Correlation Discovery for Multi-view and Multi-label Learning

Lichen Wang

Department of Electrical & Computer Engineering
College of Engineering
Northeastern University

About Me

Lichen Wang

- Fifth year PhD Candidate in Northeastern University
- Supervisor: Professor Yun Raymond Fu

Education:

- Ph.D., Computer Engineer, Northeastern University, USA, 2016 - present
- M.S., Electronic and Information, Xi'an Jiaotong University, China, 2013 – 2016
- B.S., Control Engineering, Harbin Institute of Technology, China, 2009 – 2013

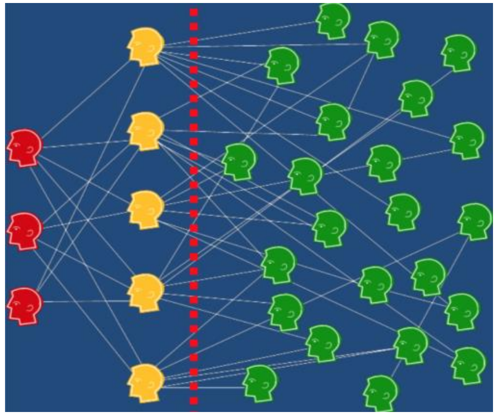
Research Topics:

- Correlation Discovery
  - Multi-label learning and multi-view learning
  - Semi-supervised learning
  - Human action recognition
  - Graph representation learning

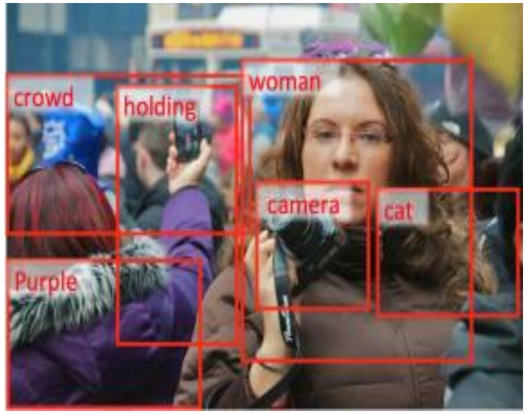
What is correlation?

Interactions/connections across different instances

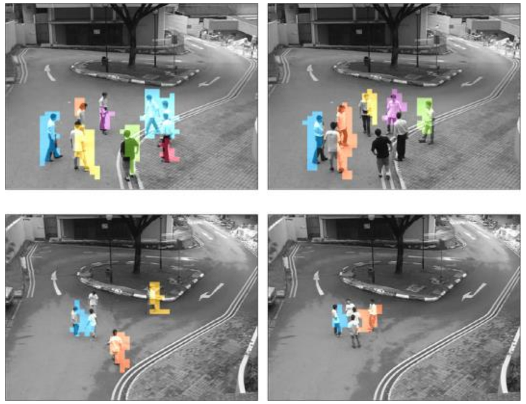
- Social Network
  - Friends connections
  - Like/unlike comments
  - Fake account
- Human action/interaction
  - Interactions of different objects
  - Intension prediction
- Time-series data
  - Latent correlations in time space
- Scene understanding
  - Relations of different objects



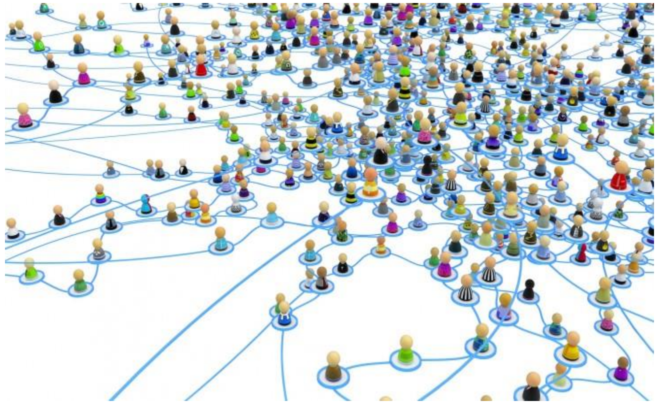
Bank account transaction



Image/scene understanding



Human interactions



Social Network

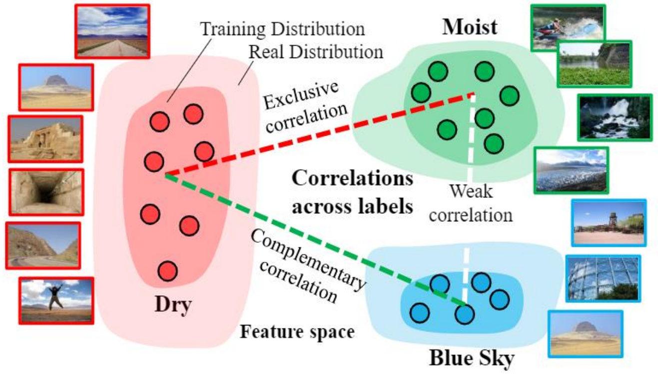
[1] Fang, Hao, et al. "From captions to visual concepts and back." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
[2] https://thenextweb.com/socialmedia/2013/11/24/facebook-grandparents-need-next-gen-social-network/

Why correlation is important?

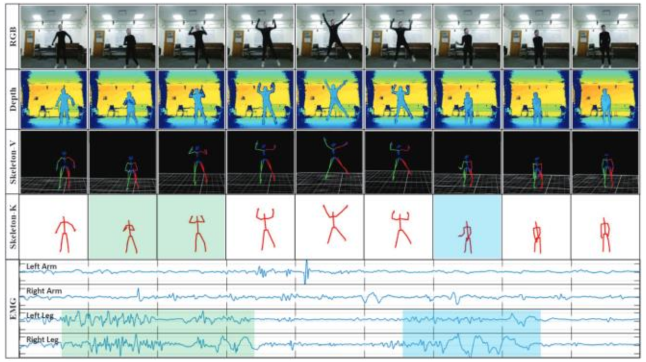
Correlation exists in a wide-range of real-world tasks

- Multi-view learning
- Multi-label learning
- Image/scene understanding
- Image captioning
- Time-series/action recognition

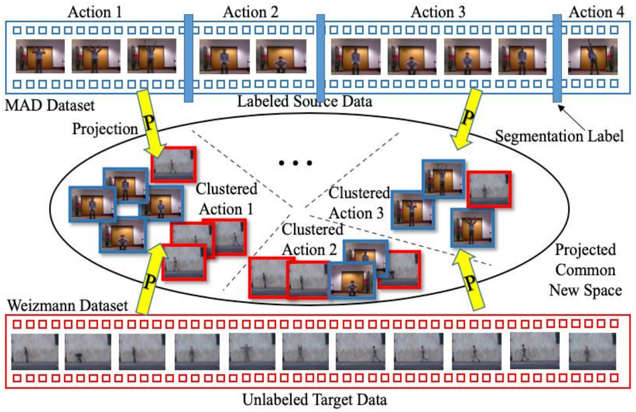
Correlation provides a unique and comprehensive view across instances



Multi-label Learning [1]



Multi-view Learning [2]



Time series data analysis [3]

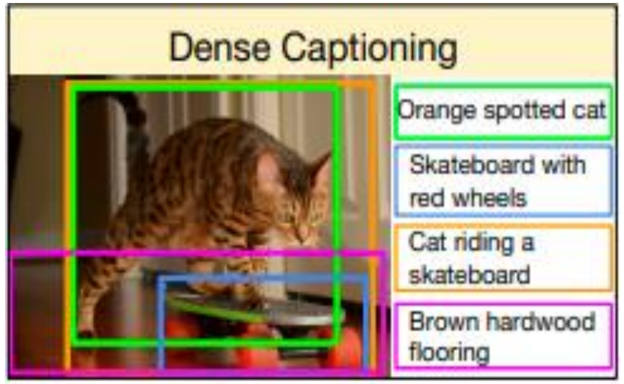


Image Captioning

[1] Wang, Lichen, et al. "Generative correlation discovery network for multi-label learning.“ ICDM 2019
[2] Wang, Lichen, et al. "EV-Action: Electromyography-Vision Multi-Modal Action Dataset." arXiv preprint arXiv:1904.12602 (2019)
[3] Wang, Lichen, et al. "Learning transferable subspace for human motion segmentation." AAAI 2018.

Challenges

Correlations are hard to define

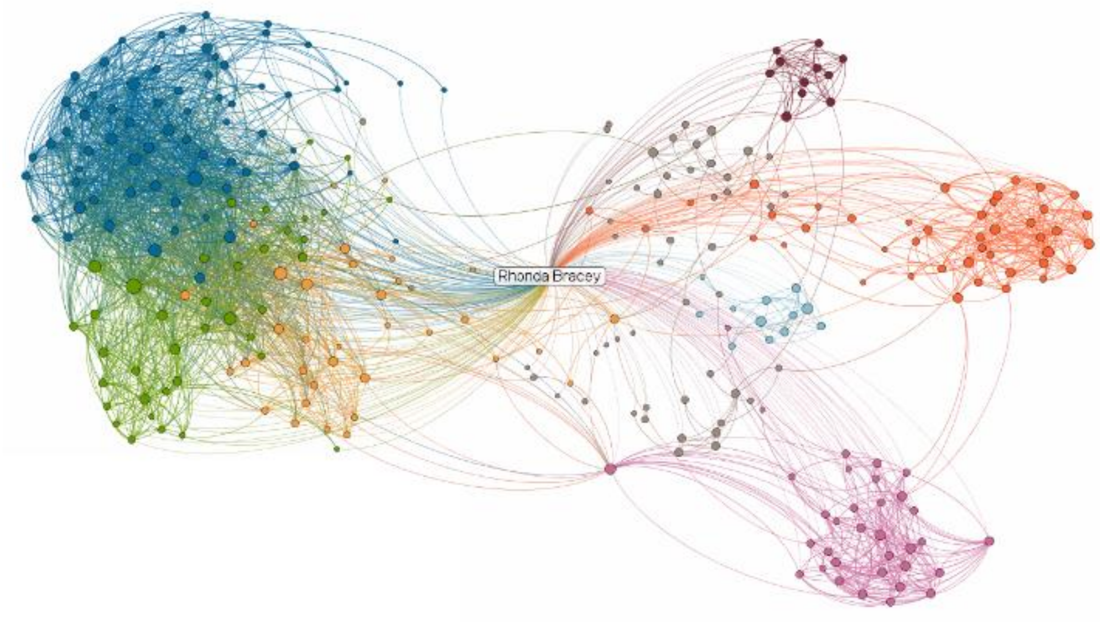
- Instance-instance correlations
- Label-label correlations (e.g., *wet* = *moist?*)
- View-correlations (e.g., RGB & depth)
- Visual-semantic correlation

No sufficient training samples

- Correlations are task specific
- Correlations are subjective and hard to define
- Difficult to obtain consistent supervision label

How to efficiently utilize the correlations?

- Sensitive to parameters and data characteristic



Correlation is hard to define and utilize

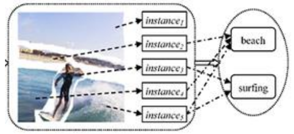
Research works

Topic

Correlation Discovery

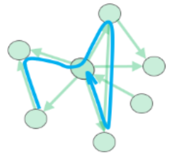
Directions

Multi-label Learning



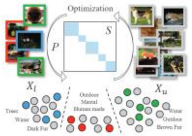
Multi-view Learning

Graph Learning

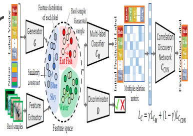


Applications

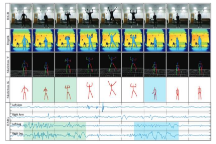
Adaptive Graph Guided Embedding, IJCAI, TIP



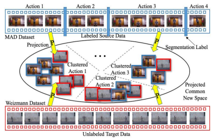
Generative Correlation Discovery Network, ICDM, AAAI, ICCV



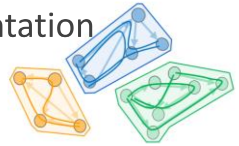
Multi-view human action recognition, ICCV, FG



