Learning Pit Pattern Characteristics For Gastroenterological Training

Roland Kwitt[1], Nikhil Rasiwasia[2], Nuno Vasconcelos[2], Andreas Uhl[1], Michael Häfner[4], Friedrich Wrba[3]

[1]Multimedia Signal Processing and Security Lab
University of Salzburg, Salzburg, Austria

[2]Statistical Visual Computing Lab (SVCL)
UCSD, San Diego, CA, USA

[3]Department of Pathology
Medical University of Vienna, Vienna, Austria

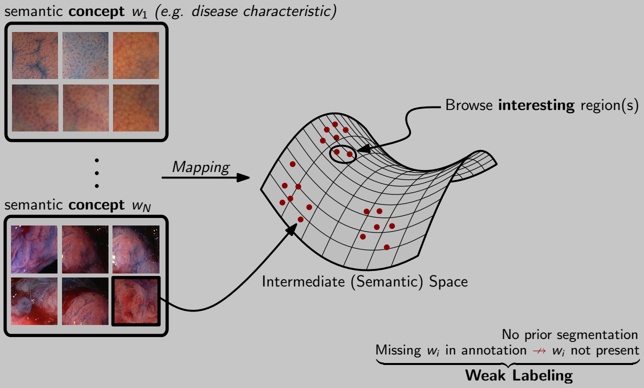
[4]Deparment of Internal Medicine
Elisabeth Hospital, Vienna, Austria

presented at **MICCAI 2011**, September 18–22, Toronto, Canada

Motivation

An introductory example

Medical images grouped by ...

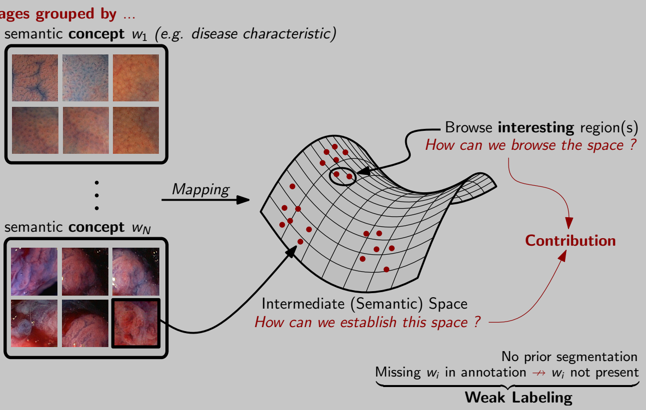


What is our objective?

"Browse those images which most-characteristically show the semantic concept $\mathcal{C}$, sorted by relevance"

Motivation

An introductory example



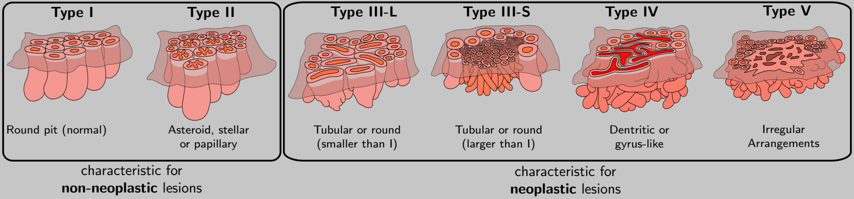
What is our objective?

"Browse those images which most-characteristically show the semantic concept $\mathcal{C}$, sorted by relevance"

Medical Data Material

Data Source

High-magnification chromo-endoscopy (HMCE) images of the colon mucosa, categorized by Kudo's [Kudo et al., 1994] pit-pattern classification criteria.



Pit-pattern analysis . . .

- is highly-predictive of the histological diagnosis [Matsuda et al., 2008]
- usually requires an experienced gastroenterologist [Chang et al., 2009]
- requires considerable (time-consuming) training effort [Togashi et al., 1999]

Related Work

In Literature [André et al., 2009, Kwitt et al., 2010, Tischendorf et al., 2010]

In vivo imagery → histological predictions

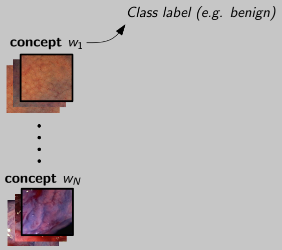
Prevalent approach: Bag-of-Visual-Words (BoW) variants (e.g. [Fei-Fei and Perona, 2005, Lazebnik et al., 2006])

