
Binary pattern dictionary learning for gene expression representation in drosophila imaginal discs

Jiri Borovec, Jan Kybic

Depart. of Cybernetics, FEE, CTU in Prague, Czech Rep.

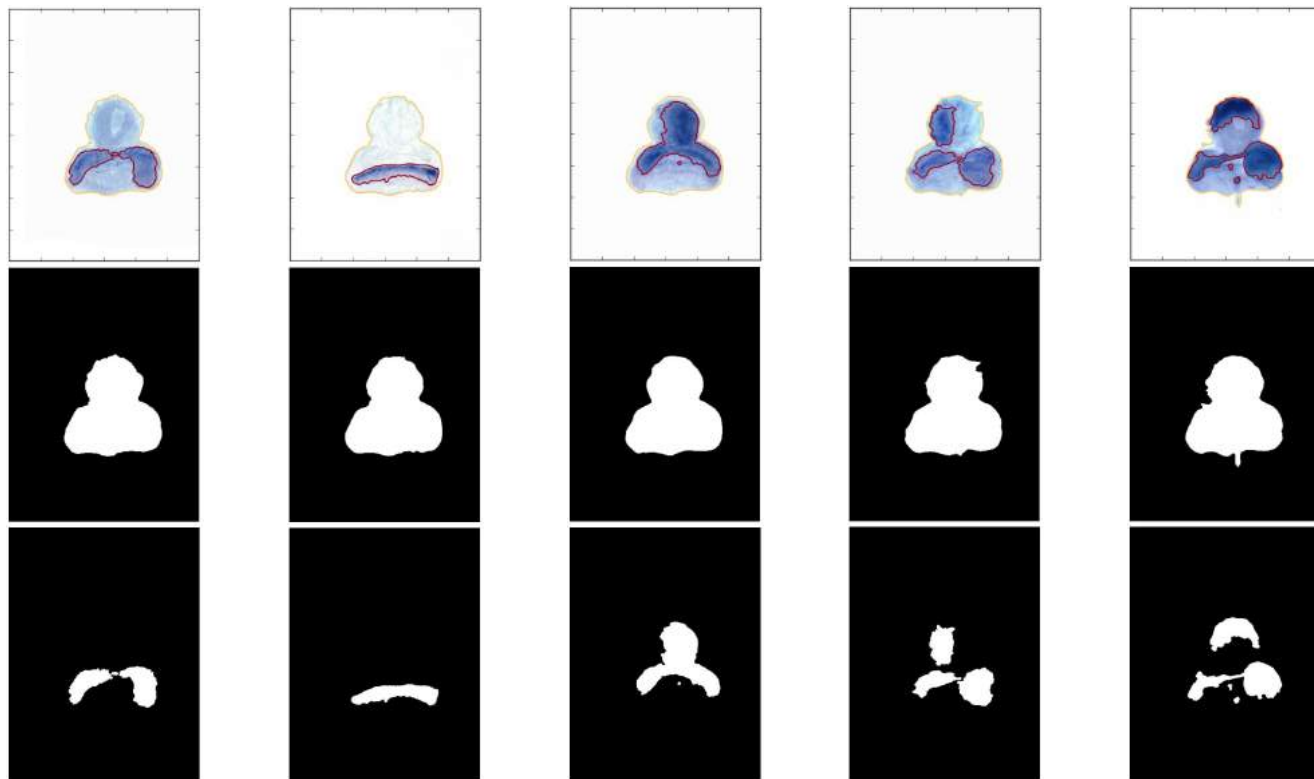
<http://cmp.felk.cvut.cz/~borovji3>

jiri.borovec@fel.cvut.cz

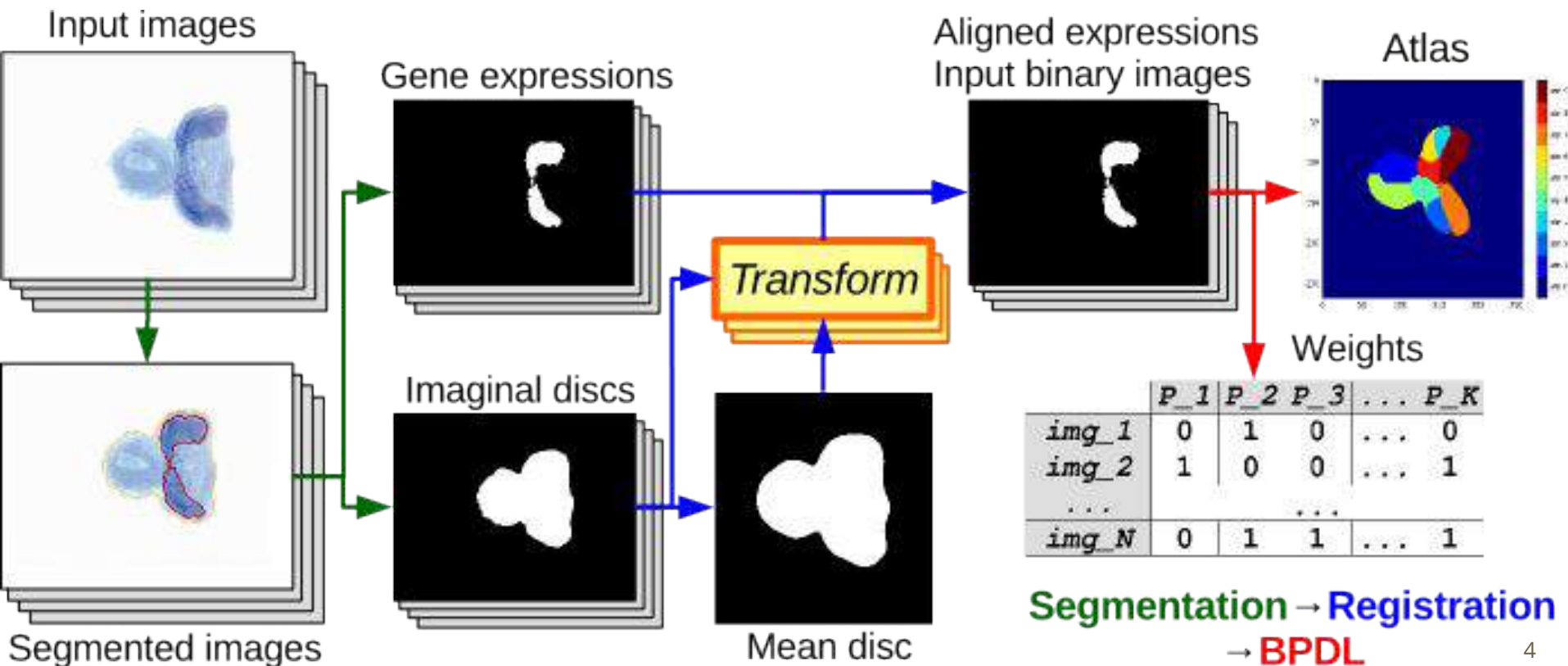
Outline

- Introduction
- Decomposition methods
- Complete pipeline
- Binary Pattern Dictionary Learning
- Experiments on synthetic & real images
- Future work: Drosophila ovary 3D

Input segmented Imaginal discs



Pipeline: segmentation - registration - BPDFL



Decomposition methods

Formulation

$$\min_{Y,W} \|X - Y \cdot W\|^2$$

- Non Negative Matrix Factorisation
- Fast Independent Component Analysis
- Sparse Principal Component Analyses
- Dictionary Learning

Task definition

- Input binary image
 - Image plain
- Atlas
 - label set
 - label 0 for background
 - labels l cannot overlap (patterns)
- Binary weight vector

$$\mathbf{g} : \Omega \rightarrow \{0, 1\}$$

$$\Omega \subseteq \mathbb{Z}^d$$

$$\mathbf{y} : \Omega \rightarrow L$$

$$L = [0, \dots, K]$$

$$\mathbf{w} : L \rightarrow \{0, 1\}$$

