Learning Discrimination Specific, Self-Collaborative and Nonlinear Model

Dimche Kostadinov, **Behrooz Razeghi**,
Slava Voloshynovskiy and Sohrab Ferdowsi

Stochastic Information Processing Group
University of Geneva
Switzerland

November 2018

UNIVERSITÉ DE GENÈVE

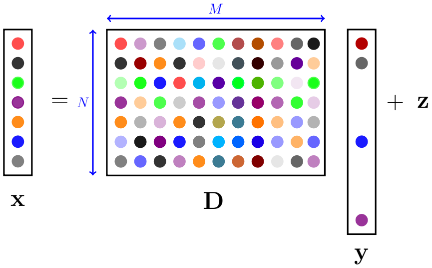
stochastic information processing

Outline

Backgrounds

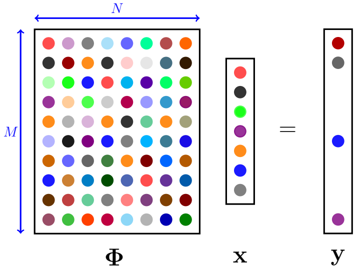
Synthesis Model for Sparse Representation

Synthesis model or **regression model with sparsity regularized penalty** synthesizes data sample $\mathbf{x} \in \mathbb{R}^N$ as an approximation by a sparse linear combination $\mathbf{y} \in \mathbb{R}^M$, $\|\mathbf{y}\|_0 \ll M$, of a few vectors $\mathbf{d}_m \in \mathbb{R}^N$, from a dictionary $\mathbf{D} = [\mathbf{d}_1, ..., \mathbf{d}_M]$, i.e., $\mathbf{x} = \mathbf{D}\mathbf{y} + \mathbf{z}$.



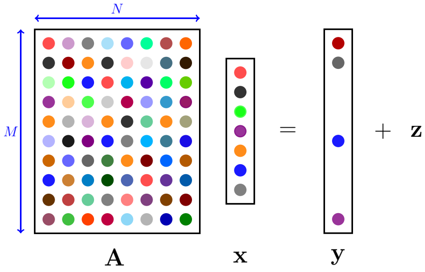
Analysis Model for Sparse Representation

Analysis model uses a dictionary $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ with $M > N$ to analyze the data sample $\mathbf{x} \in \mathbb{R}^N$. This model assumes that the product of $\mathbf{\Phi}$ and $\mathbf{x}$ is sparse, i.e., $\mathbf{\Phi x} = \mathbf{y}$ with $\|\mathbf{y}\|_0 = M - s$, $0 \leq s \leq M$.

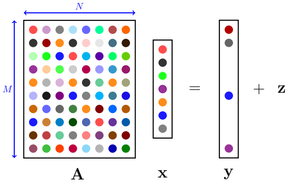


Transform Model for Sparse Representation

Transform model assumes that the data sample $\mathbf{x} \in \mathbb{R}^N$ is approximately sparsifiable under a linear transform $\mathbf{A} \in \mathbb{R}^{M \times N}$, i.e., $\mathbf{Ax} = \mathbf{y} + \mathbf{z}$, where $\mathbf{y} = \mathcal{T}(\mathbf{x})$, $\|\mathbf{y}\|_0 \ll M$ and $\mathbf{z} \in \mathbb{R}^M$ is an error vector in transform domain.



Transform Model for Sparse Representation



- Given $\mathbf{A}$, and sparsity s, **transform sparse coding** is:

$$\widehat{\mathbf{y}} = \arg\min_{\mathbf{y}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2, \quad \text{s.t.} \quad \|\mathbf{y}\|_0 \leq s$$

- $\widehat{\mathbf{y}}$ computed exactly by a thresholding $\mathbf{A}\mathbf{x}$ to the s largest magnitude elements $\Rightarrow$ Sparse coding is cheap!
- Signal recovered as $\mathbf{A}^{\dagger}\mathbf{y}$
- $\mathbf{z}$ is error term in the **transform domain**

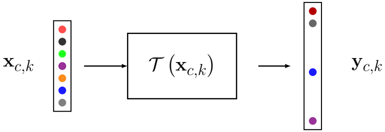
Unstructured Transform Learning

$$\left(\widehat{\mathbf{A}}, \widehat{\mathbf{Y}}\right) = \arg\min_{\mathbf{A},\mathbf{Y}} \overbrace{\|\mathbf{AX} - \mathbf{Y}\|_F^2}^{\text{Sparsification Error}} + \overbrace{\Omega\left(\mathbf{A}\right)}^{\text{Linear Map Constraint}}, \quad \text{s.t.} \quad \|\mathbf{y}_k\|_0 \leq s, \forall k$$

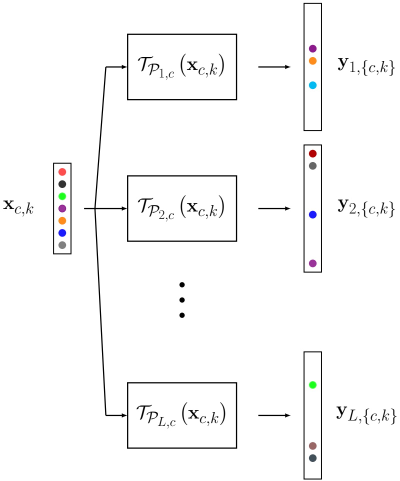
- $\mathbf{X} = [\mathbf{x}_1 \mid \mathbf{x}_2 \mid ... \mid \mathbf{x}_K] \in \mathbb{R}^{N \times K}$: matrix of training signals
- $\mathbf{Y} = [\mathbf{y}_1 \mid \mathbf{y}_2 \mid ... \mid \mathbf{y}_K] \in \mathbb{R}^{M \times K}$: matrix of sparse codes for $\mathbf{X}$
- **Sparsification Error** measures deviation of data in a transform domain
- $\Omega\left(\mathbf{A}\right)$ penalizes the information loss in order to avoid trivially unwanted matrices, e.g., matrices that have repeated or zero rows.