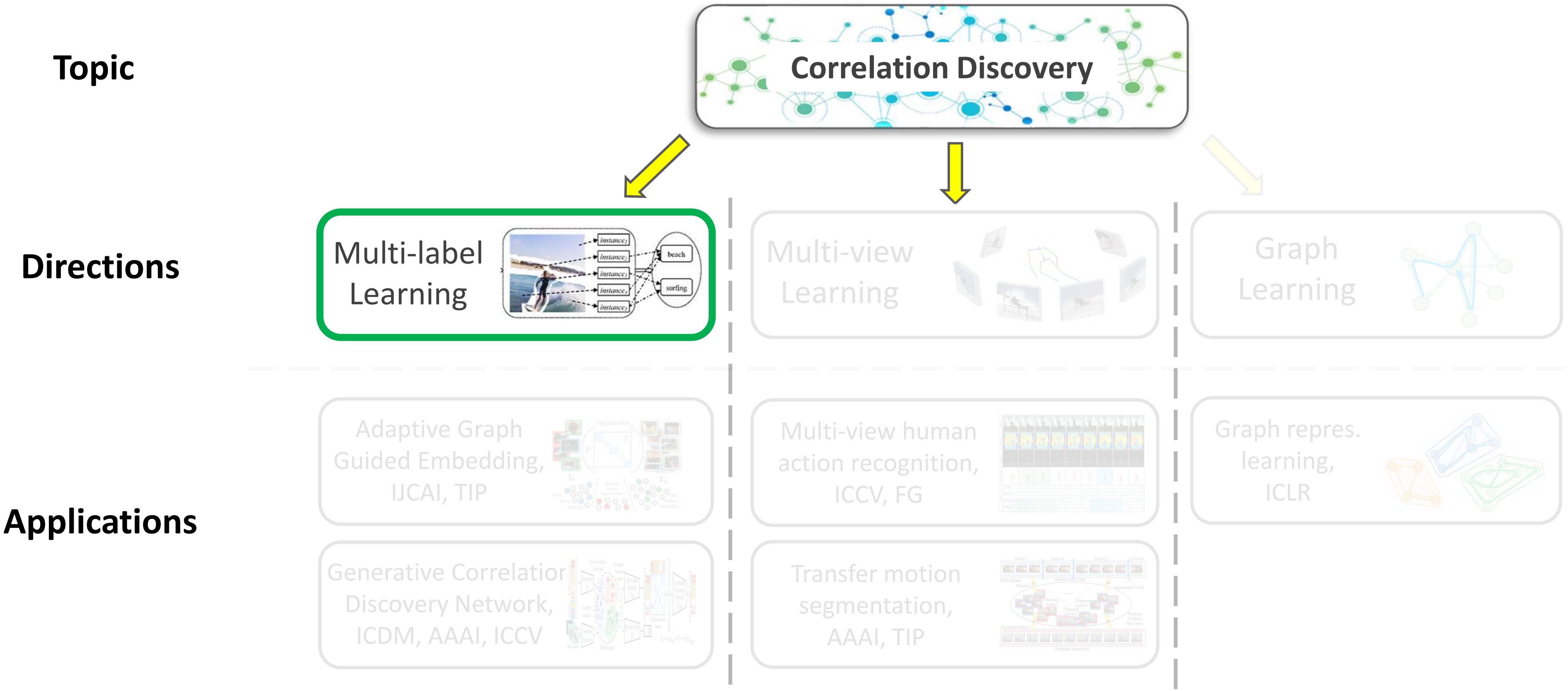
Transfer motion segmentation, AAAI, TIP



Graph representation learning, ICLR



Research works



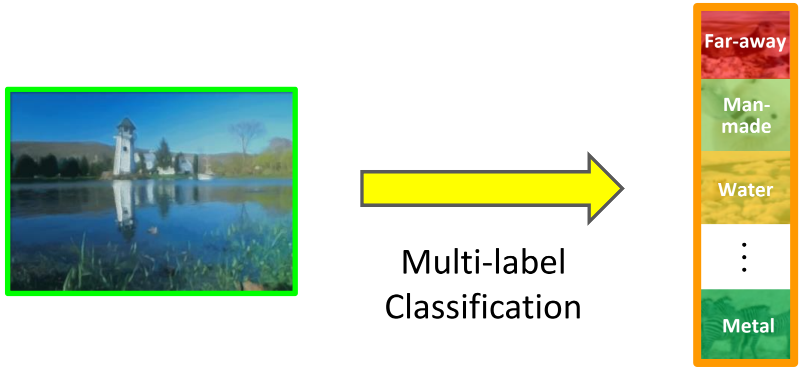
Multi-label Learning

Motivation

- One object can be described by tens or hundreds of labels. Multi-label learning corresponds to seek a mapping from the feature space to the label space.

Setting

- Input: a single instance
- Output: multiple label prediction



Input:
single instance

Output:
multiple label prediction

Multi-label learning seeks a mapping from the feature space to the label space.

[1] Patterson, Genevieve, et al. "The SUN attribute database: Beyond categories for deeper scene understanding." International Journal of Computer Vision 108.1-2 (2014): 59-81.

Multi-label Learning

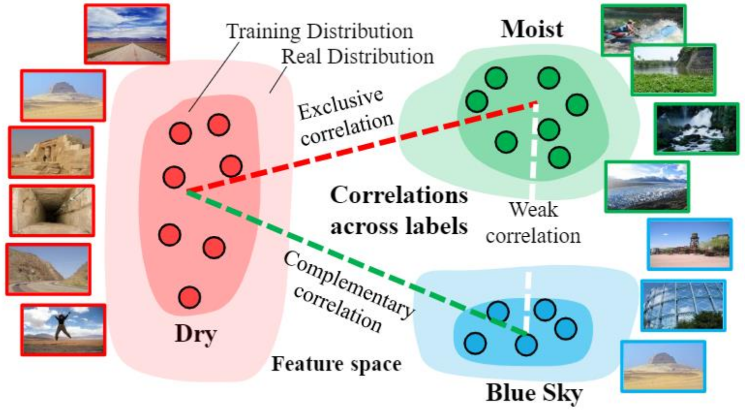
Challenges

- Long-tail label distribution
  - Some labels are extremely common (e.g., *man-made* and *outdoor light*)
  - Some labels are very rare (e.g., *fair* and *fighting*)
- Subjective Label candidates
  - Inconsistent labeling results
  - High-level label noise
- Complicated label correlations
  - e.g., *Dry-Moist, Dry-Blue Sky*

Label	Number
Man-made	8,089
Fire	73
(Total)	106,012

Long-tail label distribution in SUN [1] dataset.

Subjective labels are hard to obtain consistent label results



Complicated and latent label correlations

[1] Patterson, Genevieve, et al. "The SUN attribute database: Beyond categories for deeper scene understanding." International Journal of Computer Vision 108.1-2 (2014): 59-81.

Multi-label Learning

Related methods:

- Attention-based methods

  [1] Huynh, Dat, and Ehsan Elhamifar. "A Shared Multi-Attention Framework for Multi-Label Zero-Shot Learning." CVPR’20.

  [2] Guo, Hao, et al. "Visual attention consistency under image transforms for multi-label image classification." CVPR’19.

- Label-image or label-label correlations

  [1] Huynh, Dat, and Ehsan Elhamifar. "Interactive Multi-Label CNN Learning with Partial Labels.“ CVPR’20

  [2] Zhang, Min-Ling, and Kun Zhang. "Multi-label learning by exploiting label dependency." KDD’10.

- Semi-supervised scenario

  [1] Zhan, Wang, and Min-Ling Zhang. "Inductive semi-supervised multi-label learning with co-training.“ KDD’17.

  [2] Tan, Qiaoyu, et al. "Semi-supervised multi-label classification using incomplete label information." Neurocomputing’17.

  [3] Guo, Baolin, et al. "Semi-supervised multi-label dimensionality reduction." ICDM’16.

My research works

Topic

Correlation Discovery

Directions

Multi-label Learning

Multi-view Learning

Graph Learning

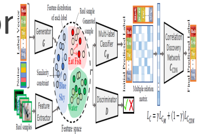
Applications

Adaptive Graph Guided Embedding, IJCAI, TIP

Multi-view human action recognition, ICCV, FG

Graph repres. learning, ICLR

Generative Correlation Discovery Network, ICDM, AAAI, ICCV

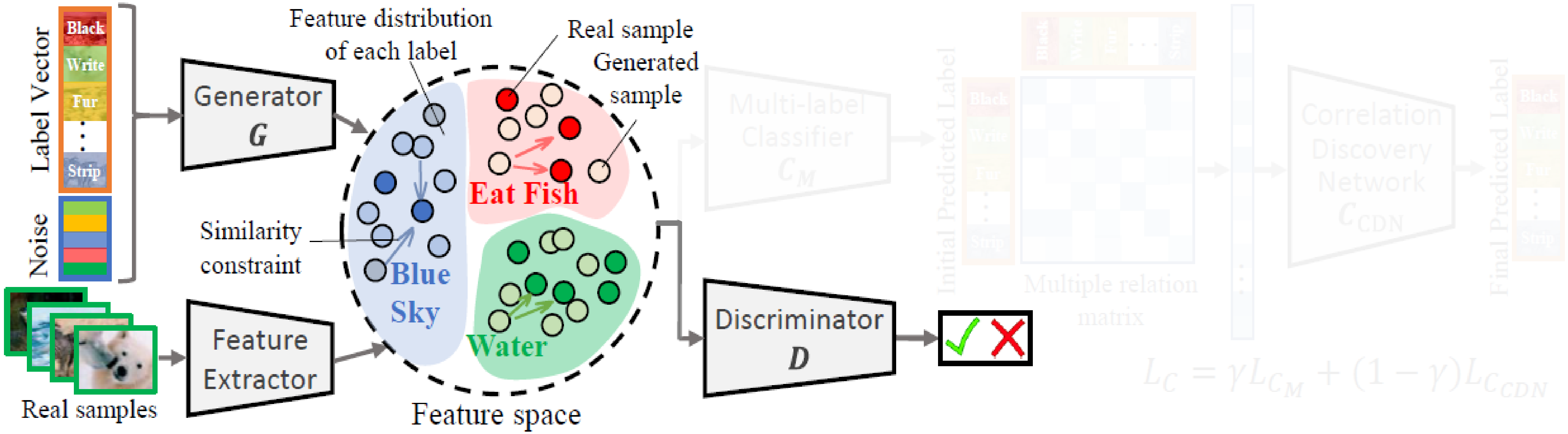


Transfer motion segmentation, AAAI, TIP

Generative Correlation Discovery Network

Generative Module

- Generate and diversify the training samples



Framework of our correlation discovery network for multi-label prediction

[1] Wang, Lichen, et al. "Generative correlation discovery network for multi-label learning." ICDM 2019

Generative Correlation Discovery Network

Correlation Discovery Network

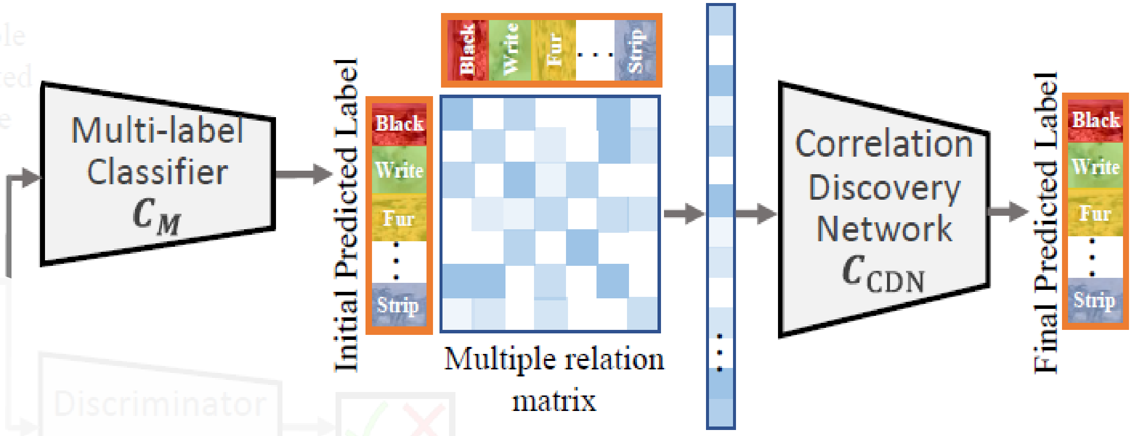
- $C_M(.)$ obtains initial (low-accurate) results first, then $C_{CDN}(.)$ further utilizes the available prediction to "*tune*" the result to high-accurate..

$$L_{C_M} = \mu\|Y - C_M(X)\|_F^2$$

$$L_{C_{CDN}} = \sum_{i=1}^{n_l} \|y_i - C_{CDN}(C_M(x_i)C_M(x_i)^\top)\|_2^2$$

- We balances the update processing between $C_M(.)$ and $C_{CDN}(.)$ to further help each other in the training stage and achieve the promising performance at last.

$$L_C = \gamma L_{C_M} + (1-\gamma)L_{C_{CDN}}$$



$$L_C = \gamma L_{C_M} + (1-\gamma)L_{C_{CDN}}$$

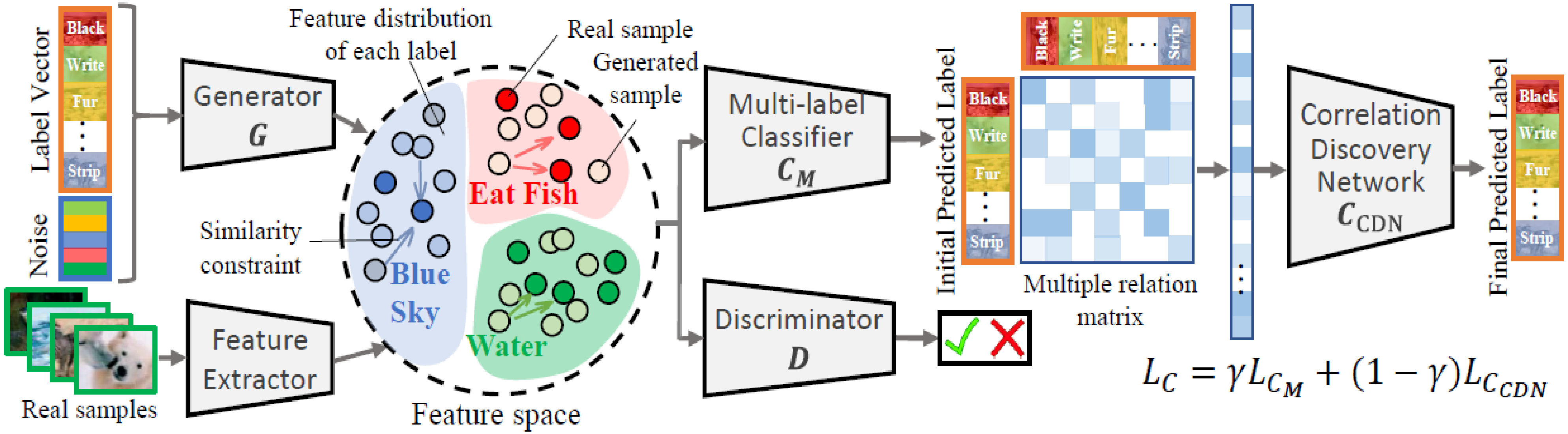
Framework of our correlation discovery network for multi-label prediction

[1] Wang, Lichen, et al. "Generative correlation discovery network for multi-label learning." ICDM 2019

Generative Correlation Discovery Network

Summary

- Generative model: generate and diversify the training samples.
- Correlation Discovery Network automatically learns the latent label correlation across different labels.
- All the networks are trained simultaneously to achieve the best performance.



