Multi-Label Classification using a Map Between Sparse Representations: Application to Retinal Blood Vessel Segmentation

Moncef Hidane
LI EA 6300, INSA CVL, Université de Tours

Joint work with **Taibou Birgui Sekou**, Julien Olivier and Hubert Cardot

27ème journée du projet CASCIMODOT
11 décembre 2017
Orléans

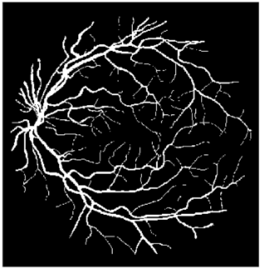
INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES CENTRE VAL DE LOIRE

LI Laboratoire d'Informatique EA 6300

UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

Retinal Blood Vessel Segmentation (RBVS)

Fundus image



Segmentation of blood vessels

Why?

- Eye vascularization is an important risk biomarker in a large number of diseases, *e.g.* diabetic retinopathy, age-related macular degeneration and glaucoma.
- The eye shares neural and vascular similarities with the brain. Eye vascularization offers a direct window to cerebral pathology.

Learning RBVS

Learning to segment

Assume that we have a training set $L = \{(I_i, S_i)\}_{1 \leq i \leq n}$, where

- $I_i \in \mathbb{R}^N$ is a (vectorized) fundus image
- $S_i \in \{0,1\}^N$ is an expert segmentation of I_i.

Learning RBVS means using L to construct a mapping $\mathcal{S} : \mathbb{R}^N \to \{0,1\}^N$ that allows to infer the segmentation of new, unseen fundus images.

Training set $\xrightarrow{\text{Learn}}$ $\underbrace{\mathcal{S} : \mathbb{R}^N \to \{0,1\}^N}_{\text{Segmentation function}}$

Learning by Feature Extraction and Empirical Risk Minimization

Patches

- We work at the level of patches, that is small $\sqrt{m} \times \sqrt{m}$ image subsets.
- The training set consists of two matrices:
  - $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{m \times n}$, whose columns are (pre-processed) image patches
  - $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_n] \in \{0, 1\}^{m \times n}$, whose columns are the corresponding expert segmentations.

Feature Extraction

- Instead of working with raw patches, we try to find good *features* that will ease segmentation.
- Feature extraction is given by a function (usually nonlinear) $F : \mathbb{R}^m \to \mathbb{R}^p$.
- The features of a test patch will serve either to infer the label of its central pixel, or the labels of all its pixels.

ERM

1. Choose a class $\mathcal{H}$ of plausible segmentation functions.
2. Choose an $S \in \mathcal{H}$ that "works well" on the training set $(\mathbf{X}, \mathbf{Y})$, that is, that minimizes an empirical risk.

This talk: feature extraction by sparse representation in a redundant learned dictionary

Sparse Dictionary Learning

Principle

Learn a redundant dictionary $(\mathbf{d}_i)_{1\geq i\geq p}$, $p > m$, such that each training patch $\mathbf{x}_j \in \mathbb{R}^m$ can be represented *sparsely* represented as a linear combination of the $\mathbf{d}_i'$s.

A Combinatorial Formulation

$$\mathbf{D}^*, \mathbf{A}^* \leftarrow \min_{\mathbf{D}\in\mathbb{R}^{m\times p}, \mathbf{A}\in\mathbb{R}^{p\times n}} \Big[\mathcal{R}(\mathbf{X}, \mathbf{D}, \mathbf{A}) = \lambda \sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{D}\mathbf{a}_i\|_2^2 + \|\mathbf{a}_i\|_0\Big],$$

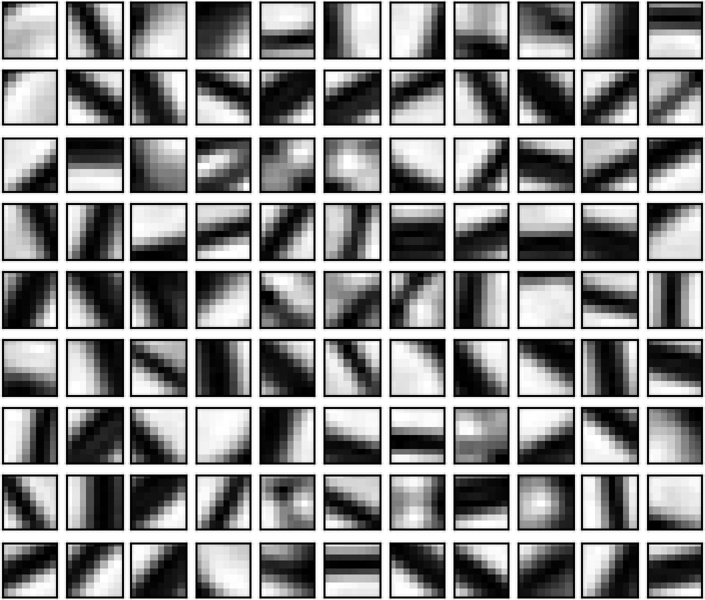
where $\|\mathbf{x}\|_0$ is the number of non-zero entries of $\mathbf{x}$.