Approach Overview

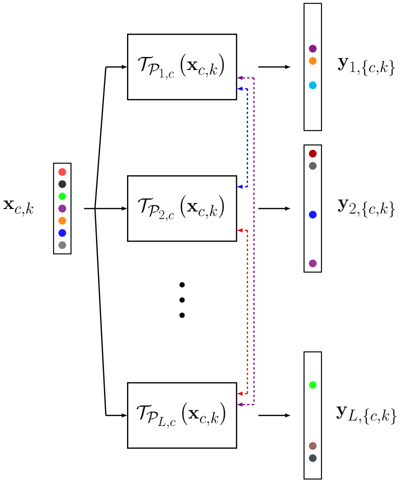
General Block Diagram



General Block Diagram

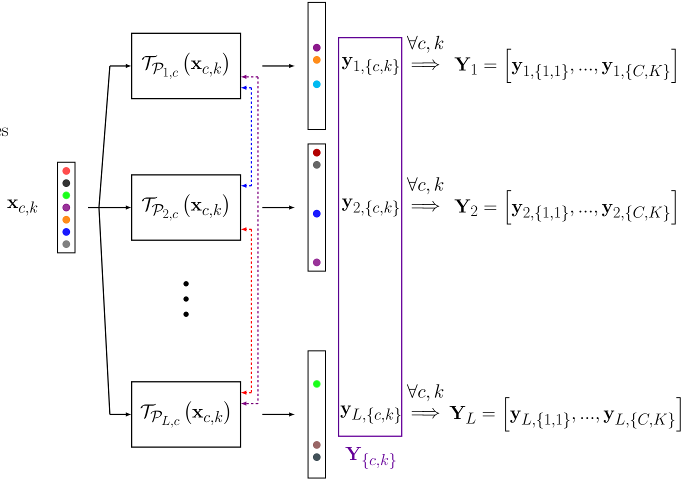


General Block Diagram



General Block Diagram

- C Classes
  $c \in \mathcal{C} = \{1, ..., C\}$
- Each class K Samples
  $k \in \mathcal{K} = \{1, ..., K\}$



Joint Modeling with Collaboration

- $p(\mathbf{x}_{c,k}, \mathbf{Y}_{\{c,k\}}, \boldsymbol{\theta}, \mathbf{A}) = p(\mathbf{x}_{c,k}, \mathbf{Y}_{\{c,k\}}, \boldsymbol{\theta}|\mathbf{A})p(\mathbf{A})$
- with $p(\mathbf{x}_{c,k}, \mathbf{Y}_{\{c,k\}}, \boldsymbol{\theta}|\mathbf{A}) = \underbrace{p(\mathbf{x}_{c,k}|\mathbf{Y}_{\{c,k\}}, \mathbf{A})}_{\substack{\text{collaboration corrective} \\ \text{nonlinear transform error}}} \underbrace{p(\boldsymbol{\theta}, \mathbf{Y}_{\{c,k\}})}_{\text{discriminative prior}}$

Joint Modeling with Collaboration

- $p(\mathbf{x}_{c,k}, \mathbf{Y}_{\{c,k\}}, \boldsymbol{\theta}, \mathbf{A}) = p(\mathbf{x}_{c,k}, \mathbf{Y}_{\{c,k\}}, \boldsymbol{\theta}|\mathbf{A})p(\mathbf{A})$
- with $p(\mathbf{x}_{c,k}, \mathbf{Y}_{\{c,k\}}, \boldsymbol{\theta}|\mathbf{A}) = \underbrace{p(\mathbf{x}_{c,k}|\mathbf{Y}_{\{c,k\}}, \mathbf{A})}_{\text{collaboration corrective nonlinear transform error}} \underbrace{p(\boldsymbol{\theta}, \mathbf{Y}_{\{c,k\}})}_{\text{discriminative prior}}$
- and $p(\mathbf{x}_{c,k}|\mathbf{Y}_{\{c,k\}}, \mathbf{A})$

$$\propto \prod_{l=1}^{L} \exp\left(-\frac{1}{\beta_0}\left(\mathbf{z}_{l,\{c,k\}}^T \mathbf{z}_{l,\{c,k\}} + \overbrace{f_{TSC}\left(\mathbf{z}_{l,\{c,k\}}, g_A(\mathbf{Z}_{\{c,k\}\setminus l})\right)}^{\text{self collaborative component}}\right)\right)$$

