort them and select the k samples with higher EGL values.
In the first iteration, a small portion of labelled samples $\mathcal{L}' \subset \mathcal{L}$ is used to train an initial model $\mathbf{M}$, and then incrementally adding the k samples to $\mathcal{L}'$ to update $\mathbf{M}$ parameters. The steps of the algorithm are depicted in Algorithm 3.3.1.

Algorithm 3.1 EGL for Active Selection of patches in a Convolutional Neural Network

Require: Patches Dataset $\mathcal{L}$, Initial Trained Model $\mathbf{M}$, Number k of most informative patches

```
while not converged do
    Create and shuffle batches from L
    for each batch do
        Compute Φ(x) using M, ∀x ∈ batch
    end for
    Sort all the Φ Values and return the higher k corresponding samples L_k
    Add the samples in L_k to L'. Update M using L'
end while
```

3.3.2 EGL for Image Selection in Convolutional Neural Networks

Since we are able to compute the most significant patches it is straightforward to extend the procedure to select not only the most informative patches but also the most informative images within training set, the modification is that instead of computing the EGL values for all the ground truth exudate and healthy patches we compute the *interestingness* of an image by patchifying the image with a given stride and then densely computing Φ, then sorting the images by their top EGL values and finally adding the labels and patches that belongs to the more interesting image to the training set for further parameter updates using Algorithm 3.3.1 until convergence, we believe that this is a more realistic scenario where medical specialists does not have the time to manually annotate all the images but only the ones that contains more information to train a computer aided system. The full algorithm is described in Algorithm 2.

Algorithm 3.2 EGL for active selection of images in a convolutional neural network.

Require: Training Image Set $\mathcal{T}$, Patch Dataset $\mathcal{L}$, Number μ of initial images to look at
  Select an initial set $\mathcal{T}_\mu$ of images randomly
  Train initial model $\mathbf{M}$ using the ground truth patches from the μ images
  while not converged **do**
    for each image in $\mathcal{T} \setminus \mathcal{T}_\mu$ **do**
      Patchify image and compute $\sigma_{image} = \sum_{patch \in image} \Phi(patch)$, using $\mathbf{M}$
    end for
    Sort all the σ_{image} values and return $\mathcal{I}_{max}$, the image with higher sum
    $\mathcal{T}_\mu = \mathcal{T}_\mu \cup \mathcal{I}_{max}$
    $\mathcal{L}_\mu = \{$ patch $\in \mathcal{L}_\mathcal{I}, \forall \mathcal{I} \in \mathcal{T}_\mu\}$
    Update $\mathbf{M}$ with k selected patches using Algorithm 3.3.1 and the patches in $\mathcal{L}_\mu$
  end while

We can also plot an interestingness mask based on the computed EGL values of the images simply by reshaping the EGL values of all the patches to the image size, this is illustrated for the application of this algorithms to the problem of exudate detection in the next chapter, where we plot what regions of the image will be more useful to train a CNN model.4.4

Chapter 4

Training Deep Convolutional Neural Networks with Active Learning for exudate classification in eye fundus images

4.1 Introduction

In this chapter the evaluation of the EGL for CNN technique described in the previous chapter will be performed on the problem of detecting exudates in eye fundus images.

Diabetes Mellitus is one of the leading causes of death according to the World Health Organization[1]. Diabetic Retinopathy (DR) is a condition caused by prolonged diabetes, causing blindness worldwide in the productive age (20-69 years). The main problem is that most people have no symptoms and suffer the disease without have been timely diagnosis. Because the retina is vulnerable to microvascular changes of diabetes, diabetic retinopathy is the most common complication of diabetes. Eye fundus imaging is considered a non-invasive and painless route to screen and monitor DR[48]. Diabetic retinopathy has four phases: I) Nonproliferative diabetic retinopathy (NPDR) - in this earliest stage, exudates and microaneurysms occur, which are small areas of inflammation in balloon shape in the tiny blood vessels of the retina; II) moderate nonproliferative diabetic retinopathy (MNDR) - as the disease progresses, a few of the blood vessels that nourish the retina are blocked; III) severe non-proliferative diabetic retinopathy (SNDR) - blocking many more blood vessels occurs, which prevents the blood supply to various areas of the retina, and IV) proliferative diabetic retinopathy (PDR) - at this late stage, the signals sent by the retina for nourishment trigger the growth of new blood vessels. These new blood vessels are abnormal and fragile, growing along the retina surface with transparent vitreous gel

