Learning the Dependency Structure of Latent Factors

Yunlong He*, Yanjun Qi†, Koray Kavukcuoglu†, Haesun Park*

*Georgia Institute of Technology   † NEC Labs America

NEC Laboratories America
Relentless passion for innovation

Background:

Goals

1. Learn interpretable latent factors with hidden mutual dependency.
2. Learn a concise dependency structure between latent factors.

Linear latent factor model

$x = s_1 B_1 + \cdots + s_k B_k + \epsilon$

where $B = [B_1, \ldots, B_k]$ is fixed but unknown,

$s = (s_1, \ldots, s_k) \in R^k$: latent factors

Sparse Undirected Graphical Model

Example: Gaussian random vector $s \sim N(0, \Phi^{-1})$

- $\Phi_{i,j} = 0$: s_i and s_j are conditionally independent.
- $\Phi_{i,j} > 0$: s_i and s_j are negatively correlated.
- $\Phi_{i,j} < 0$: s_i and s_j are positively correlated.

Structured Latent Factor Analysis (SLFA)

- **Basic idea**: Gaussian prior distribution for the latent factor vector s:

$$s \sim N(0, \Phi^{-1})$$

where Φ is a sparse precision matrix.

- Use Φ to model the dependency structure between components of s.

Why sparse precision matrix Φ?

1. Occam's razor: law of parsimony
2. Sparse graph to ease the analysis.
3. To avoid over-fitting to the data.
4. To avoid singular solutions for the optimization problem.
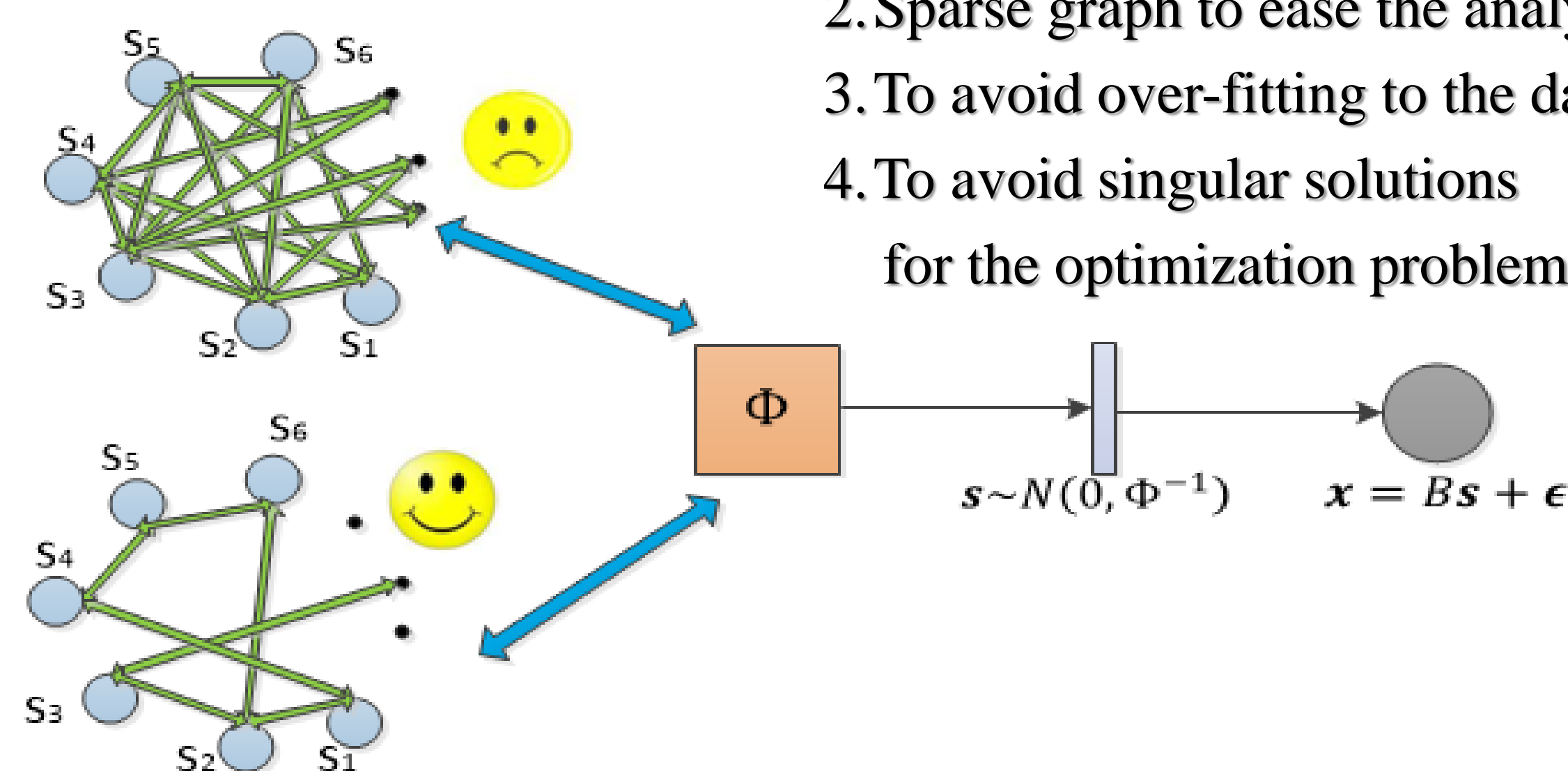
Gaussian case: SLFA