Joint Modeling with Collaboration

- $p(\mathbf{x}_{c,k}, \mathbf{Y}_{\{c,k\}}, \boldsymbol{\theta}, \mathbf{A}) = p(\mathbf{x}_{c,k}, \mathbf{Y}_{\{c,k\}}, \boldsymbol{\theta}|\mathbf{A})p(\mathbf{A})$
- with $p(\mathbf{x}_{c,k}, \mathbf{Y}_{\{c,k\}}, \boldsymbol{\theta}|\mathbf{A}) = \underbrace{p(\mathbf{x}_{c,k}|\mathbf{Y}_{\{c,k\}}, \mathbf{A})}_{\text{collaboration corrective nonlinear transform error}} \underbrace{p(\boldsymbol{\theta}, \mathbf{Y}_{\{c,k\}})}_{\text{discriminative prior}}$
- and $p(\mathbf{x}_{c,k}|\mathbf{Y}_{\{c,k\}}, \mathbf{A})$

$$\propto \prod_{l=1}^{L} \exp\left(-\frac{1}{\beta_0}\left(\mathbf{z}_{l,\{c,k\}}^T \mathbf{z}_{l,\{c,k\}} + \overbrace{f_{TSC}\left(\mathbf{z}_{l,\{c,k\}}, g_A(\mathbf{Z}_{\{c,k\}\setminus l})\right)}^{\text{self collaborative component}}\right)\right)$$

  - $\mathbf{z}_{l,\{c,k\}} = \mathbf{A}_l \mathbf{x}_{c,k} - \mathbf{y}_{l,\{c,k\}}$
  - $f_{TSC}(.) : \mathbb{R}^M \times \mathbb{R}^M \rightarrow \mathbb{R}$: **Target Specific Collaboration Function**
  - $g_A(.) : \mathbb{R}^M \times \cdots \times \mathbb{R}^M \rightarrow \mathbb{R}$: **Collaboration Aggregation Function**

Joint Modeling with Collaboration

- $p(\mathbf{x}_{c,k}, \mathbf{Y}_{\{c,k\}}, \boldsymbol{\theta}, \mathbf{A}) = p(\mathbf{x}_{c,k}, \mathbf{Y}_{\{c,k\}}, \boldsymbol{\theta}|\mathbf{A})p(\mathbf{A})$
- with $p(\mathbf{x}_{c,k}, \mathbf{Y}_{\{c,k\}}, \boldsymbol{\theta}|\mathbf{A}) = \underbrace{p(\mathbf{x}_{c,k}|\mathbf{Y}_{\{c,k\}}, \mathbf{A})}_{\text{collaboration corrective nonlinear transform error}} \underbrace{p(\boldsymbol{\theta}, \mathbf{Y}_{\{c,k\}})}_{\text{discriminative prior}}$
- and $p(\mathbf{x}_{c,k}|\mathbf{Y}_{\{c,k\}}, \mathbf{A})$

$$\propto \prod_{l=1}^{L} \exp\left(-\frac{1}{\beta_0}\left(\mathbf{z}_{l,\{c,k\}}^T \mathbf{z}_{l,\{c,k\}} + \overbrace{f_{TSC}\left(\mathbf{z}_{l,\{c,k\}}, g_A(\mathbf{Z}_{\{c,k\}\setminus l})\right)}^{\text{self collaborative component}}\right)\right)$$

  - $\mathbf{z}_{l,\{c,k\}} = \mathbf{A}_l \mathbf{x}_{c,k} - \mathbf{y}_{l,\{c,k\}}$
  - $f_{TSC}(.) : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}$: **Target Specific Collaboration Function**
  - $g_A(.) : \mathbb{R}^M \times \cdots \times \mathbb{R}^M \to \mathbb{R}$: **Collaboration Aggregation Function**

- $p(\boldsymbol{\theta}, \mathbf{Y}_{\{c,k\}}) = \prod_{l=1}^{L} p(\boldsymbol{\theta}_l|\mathbf{y}_{l,\{c,k\}})p(\mathbf{y}_{l,\{c,k\}})$

Joint Modeling with Collaboration

- $p(\mathbf{x}_{c,k}, \mathbf{Y}_{\{c,k\}}, \boldsymbol{\theta}, \mathbf{A}) = p(\mathbf{x}_{c,k}, \mathbf{Y}_{\{c,k\}}, \boldsymbol{\theta}|\mathbf{A})p(\mathbf{A})$
- with $p(\mathbf{x}_{c,k}, \mathbf{Y}_{\{c,k\}}, \boldsymbol{\theta}|\mathbf{A}) = \underbrace{p(\mathbf{x}_{c,k}|\mathbf{Y}_{\{c,k\}}, \mathbf{A})}_{\substack{\text{collaboration corrective} \\ \text{nonlinear transform error}}} \underbrace{p(\boldsymbol{\theta}, \mathbf{Y}_{\{c,k\}})}_{\text{discriminative prior}}$
- and $p(\mathbf{x}_{c,k}|\mathbf{Y}_{\{c,k\}}, \mathbf{A})$

$$\propto \prod_{l=1}^{L} \exp\left(-\frac{1}{\beta_0}\left(\mathbf{z}_{l,\{c,k\}}^T \mathbf{z}_{l,\{c,k\}} + \overbrace{f_{TSC}\left(\mathbf{z}_{l,\{c,k\}}, g_A(\mathbf{Z}_{\{c,k\}\setminus l})\right)}^{\text{self collaborative component}}\right)\right)$$

  - $\mathbf{z}_{l,\{c,k\}} = \mathbf{A}_l \mathbf{x}_{c,k} - \mathbf{y}_{l,\{c,k\}}$
  - $f_{TSC}(.) : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}$: **Target Specific Collaboration Function**
  - $g_A(.) : \mathbb{R}^M \times \cdots \times \mathbb{R}^M \to \mathbb{R}$: **Collaboration Aggregation Function**
- $p(\boldsymbol{\theta}, \mathbf{Y}_{\{c,k\}}) = \prod_{l=1}^{L} p(\boldsymbol{\theta}_l|\mathbf{y}_{l,\{c,k\}})p(\mathbf{y}_{l,\{c,k\}})$

