The proposed method

TABLE I: Basic notations for the proposed method.

Notations	Meaning
$\mathbf{X}^{(v)} \in \mathbb{R}^{d_v \times n}$	Feature matrix of the i-th view
$\mathbf{Z}_1^{(v)} \in \mathbb{R}^{d_v \times l_i}$	1-th layer cluster centroid matrix of the i-th view
$\mathbf{Z}_i^{(v)} \in \mathbb{R}^{l_{i-1} \times l_i}$	i-th layer cluster centroid matrix of the i-th view
$\mathbf{H}_i^{(v)} \in \mathbb{R}^{l_i \times n}$	i-th layer feature representation of the i-th view
$\mathbf{S}^{(v)} \in \mathbb{R}^{n \times n}$	the similarity matrix of the v-th view
$\mathbf{S} \in \mathbb{R}^{n \times n}$	Consensus similarity matrix
$\phi \in \mathbb{R}^{d_v \times l_{i-1}}$	$\mathbf{Z}_1^{(v)}\mathbf{Z}_2^{(v)} \dots \mathbf{Z}_{i-1}^{(v)}$
$\Phi \in \mathbb{R}^{d_v \times l_i}$	$\mathbf{Z}_1^{(v)}\mathbf{Z}_2^{(v)} \dots \mathbf{Z}_i^{(v)}$
$\hat{\mathbf{H}}_i^{(v)} \in \mathbb{R}^{l_i \times n}$	$\mathbf{Z}_{i+1}^{(v)} \dots \mathbf{Z}_m^{(v)}\mathbf{H}_m^{(v)}$
$\mathbf{G} \in \mathbb{R}^{n \times n}$	$\sum_{o=1,o\neq v}^{V} \alpha^{(o)}\mathbf{H_m}^{(o)\mathrm{T}}\mathbf{H_m}^{(o)}$
$\mathbf{Q} \in \mathbb{R}^{n \times n}$	$\sum_{v=1}^{V} \alpha^{(v)}\mathbf{H}_m^{(v)\mathrm{T}}\mathbf{H}_m^{(v)}$

I. OBJECT FUNCTION

We introduce some basic notations of our method firstly as described in Table I.

We assume that the feature representations of the last layer in different views are different and a consensus structure matrix $\mathbf{S}$ should be fused with individual structures. The idea can be mathematically expressed as follows,

$$\begin{aligned} \min_{\substack{\mathbf{Z}_i^{(v)},\mathbf{H}_i^{(v)}\\ \alpha^{(v)},\mathbf{S}}} & \sum_{v=1}^{V} \|\mathbf{X}^{(v)} - \mathbf{Z}_1^{(v)}\mathbf{Z}_2^{(v)} \dots \mathbf{Z}_m^{(v)}\mathbf{H}_m^{(v)}\|_F^2 \\ & + \beta\|\mathbf{S} - \sum_{v=1}^{V} \alpha^{(v)}\mathbf{H}_m^{(v)\mathrm{T}}\mathbf{H}_m^{(v)}\|_F^2, \\ \text{s.t.}\mathbf{H}_i^{(v)} \geq 0, & \sum_{v=1}^{V} \alpha^{(v)}=1, \alpha^{(v)} \geq 0, \mathbf{S1}=\mathbf{1}, \mathbf{S}\geq 0, \mathrm{diag}(\mathbf{S})=0. \end{aligned} \tag{1}$$

$\mathbf{H}_m^{(v)}$ denotes the m-th layer of the v-th view. $\mathbf{H}_m^{(v)\,\mathrm{T}}\mathbf{H}_m^{(v)}$ constructs the similarity matrix $\mathbf{S}^{(v)}$ in different layer. $\alpha^{(v)}$ is the weight coefficient of the v-th view for $\mathbf{S}^{(v)}$. $\mathbf{S}$ denotes the consensus similarity matrix. $\mathbf{S}_{i,j}$ denotes the similarity score between i-th and j-th sample so we need to add the constraints $\mathbf{S} \geq 0$ and $\mathrm{diag}(\mathbf{S}) = 0$ for $\mathbf{S}$. The larger value $\mathbf{S}_{i,j}$ is, the more likely two samples belong to the same cluster. We hope to obtain normalized solution, so we add the constraint $\mathbf{S1} = \mathbf{1}$.