Framework of our correlation discovery network for multi-label prediction

[1] Wang, Lichen, et al. "Generative correlation discovery network for multi-label learning." ICDM 2019

Experiments (1)

Setting:

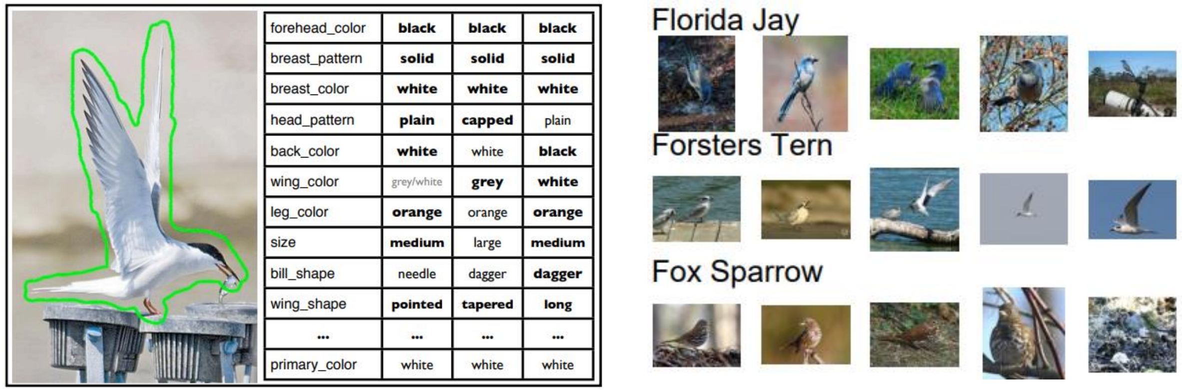
Conventional MLL, Zero-shot MLL

Image annotation, image retrieval

Datasets:

Six fine-grained image datasets:

- Corel5K Dataset [1]
- ESP Game Dataset [2]
- IAPRTC-12 Dataset [3]
- SUN Dataset [4]
- CUB Dataset [5]
- AWA Dataset [6]



forehead_color	black	black	black
breast_pattern	solid	solid	solid
breast_color	white	white	white
head_pattern	plain	capped	plain
back_color	white	white	black
wing_color	grey/white	grey	white
leg_color	orange	orange	orange
size	medium	large	medium
bill_shape	needle	dagger	dagger
wing_shape	pointed	tapered	long
...	...	...	...
primary_color	white	white	white

Samples of CUB dataset

Open area
Natural light
Trees
Grass
Foliage
Sunbathing
Vacationing
Leaves

Far-away Horizon
Man-made
Sailing
Open-area
Swimming
Still water
Concrete

Samples of SUN dataset

otter

black: yes
white: no
brown: yes
stripes: no
water: yes
eats fish: yes

polar bear

black: no
white: yes
brown: no
stripes: no
water: yes
eats fish: yes

Samples of AWA dataset

[1] "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary." ECCV, 2002. [2] "Labeling images with a computer game." SIGCHI 2004. [3] Grubinger, Michael, et al. "The iapr tc-12 benchmark: A new evaluation resource for visual information systems." OntoImage. 2006. [4] "The SUN attribute database: Beyond categories for deeper scene understanding." IJCV 2014. Wah, [5] "The caltech-ucsd birds-200-2011 dataset." 2011. [6] "Learning To Detect Unseen Object Classes by Between-Class Attribute Transfer". CVPR, 2009

Experiments (2)

Multi-label prediction performance:

- Conventional setting.
- Five metrics
  - Precision
  - Recall
  - F1
  - Non-zero recall
  - Mean average precision
  - Five metrics
- Our approach significantly outperform other baselines

Data	Method	Pre	Rec	F1	N-R	mAP
Corel	LR	0.2859	0.3211	0.3025	128	0.3630
	SSMLDR	0.2741	0.3366	0.3022	143	0.3410
	FastTag	0.3123	0.3657	0.3369	143	0.3871
	ML-PGD	0.2575	0.2911	0.2732	122	0.3727
	SAE	0.2962	0.3442	0.3184	141	0.3823
	AG2E	0.3011	0.3520	0.3245	**157**	0.3568
	Ours	**0.3335**	**0.3714**	**0.3514**	148	**0.4417**
ESP	LR	0.3793	0.2038	0.2653	215	0.3440
	SSMLDR	0.3298	0.1885	0.2399	226	0.3156
	FastTag	0.4011	0.1927	0.2617	208	0.3904
	ML-PGD	0.3239	0.2012	0.2482	210	0.4077
	SAE	0.3861	0.1743	0.2402	194	0.3842
	AG2E	0.3548	0.1525	0.2133	213	0.3730
	Ours	**0.4032**	**0.2178**	**0.2828**	**239**	**0.4327**
IAP	LR	0.4287	0.2041	0.2765	199	0.4211
	SSMLDR	0.3491	0.2520	0.2927	229	0.3981
	FastTag	0.4346	0.2267	0.2980	227	0.4596
	ML-PGD	0.4132	0.2441	0.3011	230	0.4674
	SAE	0.3537	0.2282	0.2774	213	0.4309
	AG2E	0.3829	0.2330	0.2897	229	0.4353
	Ours	**0.4732**	**0.2648**	**0.3396**	**237**	**0.5295**

Data	Method	Pre	Rec	F1	N-R	mAP
SUN	LR	0.6209	0.1473	0.2457	102	0.6807
	SSMLDR	0.6879	0.1700	0.2726	102	0.6723
	FastTag	0.6816	0.1473	0.2457	102	0.6914
	ML-PGD	0.7110	0.1614	0.2631	101	0.7087
	SAE	0.7183	0.1638	0.2668	98	0.7012
	AG2E	0.7685	0.1765	0.2871	99	0.6778
	Ours	**0.7985**	**0.1835**	**0.2985**	**102**	**0.7093**
CUB	LR	0.2010	0.0239	0.0428	157	0.0638
	SSMLDR	0.3410	0.0473	0.0832	178	0.2329
	FastTag	0.2147	0.0359	0.0615	167	0.3144
	ML-PGD	0.3334	0.0451	0.0794	155	0.3288
	SAE	0.3383	0.0514	0.0908	196	0.3255
	AG2E	0.3409	0.0531	0.0911	190	0.3106
	Ours	**0.3718**	**0.0541**	**0.0944**	**214**	**0.3561**
AWA	LR	0.8798	0.0821	0.1500	75	0.8626
	SSMLDR	0.7812	0.0858	0.1546	67	0.8346
	FastTag	0.7861	0.0949	0.1694	72	0.8791
	ML-PGD	0.5395	0.0635	0.1136	57	0.9121
	SAE	0.9683	**0.0957**	**0.1742**	73	**0.9397**
	AG2E	0.8483	0.0827	0.1507	73	0.9033
	Ours	**0.9716**	0.0871	0.1599	**83**	0.9291

Multi-label prediction results on six datasets

Experiments (3)

Multi-label prediction performance:

- Augmented multi-label datasets
  - With more labels
- Zero-shot Multi-label Learning
  - No overlapped between training and testing samples (e.g., Horse and Zebra)

Data	Methods	Pre	Rec	F1	N-R	mAP
Corel-A	LR	0.2842	0.2304	0.2545	103	0.3762
	SSMLDR	0.3036	0.2791	0.2908	134	0.3660
	FastTag	0.3329	0.3145	0.3234	136	0.4127
	ML-PGD	0.3245	0.3011	0.3124	140	0.4275
	SAE	0.3168	0.3037	0.3101	128	0.4192
	AG2E	0.3273	0.3172	0.3221	**143**	0.3985
	Ours	**0.3438**	**0.3219**	**0.3325**	138	**0.4773**
ESP-A	LR	0.3848	0.1256	0.1894	178	0.3913
	SSMLDR	0.3253	0.1697	0.2231	202	0.3357
	FastTag	0.3886	0.1531	0.2197	196	0.4254
	ML-PGD	0.3713	0.1184	0.1795	162	0.4211
	SAE	0.3153	0.1425	0.1966	156	0.4050
	AG2E	0.3518	0.1492	0.2095	196	0.4030
	Ours	**0.4772**	**0.1944**	**0.2763**	**225**	**0.4436**

Performance based on augmented datasets

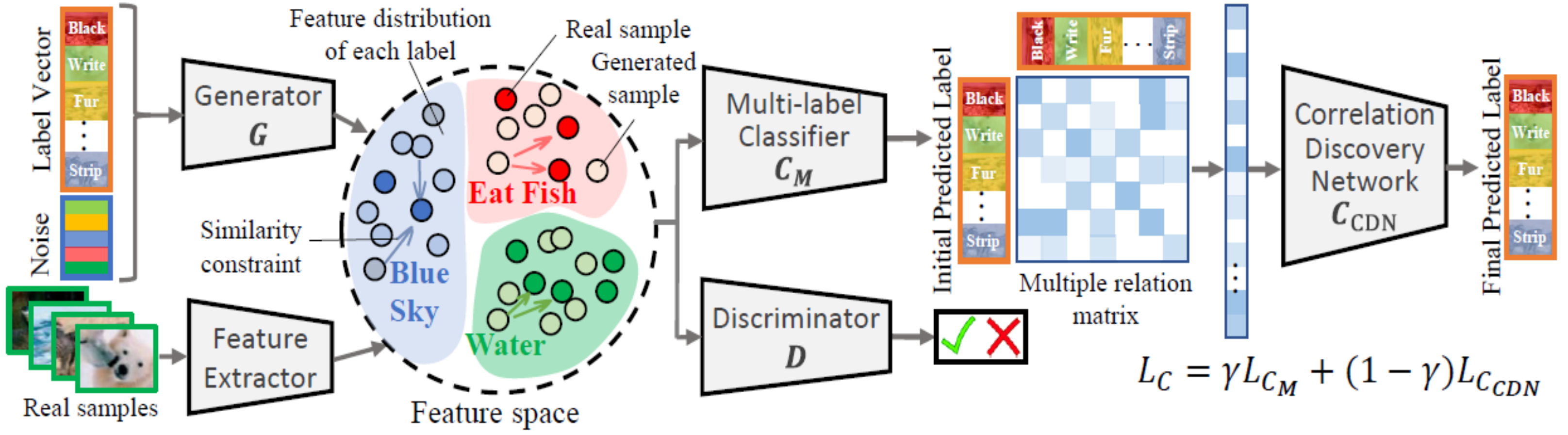
Data	Method	Pre	Rec	F1	N-R	mAP
SUN	LR	0.7047	0.1548	0.2539	97	0.6616
	SSMLDR	0.6637	0.1481	0.2422	95	0.6581
	FastTag	0.6906	0.1522	0.2494	90	0.6706
	ML-PGD	0.7037	0.1471	0.2433	95	0.6829
	SAE	0.6978	0.1710	0.2747	100	0.6513
	AG2E	0.7125	0.1618	0.2637	88	0.6693
	Ours	**0.7531**	**0.1857**	**0.2979**	**101**	**0.6911**
CUB	LR	0.2600	0.0307	0.0549	160	0.2693
	SSMLDR	0.2926	0.0383	0.0677	166	0.2329
	FastTag	0.2231	0.0434	0.0726	143	0.2967
	ML-PGD	0.2392	0.0365	0.0635	117	0.3178
	SAE	0.2552	0.0469	0.0798	167	0.3102
	AG2E	0.2808	0.0481	0.0821	163	0.2693
	Ours	**0.3091**	**0.0488**	**0.0843**	**179**	**0.3264**
AWA	LR	0.7555	0.0766	0.1392	66	0.8809
	SSMLDR	0.7017	0.0764	0.1378	66	0.7858
	FastTag	0.8610	0.0912	0.1649	81	0.8918
	ML-PGD	0.4338	0.0623	0.1091	49	0.8677
	SAE	0.9015	**0.0926**	**0.1679**	78	**0.8918**
	AG2E	0.8247	0.0811	0.1476	71	0.8874
	Ours	**0.9249**	0.0804	0.1480	**83**	0.8784

