Multi-Scale Cascading Network with Compact Feature Learning for RGB-Infrared Person Re-Identification

Can Zhang[1], Hong Liu[1], Wei Guo[2], Mang Ye[3]

1 北京大学 PEKING UNIVERSITY

2 HUAWEI

3

Problem Definition

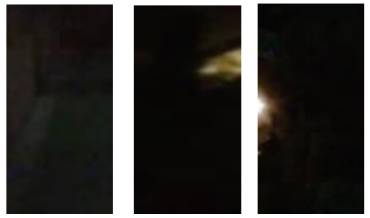
- RGB-Infrared Person Re-Identification

Query:
RGB

Gallery:
RGB

Example RGB Images

Under poor lighting condition?



(a) Persons captured by visible camera at night

(b) Persons captured by infrared camera at night

Query:
Infrared

Gallery:
RGB

Motivations

Challenges:

❖ Inter-Modality Discrepancy

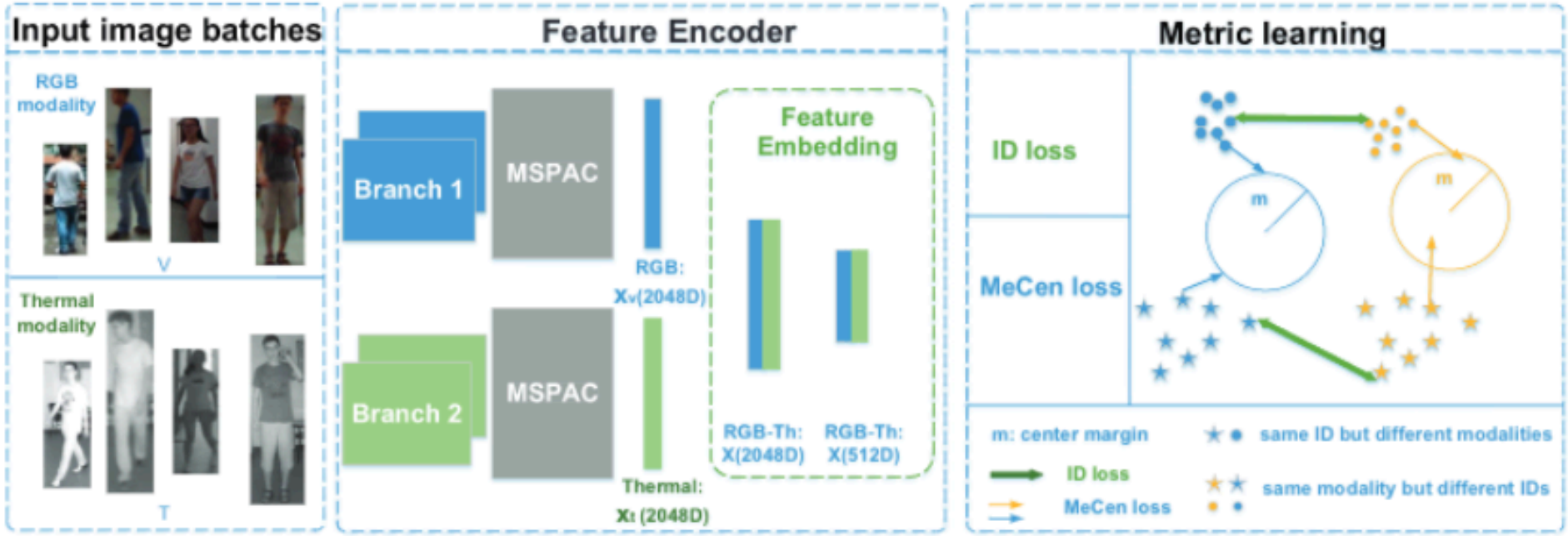
❖ Intra-Modality Discrepancy

❖ Abundant Noise

❖ Cross-Camera Variations

❖ ...