Related Work

In Literature [André et al., 2009, Kwitt et al., 2010, Tischendorf et al., 2010]

In vivo imagery → histological predictions

Prevalent approach: Bag-of-Visual-Words (BoW) variants (e.g. [Fei-Fei and Perona, 2005, Lazebnik et al., 2006])

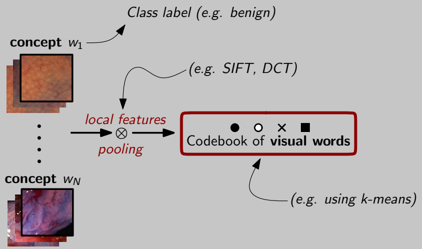


Related Work

In Literature [André et al., 2009, Kwitt et al., 2010, Tischendorf et al., 2010]

In vivo imagery → histological predictions

Prevalent approach: Bag-of-Visual-Words (BoW) variants (e.g. [Fei-Fei and Perona, 2005, Lazebnik et al., 2006])

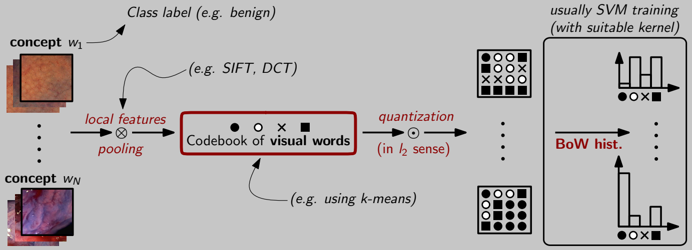


Related Work

In Literature [André et al., 2009, Kwitt et al., 2010, Tischendorf et al., 2010]

In vivo imagery → histological predictions

Prevalent approach: Bag-of-Visual-Words (BoW) variants (e.g. [Fei-Fei and Perona, 2005, Lazebnik et al., 2006])

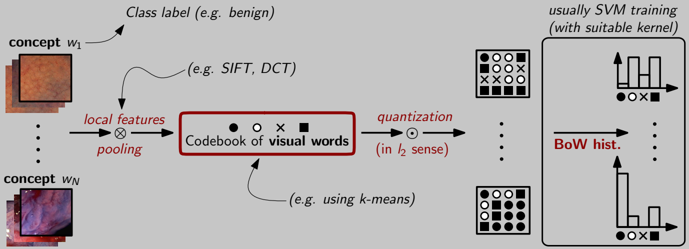


Related Work

In Literature [André et al., 2009, Kwitt et al., 2010, Tischendorf et al., 2010]

In vivo imagery → histological predictions

Prevalent approach: Bag-of-Visual-Words (BoW) variants (e.g. [Fei-Fei and Perona, 2005, Lazebnik et al., 2006])



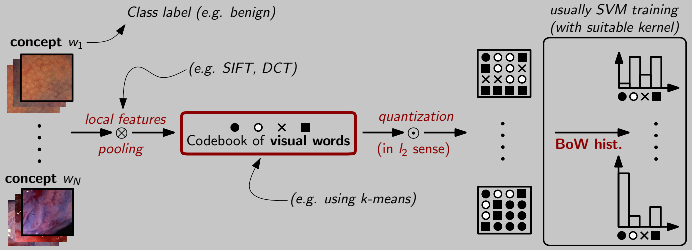
✓ Unsupervised codebook generation

Related Work

In Literature [André et al., 2009, Kwitt et al., 2010, Tischendorf et al., 2010]

In vivo imagery → histological predictions

Prevalent approach: Bag-of-Visual-Words (BoW) variants (e.g. [Fei-Fei and Perona, 2005, Lazebnik et al., 2006])



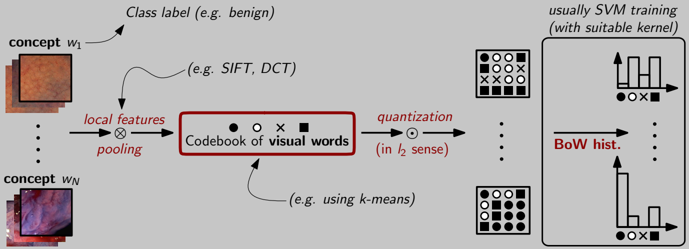
- ✓ Unsupervised codebook generation
- ✓ State-of-the-Art in computer vision