Self-Collaboration Discriminative Prior and its Measure

Unsupervised Discriminative Prior:

$$p(\boldsymbol{\theta}, \mathbf{Y}_{\{c,k\}}) = \prod_{l} p(\boldsymbol{\theta}_l | \mathbf{y}_{l,\{c,k\}}) p(\mathbf{y}_{l,\{c,k\}})$$

where

- $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_L\}, \ \boldsymbol{\theta}_l = \{\boldsymbol{\theta}_{l,1}, \boldsymbol{\theta}_{l,2}\} = \{ \overbrace{\{\boldsymbol{\tau}_{l,1}, ..., \boldsymbol{\tau}_{l,C1}\}}^{\text{Dissimilarity Parameters}}, \underbrace{\{\boldsymbol{\nu}_{l,1}, ..., \boldsymbol{\nu}_{l,C2}\}}_{\text{Similarity Parameters}} \}$
- $p(\boldsymbol{\theta}_l | \mathbf{y}_{l,\{c,k\}}) \propto \exp(-\frac{1}{\beta_I} \overbrace{l_I(\boldsymbol{\theta}_l, \mathbf{y}_{l,\{c,k\}})}^{\text{discriminative measure}})$
- $p(\mathbf{y}_{l,\{c,k\}}) \propto \exp(-\frac{\|\mathbf{y}_{l,\{c,k\}}\|_1}{\beta_{l,1}}) \quad \Rightarrow \text{sparsity inducing prior}$
- $l_I(\boldsymbol{\theta}_l, \mathbf{y}_{l,\{c,k\}}) = \min_{1 \le c1 \le C1} \max_{1 \le c2 \le C2} \Big(\mathrm{Sim}(\mathbf{y}_{l,\{c,k\}}, \boldsymbol{\tau}_{l,c1}) + \mathrm{Sim}(\mathbf{y}_{l,\{c,k\}}, \boldsymbol{\nu}_{l,c2}) + \mathrm{Stg}(\mathbf{y}_{l,\{c,k\}}, \boldsymbol{\tau}_{l,c1}) \Big)$

Similarity and Strength Measures

$$l_I(\boldsymbol{\theta}_l, \mathbf{y}_{l,\{c,k\}}) = \min_{1\le c1\le C1} \max_{1\le c2\le C2} \Big(\mathrm{Sim}(\mathbf{y}_{l,\{c,k\}}, \boldsymbol{\tau}_{l,c1}) + \mathrm{Sim}(\mathbf{y}_{l,\{c,k\}}, \boldsymbol{\nu}_{l,c2}) + \mathrm{Stg}(\mathbf{y}_{l,\{c,k\}}, \boldsymbol{\tau}_{l,c1})\Big)$$

$$\mathrm{Sim}(\mathbf{y}_{l,\{c,k\}}, \mathbf{y}_{l,\{c1,k1\}}) = \|\mathbf{y}^-_{l,\{c,k\}} \odot \mathbf{y}^-_{l,\{c1,k1\}}\|_1 + \|\mathbf{y}^+_{l,\{c,k\}} \odot \mathbf{y}^+_{l,\{c1,k1\}}\|_1$$
$$\mathrm{Stg}(\mathbf{y}_{l,\{c,k\}}, \mathbf{y}_{l,\{c1,k1\}}) = \|\mathbf{y}_{l,\{c,k\}} \odot \mathbf{y}_{l,\{c1,k1\}}\|_2^2$$

where

- $\odot$ denotes Hadamard product
- $\mathbf{y}_{l,\{c,k\}} = \mathbf{y}^+_{l,\{c,k\}} - \mathbf{y}^-_{l,\{c,k\}} \Rightarrow \mathbf{y}^+_{l,\{c,k\}} = \max(\mathbf{y}_{l,\{c,k\}}, \mathbf{0})$, $\mathbf{y}^-_{l,\{c,k\}} = \max(-\mathbf{y}_{l,\{c,k\}}, \mathbf{0})$
- $\mathbf{y}_{l,\{c1,k1\}} = \mathbf{y}^+_{l,\{c1,k1\}} - \mathbf{y}^-_{l,\{c1,k1\}}$

Illustration of Similarity and Dissimilarity Measures

$$\begin{aligned} \mathrm{Sim}(\mathbf{y}_{l,\{c,k\}}, \mathbf{y}_{l,\{c1,k1\}}) &= \|\mathbf{y}^{-}_{l,\{c,k\}} \odot \mathbf{y}^{-}_{l,\{c1,k1\}}\|_1 + \|\mathbf{y}^{+}_{l,\{c,k\}} \odot \mathbf{y}^{+}_{l,\{c1,k1\}}\|_1 \\ \mathrm{Dis}(\mathbf{y}_{l,\{c,k\}}, \mathbf{y}_{l,\{c1,k1\}}) &= \|\mathbf{y}^{+}_{l,\{c,k\}} \odot \mathbf{y}^{-}_{l,\{c1,k1\}}\|_1 + \|\mathbf{y}^{-}_{l,\{c,k\}} \odot \mathbf{y}^{+}_{l,\{c1,k1\}}\|_1 \end{aligned}$$