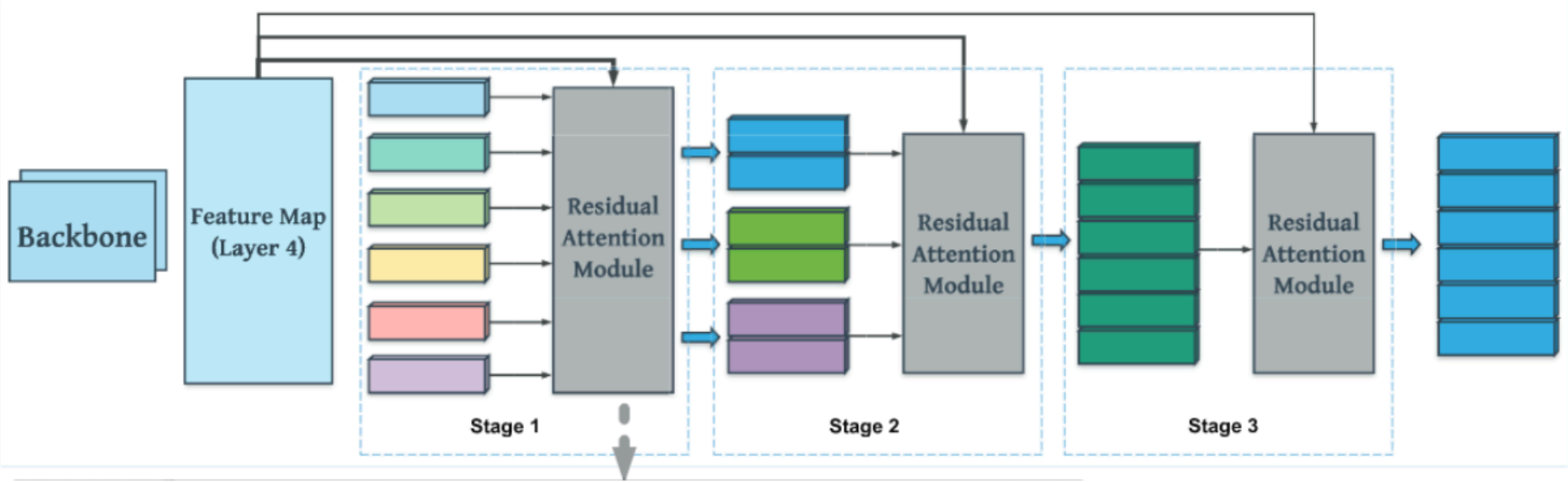
Main Idea



- Multi-Scale Part Aware Attention module (MSPAC)
  It incorporates structured semantic information and local salient information into a unified global representations
- Marginal Exponential Center loss (MeCen)
  It learns modality-invariant correlations which enables images of the same identities to cluster compactly in feature space regardless of the modality form

Multi-Scale Part-Aware Cascading Framework (MSPAC)

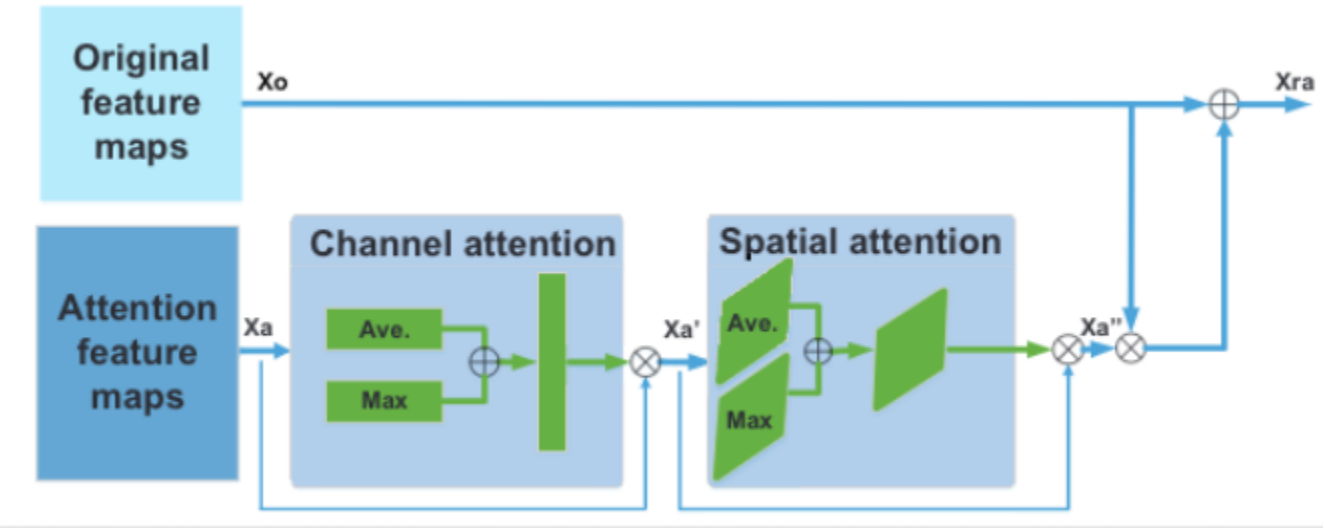
❖ Part Feature Aggregation in Cascading Framework