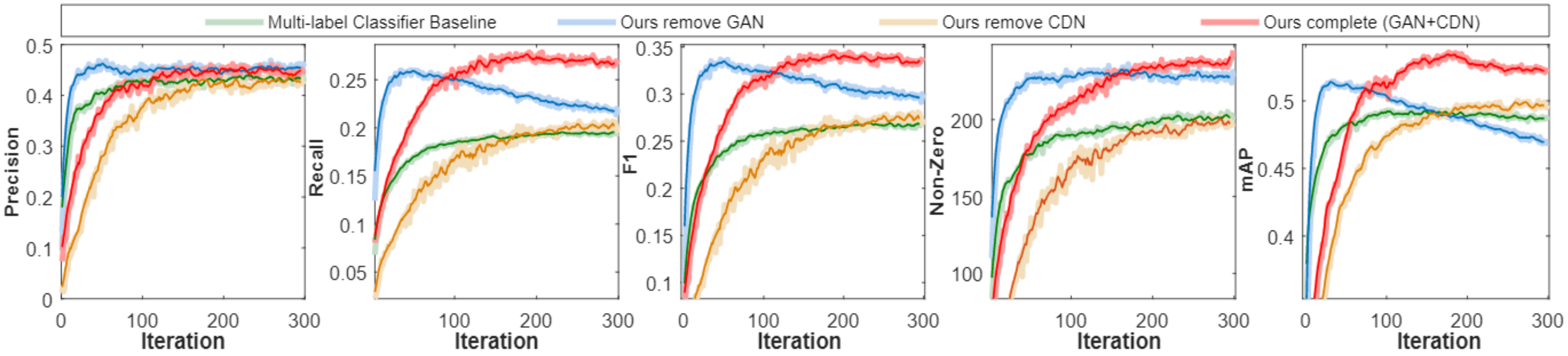
Performance of Zero-shot Learning Multi-label Learning

Experiments (4)

Ablation Study:

- Four modifications:
  - Basic model
  - Only GAN model
  - Only Correlation Discovery Network
  - Our complete model
- Conclusion
  - Each module is effective
  - Their combinations further improve the performance

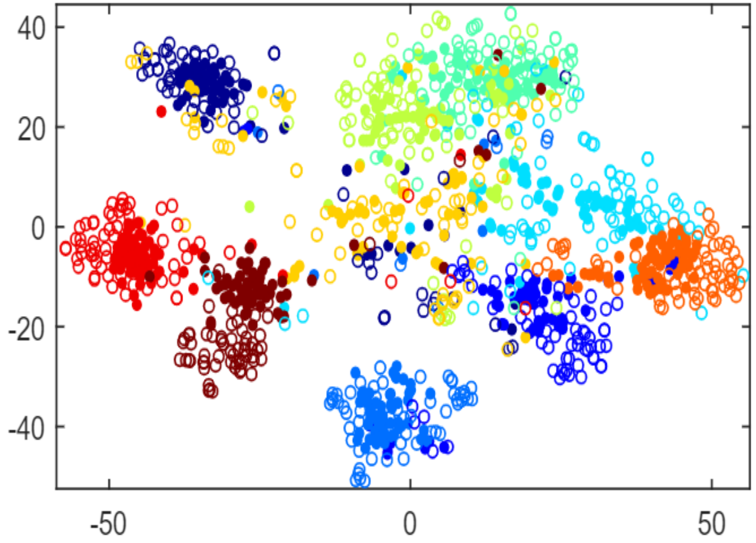




Experiments (5)

Ablation Study:

- Generative model:
  - t-SNE [1] visualization of real and generated samples
  - Performance when noise is deployed for data augmentation
- Conclusion
  - Generated samples are similar compared with real samples. Generative module is effectively in our model
  - Adding noise is not an effective strategy



Real and generated samples in visual feature space

Noise	Pre	Rec	F-1	N-R	mAP
0.00	**0.3718**	**0.0541**	**0.0944**	**214**	**0.3561**
0.05	0.3711	0.0540	0.0941	214	0.3561
0.10	0.3692	0.0538	0.0943	214	0.3537
0.15	0.3668	0.0537	0.0941	214	0.3511
0.20	0.3647	0.0534	0.0938	212	0.3482
0.25	0.3612	0.0533	0.0936	211	0.3467
0.30	0.3591	0.0531	0.0932	209	0.3416
0.35	0.3505	0.0530	0.0930	208	0.3389
0.40	0.3393	0.0529	0.0929	206	0.3351
0.45	0.3314	0.0528	0.0927	204	0.3232
0.50	0.3248	0.0526	0.0926	202	0.3215

Multi-label performance when different level of Gaussian noise is added into the visual feature

[1] SL.J.P. van der Maaten. Accelerating t-SNE using Tree-Based Algorithms. Journal of Machine Learning Research 15(Oct):3221-3245, 2014.

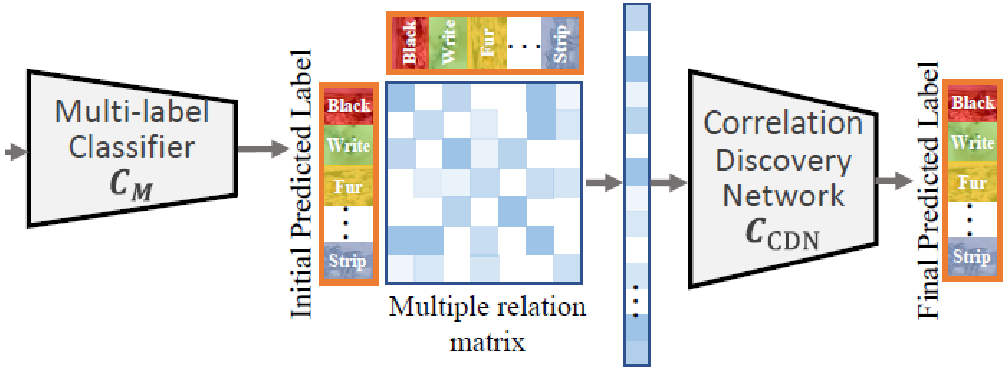
Experiments (6)

Parameter analysis

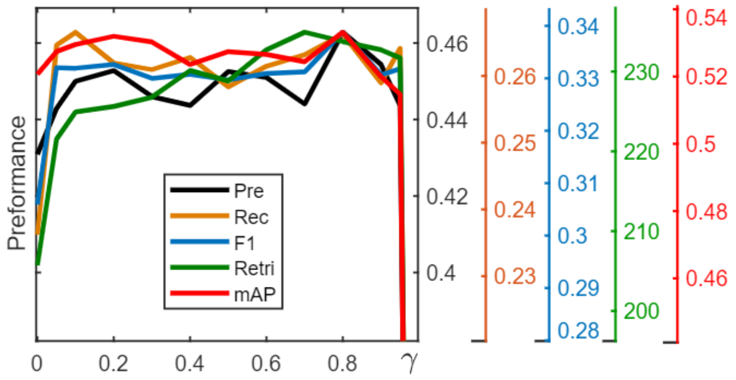
- Trade-off between C_M and C_{CDN}
  - Parameter insensitive

Time consumption

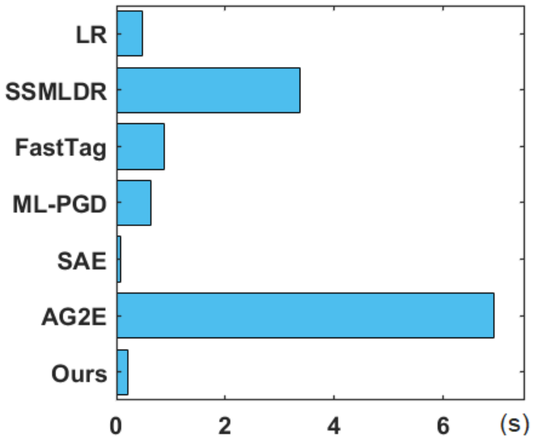
  - Efficient for large-scale applications



$$L_C = \gamma L_{C_M} + (1-\gamma) L_{C_{CDN}}$$



Parameter sensitivity of the proposed model



Time consumption in testing stage

Experiments (7)

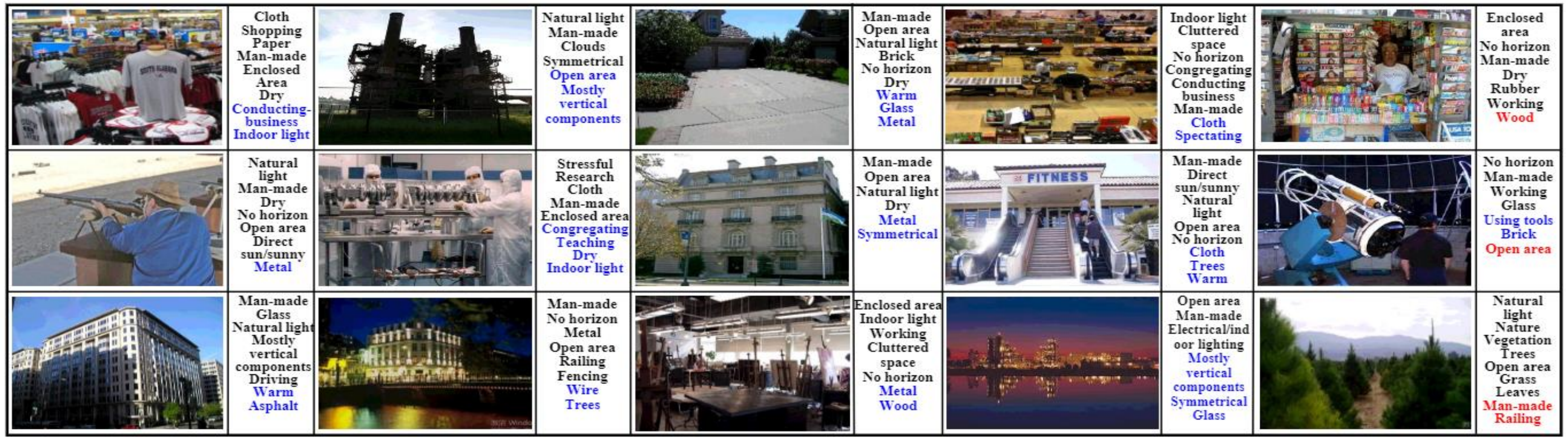
Image annotation

Zero-shot image annotation

- Given an image, predict all the positive labels.
- The image categories are not overlapped in training stage.

Image retrieval

- Given a target label, retrieve all candidate images.



Multi-label image annotation results in SUN dataset

Image retrieval result of SUN dataset. Green and red boxes indicate correct and incorrect retrieval.

[1] Patterson, Genevieve, et al. "The SUN attribute database: Beyond categories for deeper scene understanding." International Journal of Computer Vision 108.1-2 (2014): 59-81.

Related works

Dual Relation Semi-Supervised Multi-Label Learning

Lichen Wang, Yunyu Liu, Can Qin, Gan Sun, and Yun Fu. In *AAAI'*2020.

- Jointly consider feature correlation and label correlation

Low-Rank Transfer Human Motion Segmentation

Lichen Wang, Zhengming Ding, and Yun Fu. *TIP*.

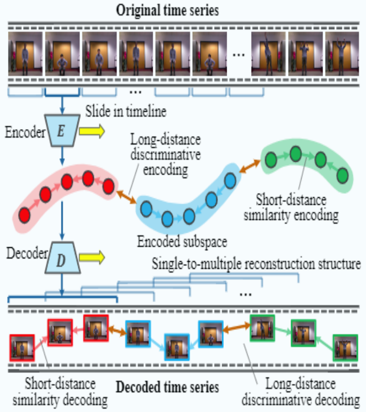
Learning Transferable Subspace for Human Motion Segmentation

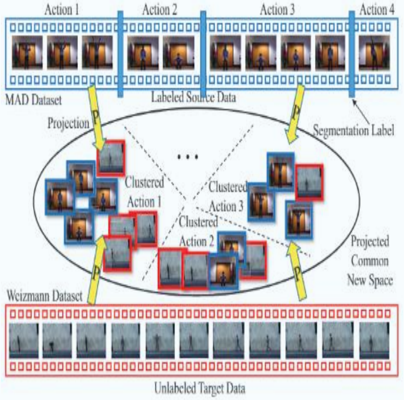
Lichen Wang, Zhengming Ding, and Yun Fu. *In IJCAI'2018*.