[1]http://www.who.int/diabetes/en/

which fills the interior of the eye.

The manual segmentation of exudates in eye fundus images, a key step to classify the grade of DR, is very time consuming and repetitive for clinical personnel[48]. Moreover, the analysis of many images without any pathologies increases the work time and leaves less time to analyze images with pathologies in detailed form. Computer-aided diagnosis systems (CADx) are a good solution to optimize the work of ophthalmologist, giving the proper time for patients with progressing disease or critical eye conditions, that require all of the ophthamologist attention. CADx helps to perform automatic grading of the disease, increasing the number of patients diagnosed and supporting early detection. This helps to reduce the cost of manual labelling [51], most CADx systems rely on techniques from computer vision and have evolved along with the advances in this field[30], which is shown by the low specificity performance of earlier systems [55][49] which are based on morphological and appearance features in combination with classical machine learning algorithms[26][41].

In recent years, deep learning techniques have greatly surpassed the performance of computer vision systems[33], such as deep convolutional neural networks (CNN), firstly used for classify natural images and recognize digits and now have start to being used in the biomedical imaging workflows and in particular to play an important role in DR grading showing superior performance in several settings and datasets, for example, in 2015 the data science competitions website Kagle[2] launched a DR Detection competition were both the winner and top entries won using CNN in more than 35000 labeled images, demonstrating that for a succesful training of such algorithms a significative amount of labeled data is required. This presents the problem where the algorithms involved in CADx for DR have to be feed with the order of thousand of samples which in practice is really hard in both time and money, this impose the challenge of how to transform the good performance algorithms such as CNN to be less data intensive and thus able to learn only with a few selected number of samples.

The rest of this chapter is organized as follows: First, in section 4.2, we give an overview of the architecture for the deep neural network model and the preprocessing steps to handle the eye fundus images that fit the model. The active learning strategy is explained in section **??**. In section 4.3 we describe the experimental setup using the reference baseline [**?**] for dataset setup and parameters of architecture and the performance reported on it. In Section 4.4, the experimental results are presented and discussed. Finally, in section **??** we discusse how to interactively use the EGL algorithm to propose masks of interest regions for further reduction labeling effort.

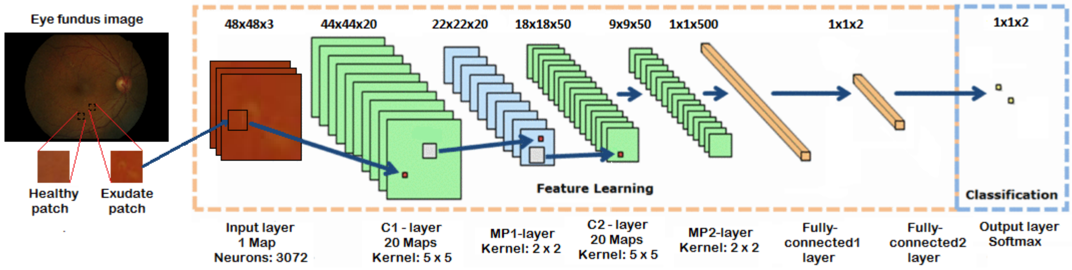

Figure **4-1**: Deep CNN architecture to classify between healty and exudate patches