Related Work

In Literature [André et al., 2009, Kwitt et al., 2010, Tischendorf et al., 2010]

In vivo imagery → histological predictions

Prevalent approach: Bag-of-Visual-Words (BoW) variants (e.g. [Fei-Fei and Perona, 2005, Lazebnik et al., 2006])



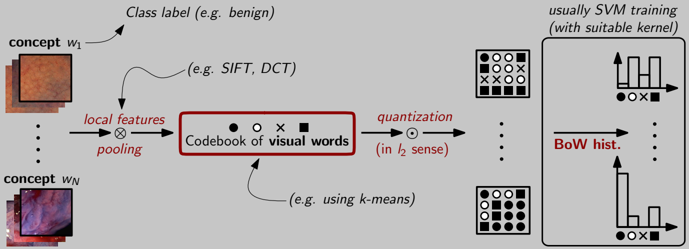
- ✓ Unsupervised codebook generation
- ✓ State-of-the-Art in computer vision
- ✓ Many extensions & efficient impl.

Related Work

In Literature [André et al., 2009, Kwitt et al., 2010, Tischendorf et al., 2010]

In vivo imagery → histological predictions

Prevalent approach: Bag-of-Visual-Words (BoW) variants (e.g. [Fei-Fei and Perona, 2005, Lazebnik et al., 2006])



✓ Unsupervised codebook generation

✓ State-of-the-Art in computer vision

✓ Many extensions & efficient impl.

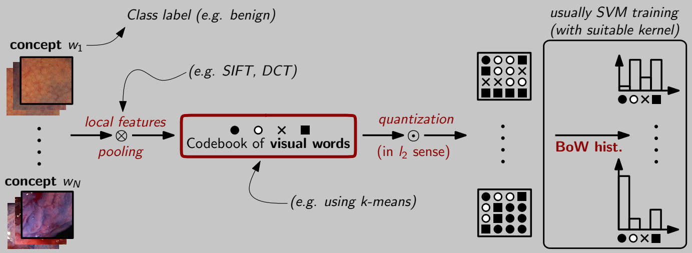
✗ No **explicit** semantic modeling step

Related Work

In Literature [André et al., 2009, Kwitt et al., 2010, Tischendorf et al., 2010]

In vivo imagery → histological predictions

Prevalent approach: Bag-of-Visual-Words (BoW) variants (e.g. [Fei-Fei and Perona, 2005, Lazebnik et al., 2006])



- ✓ Unsupervised codebook generation
- ✓ State-of-the-Art in computer vision
- ✓ Many extensions & efficient impl.

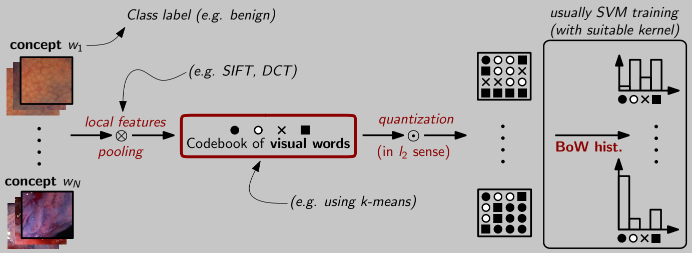
- ✗ No **explicit** semantic modeling step
- ✗ ∘, •, etc. are not semantic concepts

Related Work

In Literature [André et al., 2009, Kwitt et al., 2010, Tischendorf et al., 2010]

In vivo imagery → histological predictions

Prevalent approach: Bag-of-Visual-Words (BoW) variants (e.g. [Fei-Fei and Perona, 2005, Lazebnik et al., 2006])



- ✓ Unsupervised codebook generation
- ✓ State-of-the-Art in computer vision
- ✓ Many extensions & efficient impl.

- ✗ No **explicit** semantic modeling step
- ✗ ○, ●, etc. are not semantic concepts
- ✗ What about codebook size ?

