A. Self-Supervised Bidirectional of GM algorithm

Algorithm 1: Self-Supervised Bidirectional of GM algorithm

Input: $\mathcal{G}_s := \{\boldsymbol{X}_s, \boldsymbol{E}_s, \boldsymbol{A}_s\}$, $\mathcal{G}_t := \{\boldsymbol{X}_t, \boldsymbol{E}_t, \boldsymbol{A}_t\}$, GNN encoder $\boldsymbol{\psi}_\theta$, transformation functions $\mathcal{U}$.
Output: matching $\boldsymbol{Z}$.
repeat
  Obtain the affinity matrix $\boldsymbol{M}$ using $\boldsymbol{h}_s, \boldsymbol{h}_t$ of input data by Eq. (5). ▷ Pseudo Correspondence Generation
  Compute the predictive probability score $\boldsymbol{P}$ by employing row-wise *softmax* on $\boldsymbol{M}$.
  Generate pseudo correspondence labels $\boldsymbol{Z}$ using Eq. (6).
  Generate pseudo samples $\widehat{\boldsymbol{X}_s}, \widehat{\boldsymbol{X}_t}$ using Eq. (9). ▷ Bidirectional recycling consistency
  Obtain $\widehat{\boldsymbol{P}}$ based on Eq. (2) & Eq. (5) with *softmax*.
  Generate the augmented views $\mathcal{G}_s^{(k)}, \mathcal{G}_t^{(k)}$ using $\mathcal{U}$. ▷ GCL to enhance graph representation learning
  Obtain $\boldsymbol{y}_s^{(k)}, \boldsymbol{y}_t^{(k)}$ based on Eq. (12).
  Update $\boldsymbol{\psi}_\theta$ with loss $\mathcal{L}_{overall}$ given $\boldsymbol{y}_s^{(k)}, \boldsymbol{y}_t^{(k)}, \boldsymbol{P}, \boldsymbol{Z}, \hat{\boldsymbol{P}}$ according to Eq. (14).
until converge
Predict matching $\boldsymbol{Z}$.

B. Proof of Proposition 2

Proof. We first derive the isomorphic relationship between the source graph $\mathcal{G}_s = (\boldsymbol{X}_s, \boldsymbol{A}_s)$ and target graph $\mathcal{G}_t = (\boldsymbol{X}_t, \boldsymbol{A}_t))$ in the following equations,

$$\boldsymbol{X}_s = \boldsymbol{Z}^* \boldsymbol{X}_t, \boldsymbol{A}_s = \boldsymbol{Z}^* \boldsymbol{X}_t \boldsymbol{Z}^{*\top}. \tag{15}$$

Since $\boldsymbol{\Psi}_\theta$ is a permutation-equivariant GNN, it holds that $\boldsymbol{P}^\top [\boldsymbol{\Psi}_\theta(\boldsymbol{X}, \boldsymbol{A})] = \left[\boldsymbol{\Psi}_\theta \left(\boldsymbol{P}^\top \boldsymbol{X}, \boldsymbol{P}^\top \boldsymbol{A}\boldsymbol{P}\right)\right]$ for any permutation matrix $\boldsymbol{P} \in \{0,1\}^{|\mathcal{V}|\times|\mathcal{V}|}$, where $[\cdot]$ indicates that $\cdot$ is a matrix. We therefore have

$$\boldsymbol{H}_s = [\boldsymbol{\Psi}_\theta (\mathcal{G}_s)] = [\boldsymbol{\Psi}_\theta (\boldsymbol{X}_s, \boldsymbol{A}_s)] = \left[\boldsymbol{\Psi}_\theta \left(\boldsymbol{Z}^* \boldsymbol{X}_t, \boldsymbol{Z}^* \boldsymbol{A}_t \boldsymbol{Z}^{*\top}\right)\right] = \boldsymbol{Z}^* [\boldsymbol{\Psi}_\theta (\boldsymbol{X}_t, \boldsymbol{A}_t)] = \boldsymbol{Z}^* [\boldsymbol{\Psi}_\theta (\mathcal{G}_t)] = \boldsymbol{Z}^* \boldsymbol{H}_t. \tag{16}$$

