# General Multi-view Neural Network

## Current Research

Jiyuan

liujiyuan13@nudt.edu.cn

March 27, 2021

# Current research

# Latent Multi-view Subspace Clustering [1]

Overview:

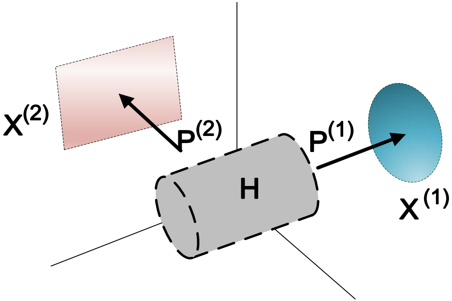


Figure 1: Overview of LMSC. Observations $\{\mathbf{X}_v\}_{v=1}^{V}$ ($V \geq 2$) corresponding to different views are partially projected by $\{\mathbf{P}_v\}_{v=1}^{V}$ from one underlying latent representation H.

[1] Zhang et. al. Latent Multi-view Subspace Clustering. IEEE CVPR, 2017.

# Latent Multi-view Subspace Clustering

Formulation:

$$\begin{aligned} \min_{\mathbf{P},\mathbf{H},\mathbf{Z},\mathbf{E}_h,\mathbf{E}_r} & \|\mathbf{E}\|_{2,1} + \lambda\|\mathbf{Z}\|_* \\ s.t.\ & \mathbf{X} = \mathbf{P}\mathbf{H} + \mathbf{E}_h,\ \mathbf{H} = \mathbf{H}\mathbf{Z} + \mathbf{E}_r, \\ & \mathbf{E} = [\mathbf{E}_h; \mathbf{E}_r], \mathbf{P}\mathbf{P}^\top = \mathbf{I}, \end{aligned} \tag{1}$$

in which

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \cdots \\ \mathbf{X}_V \end{bmatrix},\ \mathbf{P} = \begin{bmatrix} \mathbf{P}_1 \\ \cdots \\ \mathbf{P}_V \end{bmatrix}. \tag{2}$$

# Generalized Latent Multi-view Subspace Clustering [2]

Overview:

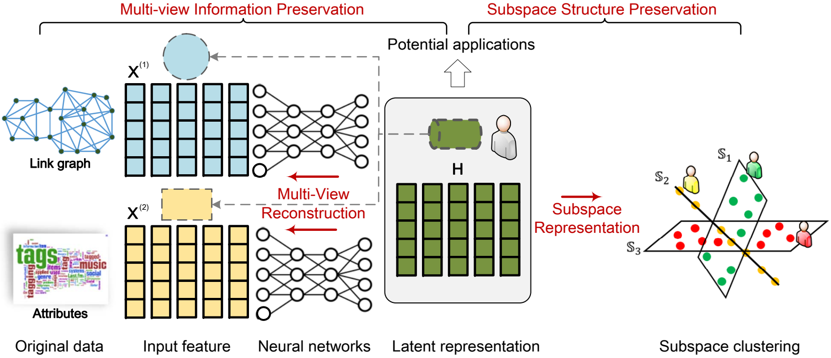


Figure 2: Overview of gLMSC. The latent representation non-linearly encodes the information from multiple views with neural networks for uncovering the data distribution in subspaces.

[2] Zhang et. al. Generalized Latent Multi-view Subspace Clustering. IEEE TPAMI, 2019.

# Generalized Latent Multi-view Subspace Clustering

Formulation:

$$\begin{aligned} \min_{\{\boldsymbol{\theta_v}\}_{v=1}^{V}, \mathbf{H}, \mathbf{Z}} \ & \ell(\mathbf{H}, \mathbf{HZ}) + \sum_{v=1}^{V} \alpha_v d_v(\mathbf{X}_v, g_{\boldsymbol{\theta}_v}(\mathbf{H})) + \lambda \Omega(\mathbf{Z}) \\ s.t. \ & g_{\boldsymbol{\theta}_v}(\mathbf{H}) = \mathbf{W}_{(k,v)} f(\mathbf{W}_{(k-1,v)} \cdots f(\mathbf{W}_{(1,v)} \mathbf{H})). \end{aligned} \tag{3}$$