Our Approach

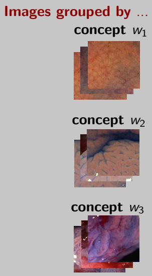
Building the Intermediate (Semantic) Space - Part I

- We exploit the generative approach of [Rasiwasia and Vasconcelos, 2008]
- Originally introduced in the context of **natural scene categorization**
- Inherently based on the image annotation approach of [Carneiro et al., 2007] (which relies on a multiple-instance learning argument)

Our Approach

Building the Intermediate (Semantic) Space - Part I

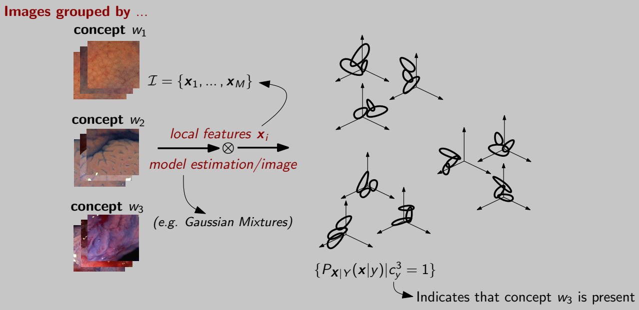
- We exploit the generative approach of [Rasiwasia and Vasconcelos, 2008]
- Originally introduced in the context of **natural scene categorization**
- Inherently based on the image annotation approach of [Carneiro et al., 2007] (which relies on a multiple-instance learning argument)



Our Approach

Building the Intermediate (Semantic) Space - Part I

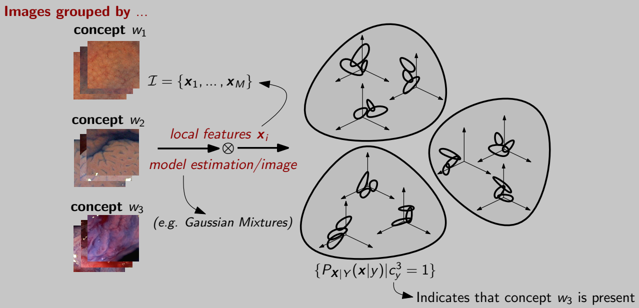
- We exploit the generative approach of [Rasiwasia and Vasconcelos, 2008]
- Originally introduced in the context of **natural scene categorization**
- Inherently based on the image annotation approach of [Carneiro et al., 2007] (which relies on a multiple-instance learning argument)



Our Approach

Building the Intermediate (Semantic) Space - Part I

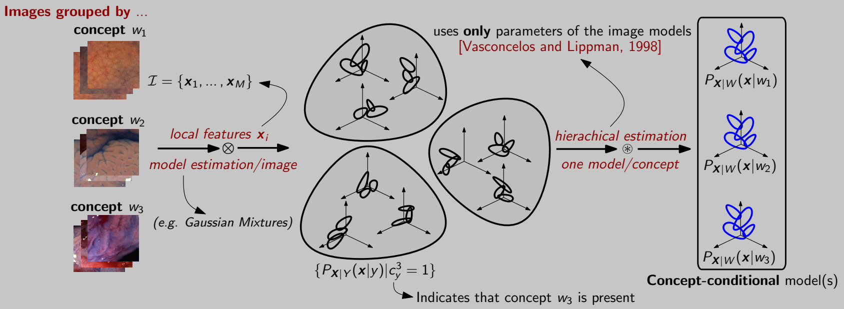
- We exploit the generative approach of [Rasiwasia and Vasconcelos, 2008]
- Originally introduced in the context of **natural scene categorization**
- Inherently based on the image annotation approach of [Carneiro et al., 2007] (which relies on a multiple-instance learning argument)



Our Approach

Building the Intermediate (Semantic) Space - Part I

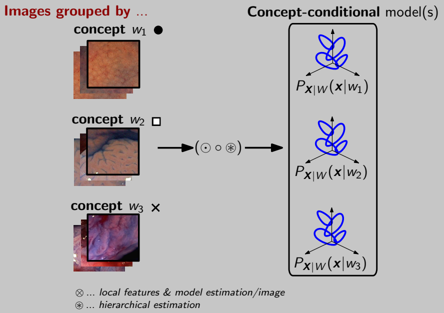
- We exploit the generative approach of [Rasiwasia and Vasconcelos, 2008]
- Originally introduced in the context of **natural scene categorization**
- Inherently based on the image annotation approach of [Carneiro et al., 2007] (which relies on a multiple-instance learning argument)