- Stage1: Fine-grained feature partition
- Stage2: Hierarchical part aggregation
- Stage3: Global representation unification

Multi-Scale Part-Aware Cascading Framework (MSPAC)

❖ Attention Mechanism with Two-branch Structure



- Trunk branch: original global feature
- Attention branch: Spatial attention and Channel attention

Multi-Scale Part-Aware Cascading Framework (MSPAC)

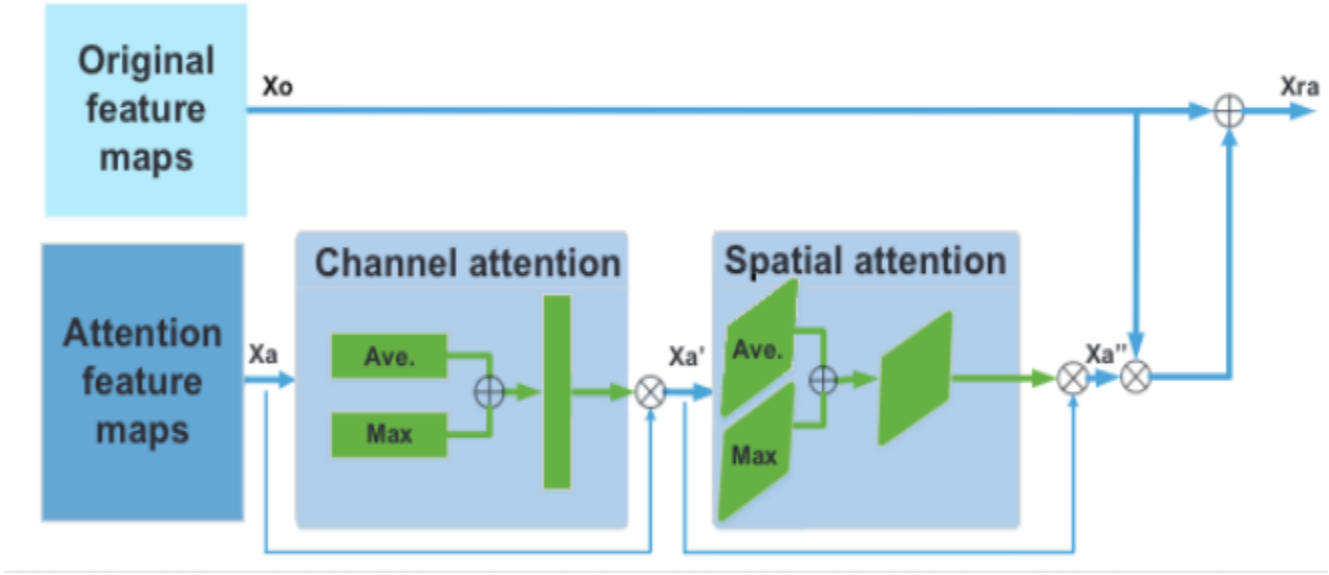
❖ Attention Mechanism with Two-branch Structure

