Cross Diffusion on Multi-Hypergraph for Multi-Modal 3D Object Recognition

Contents

3D Object Recognition

3D Movie

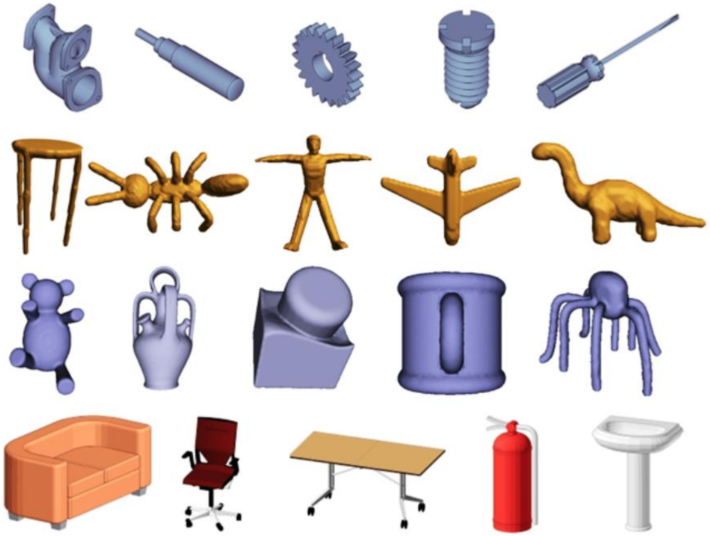
Virtual Reality

Augmented Reality

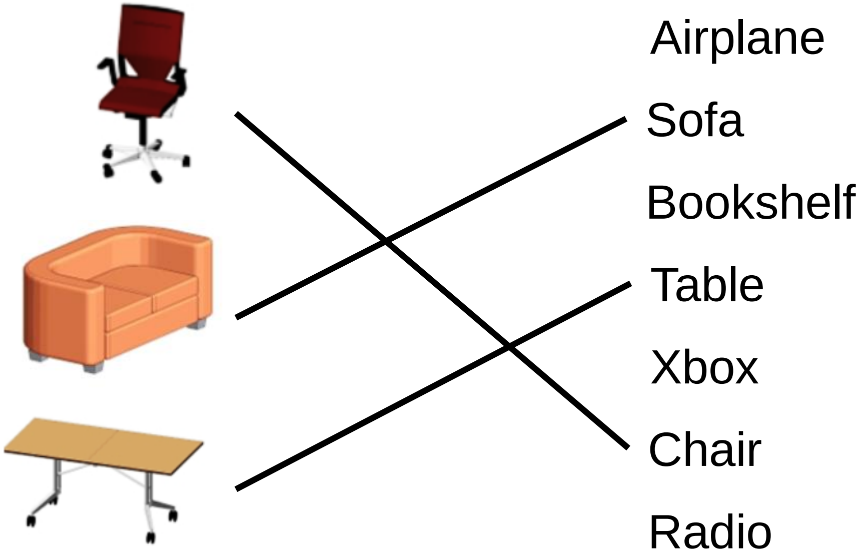
Automatic Drive

rapid increasing of 3D object data

3D Object

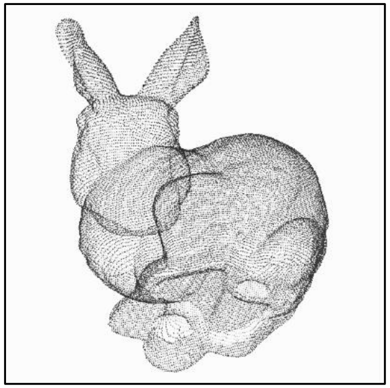


Recognition



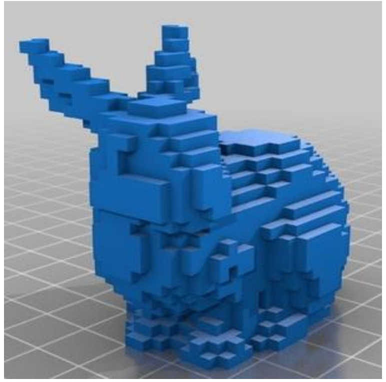
3D Object Representation

Multi-modal Data

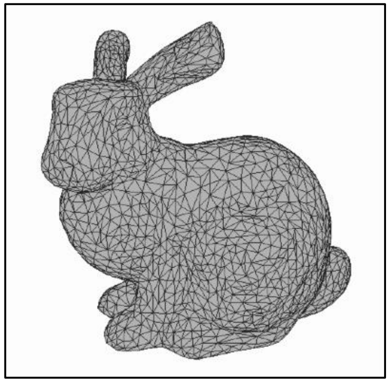


Point Cloud

View



Volumetric



Mesh

View-based Representation

MVCNN

GVCNN

3D Object Representation