4.2 Deep Learning Model

Supervised CNN models are one of the most successful deep learning models for computer vision and medical imaging field is rapidly adapting this models to solve and improve in a plethora of applications[25]. Our deep learning model is based in the LeNet CNN architecture [36] with 7 layers as shown in the inner block of Figure 1, this architecture is composed of a patch input layer followed by two convolutions and max pooling operations to finalize in a softmax classification layer that outputs the probability of a patch being healty or exudate.
The first stage of the block diagram shown at Figure 1 is the cropping of the eye fundus image with size ranging from 1440×960 to 2540×1690 pixels. The extraction of healthy and exudate patches of 48×48 pixels were made as follows, for healty a stratified set of patches were selected in which the borders and internal sections of the eye were both considered in order to train a more robust model, for exudates patches, were considered positives just the ones that exceed on a threshold of a 60% the exudate area.

4.2.1 Preprocessing

Preprocessing is a usual step in the medical image processing pipeline to enhance the characteristics of the image. The application of a set of transformations may improve the performance in the following stages. We enclose the exudate in a bounding box in order to extract the Region of interest (ROI) from the eye fundus image. Computer-aided diagnosis (CADx) systems aim at classifying a previously identified ROI in the whole film image. This ROI can be obtained by a manual segmentation or automatically detected by a computer aided detection system. Because of lesions in e-ophtha dataset 4.3.1 were manually segmented, we fixed the input size to ROIs of 48×48 pixels according to the mean of the lesion's size. With this, ROIs can be easily extracted by taking a bounding box of the segmented region. Specifically, images were cropped to the bounding box of the lesions, where the lesion is centered without scaling and preserving the surrounding region. The condition to label a patch as a true exudate is that the intersection of the

[2]https://www.kaggle.com/c/diabetic-retinopathy-detection

patch with an exudate region be greater than the 60% of the ROI. Otherwise, the patch was labelled as healthy patch.

Data augmentation The expressiveness of neural network models, and particularly deep ones, comes mainly from the large number of parameters that are learnt. However, more complex models also increase the chance of overfitting the training data. Data augmentation is a good way that helps to prevent this behaviour [33]. Data augmentation is the process of artificially create new samples by applying transformations to the original data. In a classification problem, data augmentation makes sense because an exudate can be presented in any particular orientation. Thus, the model also should be able to learn from such transformations. In particular, for each training image, we have artificially generated 7 new label-preserving samples using a combination of flipping and $90, 180$ and 270 degrees rotation transformations. 4.4

4.3 Experimental setup

4.3.1 Ophtha Dataset

The e-ophtha database with color fundus images was used in this work. The database contains 315 images with size ranging from 1440×960 to 2540×1690 pixels, 268 images with no lesion and 47 with exudates which were segmented by ophthalmologists from the OPHDIAT Tele-medical network under the the French Research Agency (ANR) project [14].
The labeled patch dataset was created with $48{\times}48$ pixel patches that contain both exudate and healthy examples. After the preprocessing steps of cropping and data augmentation, the dataset splits were built with randomly selected patches of each class as follows: a training split with 8760 patches for each class, a validation split with 328 per class and a test split with 986. Images of a given patient could only belong to a single group according to the described dataset distribution. At test time, only patches of unseen patient images in trainng are forward propagated in the trained network to obtain their class probabilities.

4.3.2 Evaluation

The technique of Decencieriere et al[14] was chosen as our baseline. A method based on machine learning and image processing techniques is proposed to detect exudates in eye fundus images reporting specificity and sensitivity in a patch-wise experimental setup using the e-ophtha dataset. The base LeNet model was trained using stochastic gradient descent (SGD) from scratch without any trans- fer learning from other datasets. The learning rate and batch size were explored in a grid search and showed robustness in the