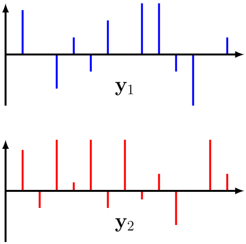


Illustration of Similarity and Dissimilarity Measures

$$\mathrm{Sim}(\mathbf{y}_{l,\{c,k\}}, \mathbf{y}_{l,\{c1,k1\}}) = \|\mathbf{y}^{-}_{l,\{c,k\}} \odot \mathbf{y}^{-}_{l,\{c1,k1\}}\|_1 + \|\mathbf{y}^{+}_{l,\{c,k\}} \odot \mathbf{y}^{+}_{l,\{c1,k1\}}\|_1$$
$$\mathrm{Dis}(\mathbf{y}_{l,\{c,k\}}, \mathbf{y}_{l,\{c1,k1\}}) = \|\mathbf{y}^{+}_{l,\{c,k\}} \odot \mathbf{y}^{-}_{l,\{c1,k1\}}\|_1 + \|\mathbf{y}^{-}_{l,\{c,k\}} \odot \mathbf{y}^{+}_{l,\{c1,k1\}}\|_1$$

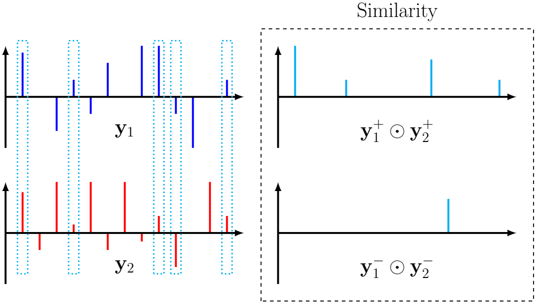
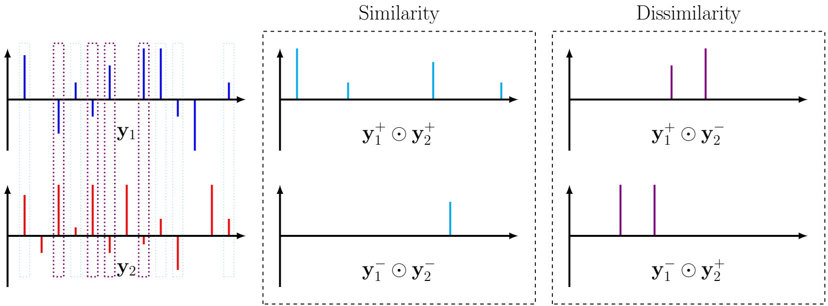


Illustration of Similarity and Dissimilarity Measures

$$\mathrm{Sim}(\mathbf{y}_{l,\{c,k\}},\mathbf{y}_{l,\{c1,k1\}}) = \|\mathbf{y}^{-}_{l,\{c,k\}} \odot \mathbf{y}^{-}_{l,\{c1,k1\}}\|_1 + \|\mathbf{y}^{+}_{l,\{c,k\}} \odot \mathbf{y}^{+}_{l,\{c1,k1\}}\|_1$$
$$\mathrm{Dis}(\mathbf{y}_{l,\{c,k\}},\mathbf{y}_{l,\{c1,k1\}}) = \|\mathbf{y}^{+}_{l,\{c,k\}} \odot \mathbf{y}^{-}_{l,\{c1,k1\}}\|_1 + \|\mathbf{y}^{-}_{l,\{c,k\}} \odot \mathbf{y}^{+}_{l,\{c1,k1\}}\|_1$$



Problem Formulation

$$\min_{\mathbf{Y},\boldsymbol{\theta},\mathbf{A}} \sum_{l=1}^{L} \Bigg(\overbrace{\frac{1}{2}\|\mathbf{A}_l\mathbf{X} - \mathbf{Y}_l\|_F^2}^{\text{Nonlinear Transform Error}}$$

$$+ \sum_{c=1}^{C}\sum_{k=1}^{K} \Big(\overbrace{\lambda_{l,I} l_I(\boldsymbol{\theta}_l, \mathbf{y}_{l,\{c,k\}})}^{\text{Discrimination Constraint}} + \overbrace{\lambda_{l,1}\|\mathbf{y}_{l,\{c,k\}}\|_1}^{\text{Sparsity Constraint}} \Big)$$

$$+ \overbrace{\frac{1}{L}\mathrm{Tr}\{(\mathbf{A}_l\mathbf{X}-\mathbf{Y}_l)^T \sum_{l1\in\{1,...,L\}\setminus l} (\mathbf{A}_{l1}\mathbf{X}-\mathbf{Y}_{l1})\}}^{\text{Target Specific Collaboration Error}} + \overbrace{\Omega(\mathbf{A}_l)}^{\text{Linear Map Constraint}} \Bigg)$$

- $\mathbf{Y} = [\mathbf{Y}_1, ..., \mathbf{Y}_L]$, $\mathbf{A} = [\mathbf{A}_1, ..., \mathbf{A}_L]$, $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_L\}$

Learning Algorithm

We propose an iterative, alternating algorithm with three distinct stages:

- representation $\mathbf{y}_{l,\{c,k\}}$ estimation with discriminative assignment
- discrimination parameters' $\boldsymbol{\theta}$ estimation
- linear map $\mathbf{A}_l$ estimation

We show that the problems at all stages have an **exact** or **approximate closed-form solutions**.