Instance:

$$\begin{aligned} & \ell(\mathbf{H}, \mathbf{HZ}) = \frac{1}{2} \|\mathbf{H} - \mathbf{HZ}\|_F^2 \\ & d_v(\mathbf{X}_v, g_{\boldsymbol{\theta}_v}(\mathbf{H})) = \frac{1}{2} \|\mathbf{X}_v - \mathbf{W}_{(2,v)} f(\mathbf{W}_{(1,v)} \mathbf{H})\|_F^2 \\ & \Omega(\mathbf{Z}) = \|\mathbf{Z}\|_*. \end{aligned} \tag{4}$$

# CPM-Nets: Cross Partial Multi-View Networks [3]

Overview:

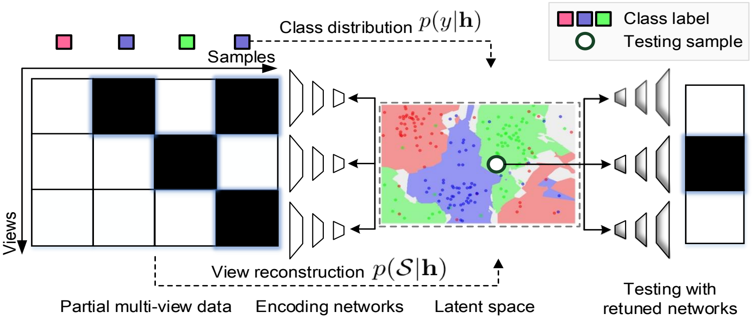


Figure 3: Overview of CPMnet. Given multi-view data with missing views (black blocks), the encoding networks degrade the complete latent representation into the available views (white blocks).

[3] Zhang et. al. CPM-Nets: Cross Partial Multi-View Networks. NeurIPS, 2020.

# CPM-Nets: Cross Partial Multi-View Networks

Formulation:

$$\min_{\{\mathbf{h}_n\}_{n=1}^{N},\boldsymbol{\Theta}_r} \frac{1}{N}\sum_{n=1}^{N} \ell_r(\mathcal{S}_n, \mathbf{h}_n; \boldsymbol{\Theta}_r) + \lambda \ell_c(y_n, y, \mathbf{h}_n) \tag{5}$$

in which

$$\begin{aligned}
&\ell_r(\mathcal{S}_n, \mathbf{h}_n; \boldsymbol{\Theta}_r) = \sum_{v=1}^{V} s_{nv} \| f_v(\mathbf{h}_n; \boldsymbol{\Theta}_r^{(v)}) - \mathbf{x}_n^{(v)} \|^2 \\
&\ell_c(y_n, y, \mathbf{h}_n) = \max_{y\in\mathcal{Y}} \Big(0, \Delta(y_n, y) + \mathbb{E}_{\mathbf{h}\sim\mathcal{T}(y)} F(\mathbf{h}, \mathbf{h}_n) - \mathbb{E}_{\mathbf{h}\sim\mathcal{T}(y_n)} F(\mathbf{h}, \mathbf{h}_n)\Big) \\
&\Delta(y_n, y) = \Delta(y_n, g(\mathbf{h}_n; \boldsymbol{\Theta}_c)) \\
&g(\mathbf{h}_n; \boldsymbol{\Theta}_c) = \arg\max_{y\in\mathcal{Y}} \mathbb{E}_{\mathbf{h}\sim\mathcal{T}(y)} F(\mathbf{h}, \mathbf{h}_n) \\
&F(\mathbf{h}, \mathbf{h}_n) = \phi(\mathbf{h}; \boldsymbol{\Theta}_c)^\top \phi(\mathbf{h}; \boldsymbol{\Theta}_c)
\end{aligned} \tag{6}$$

# CPM-Nets: Cross Partial Multi-View Networks

**Algorithm 1:** CPM-Nets: Cross Partial Multi-View Networks

\# Training

**Input:** Partial multi-view dataset: $\mathcal{D} = \{\mathcal{S}_n, y_n\}_{n=1}^N$, parameter $\lambda$.

Initialize $\{\mathbf{h}_n\}_{n=1}^N$ and $\{\mathbf{\Theta}_r^{(v)}\}_{v=1}^V$.