$$\min_{B,S,\Phi} \frac{1}{N} \|X - BS\|_F^2 + \sigma^2 \left(\frac{1}{N} tr(S^T \Phi S) - \log \det(\Phi) + \rho \|\Phi\|_1 \right)$$

$$s.t. \ \Phi \succcurlyeq 0, B \geq 0, \|B_k\|_2 \leq 1, k = 1, \ldots, K$$

General Formulation

- Considers data samples drawn from the exponential family.
- Introduce a pairwise Markov Random Field (MRF) prior on the vector of factors.
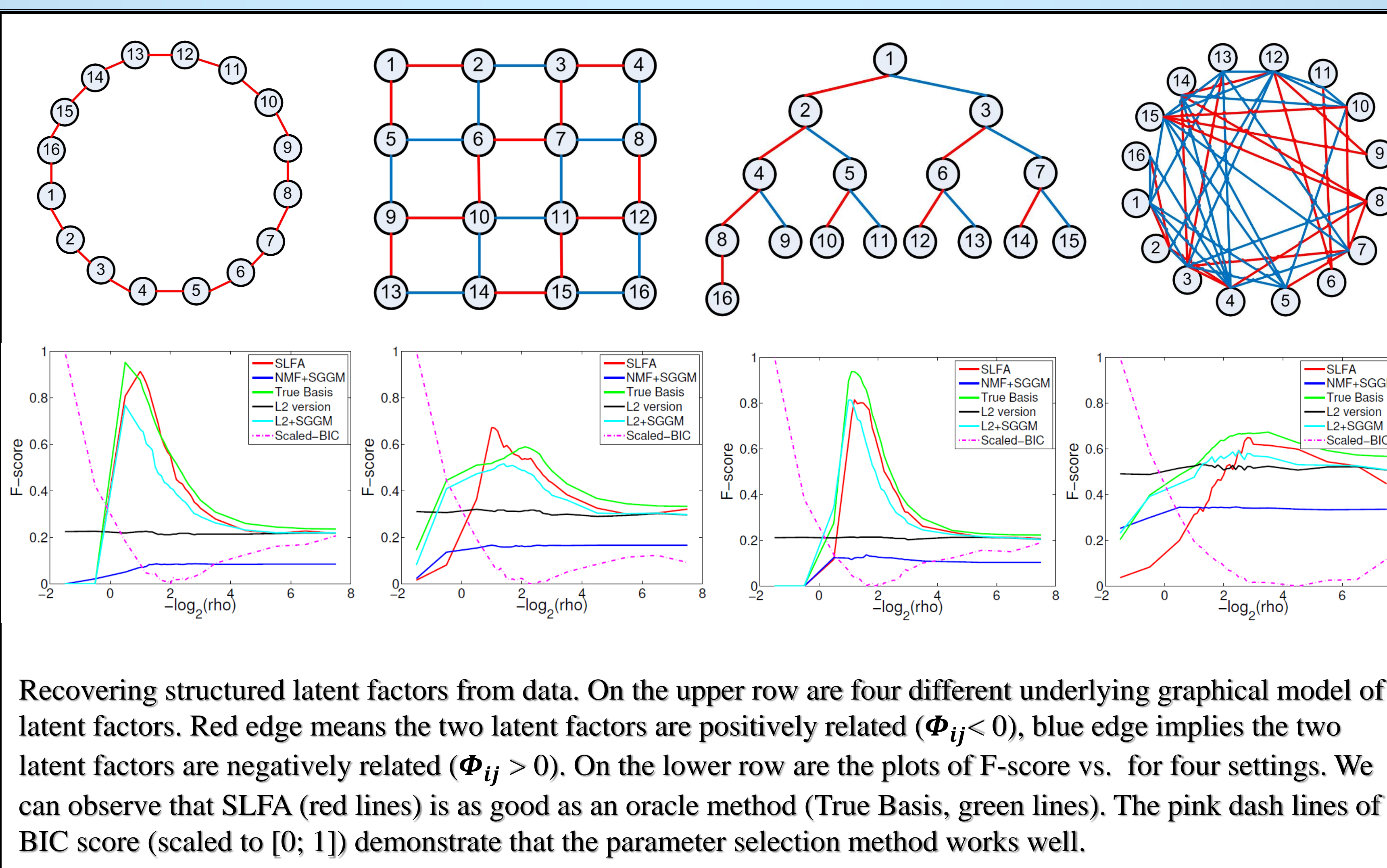- Objective