II. INITIALIZATION

Inspired by the tricks of the initialization in [?], we have pre-trained all of the layers to initialize the variables $\mathbf{Z}_i^{(v)}$ and $\mathbf{H}_i^{(v)}$ by decomposing layer by layer. Firstly, we decompose the feature matrix of the v-th view $\mathbf{X}^{(v)} \approx \mathbf{Z}_1^{(v)}\mathbf{H}_1^{(v)}$, where $\mathbf{Z}_1^{(v)} \in \mathbb{R}^{d_v \times l_1}$ and $\mathbf{H}_1^{(v)} \in \mathbb{R}^{l_1 \times n}$. Following this, we decompose the new feature matrix $\mathbf{H}_1^{(v)} \approx \mathbf{Z}_2^{(v)}\mathbf{H}_2^{(v)}$, where $\mathbf{Z}_2^{(v)} \in \mathbb{R}^{l_1 \times l_2}$ and $\mathbf{H}_2^{(v)} \in \mathbb{R}^{l_2 \times n}$. We repeat the above steps until all layers have been pre-trained. We pre-train each of the layers to have an initial approximation of the matrices $\mathbf{H}_i^{(v)}$ and $\mathbf{Z}_i^{(v)}$ which can greatly reduce the time for follow-up work. Then we use the value of $\mathbf{Z}_i^{(v)}$ and $\mathbf{H}_i^{(v)}$ to initialize $\mathbf{S}$ and $\alpha^{(v)}$ by setting $\alpha^{(v)} = \frac{1}{V}$ and $\mathbf{S} = \sum_{v=1}^{V} \alpha^{(v)}\mathbf{H}_m^{(v)\mathrm{T}}\mathbf{H}_m^{(v)}$. At the beginning, we argue that each view has the same contribution, so we initialize $\mathbf{S}$ by the construction of $\mathbf{H}_m^{(v)}$ with the same weight.

III. OPTIMIZATION

Because the objective function Eq. (1) is a non-convex problem, it seems unlikely to solve this problem in one step. So we propose a five-step alternate optimization method to address this problem. To reduce the total reconstruction error of the model, we also need to alternately minimize $\mathbf{Z_i}^{(v)}$ and $\mathbf{H_i}^{(v)}$ in each layer.

A. Update rule for matrix $\mathbf{Z}_i^{(v)}$

By fixing $\mathbf{H}_i^{(v)}$, $\mathbf{S}$, $\alpha^{(v)}$ and $\mathbf{Z}_i^{(o)}(o \neq v)$, we can update $\mathbf{Z}_i^{(v)}$ by solving the following problem without constraint,

$$\min_{\mathbf{Z}_i^{(v)}} \|\mathbf{X}^{(v)} - \phi\mathbf{Z}_i^{(v)}\hat{\mathbf{H}}_i^{(v)}\|_F^2, \tag{2}$$

where $\phi = \mathbf{Z}_1^{(v)}\mathbf{Z}_2^{(v)} \dots \mathbf{Z}_{i-1}^{(v)}$, by setting $\partial\mathcal{C}/\partial\mathbf{Z}_i^{(v)} = 0$, we can give the solutions as,

$$\mathbf{Z}_i^{(v)} = \phi^{\dagger}\mathbf{X}^{(v)}\hat{\mathbf{H}}_i^{(v)\dagger}, \tag{3}$$

where $\phi^{\dagger} = (\phi^{\mathrm{T}}\phi)^{-1}\phi^{\mathrm{T}}$ and $\hat{\mathbf{H}}_i^{(v)\dagger} = \hat{\mathbf{H}}_i^{(v)\mathrm{T}}(\hat{\mathbf{H}}_i^{(v)}\hat{\mathbf{H}}_i^{(v)\mathrm{T}})^{-1}$. $\hat{\mathbf{H}}_i^{(v)} = \mathbf{Z}_{i+1}^{(v)} \dots \mathbf{Z}_m^{(v)}\mathbf{H}_m^{(v)}$ and $\hat{\mathbf{H}}_i^{(v)}$ denotes the reconstruction of the i-th layer's representation for the v-th view.

