Person Re-identification with Deep Similarity-Guided Graph Neural Network

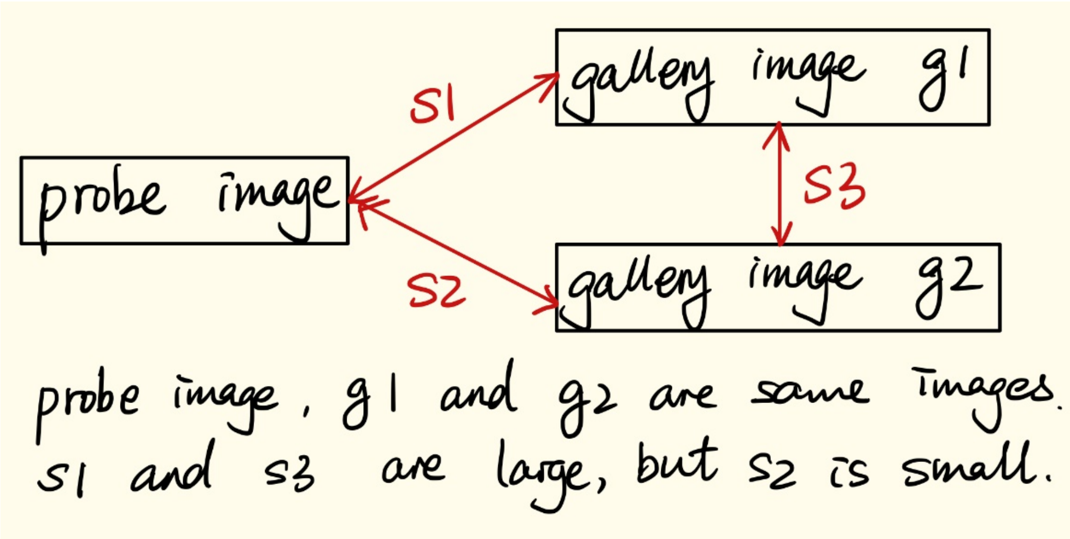
Yantao Shen, Hongsheng Li, Shuai Yi, Dapeng Chen, Xiaogang Wang