Method BPDFL

- Image representation

$$\hat{\mathbf{g}} = \sum_{l \in \mathbb{L}} \mathbf{w}_l \cdot \llbracket \mathbf{y} = l \rrbracket$$

- Similarity measure (Hamming distance)

$$F(\mathbf{g}, \mathbf{y}, \mathbf{w}) = \sum_{i \in \Omega} \llbracket \mathbf{g}_i \neq \hat{\mathbf{g}}_i \rrbracket$$

- Penalise neighbouring pixels

$$H(\mathbf{y}) = \sum_{\substack{i, j \in \Omega, i \neq j, \\ d(i, j) = 1}} \llbracket \mathbf{y}_i \neq \mathbf{y}_j \rrbracket$$

- Optimization criterion

$$\mathbf{y}^*, \mathbf{w}^* = \arg \min_{\mathbf{y}, \mathbf{W}} \frac{1}{N} \sum_n F(\mathbf{g}^n, \mathbf{y}, \mathbf{w}^n) + \beta \cdot H(\mathbf{y})$$

BPDL: Alternating minimization

- Update weights

$$w_l = \llbracket P(\mathbf{g}, \mathbf{y}, l) \geq \sigma \rrbracket \quad \text{where } \sigma = 1$$

$$\text{and } P(\mathbf{g}, \mathbf{y}, l) = \frac{\sum_{i \in \Omega, \mathbf{y}_i = l} \llbracket \mathbf{g}_i = 1 \rrbracket}{\sum_{i \in \Omega, \mathbf{y}_i = l} \llbracket \mathbf{g}_i \neq 1 \rrbracket} = \frac{\|\llbracket \mathbf{y} = l \rrbracket\|}{\sum_{i \in \Omega, \mathbf{y}_i = l} (1 - \mathbf{g}_i)} - 1$$

