

Improving Histopathological Tissue Classification with Transfer Learning and Unsupervised Pre-Training of Deep Neural Networks

**Verbesserung histopathologischer Gewebeklassifikation durch Transfer Learning
und unüberwachtes Pre-Training von Deep Neural Networks**

Bachelor Thesis

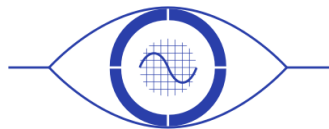
presented by

Steffen Schneider

Supervisor:

Daniel Bug M. Sc.

Updated Version of 28th November 2016



INSTITUTE OF IMAGING & COMPUTER VISION
PROF. DR.-ING. DORIT MERHOF
RWTH AACHEN UNIVERSITY
2016

Hiermit versichere ich, dass ich die vorgelegte Bachelorarbeit selbständig angefertigt habe. Es sind keine anderen als die angegebenen Quellen und Hilfsmittel benutzt worden. Zitate wurden kenntlich gemacht.

Aachen, 26. April 2016

(Steffen Schneider)

Zusammenfassung

Die Analyse von Gewebedaten in der digitalen Pathologie erfordert ein hohes Maß an Expertise und kann in der Regel nur von Experten durchgeführt werden. Die Entwicklung zuverlässiger Methoden zur automatisierten Annotation wird dabei oftmals durch die begrenzte Größe der zur Verfügung stehenden Trainingsdaten erschwert. Hinzu kommt eine stark schwankende Verteilung der vorhandenen Klassen. Oftmals sind im Datensatz stark unterrepräsentierte Klassen für die Analyse ähnlich wichtig wie häufiger auftretende. In dieser Arbeit werden eine Reihe von Methoden evaluiert, die trotz dieser Einschränkungen durch den Einsatz von Transfer Learning und unüberwachtem Pre-Training bessere Ergebnisse als herkömmliche Ansätze erzielen. Der Schwerpunkt liegt dabei auf Faltungsnetzen (Convolutional Neural Networks) mit einer großen Zahl an Layern, die so auch in der ImageNet Large Scale Recognition Challenge Verwendung finden.

Mit herkömmlichen Methoden des maschinellen Lernens ist eine Fehlerrate von 14 bis 16 Prozent erreichbar. Dabei kamen Random Forest Classifier auf Histogram-Features im LAB Farbraum zum Einsatz.

Wünschenswert wäre allerdings ein Ansatz, der unabhängig von manuell erstellten Merkmalen ein gegen Varianzen in den Daten robustes Ergebnis liefern kann. Das in den letzten Jahren verstärkt aufkommende Interesse für den Einsatz von Deep Learning hat gezeigt, dass in der digitalen Bildverarbeitung Methoden zur automatischen Extraktion von Merkmalen herkömmlichen Methoden zum Teil deutlich überlegen sind. Verbunden mit der Möglichkeit, von bereits ausgiebig getesteten Architekturen Gebrauch machen zu können, motiviert dies den Einsatz von Deep Learning auch in der digitalen Pathologie.

In der vorliegenden Ausarbeitung werden daher diverse Methoden des Deep Learning im Zusammenhang mit der automatisierten Gewebeannotation evaluiert. Durch den Einsatz von Transfer Learning mit den VGGnet und GoogLeNet Architekturen sowie mithilfe von Autoencodern und Restricted Boltzman Machines wird gezeigt, dass es auch auf begrenzten Datensätzen möglich ist, leistungsstarke State-of-the-Art Ansätze des Deep Learning zu verwenden.

Mit Methoden des unüberwachten Lernens wird gezeigt, dass die aus den Daten gewonnenen Gewichte sehr stark jenen ähneln, die bereits für die gängigen CNN Architekturen zur Verfügung stehen. Aus diesem Grund wird ein besonderes Augenmerk auf die weitere Evaluation von Transfer Learning gelegt. Auf dem ILU-224 Datensatz konnte die Genauigkeit der Klassifikation mithilfe von Transfer Learning von 84.2% auf 96.5% verbessert werden. Die Fähigkeit zur Generalisierung ist bei den herkömmlichen zum Einsatz kommenden Methoden nicht gegeben und wird erst durch den Einsatz von Deep Learning möglich. Allerdings ist es nicht eindeutig, welche Netzarchitektur für die weitere Betrachtung am sinnvollsten zu wählen ist. Eine Kombination der schlussendlich gewonnenen Modelle wird genutzt, um eine Reihe an vorher nur schwach annotierten Datensätzen zu analysieren.