B. Update rule for matrix $\mathbf{H}_i^{(v)}(i<m)$

By fixing $\mathbf{Z}_i^{(v)}$, $\mathbf{H}_m^{(v)}$, $\alpha^{(v)}$ and $\mathbf{S}$, we can update $\mathbf{H}_i^{(v)}$ by solving the following problem,

$$\min_{\mathbf{H_i}^{(v)}} \|\mathbf{X}^{(v)} - \Phi\mathbf{H}_i^{(v)}\|_F^2, \text{s.t.}\mathbf{H}_i^{(v)} \geq 0, \tag{4}$$

where $\Phi = \mathbf{Z}_1^{(v)}\mathbf{Z}_2^{(v)} \dots \mathbf{Z}_i^{(v)}$. Following the update rule in [?], the update rule for $\mathbf{H}_i^{(v)}(i<m)$ can be written as,

$$\mathbf{H}_i^{(v)} = \mathbf{H}_i^{(v)} \odot \sqrt{\frac{\left[\Phi^{\mathrm{T}}\mathbf{X}^{(v)}\right]^{+} + \left[\Phi^{\mathrm{T}}\Phi\mathbf{H}_i^{(v)}\right]^{-}}{\left[\Phi^{\mathrm{T}}\mathbf{X}^{(v)}\right]^{-} + \left[\Phi^{\mathrm{T}}\Phi\mathbf{H}_i^{(v)}\right]^{+}}}. \tag{5}$$

We also update $\mathbf{H}_m^{(v)}$ here for faster convergence and easier code writing.

C. Update rule for matrix $\mathbf{H}_m^{(v)}$

By fixing $\mathbf{Z}_i^{(v)}$, $\mathbf{H}_i^{(v)}(i<m)$, $\alpha^{(v)}$ and $\mathbf{S}$, we can update $\mathbf{H}_m^{(v)}$ by solving the following problem,

$$\begin{aligned}&\min_{\mathbf{H_m}^{(v)}} \|\mathbf{X}^{(v)} - \Phi\mathbf{H}_m^{(v)}\|_F^2 + \beta\|\mathbf{S} - \alpha^{(v)}\mathbf{H}_m^{(v)\mathrm{T}}\mathbf{H}_m^{(v)} - \mathbf{G}\|_F^2,\\ &\quad\text{s.t.}\mathbf{H}_m^{(v)} \geq 0,\end{aligned} \tag{6}$$

where the variables are defined as follows,

$$\Phi = \mathbf{Z}_1^{(v)}\mathbf{Z}_2^{(v)}\ldots\mathbf{Z}_m^{(v)}, \mathbf{G} = \sum_{o=1, o\neq v}^{V} \alpha^{(o)}\mathbf{H_m}^{(o)\mathrm{T}}\mathbf{H_m}^{(o)}. \tag{7}$$

We give the updating rule of $\mathbf{H}_m^{(v)}$ firstly, followed by the proof of it.

$$\begin{aligned}
&\mathbf{H}_m^{(v)} = \mathbf{H}_m^{(v)} \odot \sqrt{\vartheta_u(\mathbf{ZHS})/\vartheta_l(\mathbf{ZHS})},\\
&\vartheta_u(\mathbf{ZHS}) = \left[\Phi^\mathrm{T}\mathbf{X}^{(v)}\right]^+ + \left[\Phi^\mathrm{T}\Phi\mathbf{H}_m^{(v)}\right]^- + \alpha^{(v)}\beta(\left[\mathbf{H}_m^{(v)}\mathbf{S}\right]^+\\
&+ \left[\mathbf{H}_m^{(v)}\mathbf{S}^\mathrm{T}\right]^+ + \left[2\mathbf{H}_m^{(v)}\mathbf{G}\right]^- + \left[2\alpha^{(v)}\mathbf{H}_m^{(v)}\mathbf{H}_m^{(v)\mathrm{T}}\mathbf{H}_m^{(v)}\right]^-),\\
&\vartheta_u(\mathbf{ZHS}) = \left[\Phi^\mathrm{T}\mathbf{X}^{(v)}\right]^- + \left[\Phi^\mathrm{T}\Phi\mathbf{H}_m^{(v)}\right]^+ + \alpha^{(v)}\beta(\left[\mathbf{H}_m^{(v)}\mathbf{S}\right]^-\\
&+ \left[\mathbf{H}_m^{(v)}\mathbf{S}^\mathrm{T}\right]^- + \left[2\mathbf{H}_m^{(v)}\mathbf{G}\right]^+ + \left[2\alpha^{(v)}\mathbf{H}_m^{(v)}\mathbf{H}_m^{(v)\mathrm{T}}\mathbf{H}_m^{(v)}\right]^+).
\end{aligned} \tag{8}$$

Theorem 1: The limited solution of the update rule in Eq. (8) satisfies the KKT condition.