CUHK-SenseTime Joint Lab, SenseTime Research

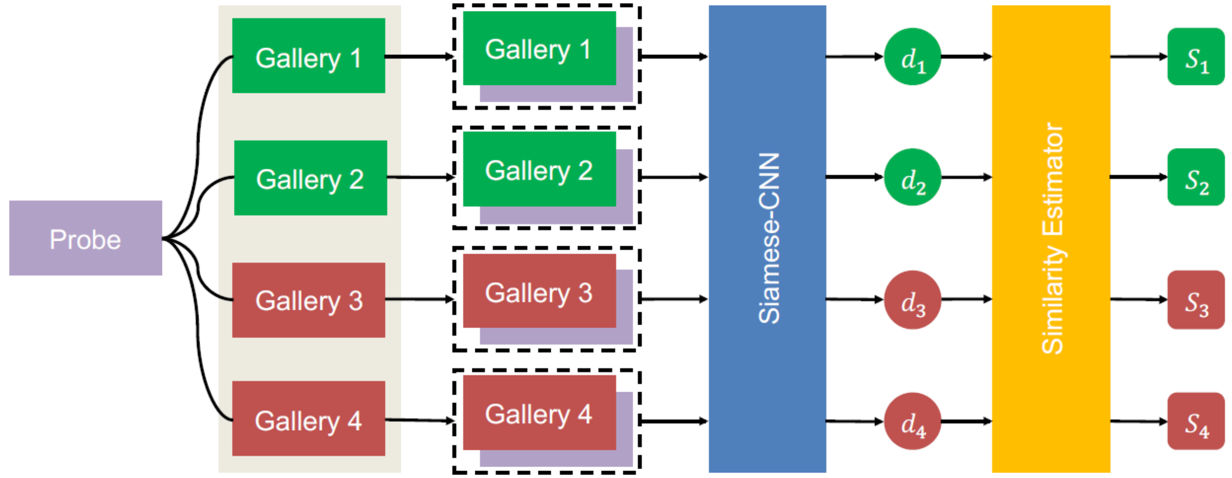
ECCV 2018

Problem

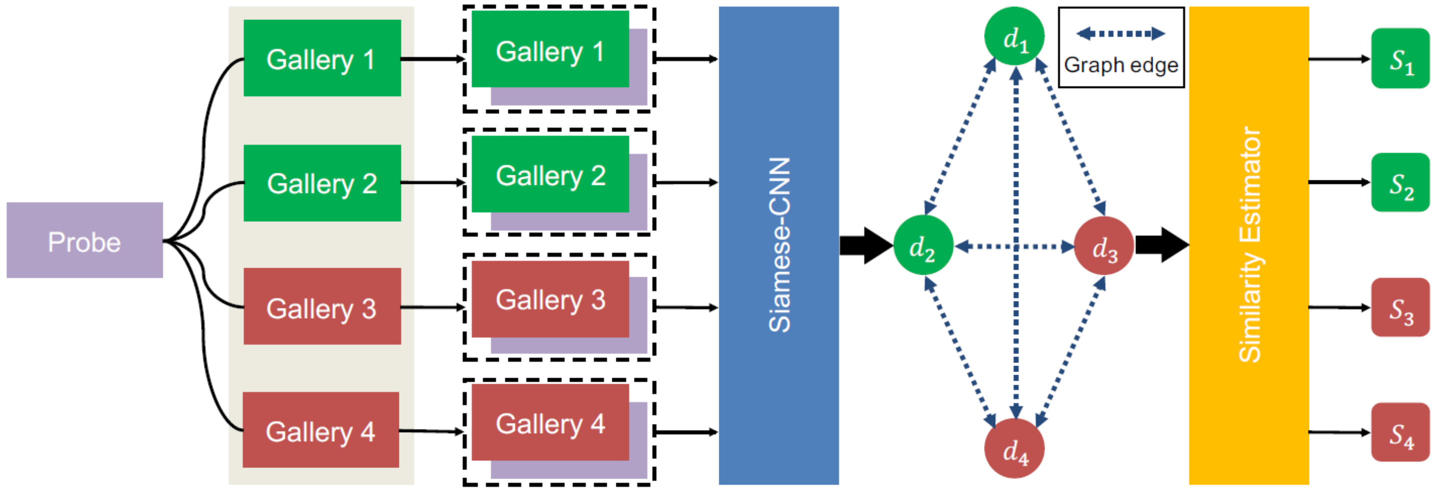
- **Weakness of the Existing Person Re-ID Models**:
  - Ignore the relationship information between different probe-gallery pairs.
  - Hard samples are difficult to get proper similarity scores.
- **Main Idea**: Update s2 by s1 and s3.