A (partially) Convex Formulation

$$\mathbf{D}^*, \mathbf{A}^* \leftarrow \min_{\mathbf{D}\in\mathbb{R}^{m\times p}, \mathbf{A}\in\mathbb{R}^{p\times n}} \Big[\mathcal{R}(\mathbf{X}, \mathbf{D}, \mathbf{A}) = \lambda \sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{D}\mathbf{a}_i\|_2^2 + \|\mathbf{a}_i\|_1\Big]. \quad (1)$$

In both versions, the columns of $\mathbf{D}$ should be constrained to have unit norm.

An Example of Learned Dictionary



RBVS by Sparse Coding then Classifier (SCTC)

Training

1. Using Eq. (1), learn a redundant dictionary $\mathbf{D}$ that sparsely encodes each training patch $\mathbf{x}_i$.
2. Train a classifier $\mathcal{C}$ (*e.g.* random forest) on the sparse codes (the $\mathbf{a}_i'$s) of each $\mathbf{x}_i$ in order to infer the label of the central pixel.

Testing

1. Extract all patches of the test image.
2. Represent each test patch $\mathbf{x}$ as sparse linear combination of the columns of $\mathbf{D}$ by solving

$$\mathbf{a}^* \leftarrow \min_{\mathbf{a} \in \mathbb{R}^p} \lambda \|\mathbf{x} - \mathbf{D}\mathbf{a}\|_2^2 + \|\mathbf{a}\|_1. \tag{2}$$

3. Compute $\mathcal{C}(\mathbf{a})$ to obtain the label of the central pixel of $\mathbf{x}$.

Both the feature extraction and the classifier are learned. But feature extraction has been learned in an unsupervised way.

RBVS by Joint Dictionary and Classifier Learning (JDCL)

Learning

- Jointly learn
  - a dictionary **D** that sparsely encodes training patches
  - and a linear classifier **W** whose inputs are the sparse codes produced by **D**:

$$\mathbf{D}^*, \mathbf{A}^*, \mathbf{W}^* \leftarrow \min_{\mathbf{D},\mathbf{A},\mathbf{W}} \alpha \|\mathbf{L} - \mathbf{W}\mathbf{A}\|_F^2 + \beta \|\mathbf{W}\|_F^2 + \mathcal{R}(\mathbf{X}, \mathbf{D}, \mathbf{A}),$$

$\mathbf{L} = [\mathbf{l}_1, ..., \mathbf{l}_n] \in \mathbb{R}^{2\times n}$ is the label matrix where the vector $\mathbf{l}_i \in \mathbb{R}^2$ has binary entries indicating whether the center of the associated patch belongs to a vessel or not.

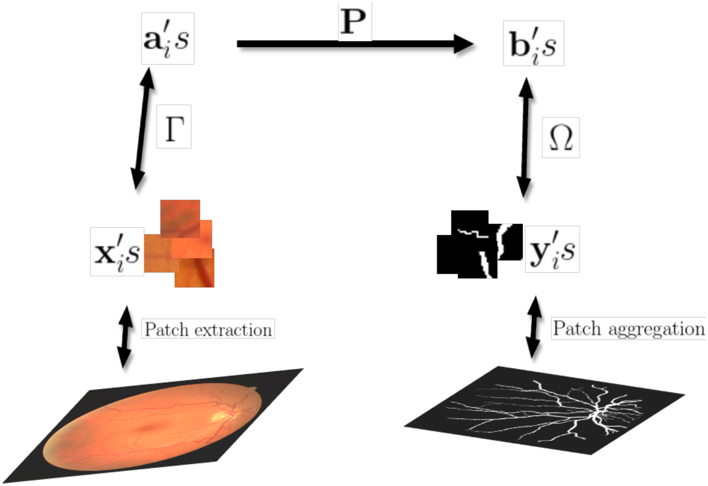
Testing

The label l of the central pixel of a query patch $\mathbf{x}$ is obtained using the following equation:

$$l = \underset{i\in\{1,2\}}{\text{argmax}} \langle \delta_i, \mathbf{W}^*\mathbf{a}^* \rangle,$$

where $\mathbf{W}^*$ is the previously learned classifier and $\mathbf{a}^*$ is the sparse code of $\mathbf{x}$ in $\mathbf{D}^*$.