Proof 1: We introduce the Lagrangian function as

$$\begin{aligned}\mathbf{L}(\mathbf{H}_m^{(v)}) =& \|\mathbf{X}^{(v)} - \Phi\mathbf{H}_m^{(v)}\|_F^2 + \beta\|\mathbf{S} - \alpha^{(v)}\mathbf{H}_m^{(v)\mathrm{T}}\mathbf{H}_m^{(v)} - \mathbf{G}\|_F^2\\ &- \mathrm{Tr}(\eta\mathbf{H}_m^{(v)}).\end{aligned} \tag{9}$$

In order to satisfy the constraint $\mathbf{H}_m^{(v)} \geq 0$, we introduce the Lagrangian multiplier η. By setting $\partial\mathbf{L}(\mathbf{H}_m^{(v)})/\partial\mathbf{H}_m^{(v)} = 0$, we can obtation:

$$\begin{aligned}&\partial\mathbf{L}(\mathbf{H}_m^{(v)})/\partial\mathbf{H}_m^{(v)} = -2(\Phi^\mathrm{T}\mathbf{X}^{(v)} - \Phi^\mathrm{T}\Phi\mathbf{H}_m^{(v)} + \alpha^{(v)}\beta\mathbf{H}_m^{(v)}\mathbf{S}\\ &+ \alpha^{(v)}\beta\mathbf{H}_m^{(v)}\mathbf{S}^\mathrm{T} - 2\alpha^{(v)}\beta\mathbf{H}_m^{(v)}\mathbf{G} - 2\alpha^{(v)2}\beta\mathbf{H}_m^{(v)}\mathbf{H}_m^{(v)\mathrm{T}}\mathbf{H}_m^{(v)})\\ &- \eta = 0.\end{aligned} \tag{10}$$

From the complementary slackness condition, we can obtain,

$$\begin{aligned}&(-\Phi^\mathrm{T}\mathbf{X}^{(v)} + \Phi^\mathrm{T}\Phi\mathbf{H}_m^{(v)} - \alpha^{(v)}\beta\mathbf{H}_m^{(v)}\mathbf{S} - \alpha^{(v)}\beta\mathbf{H}_m^{(v)}\mathbf{S}^\mathrm{T}+\\ &2\alpha^{(v)}\beta\mathbf{H}_m^{(v)}\mathbf{G} + 2\alpha^{(v)2}\beta\mathbf{H}_m^{(v)}\mathbf{H}_m^{(v)\mathrm{T}}\mathbf{H}_m^{(v)})\mathbf{H}_m^{(v)} = \eta\mathbf{H}_m^{(v)} = 0.\end{aligned} \tag{11}$$

Both the equations require that at least one of the two factors is equal to zero, so Eq. (11) and Eq. (12) have the same meaning. We multiply both sides by $\mathbf{H}_m^{(v)}$ and we can obtain:

$$\begin{aligned}&(-\Phi^\mathrm{T}\mathbf{X}^{(v)} + \Phi^\mathrm{T}\Phi\mathbf{H}_m^{(v)} - \alpha^{(v)}\beta\mathbf{H}_m^{(v)}\mathbf{S} - \alpha^{(v)}\beta\mathbf{H}_m^{(v)}\mathbf{S}^\mathrm{T}+\\ &2\alpha^{(v)}\beta\mathbf{H}_m^{(v)}\mathbf{G} + 2\alpha^{(v)2}\beta\mathbf{H}_m^{(v)}\mathbf{H}_m^{(v)\mathrm{T}}\mathbf{H}_m^{(v)})\mathbf{H}_m^{(v)2} = 0.\end{aligned} \tag{12}$$

Eq. (12) is a fixed point equation. By noting that $\Phi^\mathrm{T}\mathbf{X}^{(v)} = \left[\Phi^\mathrm{T}\mathbf{X}^{(v)}\right]^+ - \left[\Phi^\mathrm{T}\mathbf{X}^{(v)}\right]^-$, $\Phi^\mathrm{T}\Phi\mathbf{H}_m^{(v)} = \left[\Phi^\mathrm{T}\Phi\mathbf{H}_m^{(v)}\right]^+ - \left[\Phi^\mathrm{T}\Phi\mathbf{H}_m^{(v)}\right]^-$ etc, it is easy to get the update rule Eq. (8) for $\mathbf{H}_m^{(v)}$ and to see that the equation satisfies the fixed point equation. At convergence, $\mathbf{H}_m^{(v)(\infty)} = \mathbf{H}_m^{(v)(t+1)} = \mathbf{H}_m^{(v)(t)} = \mathbf{H}_m^{(v)}$.