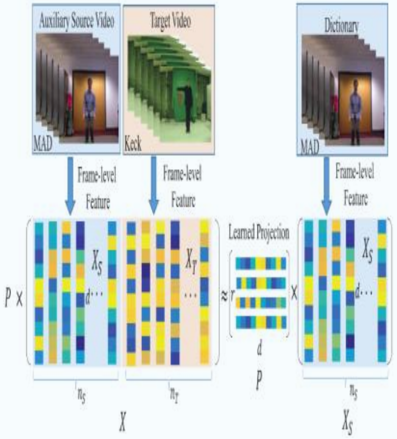
Dual-Side Auto-Encoder for High-Dimensional Time Series Segmentation

Yue Bai, Lichen Wang, Yunyu Liu, Yu Yin, Yun Fu

- Spatial-temporal correlation discovery



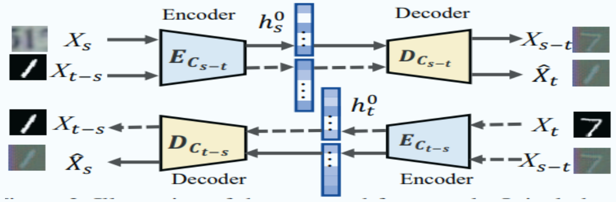




Generatively Inferential Co-Training for Unsupervised Domain Adaptation

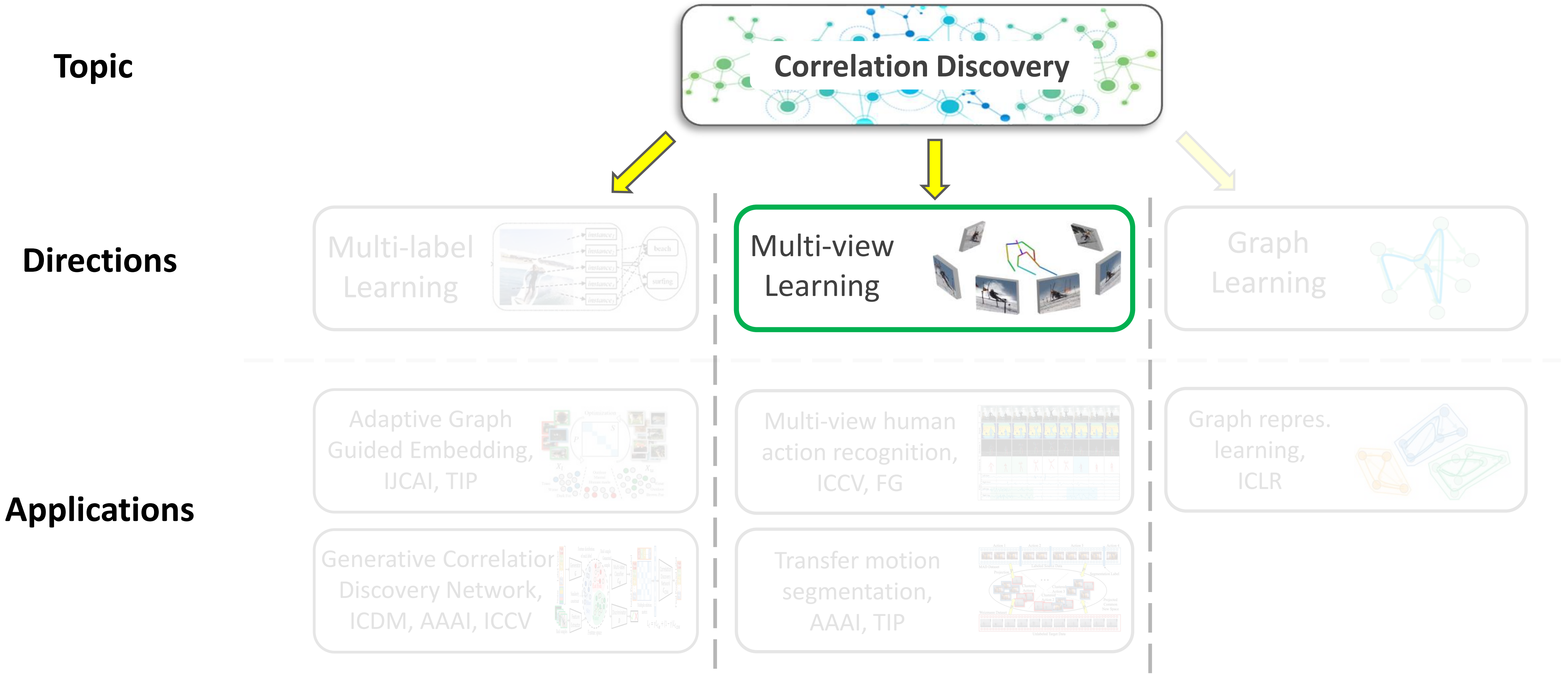
Can Qin, Lichen Wang, Yulun Zhang, Yun Fu. In *AAAI'*2020.

- Explore instance correlations across different domain



[1] Patterson, Genevieve, et al. "The SUN attribute database: Beyond categories for deeper scene understanding." International Journal of Computer Vision 108.1-2 (2014): 59-81.

Research works



Multi-view Action Recognition

Topic

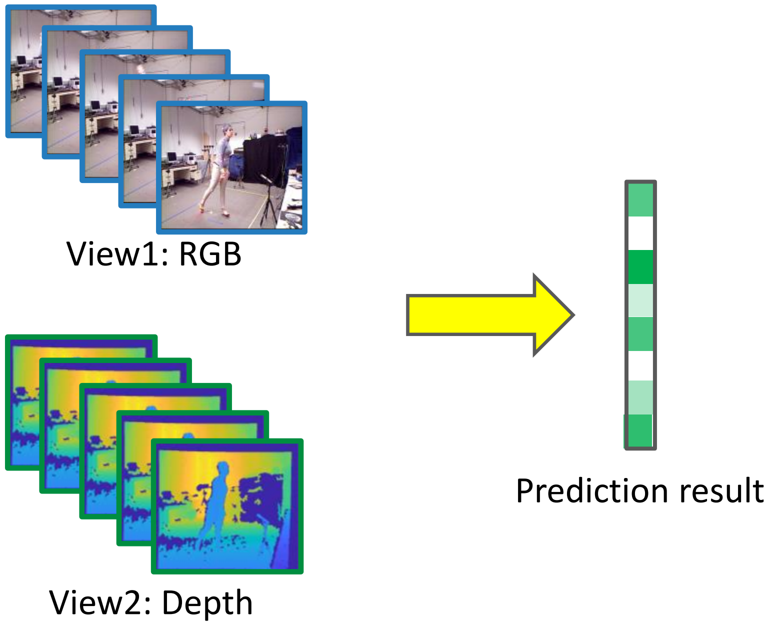
- Multi-view Action Recognition

Setting

- Input: Multi-view action sequences (e.g., RGB + Depth)
- Output: Action prediction

Challenges

- Heterogeneous multi-view feature domains
- Incomplete/missing view sequences
- Inconsistent view-specific predictions



Multi-view action recognition

Research works

Topic

Correlation Discovery

Directions

Multi-label Learning

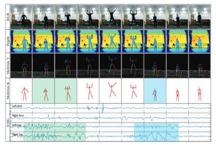
Multi-view Learning

Graph Learning

Applications

Adaptive Graph Guided Embedding, IJCAI, TIP

Multi-view human action recognition, ICCV, FG



Graph repres. learning, ICLR

Generative Correlation Discovery Network, ICDM, AAAI, ICCV

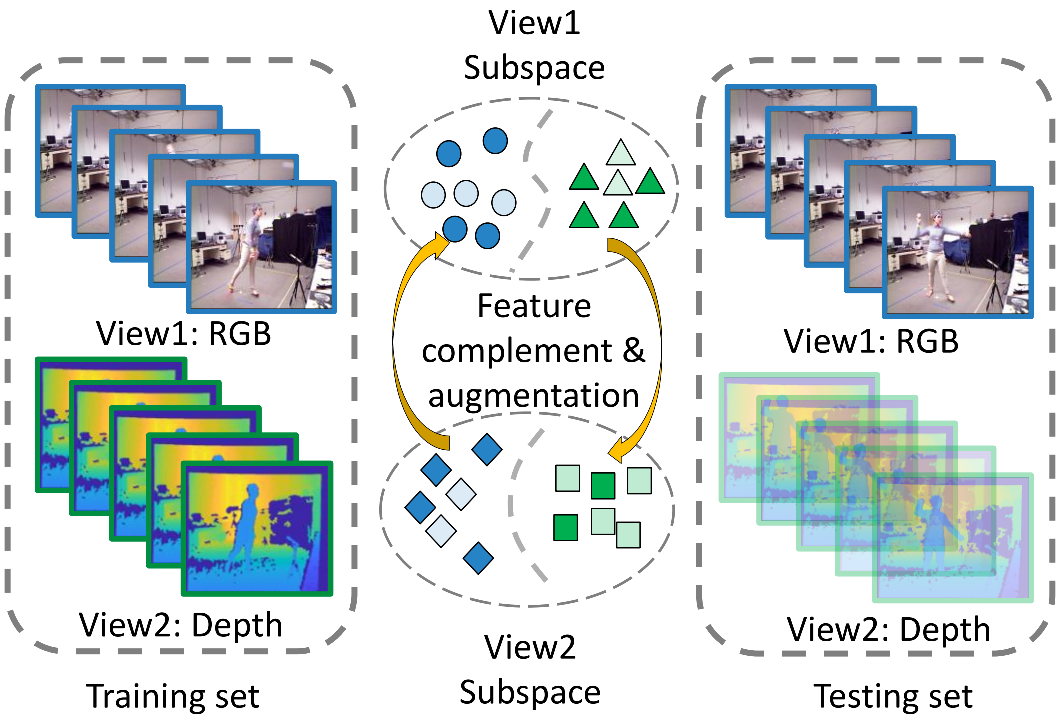
Transfer motion segmentation, AAAI, TIP

Generative Multi-view Action Recognition

Motivation

Three major components to solve the challenges:

1. View-specific Encoders
2. Cross-view Adversarial Generation
3. View Correlation Discovery Network

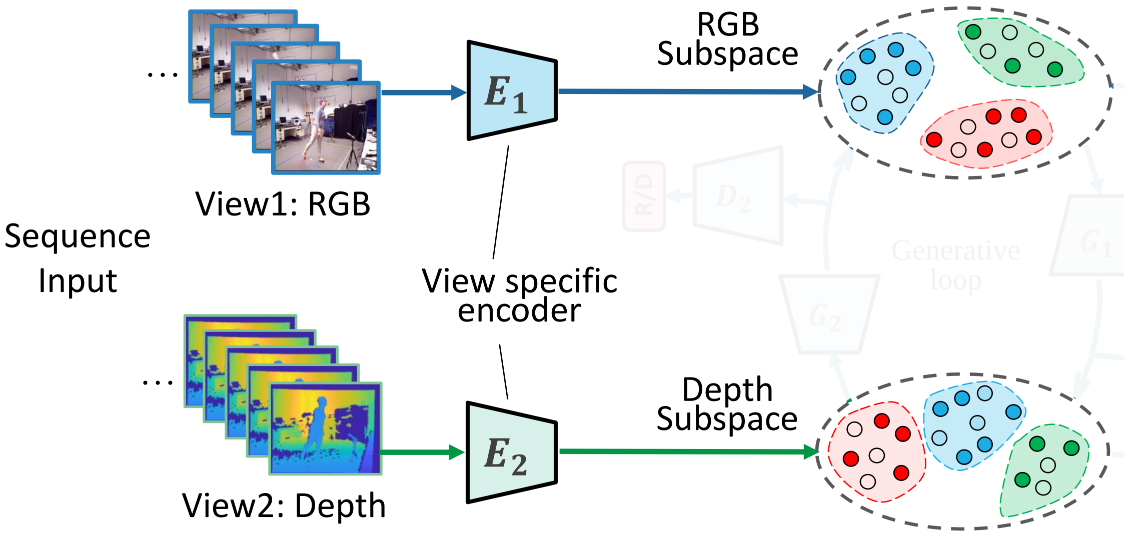


Motivation of our generative multi-view action recognition

1. View-specific Encoders

Mapping original feature to more distinctive subspaces

- Seek distinctive action representations in subspaces
- Label information + triplet loss objective：