Multi-Label Classification using a Map Between Sparse Representations



The Learning Stage

- Learning a label dictionary $\mathbf{\Omega}^*$:

$$\mathbf{\Omega}^*, \mathbf{B}^* \leftarrow \min_{\mathbf{\Omega} \in \mathcal{D}, \mathbf{B} \in \mathbb{R}^{p \times n}} \mathcal{R}(\mathbf{Y}, \mathbf{\Omega}, \mathbf{B}),$$

- Simultaneously learning a data dictionary $\mathbf{\Gamma}^*$ and a linear map $\mathbf{P}^*$:

$$\mathbf{\Gamma}^*, \mathbf{A}^*, \mathbf{P}^* \leftarrow \min_{\substack{\mathbf{\Gamma} \in \mathcal{D}, \\ \mathbf{A} \in \mathbb{R}^{p \times n}, \mathbf{P} \in \mathbb{R}^{p \times p}}} \mathcal{R}(\mathbf{X}, \mathbf{\Gamma}, \mathbf{A}) + \alpha \mathcal{C}(\mathbf{B}^*, \mathbf{P}, \mathbf{A}), \tag{3}$$

  with

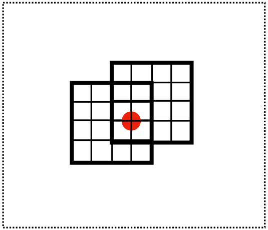
$$\mathcal{C}(\mathbf{B}^*, \mathbf{P}, \mathbf{A}) = \frac{1}{2} \|\mathbf{B}^* - \mathbf{P}\mathbf{A}\|_F^2 + \beta \|\mathbf{P}\|_F^2.$$

The Testing Stage

- Divide the input image into *overlapping* patches
- For each input patch $\mathbf{x} \in \mathbb{R}^m$, compute its segmentation map $\mathbf{s} \in \{0,1\}^m$ as follows

$$\mathbf{s} = \mathcal{T}_\kappa(\Omega^* \mathbf{P}^* \mathbf{a}^*),$$

  where $\mathbf{a}^*$ is the sparse representation of $\mathbf{x}$ over the dictionary Γ and $\mathcal{T}$ is a threshold transformation with parameter κ.
- Finally, aggregate all the estimates due to the overlap.



Quantitative Evaluation

- 2 standard datasets for evaluation: **DRIVE** and STARE
- Pre-processing by grayscale conversion and histogram equalization
- 8×8 patches
- Training with 25000 patches
- Evaluation metrics:

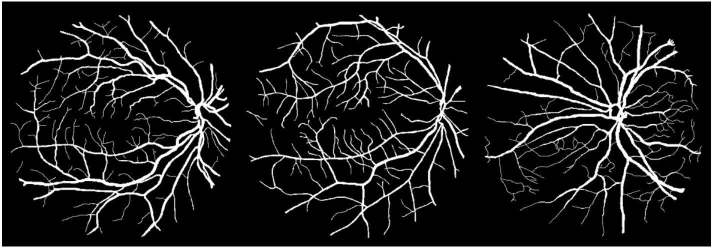
$$\mathbf{Sens} = \frac{TP}{P}, \quad \mathbf{Spec} = \frac{TN}{N}, \quad \mathbf{Acc} = \frac{TP + TN}{P + N}.$$

Methods	**Spec**	**Sens**	**Acc**
This paper	96.78	80.56	**95.34**
Javidi et al.[2]**	97.02	72.01	94.50
SCTC [3]**	95.55	83.49	94.48
Liskowski et al.[5]*	96.73	**84.60**	95.07
Dasgupta et al. [4]*	**98.01**	76.91	95.33

*deep learning methods —— ** dictionary learning methods

Qualitative Results

(a) Predictions



(b) Ground-truths

Conclusion

- We proposed a learning approach for retinal blood vessel segmentation.
- We leverage sparsity as a cue to extract features.
- We discussed a variant where we learned feature extraction and a linear classifier jointly.
- We introduced another variant where we learned data and label dictionaries and a coupling linear mapping.

Thank you for your attention!