$$\min_{B,S,\Theta} \frac{1}{N} \sum_i \{ -\log h(\mathbf{x}^{(i)}) + A(\mathbf{B}\mathbf{s}^{(i)}) - \mathbf{s}^{(i)\mathsf{T}} \mathbf{B}^\mathsf{T} T(\mathbf{x}^{(i)}) \}$$

$$+ \log Z(\mu, \Theta) + \frac{1}{N} \mu^\mathsf{T} \mathbf{S} \mathbf{1}_N + \frac{1}{2N} tr(\mathbf{S}^\mathsf{T} \Theta \mathbf{S}) + \frac{1}{2} \rho \|\Theta\|_1$$

$$s.t. \ \mathbf{B} \geq 0, \|\mathbf{B}_k\|_2 \leq 1, k = 1, \ldots, K,$$

- The algorithm is based on Block Coordinate Descent algorithm and exhibits convergence behavior to a stationary point.
- An online algorithm for learning SLFA scales to large data sets.

Synthetic Data

Recovering structured latent factors from data. On the upper row are four different underlying graphical model of latent factors. Red edge means the two latent factors are positively related ($\Phi_{ij} < 0$), blue edge implies the two latent factors are negatively related ($\Phi_{ij} > 0$). On the lower row are the plots of F-score vs. for four settings. We can observe that SLFA (red lines) is as good as an oracle method (True Basis, green lines). The pink dash lines of BIC score (scaled to [0; 1]) demonstrate that the parameter selection method works well.

Gene Microarray Analysis

| Method	SLFA	Lasso overlapped-group	Lasso	SVM	PCA
Cross-validation error rate	34.22 ±2.58	35.31 ±2.05	36.42 ±2.50	36.93 ±2.54	36:85 ±3.02