Learning Algorithm

Stage 1: Representation Estimation with Discriminative Assignment

- Given data samples $\mathbf{X}$ and current estimate $\mathbf{A}_l$
- Discriminative representation estimation problem per $\mathbf{Y}_l$ is decoupled and is formulated as:

$$\min_{\mathbf{Y}_l} \|\mathbf{A}_l\mathbf{X} - \mathbf{Y}_l\|_F^2 + \frac{1}{L}\mathrm{Tr}\{\mathbf{Y}_l^T \sum_{l1 \neq l}(\mathbf{Y}_{l1} - \mathbf{A}_{l1}\mathbf{X}))\}$$
$$+ \sum_{c=1}^{C}\sum_{k=1}^{K}\left(\lambda_{l,I} l_I(\boldsymbol{\theta}_l, \mathbf{y}_{l,\{c,k\}}) + \lambda_{l,1}\|\mathbf{y}_{l,\{c,k\}}\|_1\right)$$

Learning Algorithm

Stage 1: Representation Estimation with Discriminative Assignment

$$\min_{\mathbf{Y}_l} \|\mathbf{A}_l\mathbf{X} - \mathbf{Y}_l\|_F^2 + \frac{1}{L}\mathrm{Tr}\{\mathbf{Y}_l^T \sum_{l1\neq l}(\mathbf{Y}_{l1} - \mathbf{A}_{l1}\mathbf{X}))\} + \sum_c \sum_k \left(\lambda_{l,I} l_I(\boldsymbol{\theta}_l, \mathbf{y}_{l,\{c,k\}}) + \lambda_{l,1}\|\mathbf{y}_{l,\{c,k\}}\|_1\right)$$

- Nonlinear Transform Estimation closed-form:

$$\mathbf{y}|_{\{c1,c2\}} = \mathrm{sign}(\mathbf{b}) \odot \max\left(|\mathbf{b}| - \mathbf{p}, \mathbf{0}\right) \oslash \mathbf{n},$$

- Discriminative Assignment:
  - **Part 1:** Score Evaluation

$$l_I : s_I(c1,c2) = \mathrm{sim}(\mathbf{y}|_{\{c1,c2\}}, \boldsymbol{\tau}_{l,c1}) - \mathrm{sim}(\mathbf{y}|_{\{c1,c2\}}, \boldsymbol{\nu}_{l,c2}) + \mathrm{stg}(\mathbf{y}|_{\{c1,c2\}}, \boldsymbol{\tau}_{l,c1})$$

  - **Part 2:** Class Assignment

$$\{\widehat{c1}, \widehat{c2}\} = \arg\min_{c1,c2} s_I(c1,c2), \qquad \mathbf{y}_{l,\{c,k\}} = \mathbf{y}|_{\{\widehat{c1},\widehat{c2}\}}$$

Learning Algorithm

Stage 2: Discrimination Parameters Estimation

- Given the estimated representations $\mathbf{y}_{l,\{c,k\}}$, we update the parameters $\boldsymbol{\theta}_l, \forall l \in \{1, .., L\}$.
- Note that for each $\mathbf{y}_{l,\{c,k\}}$ the corresponding $\boldsymbol{\tau}_{l,c1}$ and $\boldsymbol{\nu}_{l,c2}$ are known from the previous stage.
- We formulate the problem associated to the update of single $\boldsymbol{\tau}_{l,c1}$ as follows:

$$\boldsymbol{\tau}_{l,c1} = \arg\min_{\boldsymbol{\tau}_{l,c1}} \frac{1}{2} \|\boldsymbol{\tau}_{l,c1}^{t-1} - \boldsymbol{\tau}_{l,c1}\|_2^2 + \lambda_{l,0} \sum_{c1} (\mathrm{Stg}(\mathbf{y}|_{\{c1,c2\}}, \boldsymbol{\tau}_{l,c1}) + \mathrm{Sim}(\mathbf{y}|_{\{c1,c2\}}, \boldsymbol{\tau}_{l,c1})).$$

- Analogous formulation for updating per single $\boldsymbol{\nu}_{l,c2}$

Learning Algorithm

Stage 3: Linear Map Estimation

- Given: data samples $\mathbf{X}$, all $\mathbf{Y} = [\mathbf{Y}_1, ..., \mathbf{Y}_L]$, and all $\mathbf{A}$ except $\mathbf{A}_l$
- Denote: $\mathbf{W}_l = \mathbf{Y}_l - \sum_{l1 \in \{1,...,L\} \setminus l} (\mathbf{A}_{l1}\mathbf{X} - \mathbf{Y}_{l1})$
- The problem related to the estimation of the linear map $\mathbf{A}_l$, reduces to:

$$\min_{\mathbf{A}_l} \frac{1}{2}\|\mathbf{A}_l\mathbf{X} - \mathbf{W}_l\|_2^2 + \frac{\lambda_{l,2}}{2}\|\mathbf{A}_l\|_F^2 + \frac{\lambda_{l,3}}{2}\|\mathbf{A}_l\mathbf{A}_l^T - \mathbf{I}\|_F^2 - \lambda_{l,4}\log|\det \mathbf{A}_l^T\mathbf{A}_l|$$

