Learning histopathological patterns

Andreas Kårsnäs

Industrial PhD Student

Centre for Image Analysis
Swedish University of Agricultural Sciences
Uppsala University

visiopharm
TURNING IMAGES INTO KNOWLEDGE

UPPSALA UNIVERSITET

Histopathology

The examination of tissue in order to study the manifestations of disease.

• Diagnosis from *immunohistochemically* stained tissue plays a major role when diagnosing cancer.

• We propose a technique for analyzing stained tissue samples for extracting features that correlate with the disease of the patient.

• More specifically, we address the problem of segmenting cell nuclei.
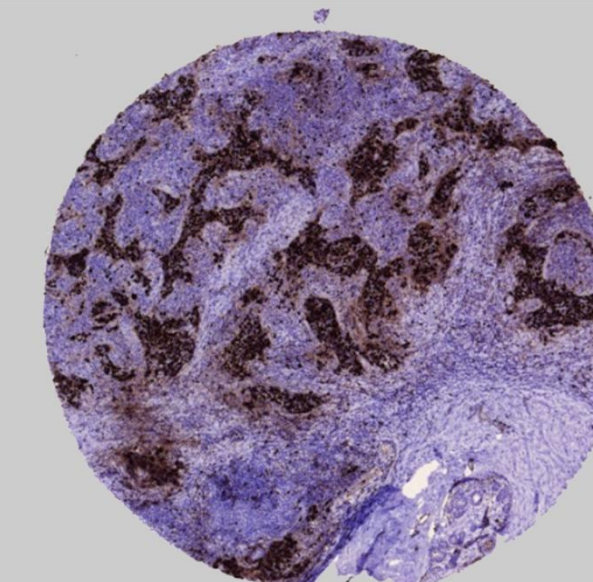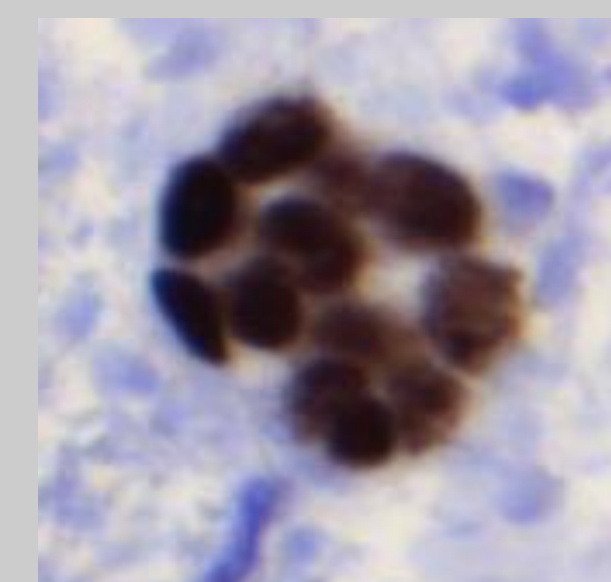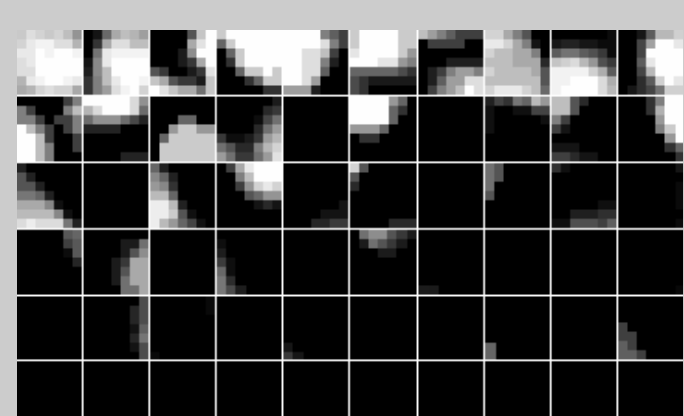
Tumor tissue   **Positive nuclei**

Learning patterns

The technique behind the segmentation is called **Sparse Label Dictionaries** and is based on Sparse coding, where the image is typically modeled as a linear combination of a few elements from a dictionary.