□

C. Proof of Proposition 3

Proof. We first consider the predictive matching probability score between the input data $\mathcal{G}_s$ and $\mathcal{G}_t$. Following the Proposition 2, we have

$$\boldsymbol{f}\left(\boldsymbol{H}_s \boldsymbol{H}_t^\top\right) = \boldsymbol{f}\left([\boldsymbol{\Psi}_\theta (\boldsymbol{X}_s, \boldsymbol{A}_s)] [\boldsymbol{\Psi}_\theta (\boldsymbol{X}_t, \boldsymbol{A}_t)]^\top\right) = \boldsymbol{f}\left([\boldsymbol{\Psi}_\theta (\boldsymbol{X}_s, \boldsymbol{A}_s)] [\boldsymbol{\Psi}_\theta (\boldsymbol{X}_s, \boldsymbol{A}_s)]^\top \boldsymbol{Z}^*\right). \tag{17}$$

Similarly, for the pseudo data $\widehat{\mathcal{G}}_s$ and $\widehat{\mathcal{G}}_t$, we have

$$\boldsymbol{f}\left(\widehat{\boldsymbol{H}}_s \widehat{\boldsymbol{H}}_t^\top\right) = \boldsymbol{f}\left(\left[\boldsymbol{\Psi}_\theta \left(\widehat{\boldsymbol{X}}_s, \boldsymbol{A}_s\right)\right] \left[\boldsymbol{\Psi}_\theta \left(\widehat{\boldsymbol{X}}_t, \boldsymbol{A}_t\right)\right]^\top\right) = \boldsymbol{f}\left(\left[\boldsymbol{\Psi}_\theta \left(\widehat{\boldsymbol{X}}_s, \boldsymbol{A}_s\right)\right] \left[\boldsymbol{\Psi}_\theta \left(\widehat{\boldsymbol{X}}_s, \boldsymbol{A}_s\right)\right]^\top \widehat{\boldsymbol{Z}}^*\right), \tag{18}$$

where $\widehat{\boldsymbol{Z}}^* \in \{0,1\}^{|\mathcal{V}|\times|\mathcal{V}|}$ denote the isomorphic relationship between $\widehat{\mathcal{G}}_s$ and $\widehat{\mathcal{G}}_t$.
According to Eq. (9), we have

$$\begin{cases} \widehat{\boldsymbol{X}}_s = \boldsymbol{Z}_\theta \boldsymbol{X}_t = \boldsymbol{Z}_\theta \boldsymbol{Z}^{*\top} \boldsymbol{X}_s \\ \widehat{\boldsymbol{X}}_t = \boldsymbol{Z}_\theta^\top \boldsymbol{X}_s, \end{cases} \tag{19}$$

then we can infer that $\widehat{\boldsymbol{X}}_s = \boldsymbol{Z}_\theta \boldsymbol{Z}^{*\top} \boldsymbol{Z}_\theta \widehat{\boldsymbol{X}}_t$. Therefore, we have

$$\widehat{\boldsymbol{Z}}^* = \boldsymbol{Z}_\theta \boldsymbol{Z}^{*\top} \boldsymbol{Z}_\theta = \boldsymbol{Z}_\theta^\epsilon \boldsymbol{Z}^* \boldsymbol{Z}^{*\top} \boldsymbol{Z}_\theta^\epsilon \boldsymbol{Z}^* = [\boldsymbol{Z}_\theta^\epsilon]^2 \boldsymbol{Z}^*. \tag{20}$$

We apply the above Eq. (20) to Eq. (19) and obtain the following equality

$$\widehat{\boldsymbol{X}}_s = \boldsymbol{Z}_\theta^\epsilon \boldsymbol{X}_s. \tag{21}$$