D. Update rule for matrix $\mathbf{S}$

By fixing $\mathbf{Z}_i^{(v)}$, $\mathbf{H}_i^{(v)}$ and $\alpha^{(v)}$, we can update $\mathbf{S}$ by solving the following problem,

$$\min_{\mathbf{S}} \|\mathbf{S} - \mathbf{Q}\|_F^2, \text{s.t.}\mathbf{S}\mathbf{1} = \mathbf{1}, \mathbf{S} \geq 0, \mathrm{diag}(\mathbf{S}) = 0, \tag{13}$$

where $\mathbf{Q} = \sum_{v=1}^{V} \alpha^{(v)}\mathbf{H}_m^{(v)\mathrm{T}}\mathbf{H}_m^{(v)}$. This problem yields a close-formed solution that,

$$\mathbf{S}_{i,:} = \max\left(\mathbf{Q}_{i,:} + \gamma\mathbf{1}^\mathrm{T}, 0\right), \mathbf{S}_{ii} = 0, \gamma = \frac{1 - \mathbf{Q}_{i,:}\mathbf{1}}{n}, \tag{14}$$

where $\mathbf{S}_{i,:}$ is the i-th row of $\mathbf{S}$, $\mathbf{Q}_{i,:}$ is the i-th row of $\mathbf{Q}$.

Theorem 2: Eq. (14) is the close-formed solution of Eq. (13).

Proof 2: The problem of Eq. (13) can be easily rewritten into n row-formed independent optimization problems as follow,

$$\min_{S_{i,:}} \|\mathbf{S}_{i,:} - \mathbf{Q}_{i,:}\|_\mathrm{F}^2, \text{s.t.}\mathbf{S}_{i,:} \geq 0, \mathbf{S}_{i,:}\mathbf{1} = 1, \mathbf{S}_{i,i} = 0. \tag{15}$$

The Lagrangian function of Eq. (15) is,

$$\mathcal{L}\left(\mathbf{S}_{i,:}, \gamma, \eta\right) = \|\mathbf{S}_{i,:} - \mathbf{Q}_{i,:}\|_F^2 - \gamma\left(\mathbf{S}_{i,:}\mathbf{1} - 1\right) - \eta\mathbf{S}_{i,:}^\mathrm{T}, \tag{16}$$

where γ and η are the Lagrangian multipliers for the constraints $\mathbf{S}_{i,:} \geq 0$ and $\mathbf{S}_{i,:}\mathbf{1} = 1$ respectively. Then the KKT condition is written as,

$$\begin{cases}\mathbf{S}_{i,:} - \mathbf{Q}_{i,:} - \gamma\mathbf{1}^\mathrm{T} - \eta = 0,\\ \eta \odot \mathbf{S}_{i,:}^\mathrm{T} = 0.\end{cases} \tag{17}$$

We can easily obtain the Eq. (14).

E. Update rule for coefficient α

By fixing $\mathbf{Z}_i^{(v)}$, $\mathbf{H}_i^{(v)}$ and $\mathbf{S}$, we can update α by solving the following problem,

$$\min_{\alpha} \|\mathbf{S} - \sum_{v=1}^{V} \alpha^{(v)}\mathbf{H}_m^{(v)\mathrm{T}}\mathbf{H}_m^{(v)}\|_F^2, \text{s.t.} \sum_{v=1}^{V} \alpha^{(v)} = 1, \alpha^{(v)} \geq 0. \tag{18}$$

Supposing $\mathbf{Q} = \sum_{v=1}^{V} \alpha^{(v)}\mathbf{H_m}^{(v)\mathrm{T}}\mathbf{H_m}^{(v)}$, we have that,

$$\|\mathbf{S} - \mathbf{Q}\|_F^2 = n - 2\,\mathrm{Tr}\left(\mathbf{S}^\top\mathbf{Q}\right) + \mathrm{Tr}\left(\mathbf{Q}^\top\mathbf{Q}\right). \tag{19}$$