Comparison



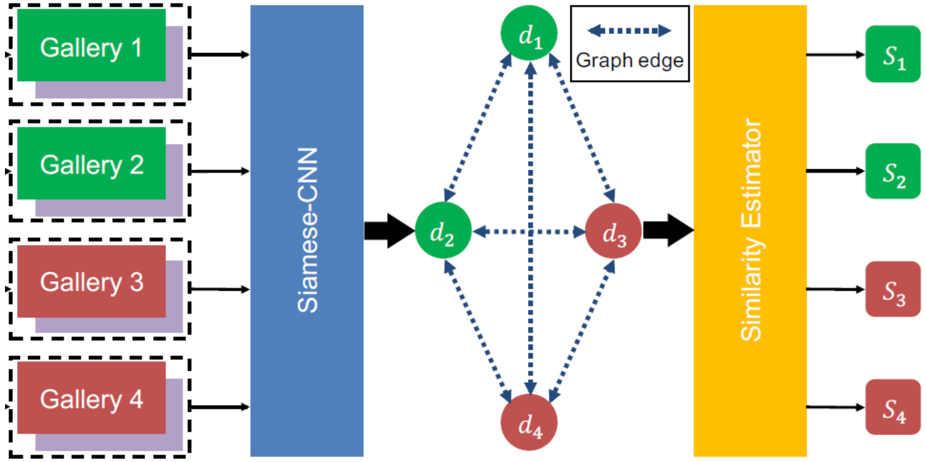
Conventional Approach



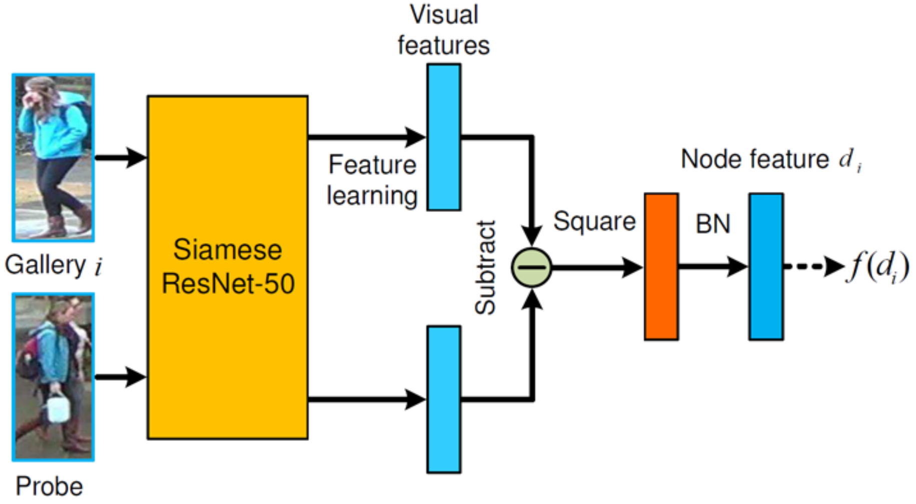
SGGNN Approach

Graph Formulation

- An undirected **complete graph** $G(V, E)$.
- **Each node** $\boldsymbol{v_i}$ represents a pair of probe-gallery images.
- **Node features** are processed difference features.