View-specific Encoder

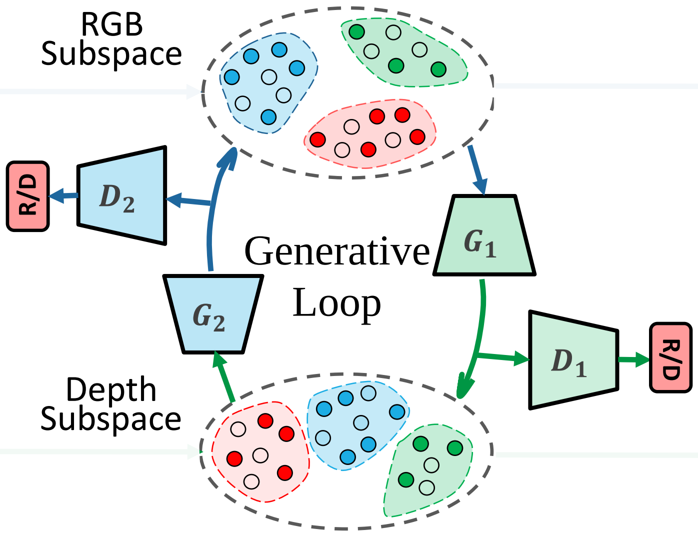
$$L_{E_m} = \sum_{i=1}^{M} \max\Big(\big[\|E_m(X^a_{tr_i}) - E_m(X^p_{tr_i})\|_2^2 - \|E_m(X^a_{tr_i}) - E_m(X^n_{tr_i})\|_2^2 + \alpha\big], 0\Big)$$

2. Cross-view Adversarial Generation

Generate one view conditioning on the other view

- Increase cross-view representation diversity
- Enhance model robustness
- Address missing/incomplete view sequences

$$L_{G_1 d} = -E_{z\sim p_z(z)} \log\Big(1 - D_1\big(G_1(z|E_1(X_{tr}^1))\big)\Big)$$
$$L_{G_1 s} = E_{z\sim p_z(z)}\Big(\|G_1(z|E_1(X_{tr}^1)) - E_2(X_{tr}^2)\|_F^2\Big)$$
$$L_{D_1} = E_{X\sim p_X(X)} \log D_1\big(E_2(X_{tr}^2)\big) + E_{z\sim p_z(z)} \log\Big(1 - D_1\big(G_1(z|E_1(X_{tr}^1))\big)\Big)$$



Cross-view Adversarial Generation

3. View Correlation Discovery Network

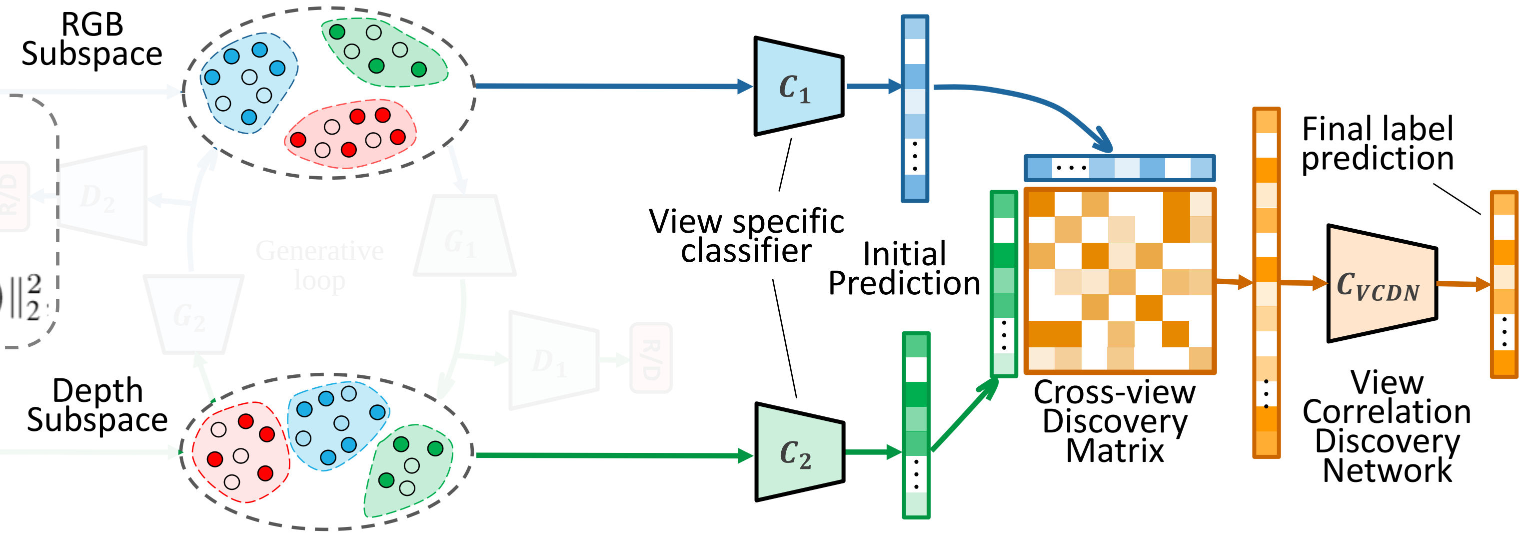
Explore high-level label correlations across different views

- View-specific initial classification is firstly obtained
- Pair-wise label correlation matrix is generated
- VCDN fully explore the latent high-level label correlation for higher performance

$$L_{C_{1_g}} = \|Y_{tr} - C_1(G_2(z|E_2(X^2_{tr}))\|^2_F$$

$$L_{C_{2_g}} = \|Y_{tr} - C_2(G_1(z|E_1(X^1_{tr}))\|^2_F$$

$$L_{C^{rr}_{VCDN}} = \sum_{i=1}^{n_{tr}} \|y_i - C_{VCDN}(y^2_{tr_i} \cdot y^{1\top}_{tr_i})\|^2_2$$

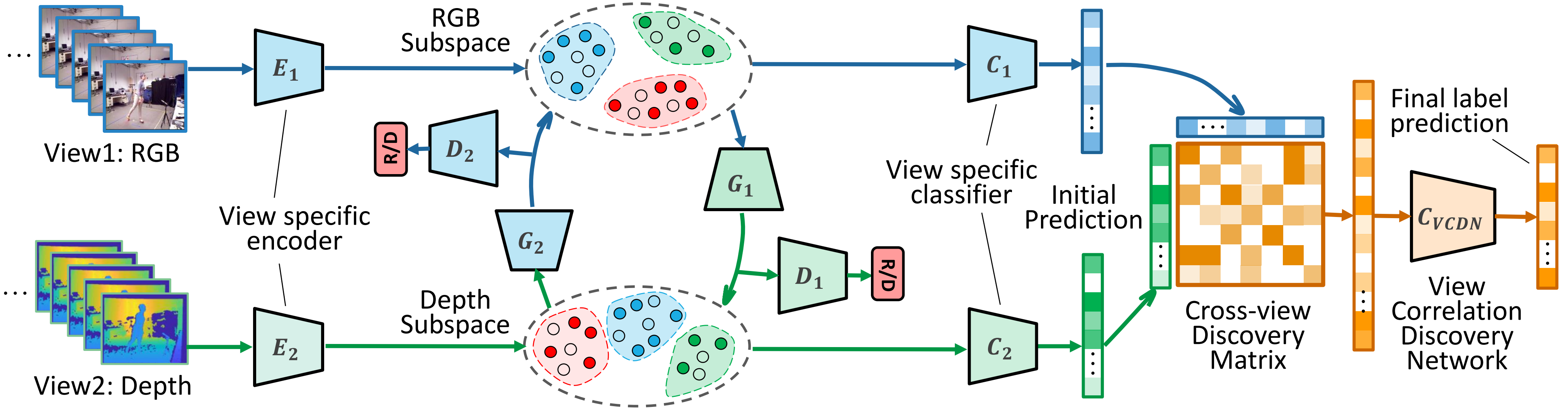


View-correlation discovery network

Our complete model

Summary

- Three components work together
- Jointly trained in end-to-end manner



Framework of Generative Multi-view Action Recognition

Experiments (1)

Action recognition:

- Datasets: UWA[1], MHAD[2], and DHA[3]
- Multi-view action recognition
- Missing/incomplete multi-view (i.e., single-view) action recognition

Method	RGB	R→D	Depth	D→R	R+D
LSR	67.59	69.17	45.45	37.73	68.77
SVM [36]	69.44	68.53	34.92	34.33	72.72
VLAD [14]	71.54	-	-	-	-
TSN [51]	71.01	-	-	-	-
WDMM [1]	-	-	46.58	-	-
AMGL [30]	69.17	71.54	39.92	35.96	68.53
MLAN [29]	67.19	67.19	33.28	33.61	66.64
PM-GANs [49]	-	71.36	-	49.01	-
Ours	-	**73.53**	-	**50.35**	**76.28**

UWA

Method	RGB	R→D	Depth	D→R	R+D
LSR	96.46	97.17	47.63	42.51	97.17
SVM [36]	96.09	96.80	45.39	45.13	96.80
VLAD [14]	97.17	-	-	-	-
TSN [51]	97.31	-	-	-	-
WDMM [1]	-	-	66.41	-	-
AMGL [30]	96.46	97.11	30.03	29.96	94.70
MLAN [29]	96.05	96.10	41.48	41.25	96.46
PM-GANs [49]	-	96.76	-	66.84	-
Ours	-	**98.23**	-	**68.32**	**98.94**

MHAD

Method	RGB	R→D	Depth	D→R	R+D
LSR	65.02	65.43	82.30	48.56	77.36
SVM [36]	66.11	**70.24**	78.92	78.18	83.47
VLAD [14]	67.13	-	-	-	-
TSN [51]	67.85	-	-	-	-
WDMM [1]	-	-	81.05	-	-
AMGL [30]	64.61	59.05	72.84	67.33	74.89
MLAN [29]	67.91	67.91	72.96	72.83	76.13
PM-GANs [49]	-	68.72	-	76.02	-
Ours	-	69.72	-	**83.48**	**88.72**

DHA

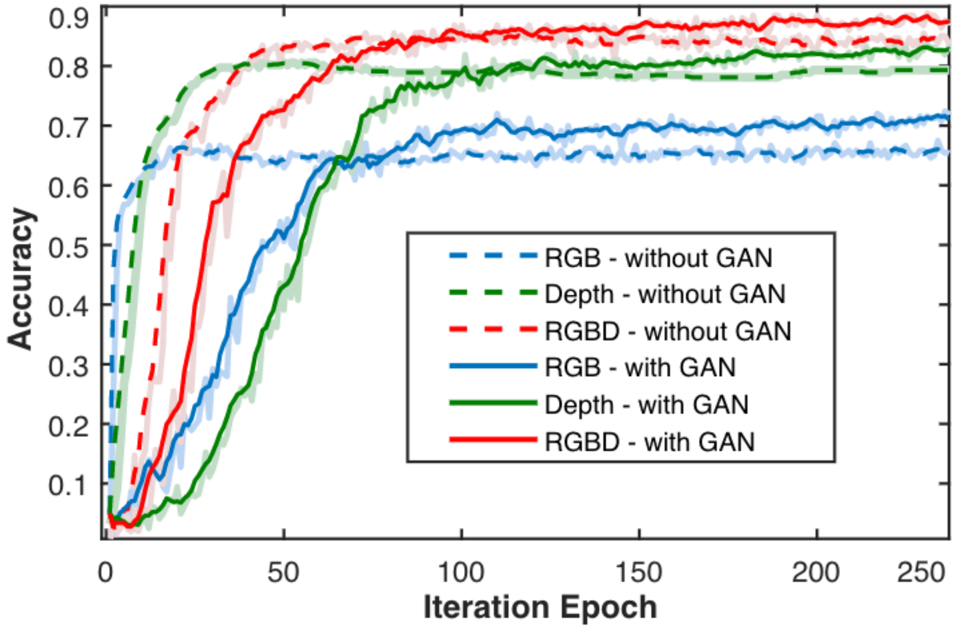
Performance on three multi-view action datasets

[1] Hossein Rahmani, et al. Histogram of oriented principal components for cross-view action recognition.IEEE Trans. PAMI, 38(12):2430–2443, 2016
[2] Ferda Ofli, et al. Berkeley mhad: A comprehensive mul-timodal human action database. In Proc. IEEE WACV, pages53–60, 2013.
[3] Yan-Ching Lin, et al. Human action recog-nition and retrieval using sole depth information. In Proc.ACM MM, pages 1053–1056, 2012.