- Update atlas

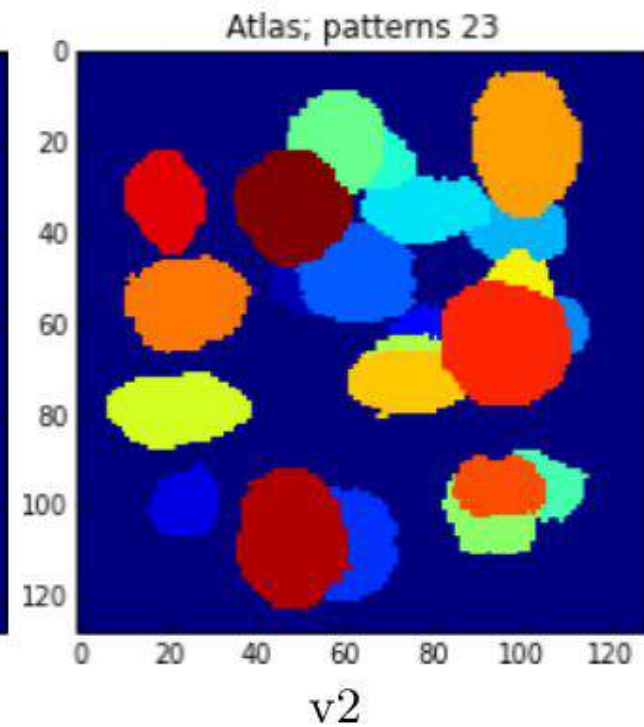
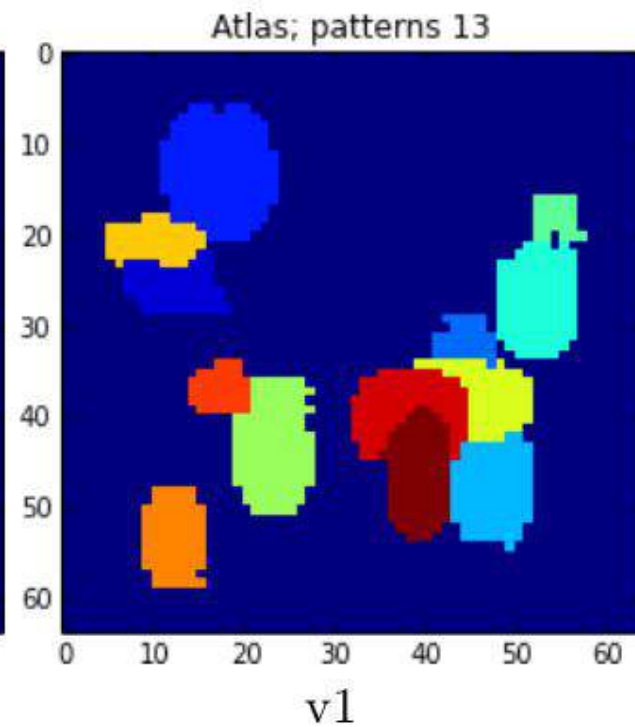
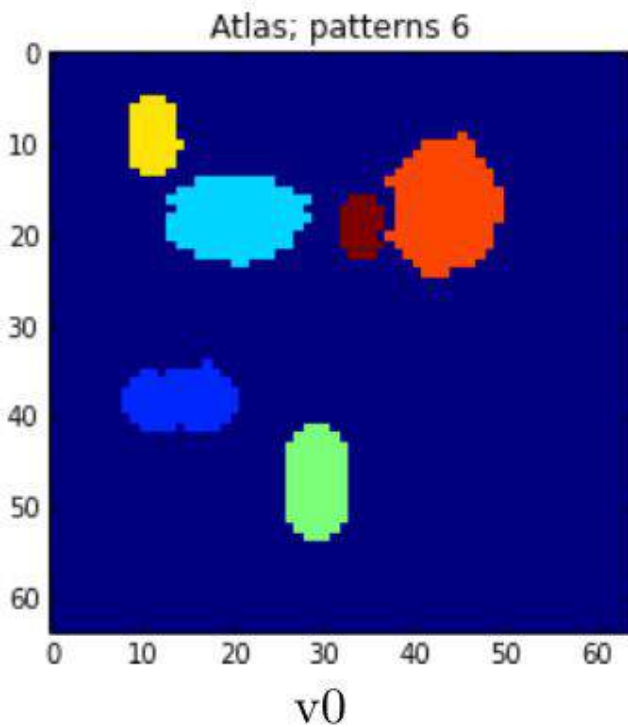
$$\frac{1}{N} \sum_{i \in \Omega} \sum_n \underbrace{\left| \mathbf{g}_i^s - \sum_{l \in \mathbb{L}} \mathbf{w}_l^s \cdot \llbracket \mathbf{y} = l \rrbracket \right|}_{U_i(\mathbf{y}_i)} + \sum_{\substack{i, j \in \Omega, i \neq j, \\ d(i, j) = 1}} \llbracket \mathbf{y}_i \neq \mathbf{y}_j \rrbracket$$

BPDL: Algorithm

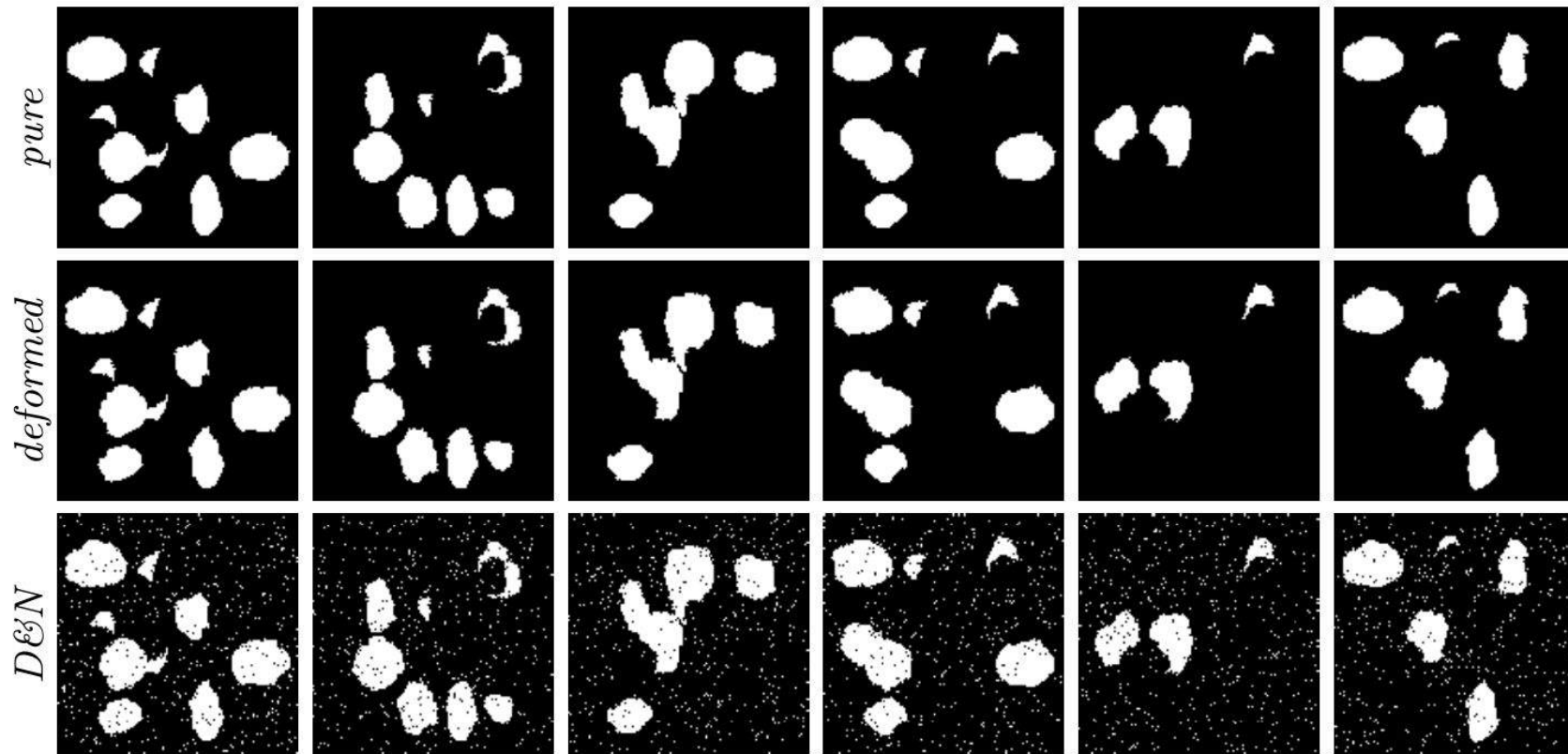
Algorithm 1 General schema of BPDL algorithm.