➢ What to focus on?

$$CH(x_a) = \sigma(W_F(x_{avg}^{ch} + x_{max}^{ch}) + b_F),$$

➢ Where to focus?

$$SP(x_a^{'}) = \sigma\{W_{conv}\{x_{avg}^{sp} \| x_{max}^{sp}\}\},$$



➢ Final feature output

$$\begin{cases} x_a^{'} = x_a \otimes CH(x_a), \\ x_a^{''} = x_a^{'} \otimes SP(x_a^{'}), \end{cases} \Rightarrow \quad x = x_o + x_o \otimes x_a^{''}.$$

Marginal Exponential Center loss (MeCen)

$$L_{MeCen} = e^{\frac{1}{2}\sum_{i=1}^{n} \max\{||x_i - c_{y_i}||_2^2 - m, 0\}} - 1,$$
$$s.t.\ ||c||_2^2 = 1.$$

- reducing acceptable variances among easy examples
- imposing strong exponential constraints on hard positive examples

Experiments

Comparison with State-of-the-Art Methods on SYSU-MM01 dataset

Datasets		SYSU															
Feature	Metric	All-search								Indoor-search							
		Single-shot				blueMulti-shot				Single-shot				blueMulti-shot			
		r=1	r=10	r=20	mAP	r=1	r=10	r=20	mAP	r=1	r=10	r=20	mAP	r=1	r=10	r=20	mAP
HOG	KISSME	2.12	16.21	29.13	3.53	2.79	18.23	31.25	1.96	3.11	25.47	46.47	7.43	4.10	29.32	50.59	3.61
	LFDA	2.33	18.58	33.38	4.35	3.82	20.48	35.84	2.20	2.44	24.13	45.50	6.87	3.42	25.27	45.11	3.19
	CCA	2.74	18.91	32.51	4.28	3.25	21.82	36.51	2.04	4.38	29.96	50.43	8.70	4.62	34.22	56.28	3.87
	CRAFT	2.59	17.93	31.50	4.24	3.58	22.90	38.59	2.06	3.03	24.07	42.89	7.07	4.16	27.75	47.16	3.17
LOMO	KISSME	2.23	18.95	32.67	4.05	2.65	20.36	34.78	2.45	3.83	31.09	52.86	8.94	4.46	34.35	58.43	4.93
	LFDA	2.89	21.11	35.36	4.81	3.86	24.01	40.54	2.61	4.81	32.16	52.50	9.56	6.27	36.29	58.11	5.15
	CCA	2.42	18.22	32.45	4.19	2.63	19.68	34.82	2.15	4.11	30.60	52.54	8.83	4.86	34.40	57.30	4.47
	CRAFT	2.34	18.70	32.93	4.22	3.03	21.70	37.05	2.13	3.89	27.55	48.16	8.37	2.45	20.20	38.15	2.69
TONE	HCML	14.32	53.16	69.17	16.16	-	-	-	-	20.82	68.86	84.46	26.38	-	-	-	-
Two-stream		11.65	47.99	65.50	12.85	16.33	58.35	74.46	8.03	15.60	61.18	81.02	21.49	22.49	72.22	88.61	13.92
One-stream		12.04	49.68	66.74	13.67	16.26	58.14	75.05	8.59	16.94	63.55	82.10	22.95	22.62	71.74	87.82	15.04
Zero-Padding		14.80	54.12	71.33	15.95	19.13	61.40	78.41	10.89	20.58	68.38	85.79	26.92	24.43	75.86	91.32	18.64
BDTR		27.32	66.96	81.07	27.32	-	-	-	-	31.92	77.18	89.28	41.86	-	-	-	-
cmGAN		26.97	67.51	80.56	27.80	31.49	72.74	85.01	22.27	31.63	77.23	89.18	42.19	37.00	80.94	92.11	32.76
D^2RL		28.90	70.60	82.40	29.20	-	-	-	-	-	-	-	-	-	-	-	-
eBDTR		27.82	67.34	81.34	28.42	-	-	-	-	32.46	77.42	89.62	42.46	-	-	-	-
MSR		37.35	83.40	93.34	38.11	43.86	86.94	95.68	30.48	39.64	89.29	97.66	50.88	46.56	93.57	98.80	40.08
AlignGAN		42.4	85.0	93.7	40.7	**51.5**	**89.4**	95.7	33.9	45.9	87.6	94.4	54.3	**57.1**	92.7	97.4	45.3
MSPAC-MeCen		**46.62**	**87.59**	**95.77**	**47.26**	47.57	87.64	**96.11**	**38.53**	**51.63**	**93.48**	**98.82**	**61.54**	52.81	**94.16**	**99.37**	**47.09**