Our Approach

Building the Intermediate (Semantic) Space - Part II

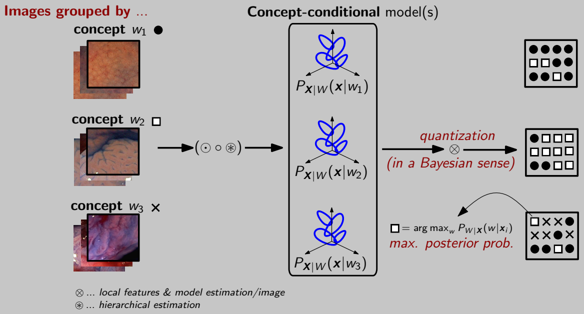
- Implement the mapping $\Pi : \mathcal{X} \to \mathcal{S}$, from visual feature space $\mathcal{X}$ to intermediate (semantic) space $\mathcal{S}$, i.e. $\Pi(\mathcal{I}) = \boldsymbol{\pi}$



Our Approach

Building the Intermediate (Semantic) Space - Part II

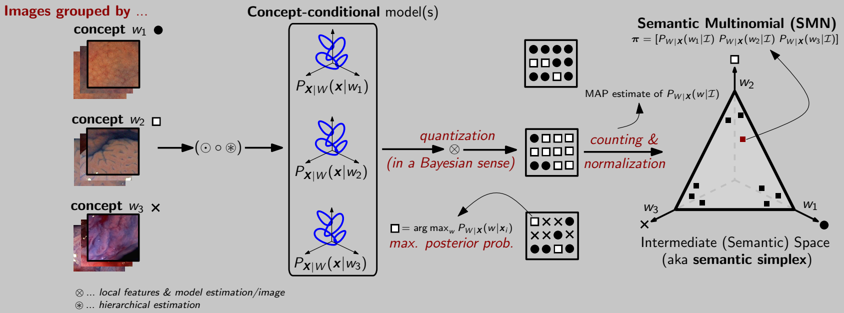
- Implement the mapping $\Pi : \mathcal{X} \rightarrow \mathcal{S}$, from visual feature space $\mathcal{X}$ to intermediate (semantic) space $\mathcal{S}$, i.e. $\Pi(\mathcal{I}) = \boldsymbol{\pi}$



Our Approach

Building the Intermediate (Semantic) Space - Part II

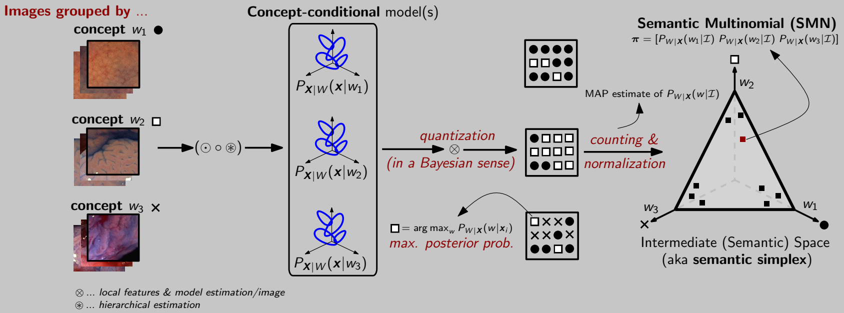
- Implement the mapping $\Pi : \mathcal{X} \to \mathcal{S}$, from visual feature space $\mathcal{X}$ to intermediate (semantic) space $\mathcal{S}$, i.e. $\Pi(\mathcal{I}) = \boldsymbol{\pi}$



Our Approach

Building the Intermediate (Semantic) Space - Part II

- Implement the mapping $\Pi : \mathcal{X} \rightarrow \mathcal{S}$, from visual feature space $\mathcal{X}$ to intermediate (semantic) space $\mathcal{S}$, i.e. $\Pi(\mathcal{I}) = \boldsymbol{\pi}$



- The axes of $\mathcal{S}$ now **do have** a semantic interpretation!

Our Approach

Browsing the Intermediate (Semantic) Space . . .

What's our goal again ?

> *"Browse those images which most-characteristically show the semantic concept $\mathcal{C}$, sorted by relevance"*

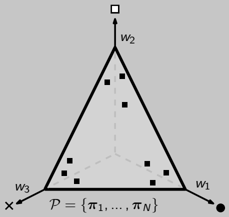
Our Approach

Browsing the Intermediate (Semantic) Space . . .

What's our goal again ?