- 1: initialise atlas \mathbf{y}
 - 2: **while** not converged **do**
 - 3: update weights $\mathbf{w} \in \mathbf{W}$
 - 4: reinitialise empty patterns in \mathbf{y}^*
 - 5: update atlas \mathbf{y}^* via Graph Cut
 - 6: **end while**
-

Synthetic datasets



Synthetic images



Comparison on synth. images

| datasets | | NMF [16] | FastICA [14] | sPCA [13] | DL [15] | BPDL * |
|----------------|----------|--|--------------|--------------|------------|--------------|
| v0 | | <i>(size 64×64 px, 6 patterns)</i> | | | | |
| <i>pure</i> | ARS | 1.0 | 1.0 | 0.961 | 1.0 | 0.999 |
| | diff. | 0.0 | 0.0 | 0.002 | 0.0 | 0.0 |
| | time | 2.780 | 168.476 | 30.842 | 304.51 | 6.658 |
| <i>deform</i> | ARS | 0.775 | 0.921 | 0.769 | 0.777 | 0.993 |
| | diff. | 0.014 | 0.004 | 0.0213 | 0.014 | 0.0 |
| | time [s] | 1.697 | 141.527 | 22.833 | 279.87 | 4.766 |
| <i>D&N</i> | ARS | 0.048 | 0.778 | 0.002 | 0.066 | 0.999 |
| | diff. | 0.033 | 0.014 | 0.033 | 0.033 | 0.0 |
| | time [s] | 2.005 | 229.47 | 24.907 | 598.83 | 6.774 |

Comparison on synth. images

| datasets | | NMF [16] | FastICA [14] | sPCA [13] | DL [15] | BPDL * |
|----------------|----------|---|--------------|-----------|--------------|--------------|
| v1 | | <i>(size 64×64 px, 13 patterns)</i> | | | | |
| <i>pure</i> | ARS | 1.0 | 1.0 | 0.992 | 0.995 | 0.999 |
| | diff. | 0.0 | 0.0 | 0.0298 | 0.019 | 0.0 |
| | time | 2.333 | 340.32 | 18.291 | 737.47 | 6.029 |
| <i>deform</i> | ARS | 0.785 | 0.948 | 0.780 | 0.779 | 0.992 |
| | diff. | 0.017 | 0.004 | 0.029 | 0.033 | 0.005 |
| | time [s] | 4.001 | 312.18 | 15.000 | 700.03 | 7.561 |
| <i>D&N</i> | ARS | 0.091 | 0.878 | 0.009 | 0.0727 | 0.951 |
| | diff. | 0.048 | 0.010 | 0.061 | 0.0499 | 0.003 |
| | time [s] | 4.490 | 439.04 | 11.420 | 697.599 | 9.562 |

Comparison on synth. images

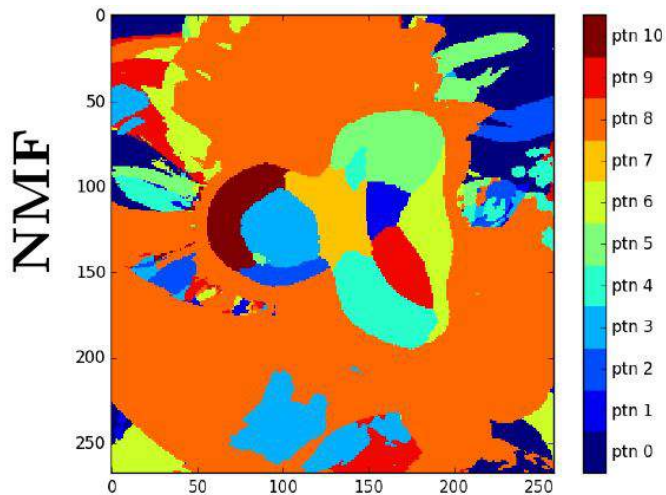
| | | | | | |
|----------|--------------------------|------------------------------|---------------------------|-------------------------|--------|
| datasets | NMF [16] | FastICA [14] | sPCA [13] | DL [15] | BPDL * |
|----------|--------------------------|------------------------------|---------------------------|-------------------------|--------|