Experiments

Comparison with State-of-the-Art Methods on RegDB dataset

Methods	Evaluation Metrics			
	r=1	r=10	r=20	mAP
LOMO	0.85	2.47	4.10	2.28
HOG	13.49	33.22	43.66	10.31
Two-stream	12.43	30.36	40.96	13.42
One-stream	13.11	32.98	42.51	14.02
Zero-Padding	17.75	34.21	44.35	18.90
TONE+HCML	24.44	47.53	56.78	20.80
BDTR	33.47	58.42	67.52	31.83
eBDTR	34.62	58.96	68.72	33.46
MSR	48.43	70.32	79.95	48.67
AlignGAN	57.9	-	-	53.6
MSPAC-MeCen	**49.61**	**72.28**	**80.63**	**53.64**

Experiments

Ablation Studies on SYSU-MM01 dataset

1. Effectiveness of Pyramid Part-aware Attention Mechanism

Methods	Evaluation Metrics				
	r=1	r=5	r=10	r=20	mAP
Baseline+gid	27.14	58.77	72.60	84.35	29.22
MSPAC+gid	38.86	68.16	79.78	89.93	40.69
Baseline+pid	37.97	73.28	84.33	92.30	41.08
MSPAC+pid	**41.41**	**70.87**	**84.09**	**93.66**	**41.98**

2. Effectiveness of Hierarchical Part Aggregation Architecture

Methods		Evaluation Metrics			
Strategy	Part number	r=1	r=10	r=20	mAP
Normal	{1}	40.94	84.22	92.74	44.86
	{3}	40.52	84.28	93.61	42.84
	{6}	42.89	84.91	93.29	44.31
Hierarchical	{1,3}	45.31	84.70	92.43	44.59
	{3,6}	33.63	81.28	92.32	36.65
	{1,3,6}	46.62	87.59	95.77	47.26

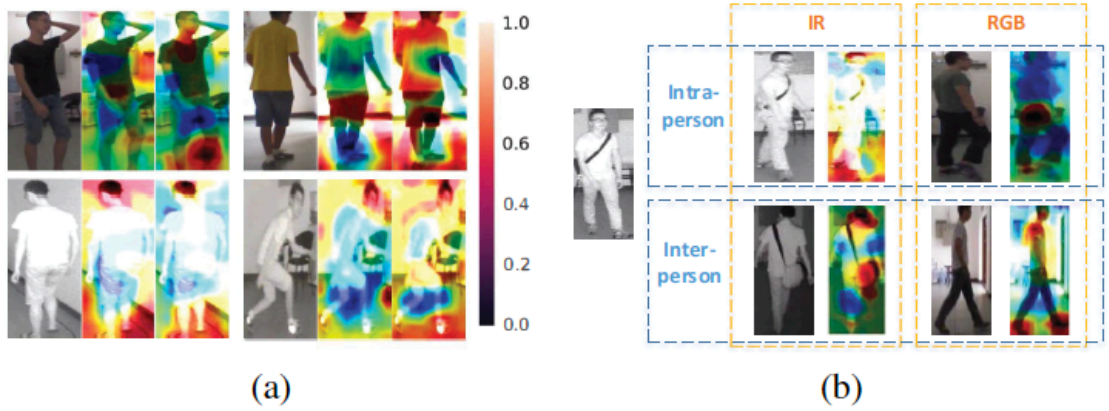
Experiments

Ablation Studies on SYSU-MM01 dataset

3. Effectiveness of Marginal Exponential Center Loss

2*Methods	Evaluation Metrics				
	r=1	r=5	r=10	r=20	mAP
MSPAC+gid	38.86	68.16	79.78	89.93	40.69
+center	40.36	72.44	83.09	91.22	41.40
+margin	42.44	72.36	82.80	90.98	43.19
+exp	**46.62**	**77.20**	**87.59**	**95.77**	**47.26**