Multi-modal Data

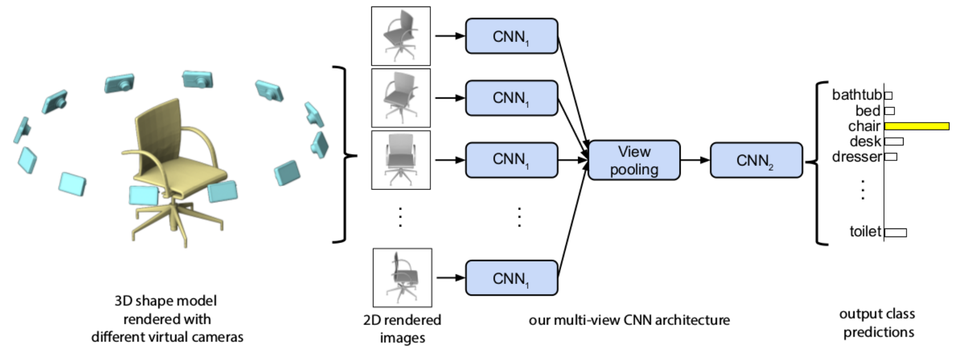
Point Cloud

View

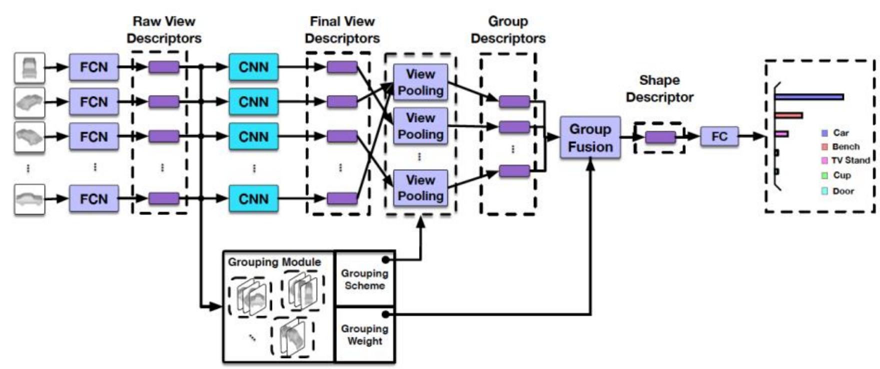
Volumetric

Mesh

View-based Representation

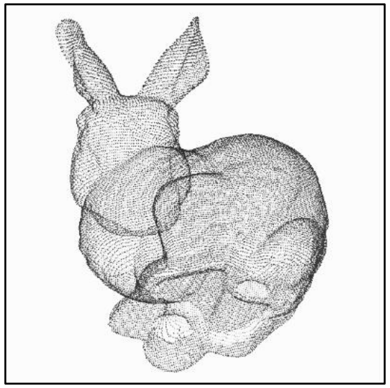


(a) MVCNN

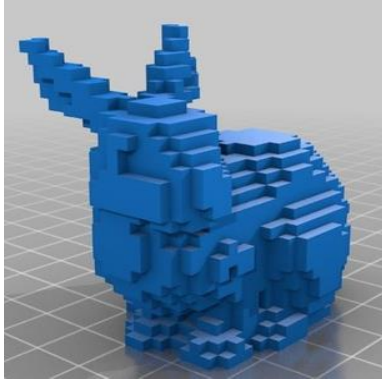


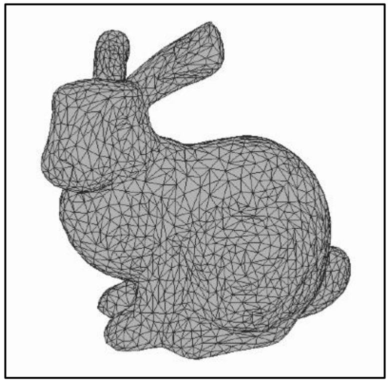
(b) GVCNN

3D Object Representation


Point Cloud


View


Volumetric


Mesh

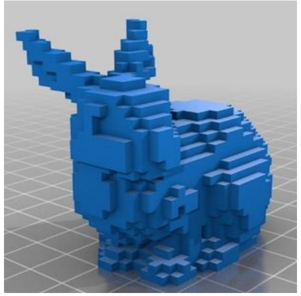
How to combine multiple 3D representations towards better 3D object recognition performance?

Challenge 1: Exploit correlation among multi-modal data