| v2 | | <i>(size 128 × 128 px, 23 patterns)</i> | | | | |
|----------------|----------|---|--------------|--------------|------------|---------------|
| <i>pure</i> | ARS | 1.0 | 1.0 | 0.989 | 1.0 | 0.999 |
| | diff. | 0.0 | 0.0 | 0.037 | 0.0 | 0.005 |
| | time [s] | 82.329 | 5533.4 | 460.82 | 14786. | 88.260 |
| <i>deform</i> | ARS | 0.818 | 0.846 | 0.801 | 0.807 | 0.970 |
| | diff. | 0.019 | 0.015 | 0.056 | 0.046 | 0.004 |
| | time [s] | 144.10 | 5683.2 | 477.47 | 13619. | 165.22 |
| <i>D&N</i> | ARS | 0.120 | 0.612 | 0.024 | 0.144 | 0.877 |
| | diff. | 0.036 | 0.036 | 0.092 | 0.039 | 0.013 |
| | time [s] | 77.399 | 6912.9 | 485.44 | 13729. | 289.51 |

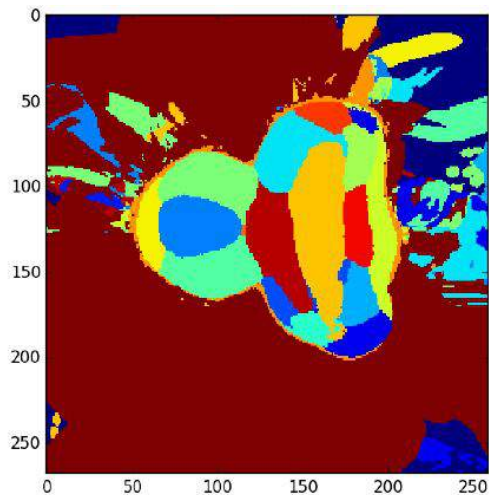
Results on Imaginal discs

Number of patterns

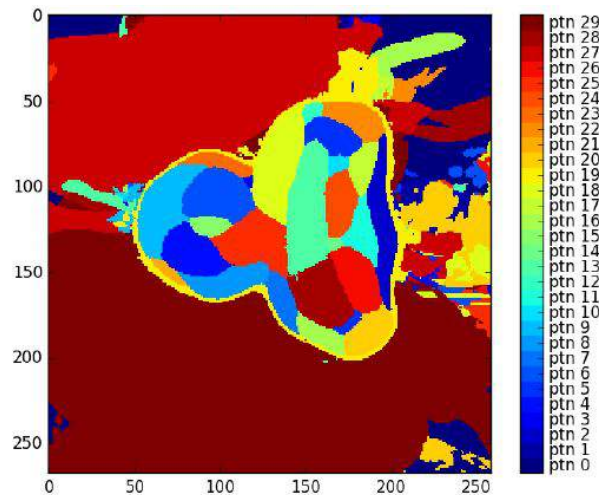
10



20



30

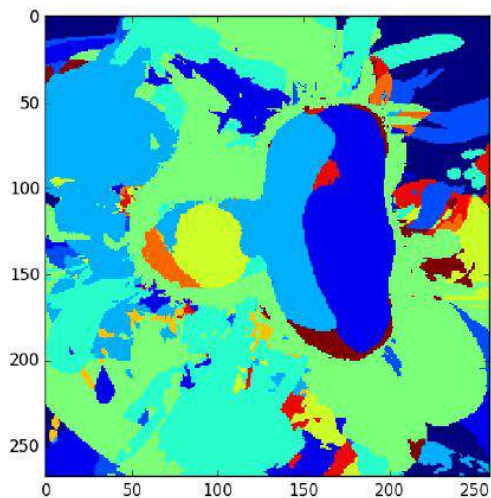


Results on Imaginal discs

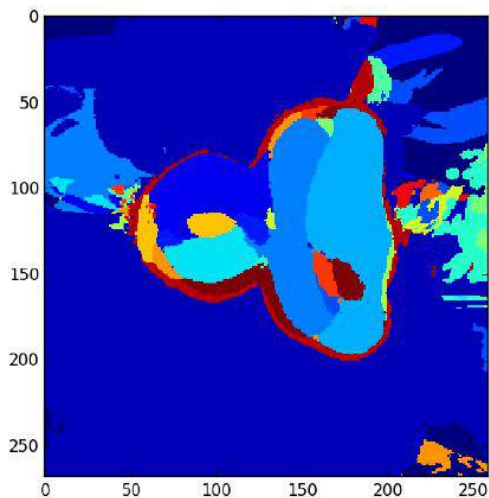
Number of patterns

10

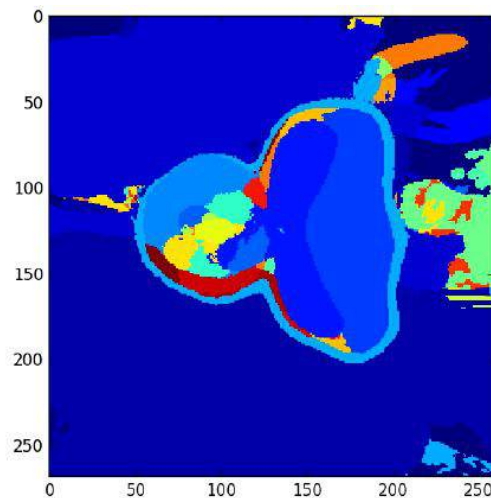
FastICA



20



30



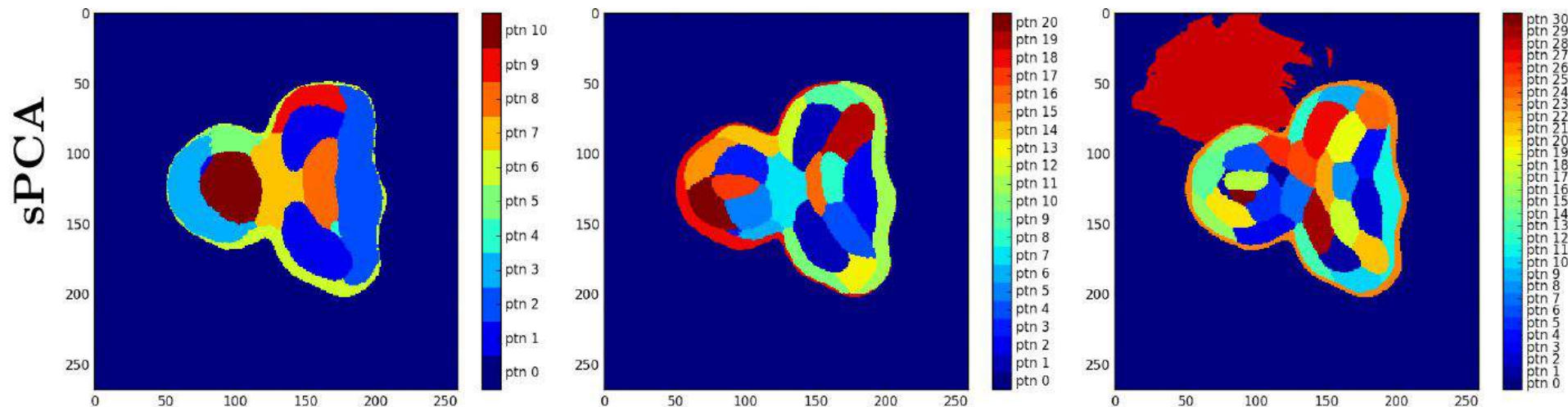
Results on Imaginal discs

Number of patterns

10

20

30



Results on Imaginal discs

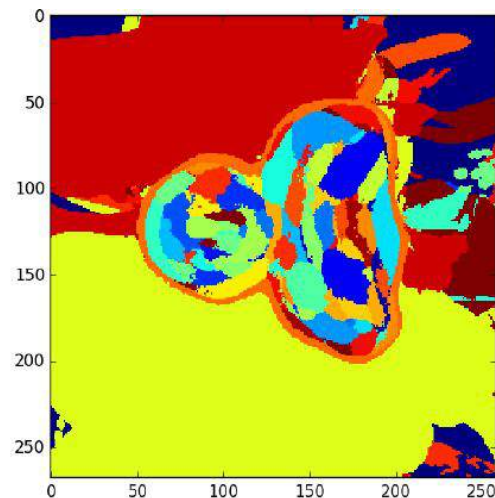
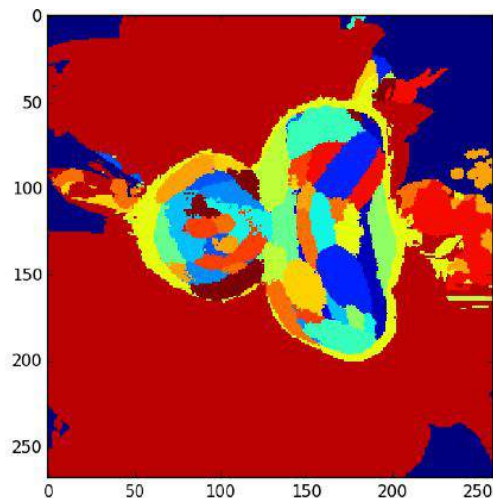
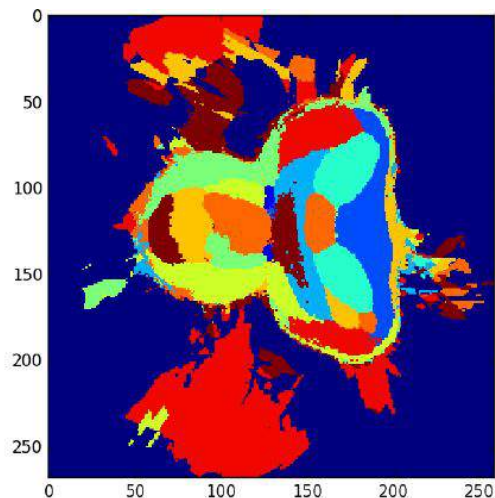
Number of patterns