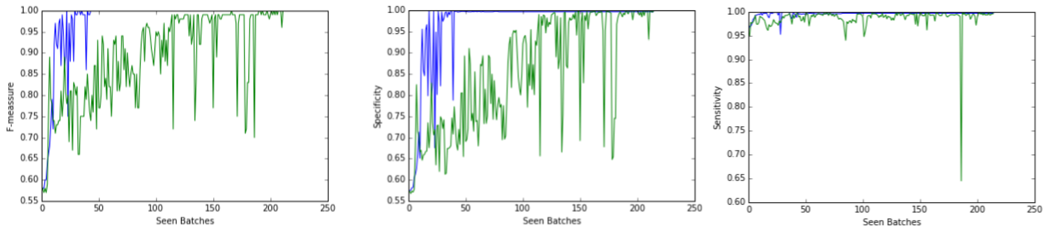

Figure **4-2**: Results for F-meassure, Sensitivity and Specificity, using Random Strategy (Green) and EGL (Blue)

range of 32-64 in terms of batch size with a learning rate of 0.01 when trained with all the training patches. In our final experiments we set the batch size to 32 and 0.01 for the learning rate, using 30 as the number of epochs to train the model.
The model **M** is the LeNet CNN model in Figure 4.2, for the patch selection algorithm we made the initial training with 5 batches of 32 samples and select for each call of EGL the 32 most informative patches, as this fits the batch size.
The proposed approach was implemented with Python 2.7 and the Caffe deep learning framework [28], which allows for efficient access to parameters and data in memory, so the . We use a NVIDIA GTX TITAN X GPU for our experiments. During all the experiments, training loss and validation loss, as well as the accuracy over the validation set were monitored.

4.4 Results

Figure **4-2** shows the results of the evaluation of the patch active selection algorithm (Algorithm 3.3.1) compared to a strategy that randomly select patches from the training data set.
When selecting the most informative patches for training the deep CNN using Algorithm 3.3.1 we can see an important improvement in terms of sample convergence. With as few as 50 batches ($50 \times 32 = 1600$ patches), the EGL approach is able to converge whereas the usual random SGD strategy takes up to 200 batches, or 6400 patches. This shows that with only 25% of the annotated dataset, our model is able to achieve the expected performance of a fully annotated strategy. We want to test our image selector algorithm (Algorithm 3.3.2) in the more realistic scenario where an ophthalmologist selects only a few important or relevant images instead of patches to annotate and train the model.
With this approach we also have interesting results, as shown in Figure 3, where the left side of the orange line is when the initial model training is performed. Then, the Algorithm 3.3.2 is used to select the most interesting image for the model and subsequently to update the model. It is interesting that the convergence is reached even at an earlier stage than when using the patch strategy (see Figure **4-3**). As few as 30 batches are enough for

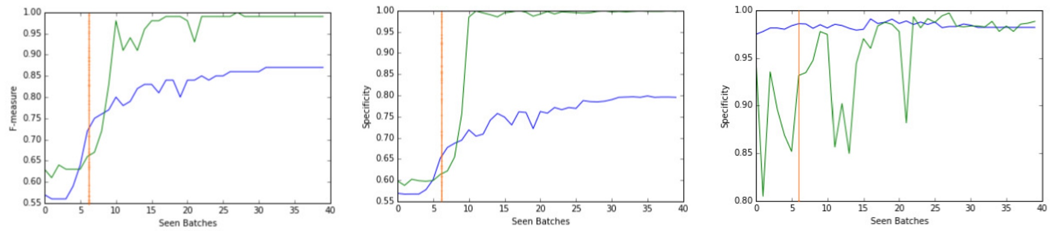

Figure **4-3**: Results for F-Measure, Sensitivity and Specificity, using a random strategy (blue) and active learning using EGL (green) for image selection. In this setup only the patches of the 4 initial training images were used for training the model in the first 6 SGD iterations, after this (orange line) we add the patches from the images with maximum EGL value to the training set.