Challenge 2: Consider multi-modal data simultaneously during multi-modal fusion process

Related Work

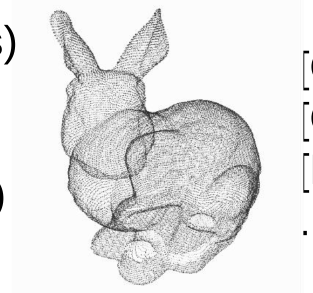
3D Object Representations

Volumetric Data



[Wu et al. 2015] (3D ShapeNets)
[Maturana et al. 2015] (VoxNet)
[Wang et al. 2017] (O-Net)
[Tatarchenko et al. 2017] (OGN)
……

Point Cloud Data



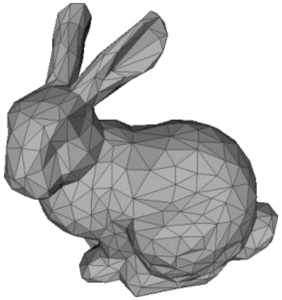
[Qi et al. 2017] (PointNet)
[Qi et al. 2017] (PointNet++)
[Fan et al. 2017] (PointSetGen)
……

How to combine them?

View Data

[Su et al. 2015] (MVCNN)
[Kalogerakis et al. 2016]
[Guo et al. 2016]
[Feng et al. 2018] (GVCNN)
……