10

20

30

DL



Results on Imaginal discs

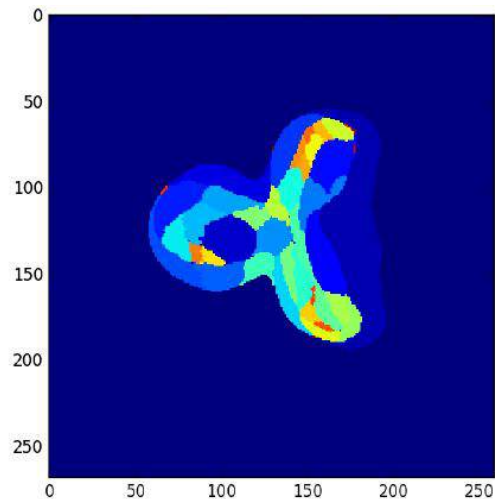
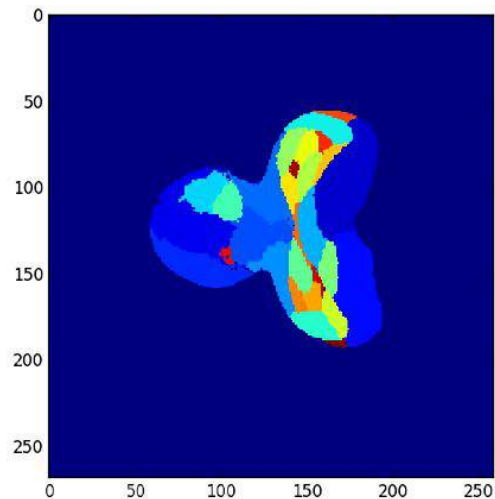
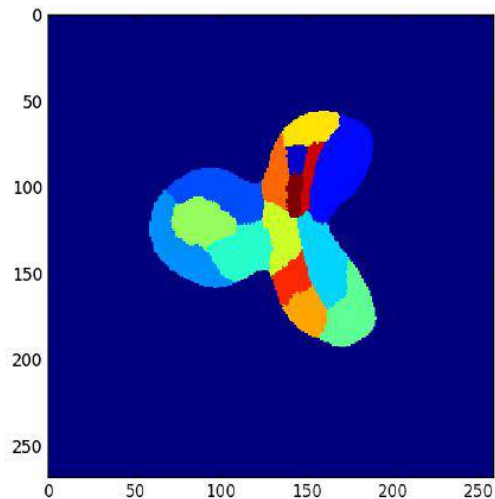
Number of patterns

10

20

30

BPDL



Results on Imaginal discs

| Method | Number of patterns K | | | Time [min] |
|--------------|------------------------|---------------|---------------|------------|
| | 10 | 20 | 30 | |
| NMF [16] | 0.0939 | 0.0823 | 0.0723 | 10 |
| FastICA [14] | 0.1197 | 0.0779 | 0.0485 | 24 |
| sPCA [13] | 0.0476 | 0.0413 | 0.0352 | 477 |
| DL [15] | 0.0939 | 0.0648 | 0.0596 | 338 |
| BPDL * | 0.0467 | 0.0395 | 0.0361 | 20 |

Results on Imaginal discs



Drosophila Imaginal discs

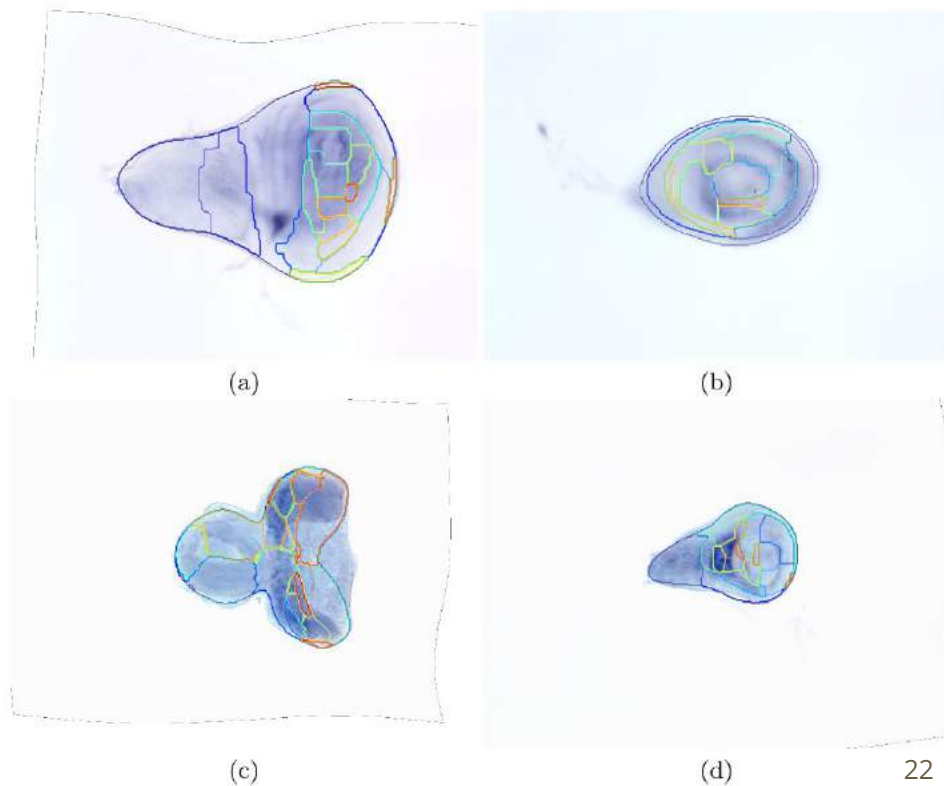
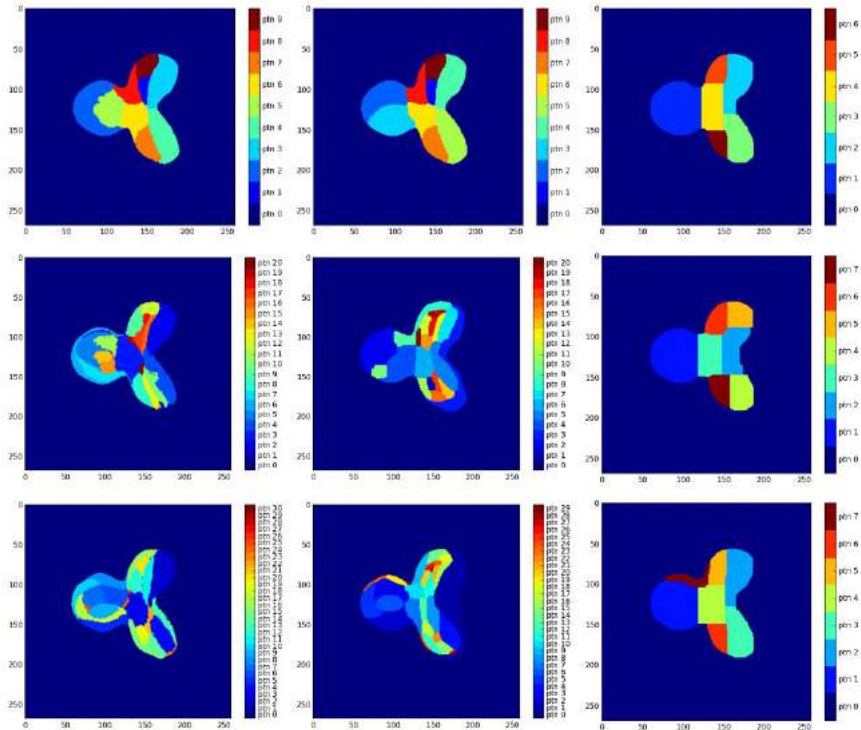
nb.
 K