Method	Sensitivity	Specificity
Decencieriere et al [14]	90	70
Full training dataset	**99.8**	99.5
Our approach with 25% Samples	98.7	**99.7**

Table **4-1**: Performance measures in the baseline model and the proposed method.

the model convergence, showing that in this more realistic scenario our strategy also outperforms the standard way of training deep CNN models.

4.4.1 Measuring and Visualizing Interestingness

Once we have an initial training of the model we can measure the interestingness of a full image computing the sum of its EGL values. This was the criteria for selecting images in algorithm 3.3.2. We can plot this value and see how this evolves as the model sees more batches.

These values are illustrated in Figure **4-5**. Here we can see how the interestingness value decays after the model has converged, when the loss function does not decrease anymore and the norm of the parameters is nearly 0. In Table **4-1** the accuracy, sensitivity and specificity of our proposed method are reported and compared with the baseline method and the CNN model trained with the full dataset. The proposed method clearly outperforms our baseline in both sensitivity and specificity and obtains almost the same performance of the classic SGD strategy that sees the entire dataset randomly. This shows that the proposed method is able to capture the visual features that characterize exudates even when there is a limited annotated dataset.

Figure **4-4**: A test image with several exudate areas for testing the interestingnes mask below

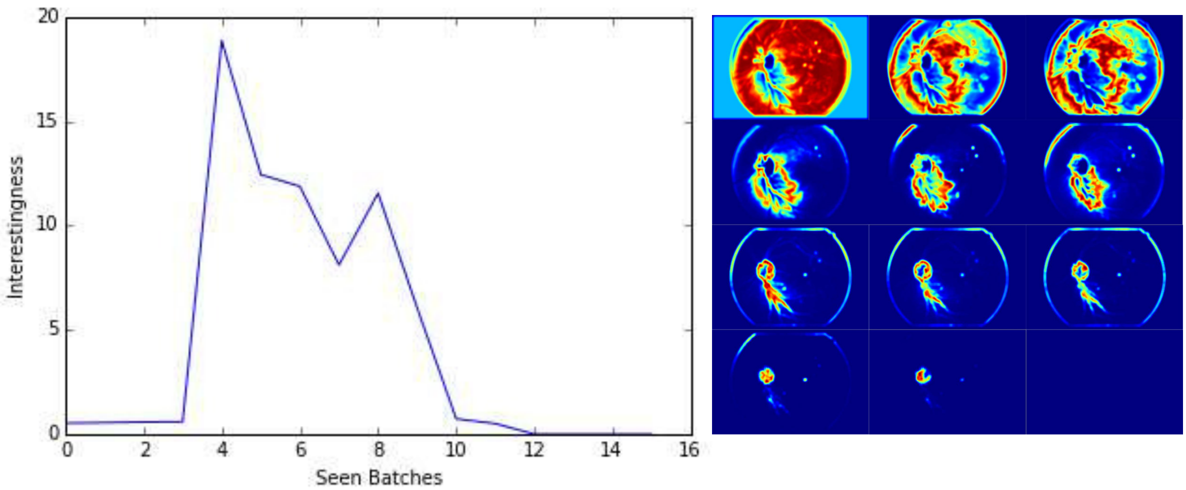

Figure **4-5**: Interestingness over training time. After the model converges the interestingness value decays to 0 because the norm of the gradient is close to 0.

Chapter 5

Conclusions and Future Work

5.1 Conclusions and discussion

In this thesis, methods for sample-efficient training of deep models using active learning were presented . The novelty of the work here presented resides in three main issues:

1. Exploring and adapting active learning algorithms to make more label-efficient deep convolutional neural networks

2. Show their feasibility to apply to medical imaging workflows where data is scarse and expensive.

3. Present a visualization for label the most interesting parts within an image.

Some of our methods show that even with an small portion of the training dataset of ~ 25 of original samples were enough to train the model with almost the same performance of the model that was trained with all the samples, showing the feasibility of active learning strategies for deep CNN training.
Our approach presents a computational drawback when the number of unlabeled datasamples to check is large, but we think that this could be overcome with traditional sampling techniques. Despite our results showing good performance using only a portion of the data, we would like to do further experimentation involving scenarios where the need for labeled data is even more critical and also in large–scale datasets where the combination of our sample selection techniques with transfer learning could lead to a performance boost.
We think that active learning techniques has a promising application landscape in the challenging tasks of medical imaging using deep learning because of their potential to relief the need for large amounts of labeled data. This will allow the usage of deep learning models in a broader set of medical imaging tasks like detection and segmentation of structures in specialized domains such as histopathology image analysis or computed tomography scans where the labels are costly

5.2 Future Work

Currently important issues exists regarding the use of deep learning technologies in the medical imaging workflow, the label cost problem was partially addressed in this thesis, giving insights with methods that helps to overcome this issue. Nevertheless the following problems are currently being investigated[25] and I strongly believe that are prommising research directions:

1. Do we need to work on getting real Big Data for each medical task, or will transfer learning be sufficient?
2. Is the fusion of image modalities and other medical information fusion approaches feassible with deep learning?
3. The creation and participation of challenges: Those events provide a precise definitions for tasks to be solved and define one or more evaluation metrics that provide a fair and standardized comparison between proposed algorithms, thus, making more reliable and traceable the progress in the field.

Bibliography

[1] John Arevalo, Angel Cruz-Roa, and Fabio A González. Hybrid image representation learning model with invariant features for basal cell carcinoma detection. In *IX international seminar on medical information processing and analysis*, pages 89220M–89220M. International Society for Optics and Photonics, 2013.

[2] Yoshua Bengio. Practical recommendations for gradient-based training of deep architectures. In *Neural networks: Tricks of the trade*, pages 437–478. Springer, 2012.

[3] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.

[4] Avrim L Blum and Pat Langley. Selection of relevant features and examples in machine learning. *Artificial intelligence*, 97(1):245–271, 1997.

[5] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*, pages 177–186. Springer, 2010.

[6] Dan Ciregan, Ueli Meier, and Jürgen Schmidhuber. Multi-column deep neural networks for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3642–3649. IEEE, 2012.

[7] David A Cohn, Zoubin Ghahramani, and Michael I Jordan. Active learning with statistical models. *Journal of artificial intelligence research*, 1996.

[8] Angel Cruz-Roa, John Arevalo, Ajay Basavanhally, Anant Madabhushi, and Fabio González. A comparative evaluation of supervised and unsupervised representation learning approaches for anaplastic medulloblastoma differentiation. In *Tenth International Symposium on Medical Information Processing and Analysis*, pages 92870G–92870G. International Society for Optics and Photonics, 2015.

[9] Angel Cruz-Roa, Ajay Basavanhally, Fabio González, Hannah Gilmore, Michael Feldman, Shridar Ganesan, Natalie Shih, John Tomaszewski, and Anant Madabhushi.

Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In *SPIE medical imaging*, pages 904103–904103. International Society for Optics and Photonics, 2014.

[10] Angel Alfonso Cruz Roa et al. *Data-driven Representation Learning from Histopathology Image Databases to Support Digital Pathology Analysis.* PhD thesis, Universidad Nacional de Colombia-Sede Bogotá.

[11] Angel Alfonso Cruz-Roa, John Edison Arevalo Ovalle, Anant Madabhushi, and Fabio Augusto González Osorio. A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 403–410. Springer, 2013.

[12] Zhengwei Cui, Xiaoming Chen, Jian Wu, Victor S Sheng, and Yujie Shi. Maximum classification optimization-based active learning for image classification. In *Image and Signal Processing (CISP), 2014 7th International Congress on*, pages 759–764. IEEE, 2014.

[13] Sanjoy Dasgupta. Two faces of active learning. *Theoretical computer science*, 412(19):1767–1781, 2011.