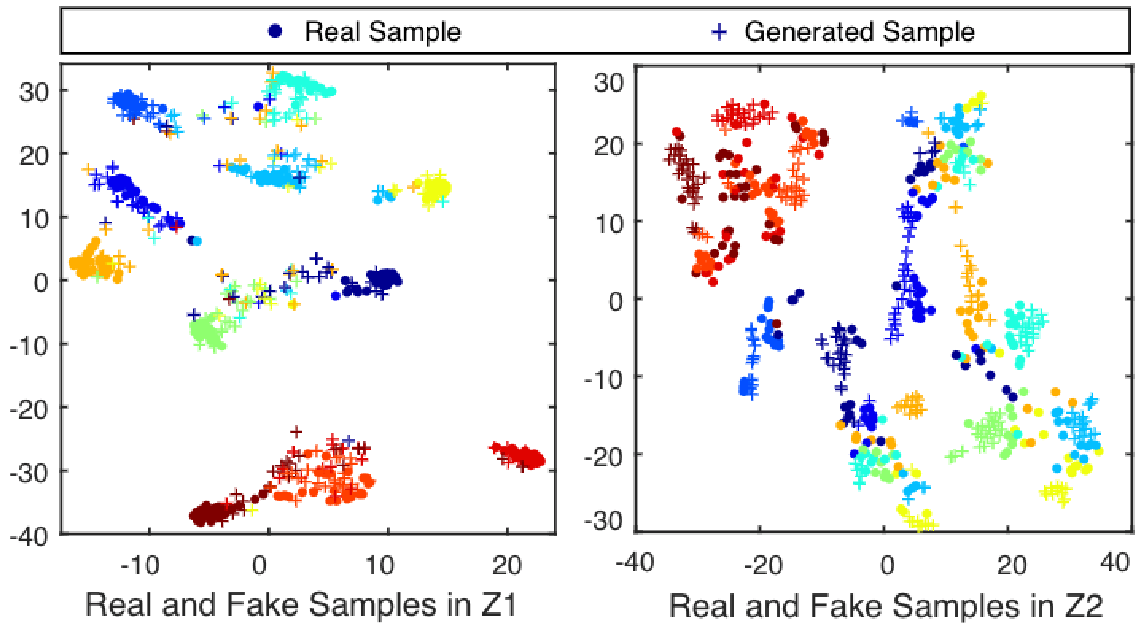
Experiments (2)

Ablation Study for generative module:

- Performance with/without generative model
- t-SNE[1] visualization of real and fake samples



Performance with & without GAN



t-SNE[1] visualization of real & generated samples

[1] SL.J.P. van der Maaten. Accelerating t-SNE using Tree-Based Algorithms. Journal of Machine Learning Research 15(Oct):3221-3245, 2014.

Experiments (3)

Ablation Study for view-correlation discovery network:

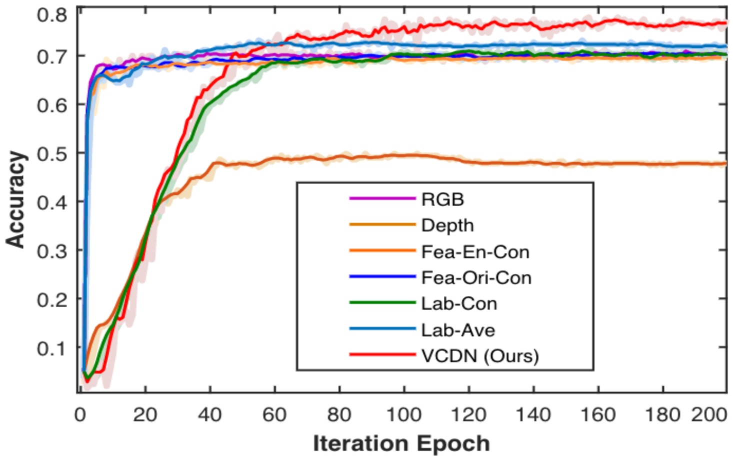
- VCDN compared with different label fusion/correlation learning models

  - Feature/label concatenation & label average/weighted fusion

- VCDN compared with baseline neural networks

Dataset	1-layer	2-layer	3-layer	4-layer	VCDN
UWA	74.31	74.70	73.52	75.10	**76.28**
MHAD	97.83	97.88	96.47	95.76	**98.94**
DHA	86.01	87.24	85.19	82.72	**88.72**

Classification performance of VCDN compared with simple NN.



Performance with different label fusion modules

Setting	UWA	MHAD	DHA
RGB-C_1	69.18	96.42	68.15
Depth-C_2	45.28	63.05	79.79
RGBD-Fea-En-Con	68.78	96.82	70.85
RGBD-Fea-Ori-Con	69.22	97.32	70.83
RGBD-Lab-Con	70.38	96.28	80.95
RGBD-Lab-Ave	71.84	97.56	83.28
RGBD-Lab-Wei	71.15	97.17	83.95
RGBD-VCDN (Ours)	**74.07**	**98.06**	**84.32**

Performance with different label fusion modules

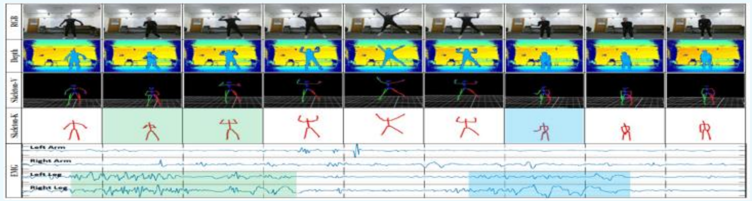
Related works

Synergetic Media Learning Lab

EV-Action: Electromyography-Vision Multi-Modal Action Dataset

Lichen Wang, Bin Sun, Joseph Robinson, Taotao Jing, Yun Fu. FG'20

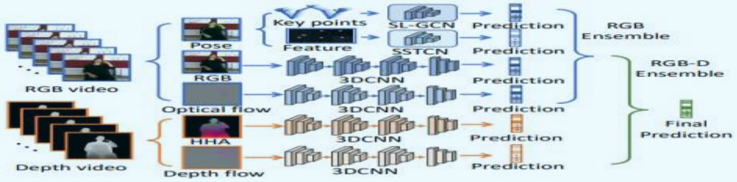
- A large-scale multi-view human action datasets



Skeleton Aware Multi-modal Sign Language Recognition

Songyao Jiang, Bin Sun, Lichen Wang, Yue Bai, Kunpeng Li, Yun Fu. CVPRW'21

- RGB, depth, and skeleton based multi-view recognition



Online Multi-task Clustering for Human Motion Segmentation

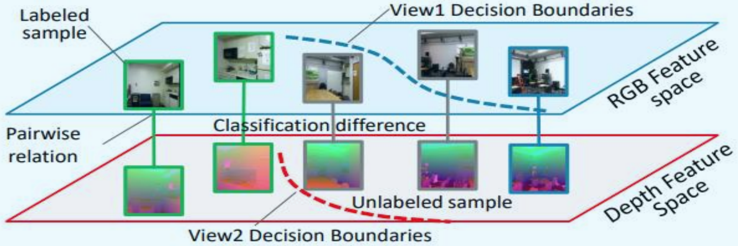
Gan Sun, Yang Cong, Lichen Wang, Zhengming Ding, Yun Fu. ICCVW'2019

- Multi-view spatial-temporal data clustering

Generative View-Correlation Adaptation for Semi-Supervised Multi-View Learning

Yunyu Liu, Lichen Wang, Yue Bai, Can Qin, Zhengming Ding, Yun Fu. In ECCV'2020.

- Explore view-correlation in semi-supervised learning scenario



Research works

Synergetic Media Learning Lab

Topic

Correlation Discovery

Directions

Multi-label Learning

Multi-view Learning

Graph Learning

Applications

Adaptive Graph Guided Embedding, IJCAI, TIP

Generative Correlation Discovery Network, ICDM, AAAI, ICCV

Multi-view human action recognition, ICCV, FG

Transfer motion segmentation, AAAI, TIP

Graph repres. learning, ICLR

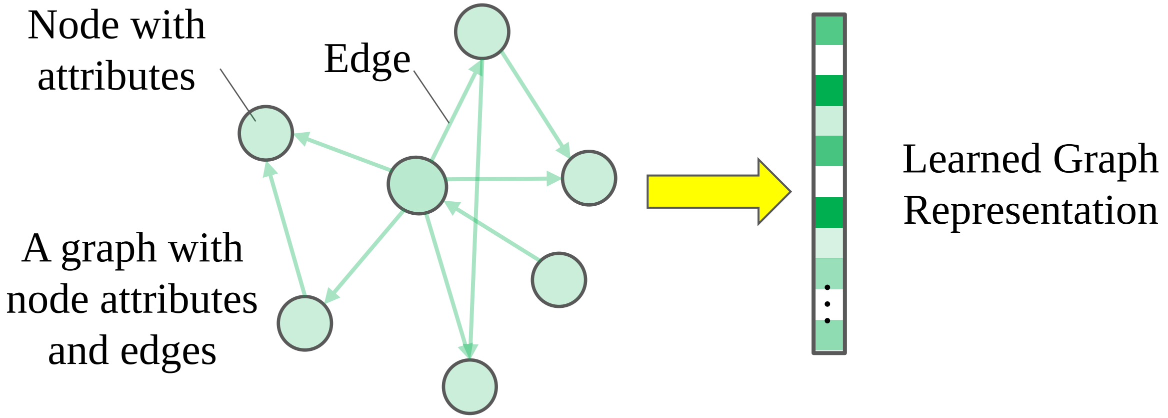
Graph Representation Learning

Topic

- Inductive and unsupervised graph representation learning

Setting

- Input: graph with node attributes and edge attributes
- Output: Dense graph representation as vectors



Unsupervised Graph Representation Learning

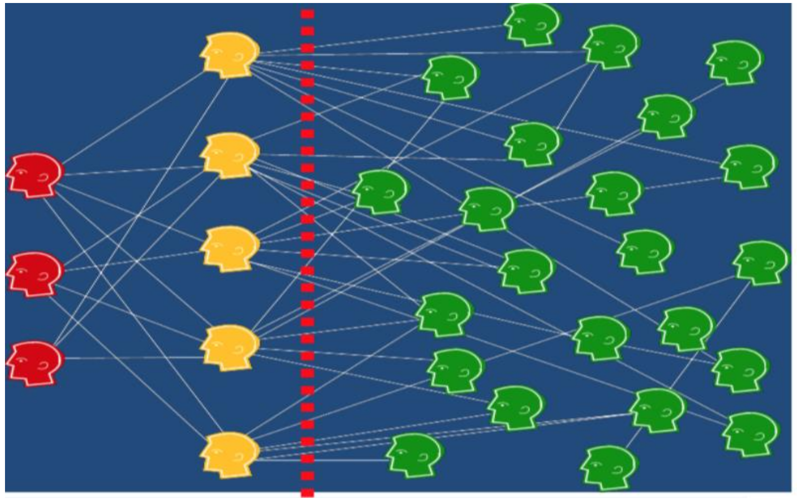
Why Inductive and Unsupervised are Important?

A wide range of potential applications [1]:

- Social Network
  - Facebook, Twitter, WhatsApp
- Finance
  - Credit card fraud, Money laundry
- Logistics Industry:
  - eBay, Amazon, FedEx

Challenges:

- Not enough labeled samples
- Learned model should be generalized to unseen data



Fake Social Account

Credit Fraud

Computer Hack

[1] Chau, Duen Horng, Shashank Pandit, and Christos Faloutsos. "Detecting fraudulent personalities in networks of online auctioneers." PKDD, 2006
[2] https://datafloq.com/read/will-analytics-technology-end-credit-card-fraud/2121

Research works

Topic

Correlation Discovery

Directions

Multi-label Learning

Multi-view Learning

Graph Learning

Applications

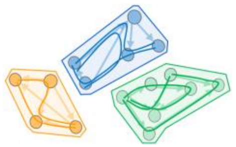
Adaptive Graph Guided Embedding, IJCAI, TIP

Generative Correlation Discovery Network, ICDM, AAAI, ICCV

Multi-view human action recognition, ICCV, FG

Transfer motion segmentation, AAAI, TIP

Graph repres. learning, ICLR



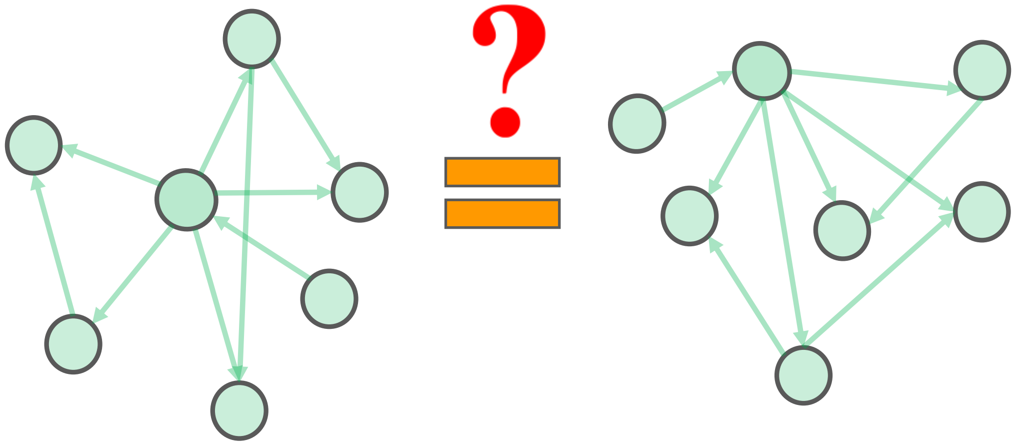
Challenges

Topic

- Inductive and unsupervised graph representation learning

Challenges:

- Existing approaches are in transductive setting
  - Difficult to handle unseen graphs
- Reconstruction-based approach
  - How similar of two graphs?
  - Graph Isomorphism is hard and rigid
  - Computational costly



Isomorphism test is a necessary but hard and computational cost in graph representation learning

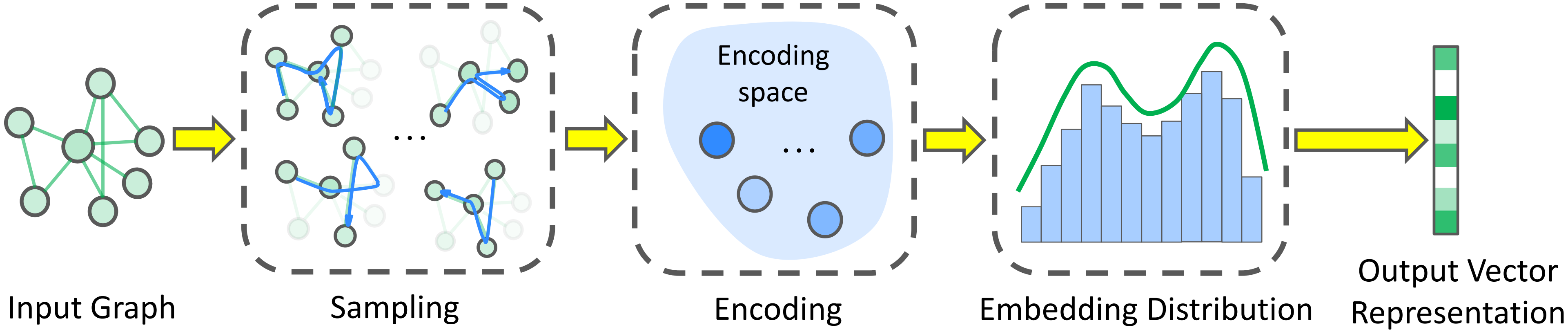
We proposed a framework that addresses the practical need for graph representation learning in real-life applications

Lichen Wang, Bo Zong, Qianqian Ma, Wei Cheng, Jingchao Ni, Wenchao Yu, Yanchi Liu, Dongjing Song, Haifeng Chen, and Yun Fu, “Inductive and Unsupervised Representation Learning on Graph Structured Objects,” 2020 International Conference on Learning Representations (ICLR), .