> *"Browse those images which most-characteristically show the semantic concept $\mathcal{C}$, sorted by relevance"*

Example: Let $\mathcal{C} = w_1$; *translate* "most-characteristically" to $\pi_{i1} > 0.8$



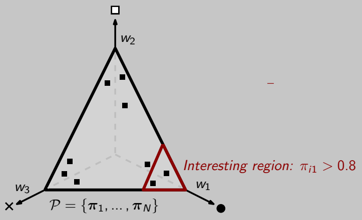
Our Approach

Browsing the Intermediate (Semantic) Space . . .

What's our goal again ?

"Browse those images which most-characteristically show the semantic concept $\mathcal{C}$, sorted by relevance"

Example: Let $\mathcal{C} = w_1$; *translate* "most-characteristically" to $\pi_{i1} > 0.8$



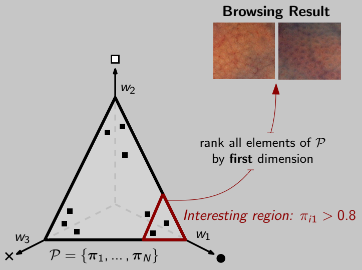
Our Approach

Browsing the Intermediate (Semantic) Space . . .

What's our goal again ?

"Browse those images which most-characteristically show the semantic concept $\mathcal{C}$, sorted by relevance"

Example: Let $\mathcal{C} = w_1$; *translate* "most-characteristically" to $\pi_{i1} > 0.8$



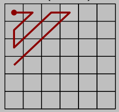
Experimental Study

Implementation & Protocol

Implementation Details

Visual Features

8 × 8 (block) **DCT**



sliding window (2 pixel)

Modeling Stage

Gaussian Mixtures

Visual Level (per image)

- 8 components
- Diagonal covariance
- EM & K-Means++

Semantic Level (per concept)

- 64 components
- Hierarchical estimation

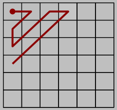
Experimental Study

Implementation & Protocol

Implementation Details

Visual Features

8×8 (block) **DCT**



sliding window (2 pixel)

Modeling Stage

Gaussian Mixtures

Visual Level (per image)

- 8 components
- Diagonal covariance
- EM & K-Means++

Semantic Level (per concept)

- 64 components
- Hierarchical estimation

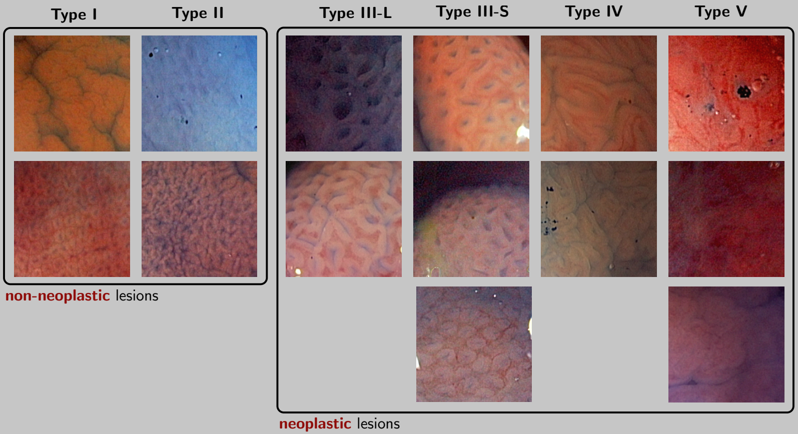
Evaluation Setup & Protocol

- 716 HMCE images, 40 patients
- Only images where pit-pattern analysis is coherent with histology
- Visual evaluation of browsing results
- Evaluate the average error rate (leave-one-patient-out protocol)

Experimental Study

Visual evaluation of browsing result

- Browse the top 10 images per concept
- **Patient Pruning:** Remove images from the same patient



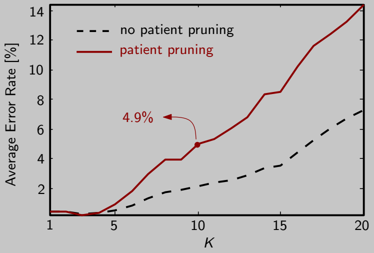
Experimental Study

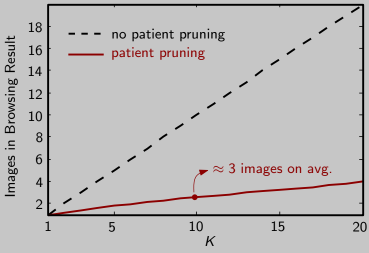
Quantitative evaluation

Average Error Rate

Iterate over all patients $p_1, \ldots, p_J$

1. Leave-out all images of patient p_i and compute SMNs
2. Browse the top K images per concept, now using **all** available images
3. Count wrong (i.e. wrong concept) images per browsing result