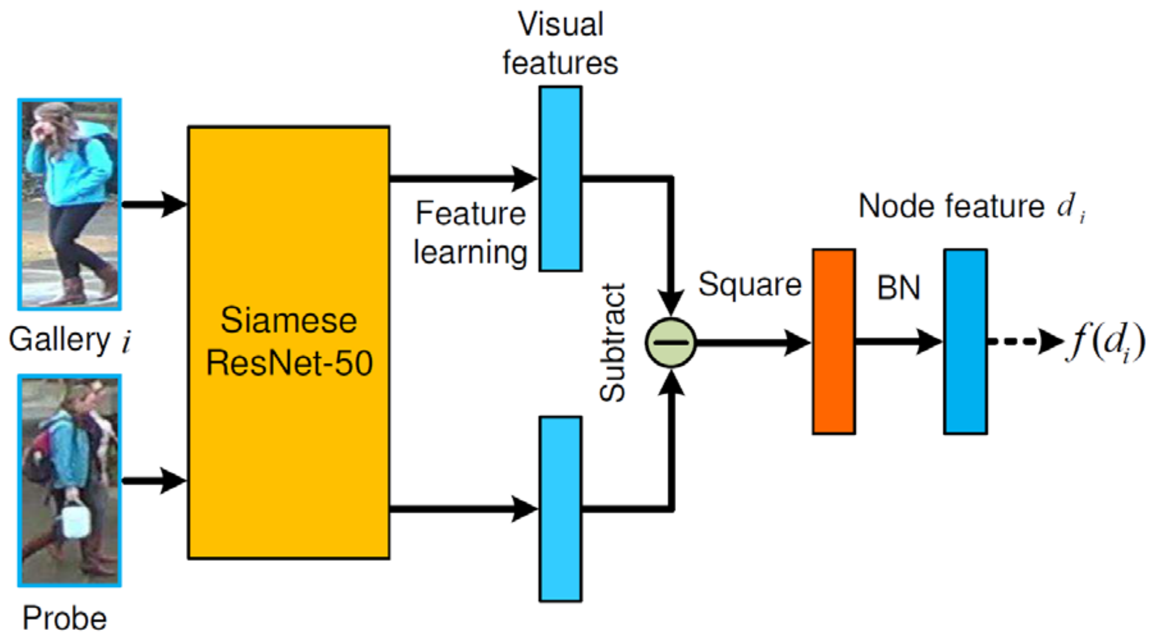
Graph Illustration



Node Feature Generating

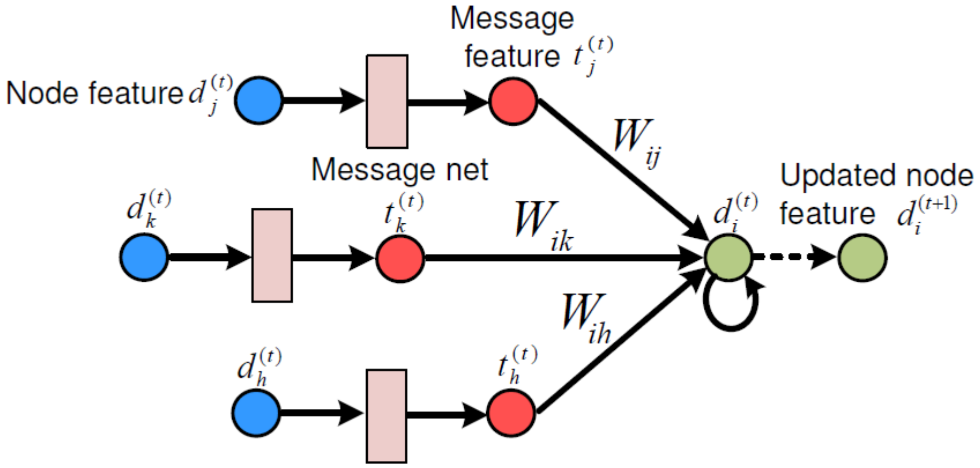
Naïve Node **Loss Function I**

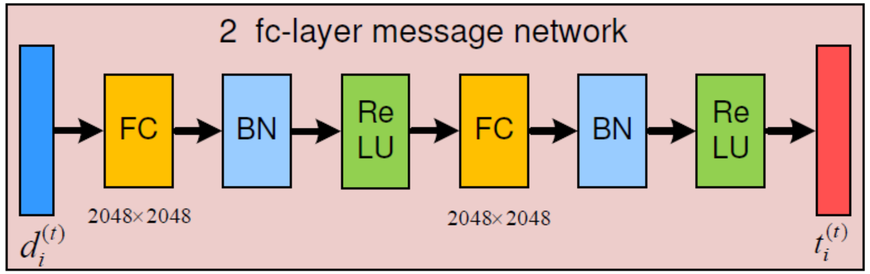
- $L = -\sum_{i=1}^{N} y_i \log\big(f(d_i)\big) + (1-y_i)\log(1-f(d_i))$
- $f()$ is a linear classifier followed by a sigmoid function.



Similarity-Guided Graph Neural Network

- **Intuition**: Using gallery-gallery similarity scores to guide the refinement of the probe-gallery relation features.
- **Updating Node Feature**: Original Feature + Fusion Feature.
- $d_i^{(t+1)} = (1-\alpha)d_i^{(t)} + \alpha \sum_{j=1}^{N} W_{ij} t_j^{(t)}$ for $i = 1,2,\dots,N$