The proposed Framework: SEED (1)

SEED: **S**ampling, **E**ncoding, and **E**mbedding **D**istributions

- **Sampling**: Random walk-based subgraph sampling from the input graph
  - Difficult to directly get whole graph representations
  - Could be easier to obtain representations for walks
- **Encoding**: Subgraph encoding via earliest visiting time
  - Make the process efficient and the representations effective



Framework of our SEED approach

The proposed Framework: SEED (2)

SEED: **S**ampling, **E**ncoding, and **E**mbedding **D**istributions

- **Embedding Distributions**:

  We encode a vector distribution into a single vector, which should preserve the similarity between vector distributions.

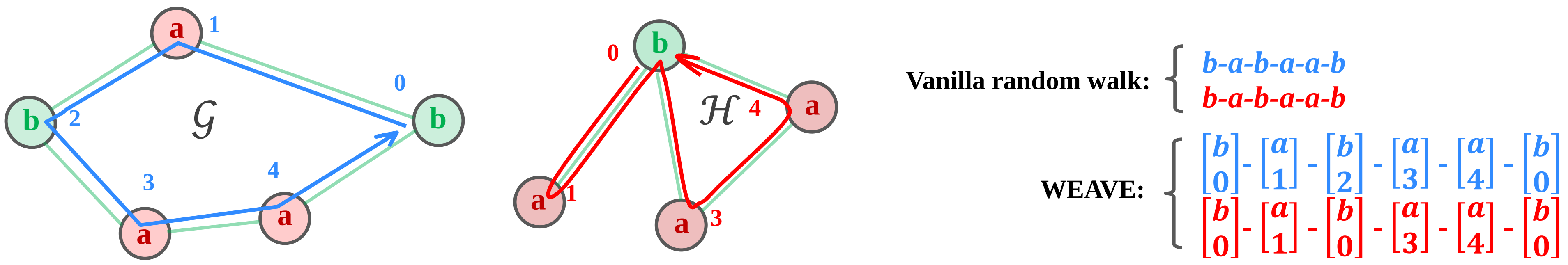
  - Each input graph is reduced into a set of vectors, each of which is the representation for a sampled subgraph.

  - Given that we have sampled a sufficient number of subgraphs, if two input graphs are similar, their vector distributions should be similar

Sampling & Encoding

WEAVE: Random **W**alk with **EA**rliest **V**isit tim**E**).

- Random walk (RW) in graphs
- Revisit information: earliest visiting time
- Advantages:
  - RW: easy to reconstruct, but no loop info preserved
  - RW + revisit: easy to reconstruct with loop info
  - RW with revisit contains more structural info



Encoding results of Vanilla random walk and WEAVE. WEAVE could distinguish the difference of the two graphs.

Embedding Distribution

- Insight: Walk distribution representation similarity $\Longrightarrow$ graph similarity
- Theoretical: as proved, distribution $R_{\mathcal{g}} = R_{\mathcal{H}}$ if graph $\mathcal{g}$ and $\mathcal{H}$ are isomorphic
- **Option 1**: Identity kernel

  \- We assume $r_{\mathcal{g}} \sim N(\mu_1, I)$ and $r_{\mathcal{H}} \sim N(\mu_2, I)$, it is simple but surprisingly effective.

$$\hat{\mu}_{\mathcal{G}} = \frac{1}{s} \sum_{i=1}^{s} \mathbf{z}_i \qquad \hat{\mu}_{\mathcal{H}} = \frac{1}{s} \sum_{i=1}^{s} \mathbf{h}_i$$

- **Option 2**: Commonly adopted kernels

$$\hat{\mu}'_{\mathcal{G}} = \frac{1}{s} \sum_{i=1}^{s} \hat{\phi}(\mathbf{z}_i; \theta_m) \quad \hat{\mu}'_{\mathcal{H}} = \frac{1}{s} \sum_{i=1}^{s} \hat{\phi}(\mathbf{h}_i; \theta_m) \quad D(P_{\mathcal{G}}, P_{\mathcal{H}}) = \|\hat{\mu}'_{\mathcal{G}} - \hat{\mu}'_{\mathcal{H}}\|_2^2$$

Theoretical Insights

Theorem: Given graphs $\mathcal{G}$ and $\mathcal{H}$, distribution $R_{\mathcal{G}} = R_{\mathcal{H}}$ if graph $\mathcal{G}$ and $\mathcal{H}$ are isomorphic

The theorem holds for the situations:

- Graphs without any attributes
- Graphs with node attributes
- Graphs with node and edge attributes

Experiments (1)

- Seven graph datasets
- Two down-stream tasks:
  - Clustering
  - Classification
- Our approach obtains the highest performance.
  - Up to 10% improvements

Setting	Datasets	Methods / Metric	SAGE	GIN	GMN	SEED	SAGE	GIN	GMN	SEED
			Node Feature **Excluded**				Node Feature **Included**			
Clustering	Dezzer	ACC	0.3853	0.4913	0.4924	**0.4927**	0.3840	**0.4930**	0.4808	0.4810
		NMI	0.0079	0.0958	0.0726	**0.1277**	0.0003	**0.0893**	0.0651	0.0566
	MUTAG	ACC	0.6649	0.4997	0.4990	**0.8014**	0.6649	0.4963	0.4910	**0.7260**
		NMI	0.0150	0.0946	0.0825	**0.3214**	0.0070	0.0933	0.0917	**0.1567**
	NCI1	ACC	0.5098	0.5221	0.5022	**0.5510**	0.5070	0.5204	0.5005	**0.5441**
		NMI	0.0003	0.0015	0.0034	**0.0073**	0.0002	0.0013	0.0042	**0.0089**
	PROTEINS	ACC	0.5657	0.5957	**0.5966**	0.5957	0.5657	0.5957	0.5957	**0.5957**
		NMI	0.0013	0.0038	0.0117	**0.0518**	0.0004	0.0034	0.0067	**0.0689**
	COLLAB	ACC	0.5208	0.5458	0.5173	**0.5973**	-	-	-	-
		NMI	0.0025	0.0729	0.0193	**0.2108**	-	-	-	-
	IMDB-BINARY	ACC	0.5069	0.6202	0.5010	**0.5776**	-	-	-	-
		NMI	0.0002	0.0459	0.0093	**0.0241**	-	-	-	-
	IMDB-MULTI	ACC	0.3550	3607	0.3348	**0.3816**	-	-	-	-
		NMI	0.0019	0.0185	0.0112	**0.0214**	-	-	-	-
Classification	Dezzer	ACC	0.3775	0.5094	0.5427	**0.6327**	0.3754	0.5270	0.5627	**0.7451**
	MUTAG	ACC	0.6778	0.6778	0.6889	**0.8112**	0.6889	0.6778	0.6889	**0.8222**
	NCI1	ACC	0.5410	0.5571	0.5123	**0.6105**	0.5328	0.5231	0.5133	**0.6151**
	PROTEINS	ACC	0.6846	**0.7387**	0.6216	0.7207	0.7027	0.7207	0.6357	**0.7462**
	COLLAB	ACC	0.5650	0.6170	0.5460	**0.6720**	-	-	-	-
	IMDB-BINARY	ACC	0.5400	0.7310	0.5140	**0.7660**	-	-	-	-
	IMDB-MULTI	ACC	0.3866	0.3843	0.3478	**0.4466**	-	-	-	-

Clustering & Classification Performance

Experiments (2)

How parameters impact the output quality?

- Subgraph extraction with different sampling number and walk length.
  - Quantitative performance
  - t-SNE[1] visualization

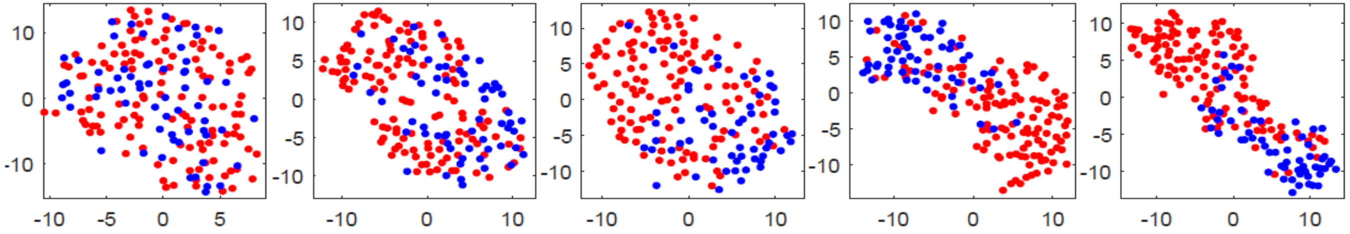
Summary

- More sampling number and walk length could improve the learned representation quality

Sampling Number	Classification Accuracy	Clustering ACC	Clustering NMI
25	0.6832	0.6649	0.0031
50	0.6778	0.6649	0.0005
100	0.7778	0.6649	0.0537
150	0.7889	0.6968	0.1081
200	0.7778	0.7633	**0.2100**
300	0.7833	0.7502	0.1995
400	**0.8389**	0.7628	0.1928
800	0.8111	**0.7660**	0.1940

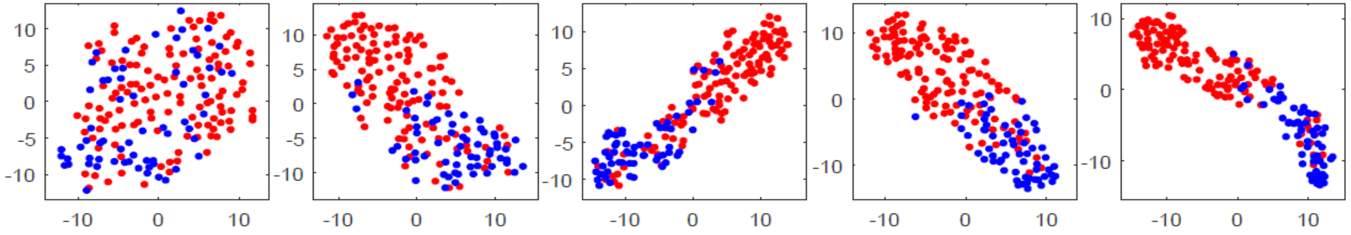
Walk Length	Classification Accuracy	Clustering ACC	Clustering NMI
5	0.7278	0.6649	0.0534
10	0.7778	0.7633	0.2100
15	0.8167	0.7723	0.2495
20	0.8778	0.8245	0.3351
25	0.8722	0.8218	**0.3380**
30	**0.8743**	**0.8285**	0.3321

Classification & clustering performance



(a) Samples = 50 (b) Samples = 100 (c) Samples = 150 (d) Samples = 200 (e) Samples = 400

t-SNE visualization with different sampling numbers



(a) Walk length = 5 (b) Walk length = 10 (c) Walk length = 15 (d) Walk length = 20 (e) Walk length = 25

t-SNE visualization with different work length

[1]] Maaten, Laurens van der, and Geoffrey Hinton. "Visualizing data using t-SNE." Journal of machine learning research 9, no. Nov (2008): 2579-2605.

Conclusion

- Correlation Discovery
  - Multi-label Learning:
    - Clustering
    - Classification
  - Multi-view learning
    - Feature space correlations
    - Label space correlation
  - Graph representation
    - Correlation representation

Thank you!

Lichen Wang

Electrical and Computer Engineering
College of Engineering
Northeastern University