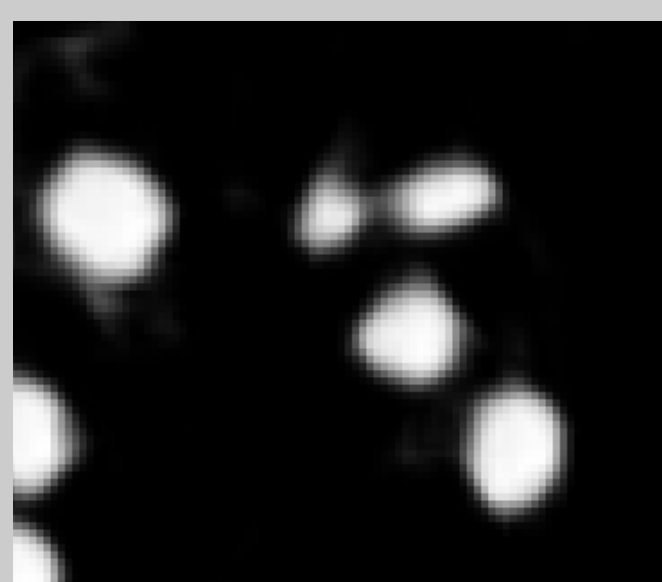Compared to more classical usages of sparse coding, Sparse Label Dictionaries use two coupled dictionaries:
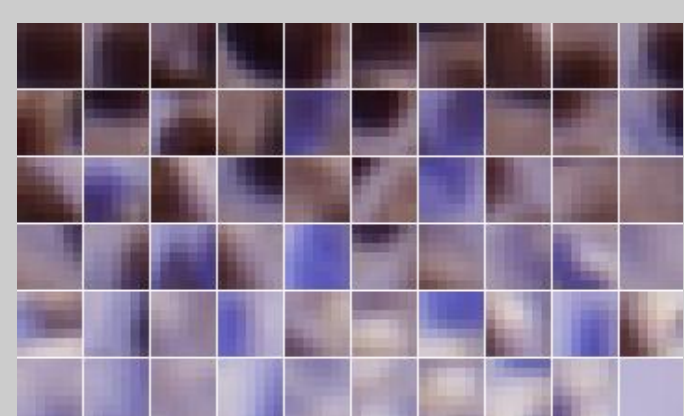
Intensity dictionary   **Class dictionary**

The **Intensity dictionary** is used for encoding the image and the corresponding **Class dictionary** for constructing class probabilities for each pixel.

Class probabilities for a sample image Light values correspond to high probability of being foreground and vice versa.

The segmentation of nuclei consists of three stages: **Dictionary training, Classification** and **Nuclei separation**

Dictionary Training

1) A training region is selected where foreground patterns are manually labeled.

2) A number of neighbourhood patches are extracted. (typically > 1000). The patches contain both intensity and class information (background/foreground).
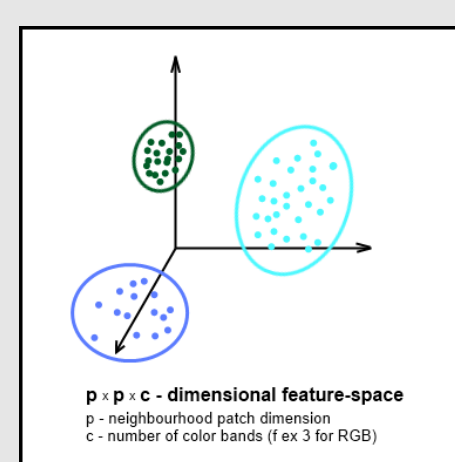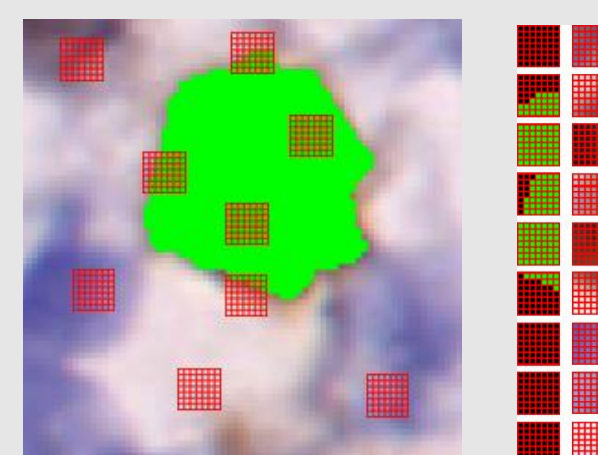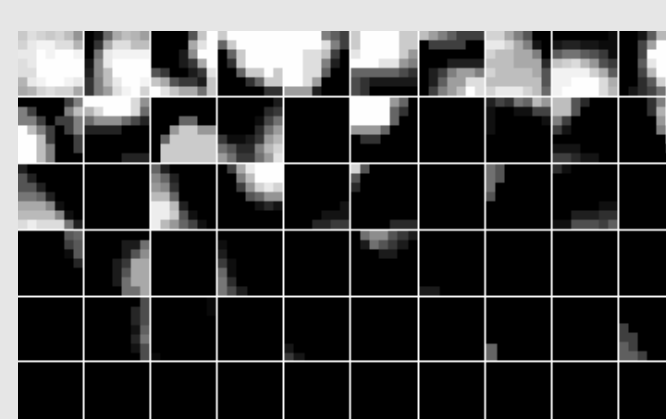
3) The samples are re-arranged to feature vectors and then clustered in feature space, using for example k-means clustering.