**while** not convergence **do**

  **for** $v = 1 : V$ **do**

    Update the network parameters $\mathbf{\Theta}_r^{(v)}$ with $\ell_r(\mathcal{S}_n, \mathbf{h}_n; \mathbf{\Theta}_r)$;

  **end**

  **for** $n = 1 : N$ **do**

    Update the representation $\mathbf{h}_n$ with $\ell_r(\mathcal{S}_n, \mathbf{h}_n; \mathbf{\Theta}_r) + \lambda \ell_c(y_n, y, \mathbf{h}_n)$;

  **end**

**end**

**Output:** network parameters $\{\mathbf{\Theta}_r^{(v)}\}_{v=1}^V$ and latent representations $\{\mathbf{h}_n\}_{n=1}^N$

\# Testing

Train the retuned networks $(\{\mathbf{\Theta}_r^{(v)}\}_{v=1}^V)$ for test;

Calculate the latent representation for test instance;

Classify the test instance with $y = \arg\max_{y \in \mathcal{Y}} \mathbb{E}_{\mathbf{h} \sim \mathcal{T}(y)} F(\mathbf{h}, \mathbf{h}_{test})$.

# Deep Partial Multi-view Learning [4]

Overview:

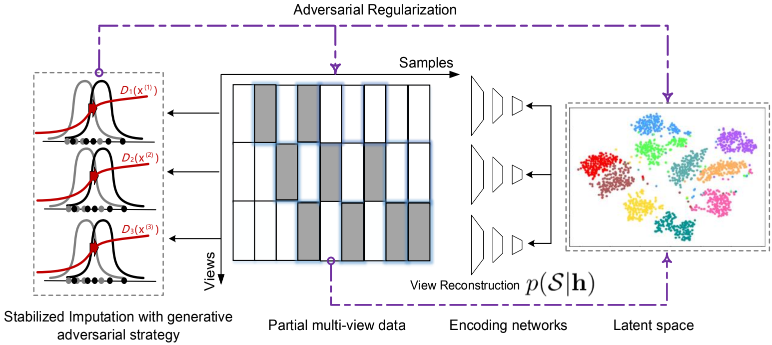


Figure 4: Overview of CPM-GAN. The latent representation learning and missing data imputation equipped with adversarial strategy are jointly conducted to improve each other.

[4] Zhang et. al. Deep Partial Multi-view Learning. IEEE TPAMI, 2021.

# Deep Partial Multi-view Learning

Formulation:

$$\mathcal{L} = \min_{G} \max_{D} \min_{\mathbf{h}} (\mathcal{L}_{adv} + \mathcal{L}_r) \tag{7}$$

in which

$$\begin{aligned} \mathcal{L}_{adv} &= \sum_{n=1}^{N} \sum_{v=1}^{V} \sum_{i=1}^{I} (1 - s_{nv}) [\log D_v(\mathbf{x}_i^{(v)}; \mathbf{\Theta}_d^{(v)}) \\ &\quad + \log(1 - D_v(G_v(\mathbf{h}_n; \mathbf{\Theta}_g^{(v)}); \mathbf{\Theta}_d^{(v)}))]; \\ \mathcal{L}_r &= \sum_{n=1}^{N} \sum_{v=1}^{V} s_{nv} \| G_v(\mathbf{h}_n; \mathbf{\Theta}_g^{(v)}) - \mathbf{x}_n^{(v)} \|^2. \end{aligned} \tag{8}$$

# Deep Partial Multi-view Learning

**Algorithm 2:** Deep Partial Multi-view Learning - CPM-GAN

**Input:** Partial multi-view dataset: $\mathcal{D} = \{\mathcal{S}_n\}_{n=1}^N$.

Initialize $\{\mathbf{h}_n\}_{n=1}^N$, $\{\mathbf{\Theta}_g^{(v)}\}_{v=1}^V$ and $\{\mathbf{\Theta}_d^{(v)}\}_{v=1}^V$.