Based on Eq. (17), Eq (18), Eq. (20) and Eq. (21), we have

$$\begin{aligned} \boldsymbol{J}(\boldsymbol{\theta}) &= \|\boldsymbol{f}\left(\widehat{\boldsymbol{H}}_s \widehat{\boldsymbol{H}}_t^\top\right) - \boldsymbol{f}\left(\boldsymbol{H}_s \boldsymbol{H}_t^\top\right)\| \\ &= \|\boldsymbol{f}\left([\boldsymbol{\Psi}_\theta (\boldsymbol{Z}_\theta^\epsilon \boldsymbol{X}_s, \boldsymbol{A}_s)] [\boldsymbol{\Psi}_\theta (\boldsymbol{Z}_\theta^\epsilon \boldsymbol{X}_s, \boldsymbol{A}_s)]^\top [\boldsymbol{Z}_\theta^\epsilon]^2 \boldsymbol{Z}^*\right) - \boldsymbol{f}\left([\boldsymbol{\Psi}_\theta (\boldsymbol{X}_s, \boldsymbol{A}_s)] [\boldsymbol{\Psi}_\theta (\boldsymbol{X}_s, \boldsymbol{A}_s)]^\top \boldsymbol{Z}^*\right)\| \geq 0, \end{aligned} \tag{22}$$

which completes the proof of Proposition 3. □

D. Dataset Statistics

Statics including the number of images, the number of classes, averaged number of nodes, partial rate and data type are summarized in Tab. 4.

dataset name	# images	# classes	avg # nodes	partial rate	data type
PascalVOC	8702	20	9.07	28.5%	RGB
WilloWObject	404	5	10	0.0%	RGB
Cub2011	11788	200	12.0	20.0%	RGB
CMU-House/Hotel	212	2	30	0.0%	gray-scale
IMC-PT-SparseGM	25061	16	21.36	57.3%	RGB

Table 4: Detailed statistics of the GM benchmark datasets.

E. Experimental Results on IMC-PT-SparseGM and Willow Object

Here, we provide the numberical value of matching accuracies (%) on the IMC-PT-SparseGM and Willow Object dataset in Tab. 5 and Tab. 6, respectively.

Method	reichstag	sacre_coeur	st_peters_square	ave
IA-GM	**98.1**	73.5	84.2	85.3
GANN	93.0	71.9	82.0	82.3
BiGM	95.2	**75.2**	**88.5**	**86.3**
IA-GM	18.9	16.0	23.0	19.3
GANN	77.2	58.4	**68.1**	67.9
BiGM	**81.0**	**59.1**	66.9	**69.0**

Table 5: Matching accuracies (%) on the IMC-PT-SparseGM testing dataset. white (up): INLIER; gray (bottom): OUTLIER.

METHOD	FACE	MBIKE	CAR	DUCK	WBOTTLE	AVE
IA-GM	91.2	84.2	100.0	**98.0**	94.0	93.5
GANN	85.4	89.8	100.0	88.6	96.4	92.0
BIGM	**100.**	**99.6**	100.0	92.1	**99.4**	**98.2**

Table 6: Matching accuracies (%) on the Willow Object dataset.

F. Comparisons with Supervised Learning Results

We compare the results by self-supervised learning with those by supervised learning methods in Tab. 7. The matching accuracies are directly collected from the literature papers. It is noted that not all the literature papers have reported results on the five benchmark datasets. We observe that our self-supervised method is even superior to GMN (Zanfir and Sminchisescu 2018), or comparable to the supervised version of IA-GM (Zhao, Tu, and Xu 2021). Although there is large gap between our self-supervised method and DLGM-G (Yu et al. 2021) in the INLIER setting of PASCALVOC, the difference becomes very small at the OUTLIER setting, indicating that our self-supervised method is robust against outliers. It deserves more work in the future to see whether self-supervised learning can further boost the performance of the supervised learning methods.

G. Visual Examples of the Matching Results.

METHOD	LEARNING	PASCALVOC		WILLOW	CUB2011		CMU	IMC-PT-SPARSEGM	
		INLIER	OUTLIER	INLIER	INLIER	OUTLIER	INLIER	INLIER	OUTLIER
GMN	supervised	40.6	-	-	86.0	-	-	-	-
PCA-GM	supervised	63.8	-	96.9	93.8	-	-	53.5	-
IA-GM	supervised	66.58	-	92.3	95.2	-	-	75.2	-
CIE-H	supervised	68.9	-	97.6	94.2	-	-	-	-
BBGM	supervised	80.1	61.4*	98.8	-	-	-	-	-
NGM-V2	supervised	80.1	-	97.5	96.3	-	-	87.1	-
GAMnet	supervised	80.7	-	99.4	-	-	-	-	-
DLGM-G	supervised	83.8	64.8*	99.3	-	-	-	-	-
IA-GM	self-sup.	56.1	41.7	93.5	62.0	46.4	98.7	85.3	19.3
GANN	self-sup.	57.2	24.3	92.0	79.0	70.8	100.0	82.3	67.9
BiGM(pretrain)	self-sup.	63.4	61.3	95.6	84.8	84.3	100.0	84.7	71.5
BiGM	self-sup.	65.2	62.8	98.2	86.5	85.6	100.0	86.3	69.0