4) The cluster centers are then used to build our dictionaries, where each center correspond to one **intensity patch** and one **class patch** in our dictionaries.
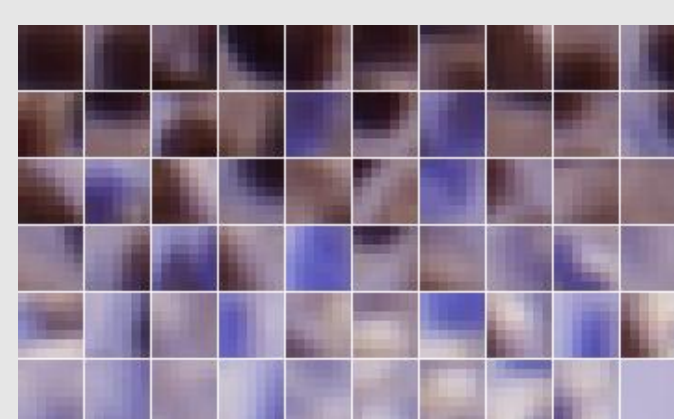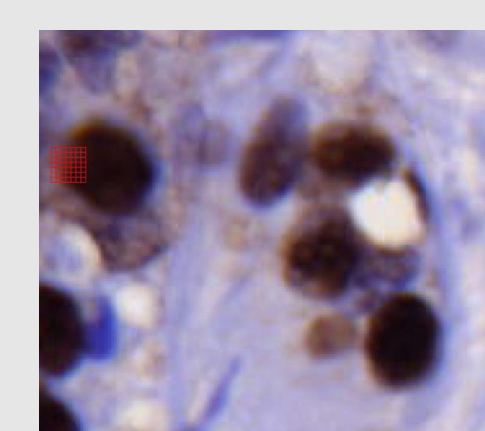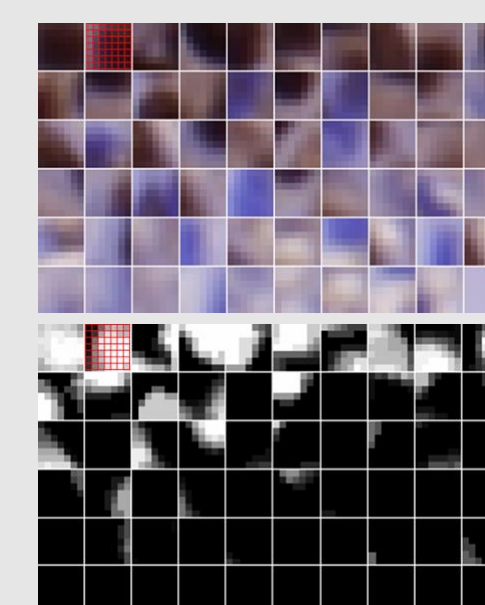
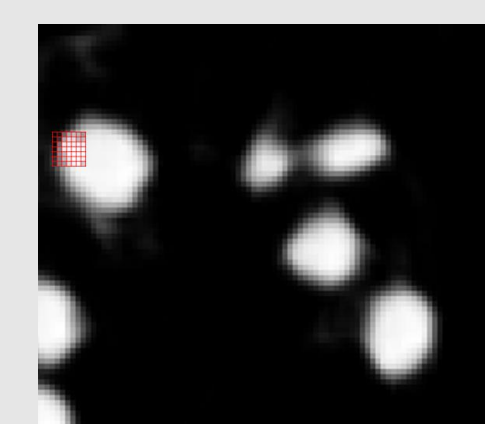**Intensity dictionary**   **Class dictionary**