Abstract

The analysis and annotation of tissue data from histopathological whole slide images is a complex task requiring knowledge of the domain when labeling is done by hand. The design of robust computer vision architectures automating this task can be tedious because of the lack of labeled data for testing and training, further complicated by the fact of notable differences in the number of training examples per class. In this thesis, methods to overcome those constraints are developed either by making use of pre-trained architectures from the current state of the art convolutional networks used in the ILSVRC classification task as well as unsupervised convolutional models for feature extraction. Misclassification rates using previous approaches reached between 14 and 16 percent accuracy the ILU-224 dataset using histogram-based features in LAB colorspace along with a random forest classifier.

Since classical approaches using hand-crafted features can become involved when it comes to the right feature design, a robust approach, easily portable to different domains of medical computer vision problems, is desirable. The recent success of deep neural networks motivates the investigation of methods to automatically construct good feature representations directly from data and thereby benefit from already established, complex architectures originally developed for the use in other domains.

In this thesis, various methods for the design of deep architectures for tissue classification are presented. By using transfer learning and unsupervised feature learning, it is shown that powerful state of the art models with millions of parameters can be finetuned to outperform previous approaches despite the lack of sufficient labeled training examples. Several models such as the 16-layer VGGnet, the GoogLeNet model with some extension, convolutional restrict boltzman machines and convolutional denoising autoencoders were trained on the ILUMINATE-9 dataset. Along with evidence provided on how the training policy for the networks should be designed, a whole model zoo, trained on the ILUMINATE-9 dataset, is provided along with this thesis.

Results from unsupervised learning techniques show that the learned low-level filters resemble the ones already incorporated in current CNN architectures. Pretraining with weights obtained with unsupervised is therefore not necessary, since weights adapted to the ImageNet dataset provide good results also for histopathological image analysis. On the ILU-224 whole slide image, classification performance is improved from from 84.2% to 96.5% in terms of accuracy when using transfer learning with weights from the VGG 16-layer model. Concerning generalization performance, the approaches discussed are able to outperform the previously tested classifiers, however no clear winner emerged during the experiments. An ensemble of the models finally derived is used to label a range of previously weakly annotated whole slide images than can now be further examined.

Contents

1	Introduction	1
1.1	Outline: Building strong classifiers through Unsupervised Learning and Feature Transfer	2
1.2	Full-Scale Deep Learning Models in Texture Classification	3
1.3	Outline of the thesis	4
2	Preliminaries	5
2.1	Notation	5
2.2	Domain Description: Texture Classification	6
2.3	The Motivation for using Deep Architectures	8
2.4	Models for unsupervised and supervised learning	10
2.5	Convolutional Neural Networks	13
2.5.1	Forward and Backward Passes Through a Network	13
2.5.2	Building blocks of neural networks	14
2.5.3	Generating training examples with data augmentation techniques	23
2.6	Unsupervised Learning for Pre-Training and Stain Normalization	24
2.6.1	Dimensionality reduction vs. Sparse, overcomplete representations	25
2.6.2	Convolutional Restricted Boltzman Machines for Image Processing	25
2.6.3	Learning in Restricted Boltzman Machines (RBMs) using Contrastive Divergence	29
2.6.4	Sparsity criterion	31
2.6.5	From Boltzman Machines to Autoencoders	31
2.7	Transfer Learning	32
2.8	Feature Evaluation	36
2.8.1	Graphical analysis using t-SNE	36
2.8.2	Incorporating class priors	36
3	Implementation	39
3.1	Processing Pipeline	40
4	Experiments	43
4.1	Unsupervised Learning	44

4.2	Transfer Learning with Deep Convolutional Networks	47
4.2.1	VGGnet	48
4.3	GoogleNet and Variants	49
4.4	Feature Evaluation	50
4.5	Classification of new whole slide images	51
4.6	Further experiments	56
5	Discussion	59
5.1	The Choice of depth	60
5.2	Learning with few training examples	61
5.3	Conclusion	62
A	Derivations	65
A.1	Probability Distributions	68
A.1.1	Computing Expectations with Sampling	69
A.2	MAP and ML	69
B	Algorithms and Software	71
B.1	Data Augmentation on PCA channels	71
B.2	Software packages	72
B.3	Practical Issues when training deep networks	72
C	Datasets and Model Evaluation	73
	List of Figures	82
	List of Acronyms	83
	Nomenclature	85
	Bibliography	87