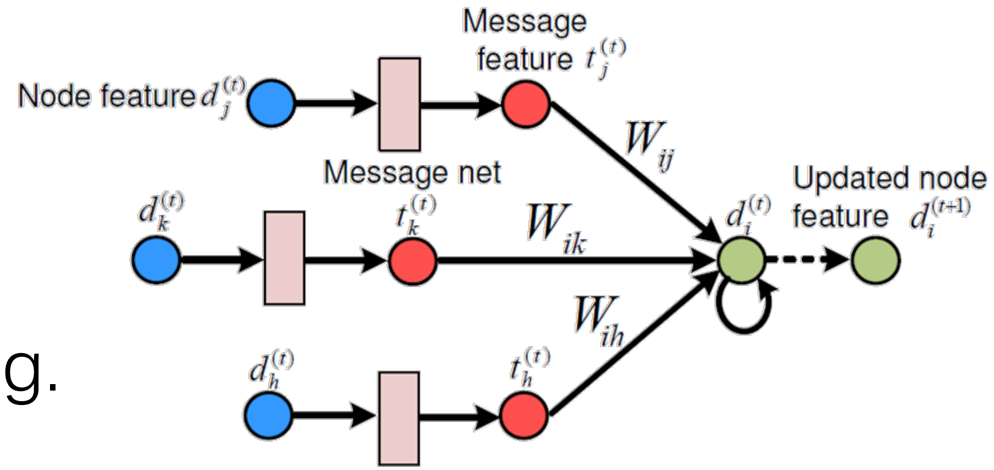


Similarity-Guided Graph Neural Network

- **Updating**: $d_i^{(t+1)} = (1-\alpha)d_i^{(t)} + \alpha \sum_{j=1}^{N} W_{ij} t_j^{(t)}$ for $i = 1,2,\dots,N$
- W_{ij} is a scalar edge weight, represents the relation importance between node i and node j.
- $W_{ij} = \begin{cases} \frac{\exp(S(g_i,g_j))}{\sum_j \exp(S(g_i,g_j))}, & i \neq j \\ \boxed{0, \quad i = j} \end{cases}$

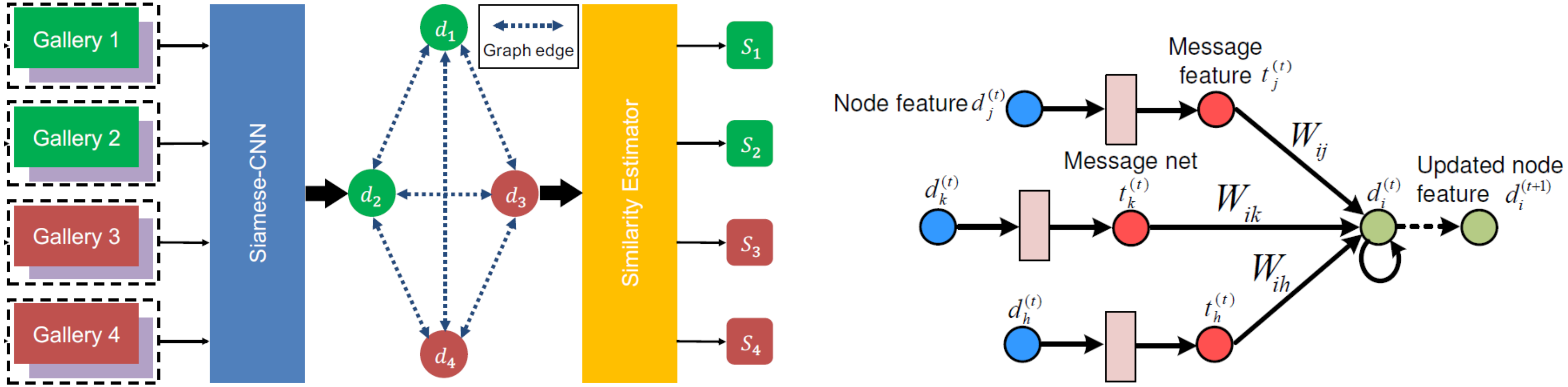
Avoid Self-Enhancing

- $S()$ is a pairwise similarity function.
- Set $t = 1$ in both training and testing.



Update Node **Loss Function II**

- $L = -\sum_{i=1}^{N} y_i \log(s_i) + (1 - y_i)\log(1 - s_i)$
- Similarity estimator is a linear classifier followed by a sigmoid function.



Datasets, Metrics, Experiments

- **Datasets**: CUHK03, Market-1501, DukeMTMC
- **Metrics**: mAP and CMC top-1, top-5, top-10 accuracies.

CUHK03

Methods	Conference	CUHK03 [28] mAP	top-1	top-5	top-10
Quadruplet Loss [9]	CVPR 2017	-	75.5	95.2	99.2
OIM Loss [65]	CVPR 2017	72.5	77.5	-	-
SpindleNet [73]	CVPR 2017	-	88.5	97.8	98.6
MSCAN [26]	CVPR 2017	-	74.2	94.3	97.5
SSM [2]	CVPR 2017	-	76.6	94.6	98.0
k-reciprocal [78]	CVPR 2017	67.6	61.6	-	-
VI+LSRO [77]	ICCV 2017	87.4	84.6	97.6	98.9
SVDNet [61]	ICCV 2017	84.8	81.8	95.2	97.2
OL-MANS [80]	ICCV 2017	-	61.7	88.4	95.2
Pose Driven [60]	ICCV 2017	-	88.7	98.6	**99.6**
Part Aligned [74]	ICCV 2017	-	85.4	97.6	99.4
HydraPlus-Net [39]	ICCV 2017	-	91.8	98.4	99.1
MuDeep [49]	ICCV 2017	-	76.3	96.0	98.4
JLML [29]	IJCAI 2017	-	83.2	98.0	99.4
MC-PPMN [43]	AAAI 2018	-	86.4	98.5	**99.6**
Proposed SGGNN		**94.3**	**95.3**	**99.1**	**99.6**