Concluding Remarks

- We propose to shift from visual $\rightarrow$ semantic modeling of medical content
- Generic approach to establish a semantic space for medical imagery
- Imaging modality $\rightarrow$ choose suitable features (e.g. SIFT, SURF, HOG, etc.)
- **Other potential tasks:** cross-modal browsing/retrieval, classification

Future Work:

- Suitable similarity measure (kernel) on the semantic space
- Incorporate spatial information (e.g. spatial pyramid [Lazebnik et al., 2006])

Concluding Remarks

- We propose to shift from visual $\rightarrow$ semantic modeling of medical content
- Generic approach to establish a semantic space for medical imagery
- Imaging modality $\rightarrow$ choose suitable features (e.g. SIFT, SURF, HOG, etc.)
- **Other potential tasks:** cross-modal browsing/retrieval, classification

Future Work:

- Suitable similarity measure (kernel) on the semantic space
- Incorporate spatial information (e.g. spatial pyramid [Lazebnik et al., 2006])

Thank You!

(come visit our poster P1-10-W)

Resources will be available at www.wavelab.at

Bibliography I

André, B., Vercauteren, T., Perchant, A., Buchner, A. M., Wallace, M. B., and Ayache, N. (2009).
Endomicroscopic image retrieval and classification using invariant visual features.
In *ISBI*.

Carneiro, G., , A. B. C., Moreno, P., and Vasconcelos, N. (2007).
Supervised learning of semantic classes for image annotation and retrieval.
IEEE Trans. Pattern Anal., 29(3).

Chang, C., Hsieh, C.-R., Lou, H.-Y., Fang, C.-L., Tiong, C., Wang, J.-J., Wei, I.-V., Wu, S.-C., and Chen, J.-N. (2009).
Comparative study of conventional colonoscopy, magnifying chromoendoscopy, and magnifying narrow-band imaging systems in the differential diagnosis of small colonic polyps between trainee and experienced endoscopist.
Int. J. Colorectal Dis., 24:1413–1419.

Fei-Fei, L. and Perona, P. (2005).
A bayesian hierarchical model for learning natural scene categories.
In *CVPR*.

Kudo, S., Hirota, S., Nakajima, T., Hosobe, S., Kusaka, H., Kobayashi, T., Himori, M., and Yagyuu, A. (1994).
Colorectal tumours and pit pattern.
J. Clin. Pathol., 47:880–885.

Kwitt, R., Uhl, A., Häfner, M., Gangl, A., Wrba, F., and Vecsei, A. (2010).
Predicting the histology of colorectal lesions in a probabilistic framework.
In *MMBIA*.

Bibliography II

Lazebnik, S., Schmid, C., and Ponce, J. (2006).
Beyond bags of features: Spatial pyramid matching for recognizing scene categories.
In *CVPR*.

Matsuda, T., Fujii, T., Saito, Y., Nakajima, T., Uraoka, T., Kobayashi, N., Ikehara, H., Ikematsu, H., Fu, K., Emura, F., Ono, A., Sano, Y., Shimoda, T., and Fujimori, T. (2008).
Efficacy of the invasive/non-invasive pattern by magnifying chromoendoscopy to estimate the depth of invasion of early colorectal neoplasms.
Am. J. Gastroenterol., 103(11):2700–2706.

Rasiwasia, N. and Vasconcelos, N. (2008).
Scene classification with low-dimensional semantic spaces and weak supervision.
In *CVPR*.

Tischendorf, J. J. W., Gross, S., Winograd, R., Hecker, H., Auer, R., Behrens, A., Trautwein, C., Aach, T., and Stehle, T. (2010).
Computer-aided classification of colorectal polyps based on vascular patterns: a pilot study.
Endoscopy, 42(3):203–207.

Togashi, K., Konishi, F., Ishizuka, T., Sato, T., Senba, S., and Kanazawa, K. (1999).
Efficacy of magnifying endoscopy in the differential diagnosis of neoplastic and non-neoplastic polyps of the large bowel.
Dis. Colon Rectum, 42(12):1602–1608.