Graph Cut regularization β

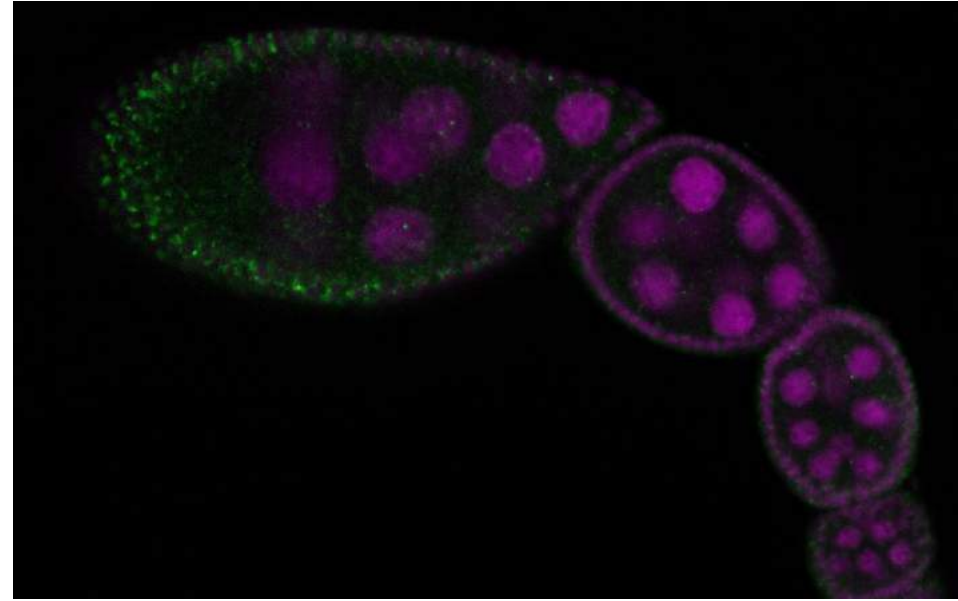
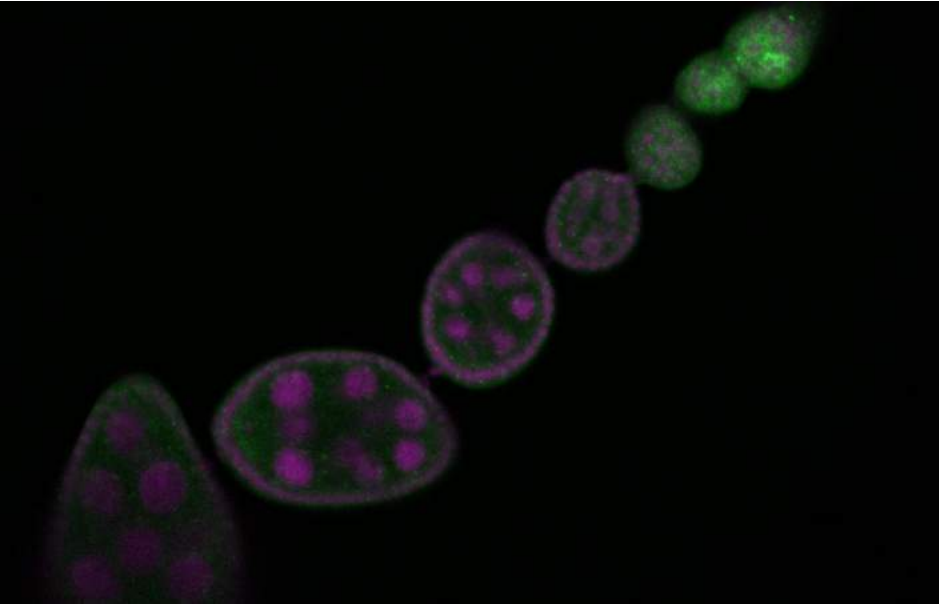
0.0

0.001

0.1



Future: Drosophila ovary 3D



Home page: <http://cmp.felk.cvut.cz/~borovji3>

Mail: jiri.borovec@fel.cvut.cz