Market-1501

Methods	Reference	Market-1501 [75] mAP	top-1	top-5	top-10
OIM Loss [65]	CVPR 2017	60.9	82.1	-	-
SpindleNet [73]	CVPR 2017	-	76.9	91.5	94.6
MSCAN [26]	CVPR 2017	53.1	76.3	-	-
SSM [2]	CVPR 2017	68.8	82.2	-	-
k-reciprocal [78]	CVPR 2017	63.6	77.1	-	-
Point 2 Set [81]	CVPR 2017	44.3	70.7	-	-
CADL [35]	CVPR 2017	47.1	73.8	-	-
VI+LSRO [77]	ICCV 2017	66.1	84.0	-	-
SVDNet [61]	ICCV 2017	62.1	82.3	92.3	95.2
OL-MANS [80]	ICCV 2017	-	60.7	-	-
Pose Driven [60]	ICCV 2017	63.4	84.1	92.7	94.9
Part Aligned [74]	ICCV 2017	63.4	81.0	92.0	94.7
HydraPlus-Net [39]	ICCV 2017	-	76.9	91.3	94.5
JLML [29]	IJCAI 2017	65.5	85.1	-	-
HA-CNN [30]	CVPR 2018	75.7	91.2	-	-
Proposed SGGNN		**82.8**	**92.3**	**96.1**	**97.4**

DukeMTMC

Methods	Reference	DukeMTMC [52] mAP	top-1	top-5	top-10
BoW+KISSME [75]	ICCV 2015	12.2	25.1	-	-
LOMO+XQDA [34]	CVPR 2015	17.0	30.8	-	-
ACRN [54]	CVPRW 2017	52.0	72.6	84.8	88.9
OIM Loss [65]	CVPR 2017	47.4	68.1	-	-
Basel.+LSRO [77]	ICCV 2017	47.1	67.7	-	-
SVDNet [61]	ICCV 2017	56.8	76.7	86.4	89.9
Proposed SGGNN		**68.2**	**81.1**	**88.4**	**91.2**

Ablation Study

- **Base Model**: Only use the naïve node loss function.
- **SGGNN w/o SG**: $d_i^{(t+1)} = (1-\alpha)d_i^{(t)} + \alpha \sum_{j=1}^{N} h(d_i, d_j)t_j^{(t)}$, where $h(*,*)$ is an inner product function.

Methods	Market-1501 [75]		CUHK03 [28]		DukeMTMC [52]	
	mAP	top-1	mAP	top-1	mAP	top-1
Base Model	76.4	91.2	88.9	91.1	61.8	78.8
Base Model + SGGNN w/o SG	81.2	90.6	92.7	93.6	67.3	80.5
Base Model + SGGNN	**82.8**	**92.3**	**94.3**	**95.3**	**68.2**	**81.1**

Normal Ablation Study.

Model	Market-1501 [75]		CUHK03 [28]		DukeMTMC [52]	
	mAP	top-1	mAP	top-1	mAP	top-1
Base Model	74.6	90.4	87.6	91.0	60.3	77.6
Base Model + SGGNN w/o SG	75.4	90.4	87.7	91.5	61.7	78.1
Base Model + SGGNN	**76.7**	**91.5**	**88.1**	**93.6**	**64.6**	**79.1**

To show SGGNN also learns better visual features.
Evaluate the performance by directly calculating the l2 distance between probe and gallery image features from ResNet-50 model

Conclusion

- Present SGGNN to incorporate the rich gallery-gallery similarity information into training process.
- + Consider the relationship between each probe-gallery pair.
- + Add directly label supervision for guidance.
- - A complete graph might be slow when the number of nodes increasing.