- We use an approximate closed-form solution

Quantifying a Discrimination Quality

- Transform parameter set: $\mathcal{P}_t = \{\mathbf{A} = [\mathbf{A}_1, ..., \mathbf{A}_L]^T \in \Re^{M \times N}, \tau\mathbf{1} \in \Re^M\}$
- **Expected similarity** of all $\mathbf{u}_{c,k} = [\mathbf{y}_{1,\{c,k\}}^T, ..., \mathbf{y}_{L,\{c,k\}}^T]^T$ across all the transform representations $\mathbf{Y}_c$ that come from the **different classes** $c1 \neq c$:

$$D_{\ell_1}^{\mathcal{P}_t}(\mathbf{X}) = \sum_{c=1}^{C} \sum_{c1 \neq c} \sum_{k=1}^{K} \sum_{k1 \neq k} (\|\mathbf{u}_{c,k}^+ \odot \mathbf{u}_{c1,k1}^+\|_1 + \|\mathbf{u}_{c,k}^- \odot \mathbf{u}_{c1,k1}^-\|_1)$$

- **Expected similarity** using the positive and negative components of all $\mathbf{u}_{c,k} = [\mathbf{y}_{1,\{c,k\}}^T, ..., \mathbf{y}_{L,\{c,k\}}^T]^T$ across all the transform representations $\mathbf{Y}_c$ that come from the **same classes** c:

$$D_{\ell_1,c}^{\mathcal{P}_t}(\mathbf{X}) = \sum_{c=1}^{C} \sum_{k=1}^{K} \sum_{k1 \neq k} (\|\mathbf{u}_{c,k}^+ \odot \mathbf{u}_{c,k1}^+\|_1 + \|\mathbf{u}_{c,k}^- \odot \mathbf{u}_{c,k1}^-\|_1)$$

Quantifying a Discrimination Quality

- Transform parameter set: $\mathcal{P}_t = \{\mathbf{A} = [\mathbf{A}_1, ..., \mathbf{A}_L]^T \in \Re^{M \times N}, \tau \mathbf{1} \in \Re^M\}$
- **Expected similarity** of all $\mathbf{u}_{c,k} = [\mathbf{y}^T_{1,\{c,k\}}, ..., \mathbf{y}^T_{L,\{c,k\}}]^T$ across all the transform representations $\mathbf{Y}_c$ that come from the **different classes** $c1 \neq c$:

$$D^{\mathcal{P}_t}_{\ell_1}(\mathbf{X}) = \sum_{c=1}^{C} \sum_{c1 \neq c} \sum_{k=1}^{K} \sum_{k1 \neq k} (\|\mathbf{u}^+_{c,k} \odot \mathbf{u}^+_{c1,k1}\|_1 + \|\mathbf{u}^-_{c,k} \odot \mathbf{u}^-_{c1,k1}\|_1)$$

- **Expected similarity** using the positive and negative components of all $\mathbf{u}_{c,k} = [\mathbf{y}^T_{1,\{c,k\}}, ..., \mathbf{y}^T_{L,\{c,k\}}]^T$ across all the transform representations $\mathbf{Y}_c$ that come from the **same classes** c:

$$D^{\mathcal{P}_t}_{\ell_1,c}(\mathbf{X}) = \sum_{c=1}^{C} \sum_{k=1}^{K} \sum_{k1 \neq k} (\|\mathbf{u}^+_{c,k} \odot \mathbf{u}^+_{c,k1}\|_1 + \|\mathbf{u}^-_{c,k} \odot \mathbf{u}^-_{c,k1}\|_1)$$

- **Discrimination Power** for any pair of labels and dataset $\mathbf{X} \in \Re^{M \times CK}$:

$$\mathcal{I}^t = \log(D^{\mathcal{P}_t}_{\ell_1,c}(\mathbf{X})) - \log(D^{\mathcal{P}_t}_{\ell_1}(\mathbf{X}) + \epsilon)$$

Quantifying a Discrimination Quality

- Transform parameter set: $\mathcal{P}_t = \{\mathbf{A} = [\mathbf{A}_1, ..., \mathbf{A}_L]^T \in \Re^{M \times N}, \tau\mathbf{1} \in \Re^M\}$
- **Expected similarity** of all $\mathbf{u}_{c,k} = [\mathbf{y}_{1,\{c,k\}}^T, ..., \mathbf{y}_{L,\{c,k\}}^T]^T$ across all the transform representations $\mathbf{Y}_c$ that come from the **different classes** $c1 \neq c$:

$$D_{\ell_1}^{\mathcal{P}_t}(\mathbf{X}) = \sum_{c=1}^{C} \sum_{c1 \neq c} \sum_{k=1}^{K} \sum_{k1 \neq k} (\|\mathbf{u}_{c,k}^+ \odot \mathbf{u}_{c1,k1}^+\|_1 + \|\mathbf{u}_{c,k}^- \odot \mathbf{u}_{c1,k1}^-\|_1)$$

- **Expected similarity** using the positive and negative components of all $\mathbf{u}_{c,k} = [\mathbf{y}_{1,\{c,k\}}^T, ..., \mathbf{y}_{L,\{c,k\}}^T]^T$ across all the transform representations $\mathbf{Y}_c$ that come from the **same classes** c:

$$D_{\ell_1,c}^{\mathcal{P}_t}(\mathbf{X}) = \sum_{c=1}^{C} \sum_{k=1}^{K} \sum_{k1 \neq k} (\|\mathbf{u}_{c,k}^+ \odot \mathbf{u}_{c,k1}^+\|_1 + \|\mathbf{u}_{c,k}^- \odot \mathbf{u}_{c,k1}^-\|_1)$$

