Lung cancer detection using Deep Learning

Fabian M. Falck*
Institute for Interdisciplinary Information Sciences
Tsinghua University
Beijing, 100084
fabian.falck@web.de

Abstract

The abstract paragraph should be indented ½ inch (3 picas) on both the left- and right-hand margins. Use 10 point type, with a vertical spacing (leading) of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

1 Introduction

2 Related Work

3 Method

4 Data

4.1 Data set

The data set contains 723 lung cancer scans. Each scan consists of on average XXX images which are the observations. Each observation has an interpolated pixel area of $40 \times 40 \times 40$, i.e. each pixel represents an area of approx. $1mm^3$. The values of the pixels are given in a linear transformation Hounsfield unit (HU) scale defined as

$$HU = 1000 \frac{\mu - \mu_{water}}{\mu_{water} - \mu air} \tag{1}$$

where μ_{water} and μair are the linear atenuation coefficients of water and air **?**. The linear transformation was applied to guarantee non-negative values which is necessary for training our model. Furthermore, each image was hand-labeled by physicians if the image contains a nodule or not. A nodule is a positive lung cancer observation which can occure in different types. The types are further discussed in section 4.2. Each observation is labeled with either a 1 (true nodule) or a 0 (false nodule). Additional information for each obervation is a nodule ID. Since a true nodule can occure in multiple observations, this ID allows to identify the unique nodules.
The data set is highly unbalanced:

4.2 Nodule types

*Website: fabianfalck.com

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.