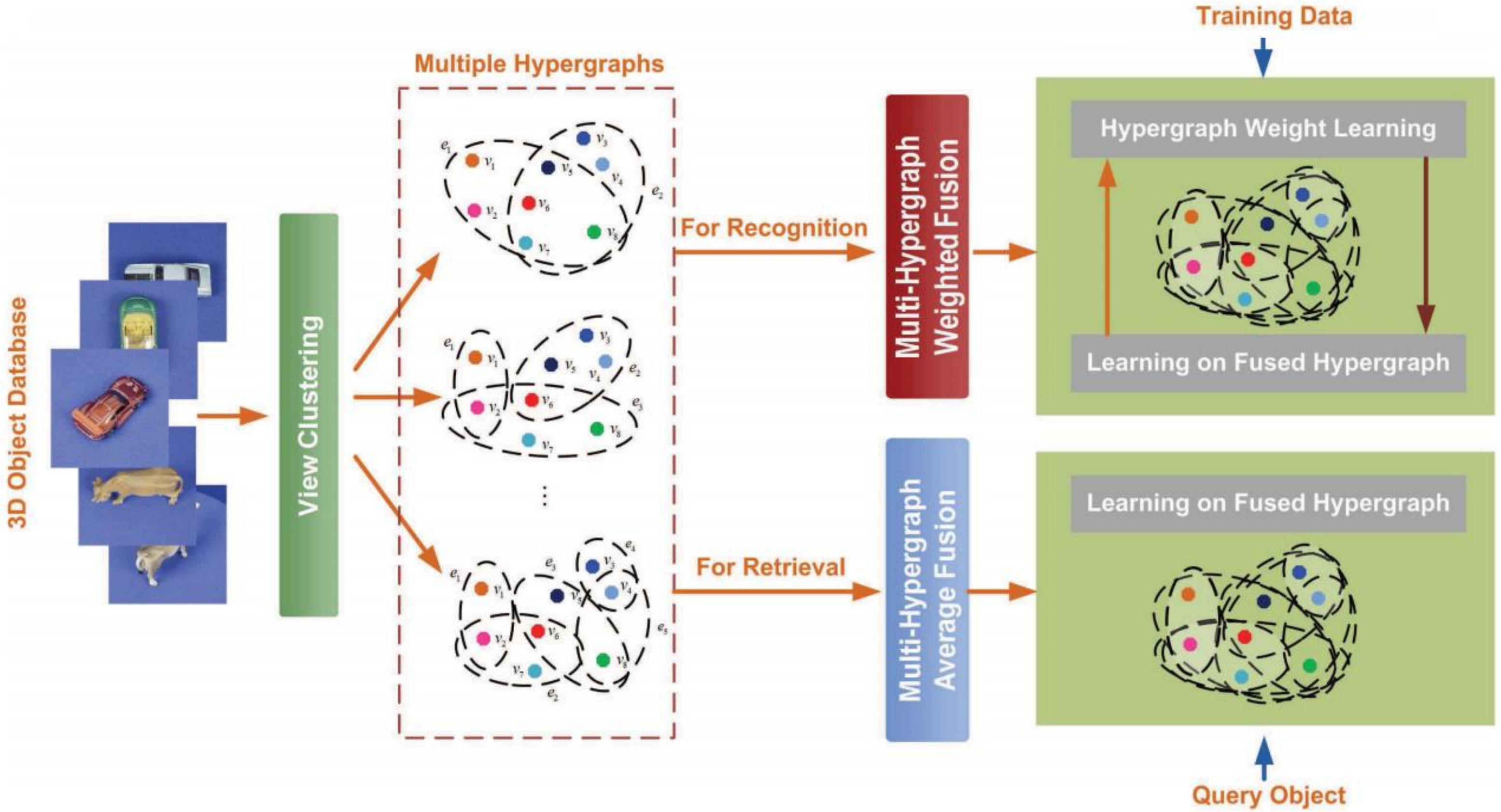
Mesh Data



[Defferard et al. 2016]
[Henaff et al. 2015]
[Yi et al. 2017]
……

Related Work

Multi-Modal Data



Multi-Hypergraph Learning *[Gao et al. TIP'12]*

Disadvantages:

1. **The computational cost is very high.**
2. **Multi-modal data are only considered in the fusion part.**