Table 7: Mean accuracy (%) of both supervised and self-supervised methods on all datasets. The results are taken from the original papers. * means both source and target graphs have outliers in their papers, which is harder than our settings.

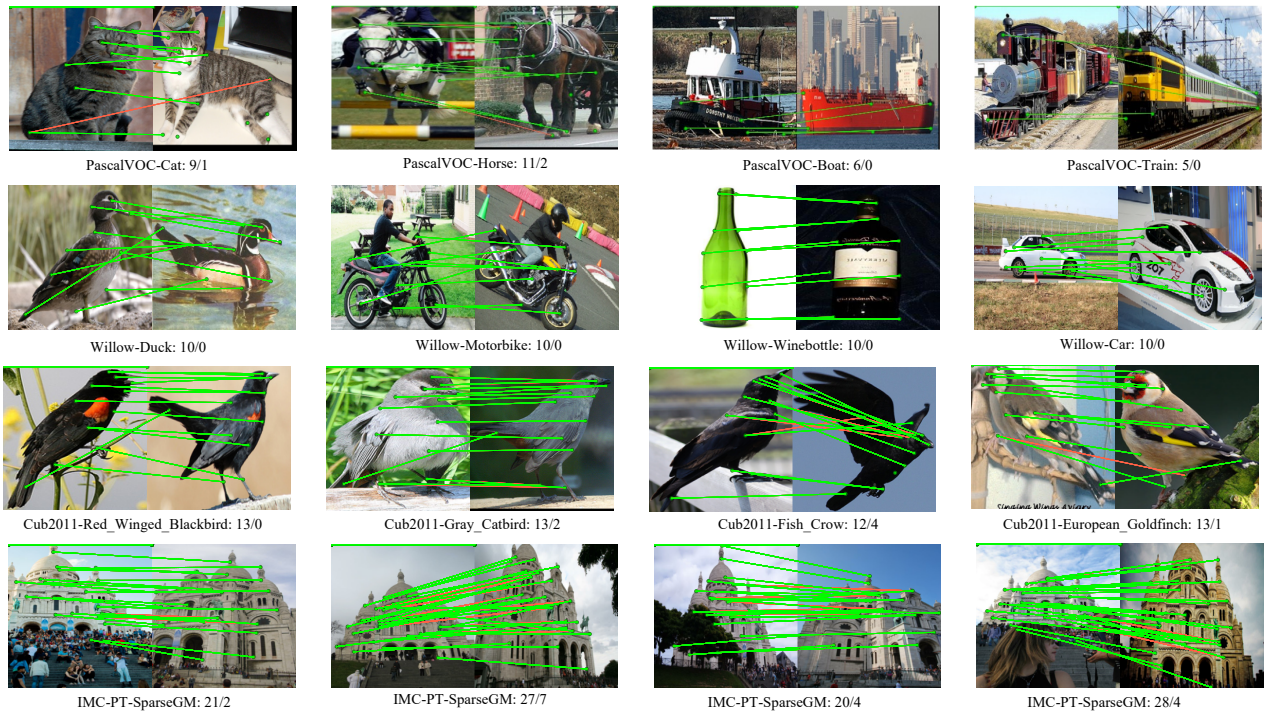

Figure 2: Visual illustration of the matching results on the four sampled categories of PASCALVOC, WILLOWOBJECT, CUB2011, IMC-PT-SPARSEGM datasets.Green and red lines represent correct and incorrect node matchings respectively. The subtitle of each figure shows the correct / incorrect matching count.