[14] Etienne Decencière, Guy Cazuguel, Xiwei Zhang, Guillaume Thibault, J-C Klein, Fernand Meyer, Beatriz Marcotegui, Gwénolé Quellec, Mathieu Lamard, Ronan Danno, et al. Teleophta: Machine learning and image processing methods for teleophthalmology. *IRBM*, 34(2):196–203, 2013.

[15] Li Deng, Geoffrey Hinton, and Brian Kingsbury. New types of deep neural network learning for speech recognition and related applications: An overview. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 8599–8603. IEEE, 2013.

[16] Pinar Donmez and Jaime G Carbonell. Proactive learning: cost-sensitive active learning with multiple imperfect oracles. In *Proceedings of the 17th ACM conference on Information and knowledge management*, pages 619–628. ACM, 2008.

[17] Scott Doyle, James Monaco, Michael Feldman, John Tomaszewski, and Anant Madabhushi. An active learning based classification strategy for the minority class problem: application to histopathology annotation. *BMC bioinformatics*, 12(1):424, 2011.

[18] Dumitru Erhan, Yoshua Bengio, Aaron Courville, and Pascal Vincent. Visualizing higher-layer features of a deep network. *University of Montreal*, 1341, 2009.

[19] Seyda Ertekin, Jian Huang, Leon Bottou, and Lee Giles. Learning on the border: active learning in imbalanced data classification. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 127–136. ACM, 2007.

[20] Lewis PG Evans, Niall M Adams, and Christoforos Anagnostopoulos. Estimating optimal active learning via model retraining improvement. *arXiv preprint arXiv:1502.01664*, 2015.

[21] Rosa L Figueroa, Qing Zeng-Treitler, Long H Ngo, Sergey Goryachev, and Eduardo P Wiechmann. Active learning for clinical text classification: is it better than random sampling? *Journal of the American Medical Informatics Association*, 19(5):809–816, 2012.

[22] Thomas J Fuchs and Joachim M Buhmann. Computational pathology: Challenges and promises for tissue analysis. *Computerized Medical Imaging and Graphics*, 35(7):515–530, 2011.

[23] Manal Y Gabril and George M Yousef. Informatics for practicing anatomical pathologists: marking a new era in pathology practice. *Modern Pathology*, 23(3):349–358, 2010.

[24] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Domain adaptation for large-scale sentiment classification: A deep learning approach. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 513–520, 2011.

[25] Hayit Greenspan, Bram van Ginneken, and Ronald 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.

[26] Nikita Gurudath, Mehmet Celenk, and H Bryan Riley. Machine learning identification of diabetic retinopathy from fundus images. In *Signal Processing in Medicine and Biology Symposium (SPMB), 2014 IEEE*, pages 1–7. IEEE, 2014.

[27] Steven CH Hoi, Rong Jin, Jianke Zhu, and Michael R Lyu. Batch mode active learning and its application to medical image classification. In *Proceedings of the 23rd international conference on Machine learning*, pages 417–424. ACM, 2006.

[28] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 675–678. ACM, 2014.

[29] Ajay J Joshi, Fatih Porikli, and Nikolaos P Papanikolopoulos. Scalable active learning for multiclass image classification. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2259–2273, 2012.

[30] Tomi Kauppi et al. *Eye fundus image analysis for automatic detection of diabetic retinopathy.* Lappeenranta University of Technology, 2010.

[31] Ksenia Konyushkova, Raphael Sznitman, and Pascal Fua. Introducing geometry in active learning for image segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2974–2982, 2015.

[32] Sonal Kothari, John H Phan, Adeboye O Osunkoya, and May D Wang. Biological interpretation of morphological patterns in histopathological whole-slide images. In *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine*, pages 218–225. ACM, 2012.

[33] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*, pages 1097–1105. 2012.

[34] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[35] Quoc V Le, Will Y Zou, Serena Y Yeung, and Andrew Y Ng. Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 3361–3368. IEEE, 2011.