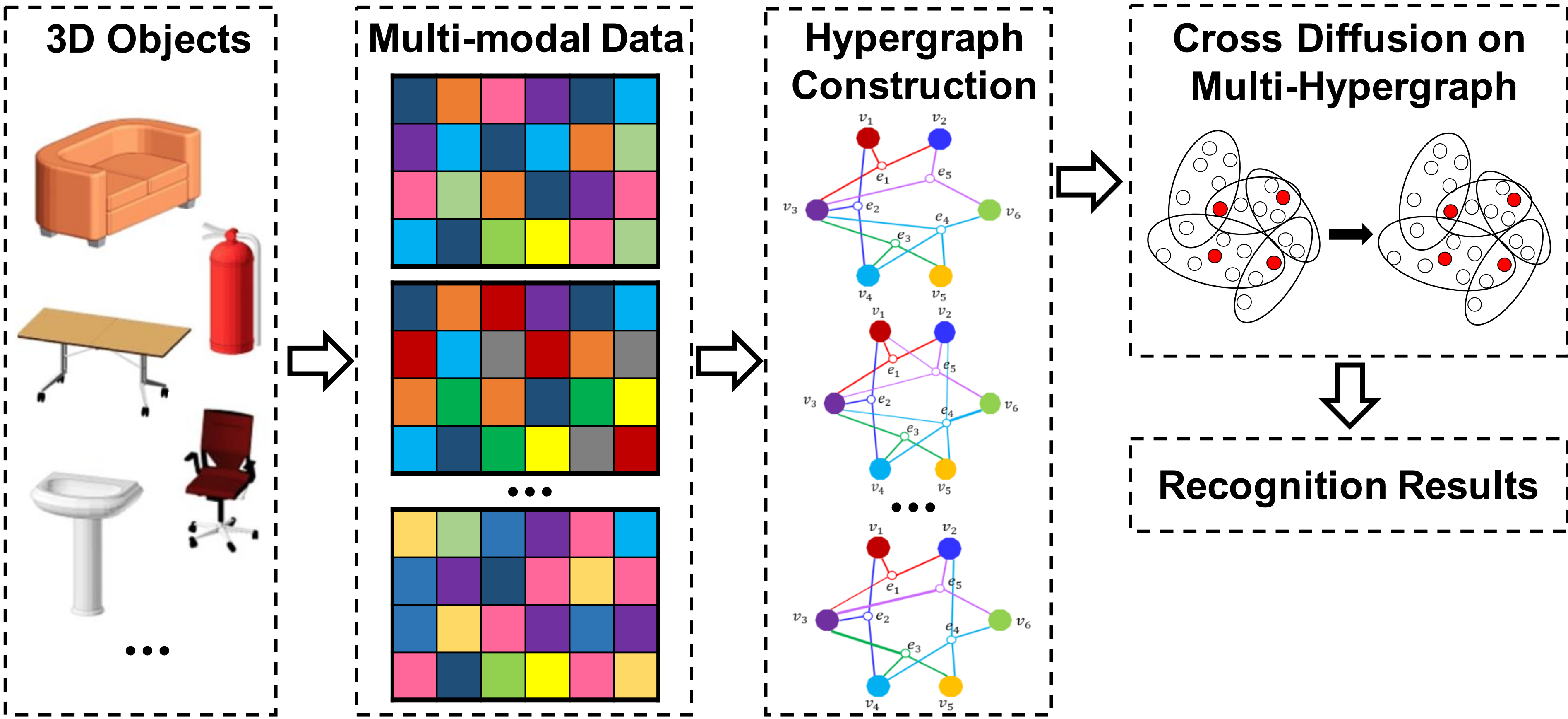
Motivations

How to combine multiple 3D representations towards better 3D object recognition performance?

Challenge 1: Exploit correlation among multi-modal data
Challenge 2: Consider multi-modal data simultaneously during multi-modal fusion process

Task 1: Employ multi-hypergraph structure to formulate the correlation among 3D objects
Task 2: Conduct cross diffusion process on the multi-hypergraph structure

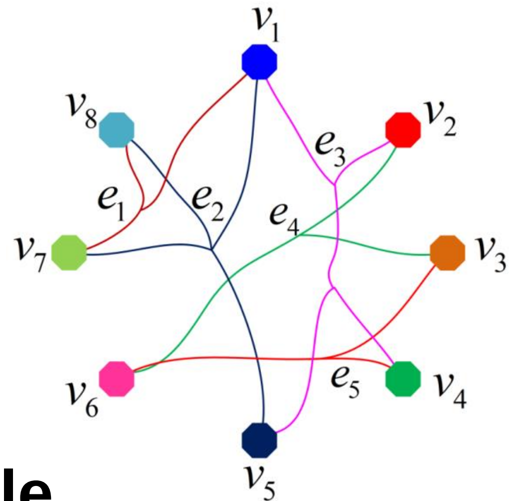
Framework



3D Objects
Multi-modal Data
Hypergraph Construction
Cross Diffusion on Multi-Hypergraph
Recognition Results