Note that $\mathbf{Q}$=$\sum_{v=1}^{V} \alpha^{(v)}\mathbf{H_m}^{(v)\mathrm{T}}\mathbf{H_m}^{(v)}$, we have $\mathbf{Q}^\mathrm{T}$=$\mathbf{Q}$. Taking them into Eq. (19), the optimization can be written as follows,

$$\min_{\alpha} \frac{1}{2}\alpha^\mathrm{T}\mathbf{A}\alpha - \mathbf{f}^\mathrm{T}\alpha, \text{s.t.}\alpha^\mathrm{T}\mathbf{1} = 1, \quad \alpha \geq 0, \tag{20}$$

where the variables are defined as follows,

$$\begin{aligned}
&\mathbf{f}^\mathrm{T} = \left[f_1, f_2, \ldots, f_V\right],\\
&\mathbf{f}_v = \mathrm{Tr}\left(\mathbf{S}^\mathrm{T}\mathbf{H}_m^{(v)\mathrm{T}}\mathbf{H}_m^{(v)}\right),\\
&\mathbf{A}_{pq} = \mathrm{Tr}\left(\mathbf{H}_m^{(p)\mathrm{T}}\mathbf{H}_m^{(p)}\mathbf{H}_m^{(q)\mathrm{T}}\mathbf{H}_m^{(q)}\right),\\
&\alpha = \left[\alpha^{(1)}, \alpha^{(2)}, \ldots, \alpha^{(V)}\right]^\mathrm{T}.
\end{aligned} \tag{21}$$

Algorithm 1 MVC-DMF-GGR

Input : Set of given Multi-view data matrices $\mathbf{X}^{(v)}(1 \leq v \leq V)$, tuning parameters β and the depth of layers p.
Initialize : Initialize $\mathbf{H}_i^{(v)}$ and $\mathbf{Z}_i^{(v)}$ according to II, and then initialize $\alpha^{(v)}$ and $\mathbf{S}$.
Output : Performing spectral clustering on $\mathbf{S}$.

1: **while** not convergence **do**
2:     **for** $i \leq$ m **do**
3:         update $\mathbf{Z}_i^{(v)}$ by solving Eq. (2).
4:         update $\mathbf{H}_i^{(v)}$ by solving Eq. (5).
5:     **end for**
6:     update $\mathbf{H}_m^{(v)}$ by solving Eq. (8).
7:     update $\mathbf{S}$ by solving Eq. (14).
8:     update α by solving Eq. (20).
9: **end while**
10: **return** Similarity matrix $\mathbf{S}$. Performing spectral clustering on $\mathbf{S}$ to get final clustering partition.

For every $x \in \mathbb{R}^m$, we have that $x^{\mathrm{T}}\mathbf{A}x = \left\|\sum_{v=1}^{V} x_v \mathbf{H}_m^{(v)} \mathbf{H}_m^{(v)\mathrm{T}}\right\|_F^2 \geq 0$. So the matrix $\mathbf{A}$ is a positive semi-definite matrix and quadratic programming could be used in Eq. (20).

The entire approach is outlined in Algorithm 1. We train the proposed algorithm at least 150 iterations until convergence, then we perform spectral clustering on $\mathbf{S}$ to obtain the clustering results.

IV. Analysis and discussions

Computational Complexity: Pre-training and fine-tuning are the two main stages of our proposed method, and we will analyze them separately. To make the analysis clearer, we assume the dimensions in all the layers are the same. So we denote l and the dimensions of the original feature for all the views are the same which denoted d. t_{pre} denotes the number of iterations to achieve convergence in pre-training process and t_{fine} denotes the number of iterations to achieve convergence in fine-tuning process. So the complexity of pre-training and fine-tuning stages are $O(Vmt_{pre}(nd^2+nl^2+ldn+dl^2))$ and $O(Vmt_{fine}(dnl+nl^2+dl^2+ln^2))$ respectively, where $l \leq d$ normally. In conclusion, the time complexity of our algorithm is $O(Vm(t_{pre}(nd^2+dl^2)+t_{fine}(dnl+dl^2+ln^2)))$.

Convergence: It is easy to obtain that the lower bound of the whole optimization function is 0. When we optimize one variable with fixing the others, the four (optimizing $\mathbf{Z}_i^{(v)}$ and $\mathbf{H}_i^{(v)}$ as one subproblem) subproblems are strictly convex and the objective of Algorithm 1 is monotonically decreased at each iteration. As a result, the proposed algorithm can be confirmed to be convergent.