Tumor classification based on gene expression values of 8141 genes for 295 breast cancer tumor samples. SLFA does not use prior knowledge like biological gene network graph.
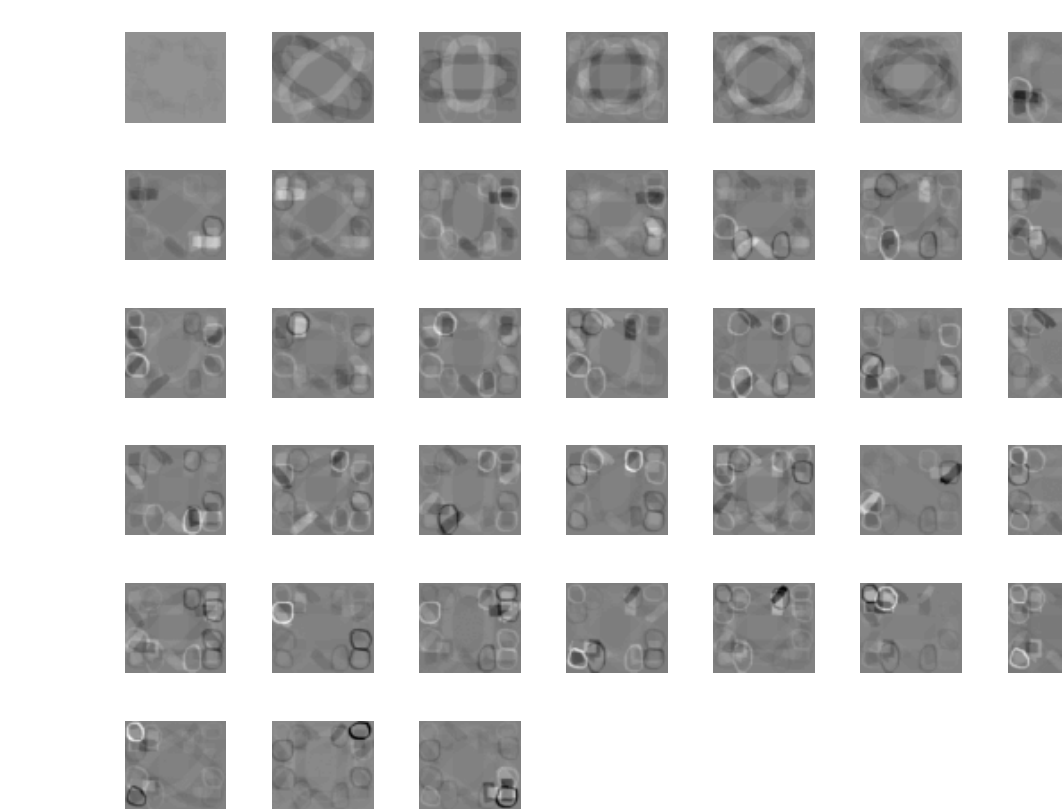
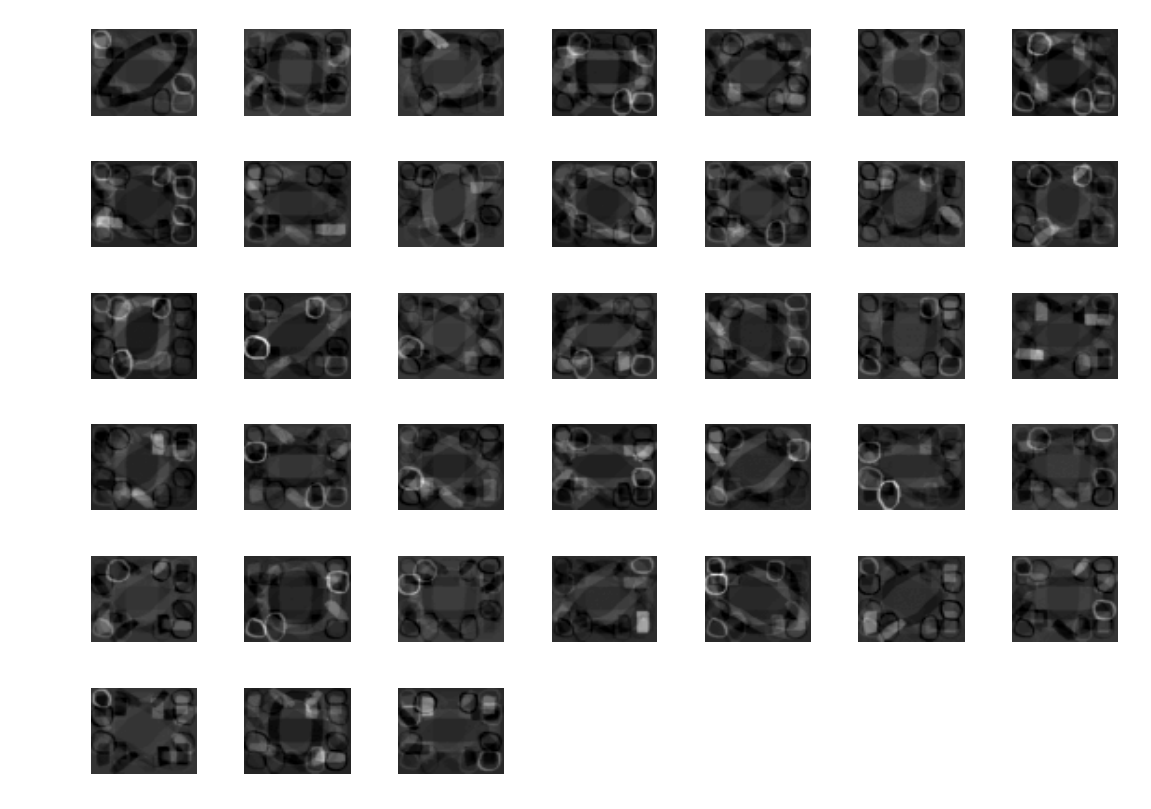## More Experiment
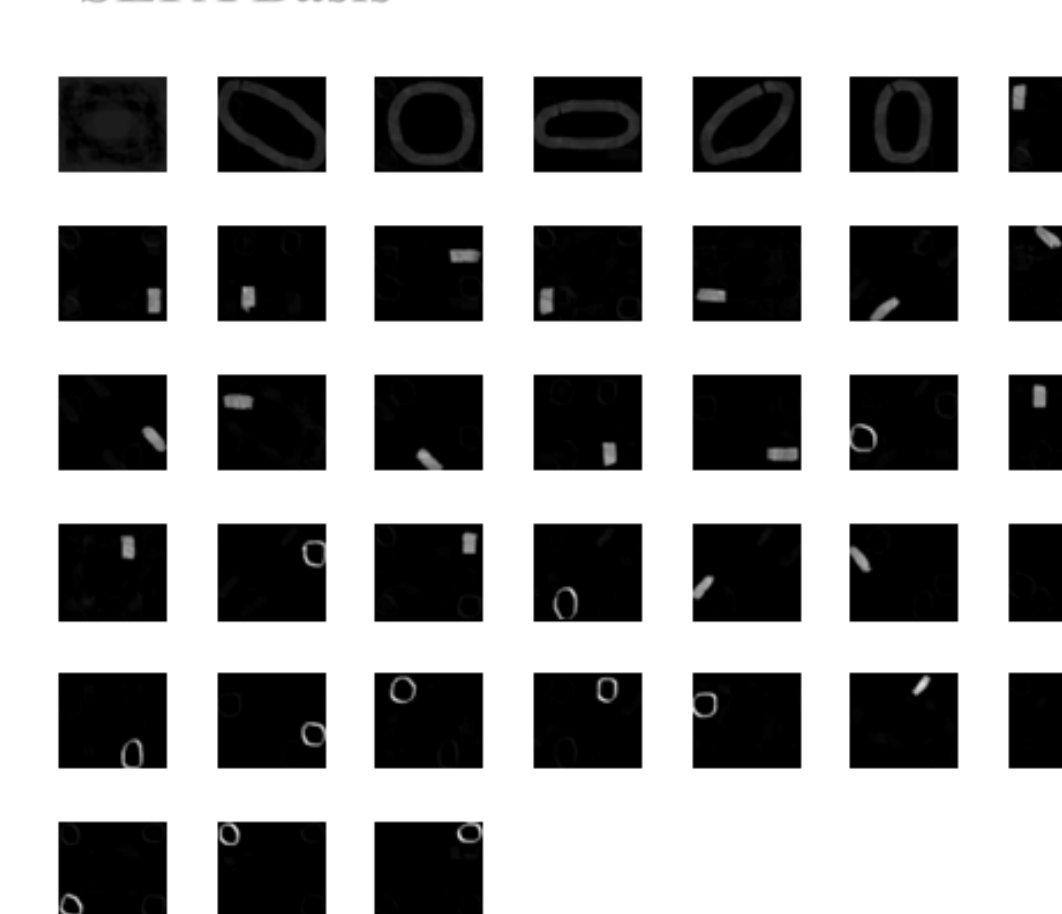
Structured Visualization of NIPS topics