Correlation Modelling

Correlation Modelling



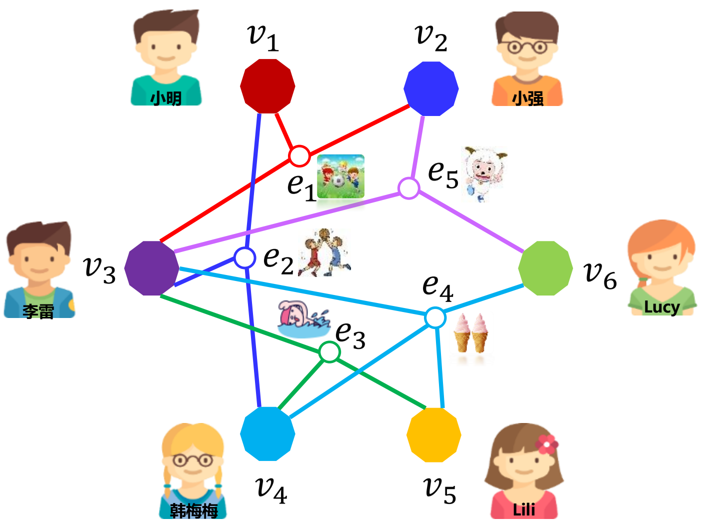
Hypergraph

$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathrm{W})$

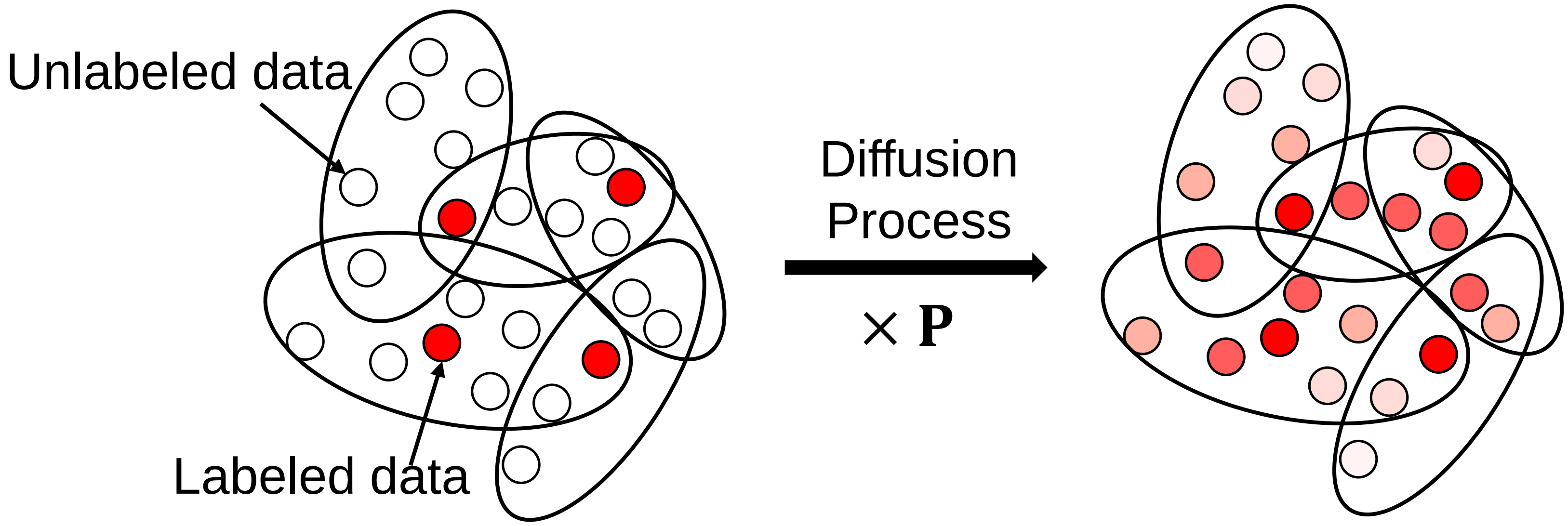
Each vertex represents an object.

Example

Name	foot-ball	baske-tball	swim-ming	sweet	carton
Ming	✓	✓	✗	✗	✗
Qiang	✓	✗	✗	✗	✓
Lei	✓	✓	✓	✓	✓
Mei	✗	✓	✓	✓	✗
Lili	✗	✗	✓	✓	✗
Lucy	✗	✗	✗	✓	✓