- **Discrimination Power** for any pair of labels and dataset $\mathbf{X} \in \Re^{M \times CK}$:

$$\mathcal{I}^t = \log(D_{\ell_1,c}^{\mathcal{P}_t}(\mathbf{X})) - \log(D_{\ell_1}^{\mathcal{P}_t}(\mathbf{X}) + \epsilon)$$

Mutual Coherence & Condition Number

	AR	YALE B	COIL20	NORB
$\frac{1}{L}\sum_l \mu(\mathbf{A}_l)$	2.1e-4	1e-4	1.9e-4	3.1e-4
$\frac{1}{L}\sum_l C_n(\mathbf{A}_l)$	16.1	26.3	18	19.1

Table: The **cumulative expected mutual coherence** $\frac{1}{L}\sum_l \mu(\mathbf{A}_l)$ and the **cumulative conditioning number** $\frac{1}{L}\sum_l C_n(\mathbf{A}_l)$ for the linear maps $\mathbf{A}_l, l \in \{1, ..., 6\}$ with dimensions $6570 \times N$, where N is the dimensionality of the input data

Discrimination Power Evaluation

	AR	YALE B	COIL20	NORB
$\mathcal{I}^o$	2.13	1.45	1.18	0.41
$\mathcal{I}^R$	2.41	1.66	1.61	0.40
$\mathcal{I}^S$	2.71	1.76	1.92	0.40
$\mathcal{I}^*$	**3.04**	**2.14**	**2.63**	**0.42**

Table: The discrimination power in the **original domain**, after **random transform**, after **learned sparsifying transform** and after **learned self-collaborating target specific nonlinear transform** with dimension $M = 6570$.

Recognition Evaluation

	AR	YALE B	COIL20	NORB
original domain [%]	96.1	95.4	96.8	97
proposed [%]	97.1	97.1	97.8	96.8

Table: The recognition results on the databases AR, YALE B, COIL20 and NORB, using k-NN on the sparse representations using our model with dimension $M = 6570$.

Discrimination Power and Recognition Comparison

a)

	YALE B $\mathcal{I}$	MNIST $\mathcal{I}$
$dlsi$	0.71	0.67
$fddl$	0.87	0.63
$copar$	0.57	0.54
$lrsdl$	0.42	0.40
$*$	**0.90**	**0.81**
$*$	**0.90**	**0.81**

b), c)

YALE B	Acc. [%]	MNIST	Acc. [%]
	96.5		98.74
	97.5		96.31
	98.3		96.41
	98.7		–
k-nn	97.1	k-nn	97.32
l-svm	98.8	l-svm	98.45

Table: a) The discrimination power for the methods $dlsi$, $fddl$, $copar$ and $lrsdl$ and the proposed method $*$, b), c) The recognition results on the Extended Yale B and MNIST

Recognition Accuracy Comparison with State-of-the-Art

MNIST

Method	Acc.
lif-cnn [1]	98.37
s-cw-a [2]	98.62
reg-l [3]	99.08
f-max [4]	99.65
$*$ k-nn	97.11
$*$ l-svm	99.10

F-MNIST

Method	Acc.
log-reg [5]	84.00
rf-c [5]	87.70
svc [5]	89.98
cnn [6]	92.10
$*$ k-nn	88.10
$*$ l-svm	92.22

SVHN

Method	Acc.
ssae [7]	89.70
c-km [7]	90.60
s-cw-a [2]	93.10
tma [8]	98.31
$*$ k-nn	86.41
$*$ l-svm	90.28

Table: Recognition accuracy comparison between sota and 1) k Nearest Neighbor (k-nn) search and 2) linear SVM (l-svm) that use the Sparsifying Nonlinear Transform (sNT) representations from our model on extracted HOG image features. We use our algorithm to learn the model on the HOG features. Then we get the sNT representations with dimensionality 9800 for the respective training and test sets. Considering the obtained result for database SVHN, we note that the unlabeled training data from the respective database was not used during the learning of the corresponding model.

Conclusions:

- We introduced a novel collaboration structured model with minimum information loss, collaboration corrective and discriminative priors for joint learning of multiple nonlinear transforms.
- An efficient solution was proposed by an iterative, coordinate descend algorithm.
- The introduced discrimination measure and the recognition accuracy on the used databases showed promising performance and advantages w.r.t. state-of-the-art methods.


Thank You!

References

[1] Eric Hunsberger and Chris Eliasmith, "Spiking deep networks with LIF neurons", 2015.

[2] Alireza Makhzani and Brendan J Frey. "Winner-take-all autoencoders", In NIPS. 2015.

[3] Priyadarshini Panda and Kaushik Roy. "Unsupervised regenerative learning of hierarchical features in spiking deep networks for object recognition", CoRR, 2016.

[4] Benjamin Graham. "Fractional max-pooling", CoRR, 2014.

[5] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion- mnist: a novel image dataset for benchmarking machine learning algorithms. CoRR, 2017.

[6] Maheshkumar H. Kolekar Shobhit Bhatnagar, Deepan- way Ghosal. Classification of fashion article images using convolutional neural networks. In IEEE ICIIP, 2017.

[7] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y. Ng. Reading digits in natural images with unsupervised feature learning. In NIPS Wor. UFL, 2011.

[8] Chen-Yu Lee, Patrick W. Gallagher, and Zhuowen Tu. Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree. In 19th ICAIS, 2016.