Classification

1) For each pixel in the input image, a neighbourhood patch is extracted

2) The patch is matched against the intensity dictionary to find the closest match. Corresponding class patch is selected from the class dictionary.

3) The class patch is added to a probability map where it contributes to the total class probability for that neighbourhood.
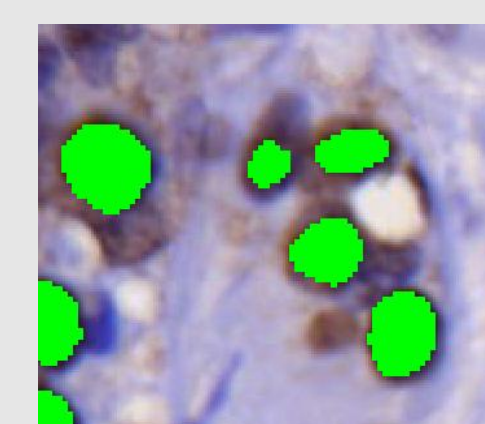
4) When all pixels are gone through, the resulting probability map is then thresholded to achieve the segmentation.

Nuclei separation

• Due to clustering and overlapping multiple nuclei are often merged and needs to be separated.

• We use an iterative algorithm that, in every iteration:
 1) Removes local minima using the **h-Minima** transform (with h=1) on a distance transformed image.
 2) Separates the regions using the **watershed** algorithm.
 3) Validates the separated regions with a score based on **ellipticity** and **size**.

• We use the Gray-Weighted Distance Transform on the probability map to include class probabilities in the separation decision.

• The separation with the highest score is selected as final segmentation.

Results & conclusion

• The algorithm was tested on 1006 nuclei from 207 images stained for the ER-marker.
• The results from our method was compared with another method using a Bayes classifier and a recently developed algorithm for separating nuclei.
• Our method had an error rate of correctly segmenting a nuclei of 4.8% compared to 17.3% of the competitive method. Using our classifier and the competitive method for separating nuclei, the error rate was 7.4%
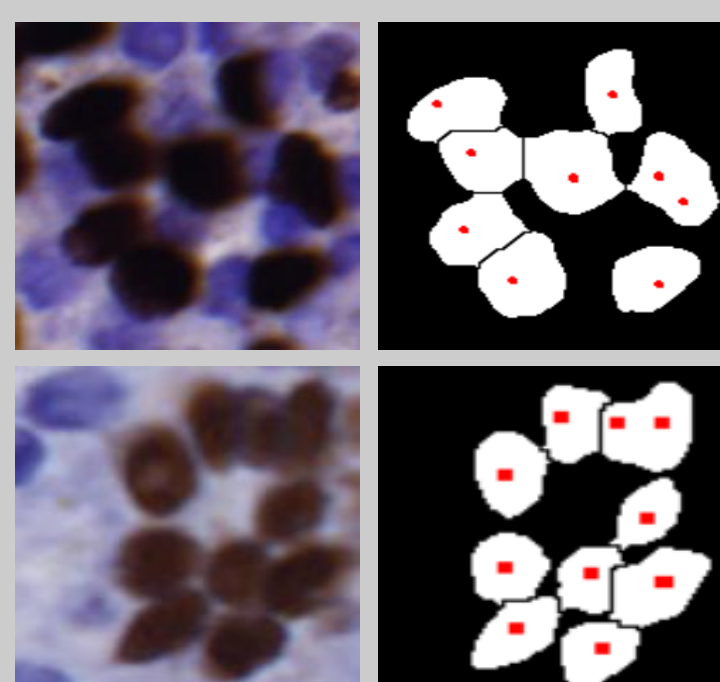
We also used the same approach to detect cancer tissue at lower resolution to show the flexibility of the method