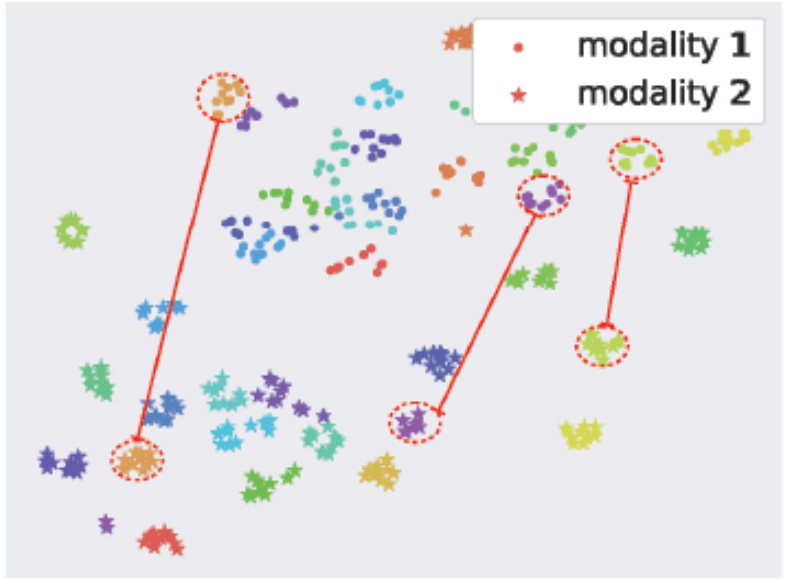
4. Visualization Results: attention score in heat maps



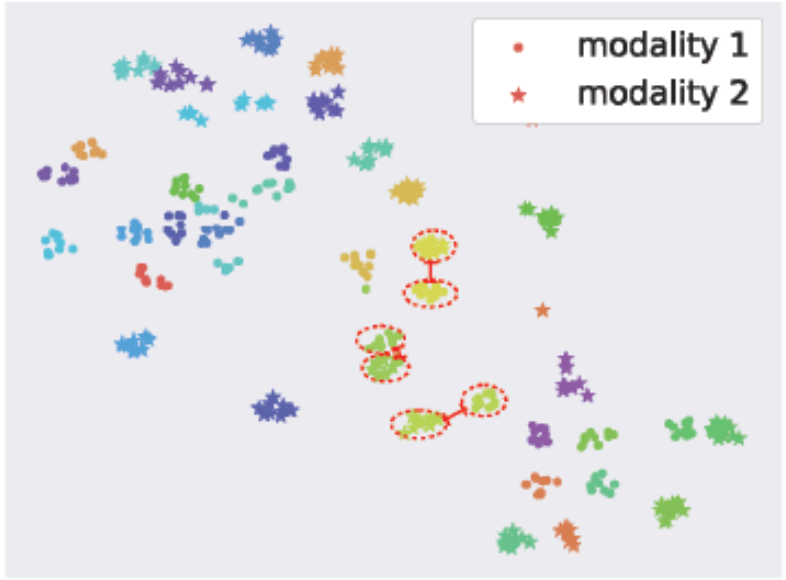
Attention maps in middle and right column are results of Baseline model and ours model respectively

Experiments

Visualization of clustering performance of MeCen in comparison with the BDTR baseline on the SYSU-MM01 dataset



BDTR



Ours

Conclusions

- A multi-scale part aware mechanism in both channel and spatial dimension.
- A hierarchical part aggregation architecture in a cascading fashion
- A novel MeCen loss to model cross-modality correlations