**while** not convergence **do**

  **for** $v = 1 : V$ **do**

    Update the discriminator parameters $\mathbf{\Theta}_d^{(v)}$ with gradient descent;

  **end**

  **for** $v = 1 : V$ **do**

    Update the generator parameters $\mathbf{\Theta}_g^{(v)}$ with gradient descent;

  **end**

  **for** $n = 1 : N$ **do**

    Update the representation $\mathbf{h}_n$ with gradient descent;

  **end**

**end**

**Output:** $\{\mathbf{h}_n\}_{n=1}^N$, $\{\mathbf{\Theta}_d^{(v)}\}_{v=1}^V$ and $\{\mathbf{\Theta}_g^{(v)}\}_{v=1}^V$

# CVOID-19 Detection with Multi-view Learning [5]

Overview:

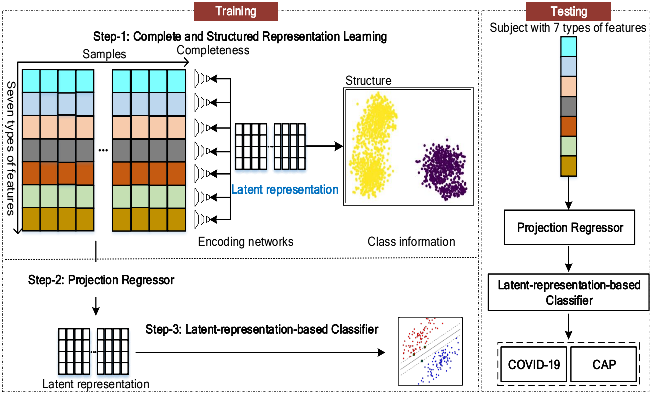


Figure 5: Overview of the latent-representation-based diagnosis framework.

[5] Kang et. al. Diagnosis of Coronavirus Disease 2019 (COVID-19) with Structured Latent Multi-View Representation Learning. IEEE TMI, 2021.

# Thought

# Advantages and Shortcomings

Advantages:

1. Native framework to integrate multi-view information;
2. Compatible with incomplete data.

Shortcomings:

1. Feeding the training data all at once (memory and time);
2. Requiring the training data in testing (storage and time);
3. Both illustrates they are not capable of large-scale problems.

# General Multi-view Neural Network

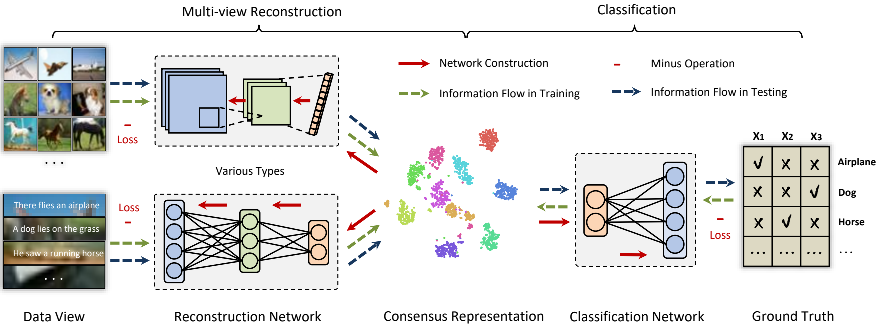


Figure 6: Overview of the proposed General Multi-view Neural Network.

# General Multi-view Neural Network

Formulation:

$$\min_{\mathbf{h},\mathbf{\Theta}_c,\mathbf{\Theta}_r} \ell_c(\mathbf{y},\mathbf{h};\mathbf{\Theta}_c) + \lambda \sum_{v=1}^{V} \ell_r(\mathbf{X}_v,\mathbf{h};\mathbf{\Theta}_r) \tag{9}$$

in which

$$\begin{aligned} \ell_c(\mathbf{y},\mathbf{h};\mathbf{\Theta}_c) &= \Delta(\mathbf{y}, g(\mathbf{h};\mathbf{\Theta}_c)) \\ \ell_r(\mathbf{X}_v,\mathbf{h};\mathbf{\Theta}_r) &= \sum_{v=1}^{V} \| f_v(\mathbf{h};\mathbf{\Theta}_r^{(v)}) - \mathbf{X}_v \|^2 \end{aligned} \tag{10}$$

# Thanks !