Positively Related Topics   Negatively Related Topics

Modeled using 40 topics
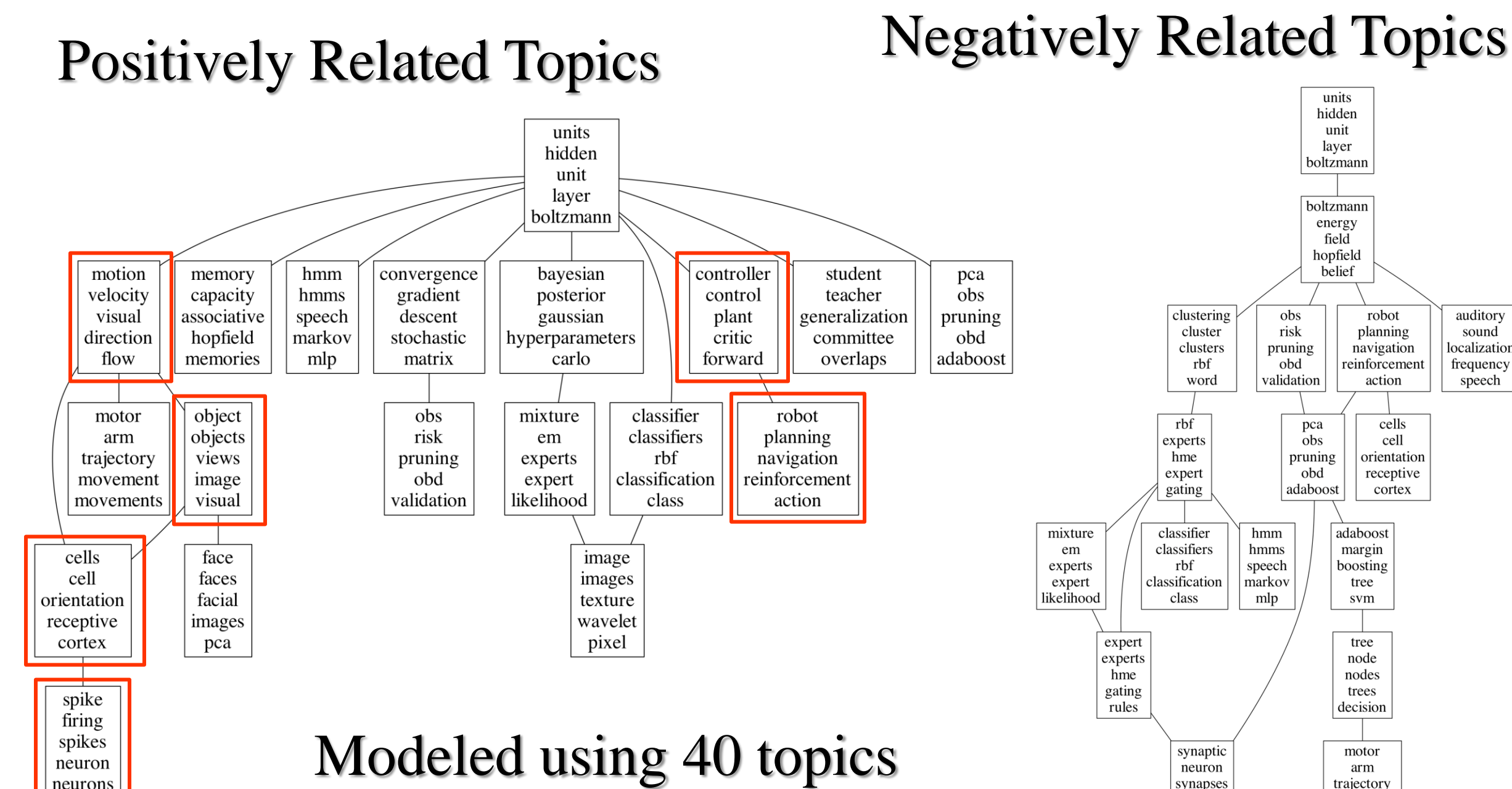
Simulated Bug Images
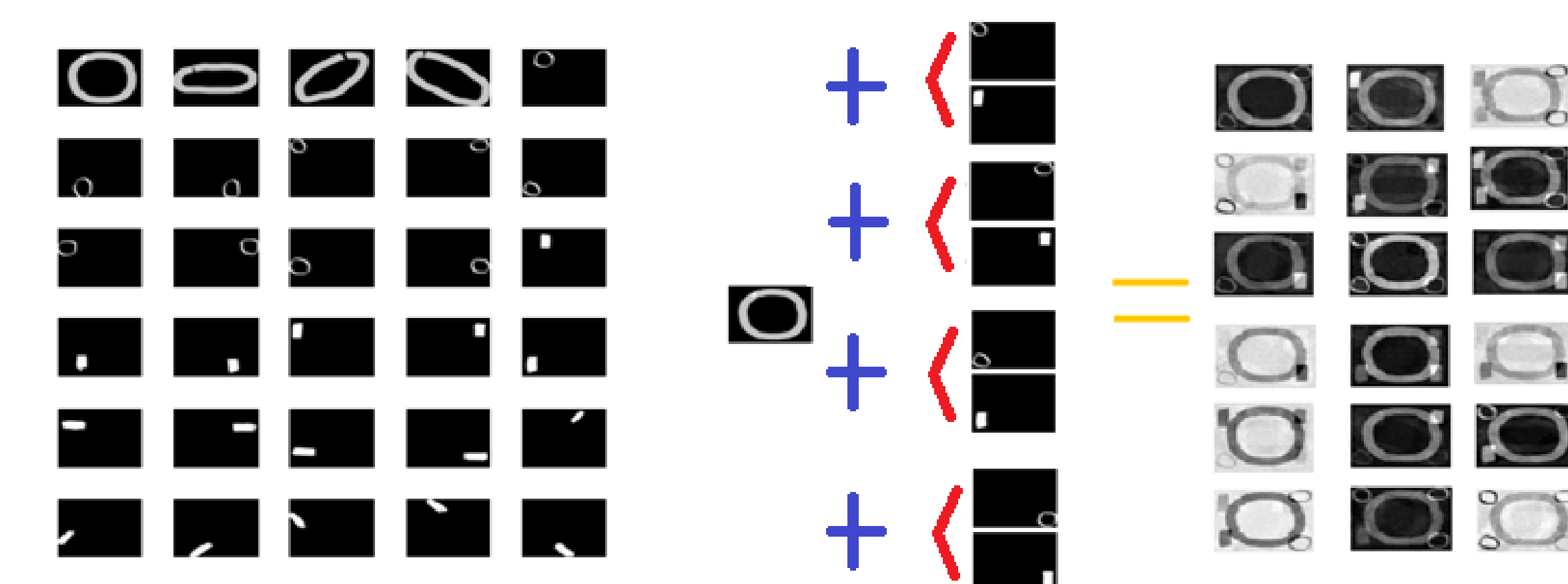
PCA Basis

NMF Basis

SLFA Basis

Φ Analysis $\Phi > 0$: negatively correlated
$\Phi < 0$: positively correlated

$\Phi(i,j)$	Image Pair	$\Phi(i,j)$
0.030		−0.016
0.020		−0.015
0.015		−0.013
0.015		−0.012
0.014		−0.011
0.013		−0.011

Contact: he.yunlong@gatech.edu