[36] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[37] Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y Ng. Unsupervised learning of hierarchical representations with convolutional deep belief networks. *Communications of the ACM*, 54(10):95–103, 2011.

[38] Claire E Monteleoni. Learning with online constraints: shifting concepts and active learning. 2006.

[39] Nick Pawlowski, Juan C Caicedo, Shantanu Singh, Anne E Carpenter, and Amos Storkey. Automating morphological profiling with generic deep convolutional networks. *bioRxiv*, page 085118, 2016.

[40] Thierry Pun, Guido Gerig, and Osman Ratib. Image analysis and computer vision in medicine. *Computerized Medical Imaging and Graphics*, 18(2):85–96, 1994.

[41] Gwénolé Quellec, Mathieu Lamard, Guy Cazuguel, Lynda Bekri, Wissam Daccache, Christian Roux, and Béatrice Cochener. Automated assessment of diabetic retinopathy severity using content-based image retrieval in multimodal fundus photographs. *Investigative ophthalmology & visual science*, 52(11):8342–8348, 2011.

[42] Roberts Rosebud O, Lynch Charles F, Jones Michael P, and Michael N. Hart. Medulloblastoma: A population-based study of 532 cases. *J Neuropath Exp Neur*, 50(2):134–144, 1991.

[43] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015.

[44] Burr Settles. Active learning literature survey. *University of Wisconsin, Madison*, 52(55-66):11, 2010.

[45] Burr Settles. Active learning literature survey. *University of Wisconsin, Madison*, 52(55-66):11, 2010.

[46] Burr Settles and Mark Craven. An analysis of active learning strategies for sequence labeling tasks. In *Proceedings of the conference on empirical methods in natural language processing*, pages 1070–1079. Association for Computational Linguistics, 2008.

[47] Burr Settles, Mark Craven, and Soumya Ray. Multiple-instance active learning. In *Advances in neural information processing systems*, pages 1289–1296, 2008.

[48] Alan W Stitt, Noemi Lois, Reinhold J Medina, Peter Adamson, and Timothy M Curtis. Advances in our understanding of diabetic retinopathy. *Clinical science*, 125(1):1–17, 2013.

[49] Kenneth W Tobin, Edward Chaum, V Priya Govindasamy, and Thomas P Karnowski. Detection of anatomic structures in human retinal imagery. *IEEE transactions on medical imaging*, 26(12):1729–1739, 2007.

[50] Andrew Top, Ghassan Hamarneh, and Rafeef Abugharbieh. Active learning for interactive 3d image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 603–610. Springer, 2011.

[51] Mark JJP van Grinsven, Bram van Ginneken, Carel B Hoyng, Thomas Theelen, and Clara I Sánchez. Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images. *IEEE transactions on medical imaging*, 35(5):1273–1284, 2016.

[52] Harini Veeraraghavan and James V Miller. Active learning guided interactions for consistent image segmentation with reduced user interactions. In *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1645–1648. IEEE, 2011.

[53] Dan Wang, Canxiang Yan, Shiguang Shan, and Xilin Chen. Active learning for interactive segmentation with expected confidence change. In *Asian Conference on Computer Vision*, pages 790–802. Springer, 2012.

[54] Haibo Wang, Angel Cruz-Roa, Ajay Basavanhally, Hannah Gilmore, Natalie Shih, Mike Feldman, John Tomaszewski, Fabio Gonzalez, and Anant Madabhushi. Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. *Journal of Medical Imaging*, 1(3):034003–034003, 2014.

[55] Tae Keun Yoo and Eun-Cheol Park. Diabetic retinopathy risk prediction for fundus examination using sparse learning: a cross-sectional study. *BMC medical informatics and decision making*, 13(1):106, 2013.

[56] Xingquan Zhu, Peng Zhang, Xiaodong Lin, and Yong Shi. Active learning from data streams. In *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, pages 757–762. IEEE, 2007.