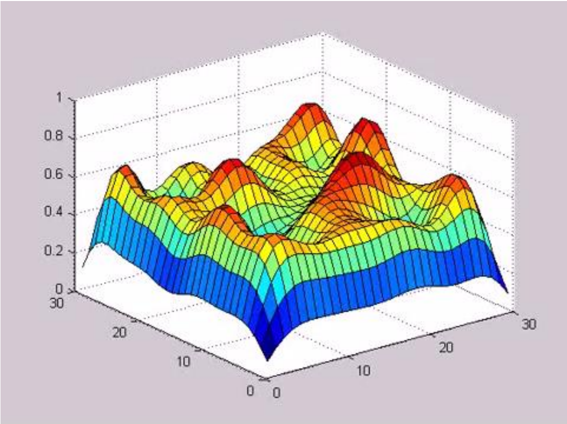


Diffusion Process on Single Hypergraph



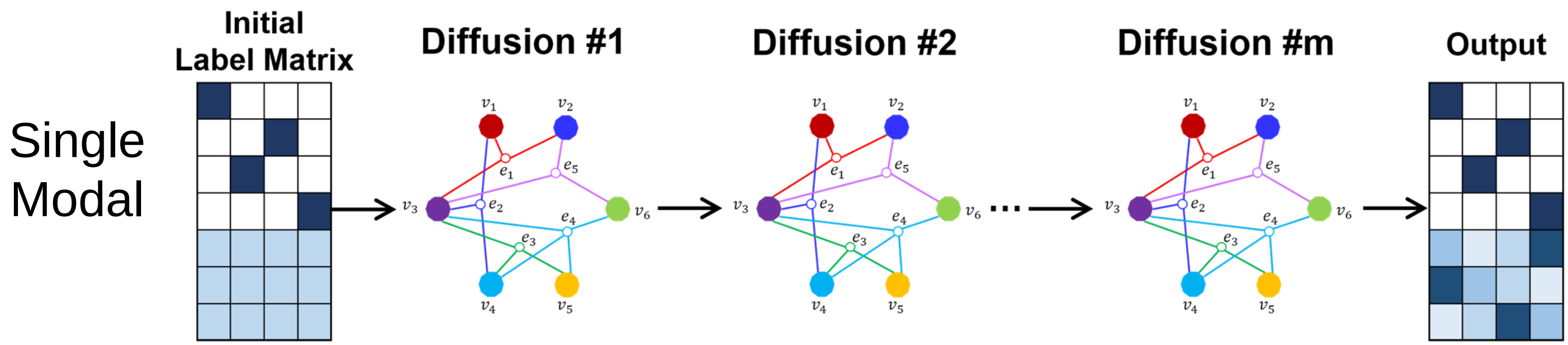
Similarity Matrix $\mathbf{\Lambda^t} = \mathbf{H^t W^t {D_e^t}^{-1} {H^t}^T}$

Transition Matrix $\mathbf{P}^t(i,j) = \dfrac{\mathbf{\Lambda^t}(i,j)}{\sum_{w \in \mathcal{V}^t} \mathbf{\Lambda^t}(i,w)}$



The diffusion process is much faster than traditional methods.

Cross Diffusion Process on Multi-Hypergraph

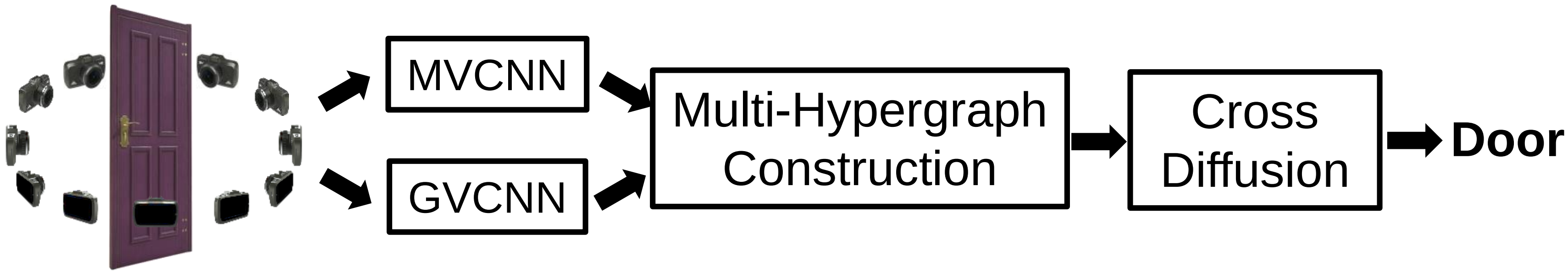


Diffusion on single hypergraph

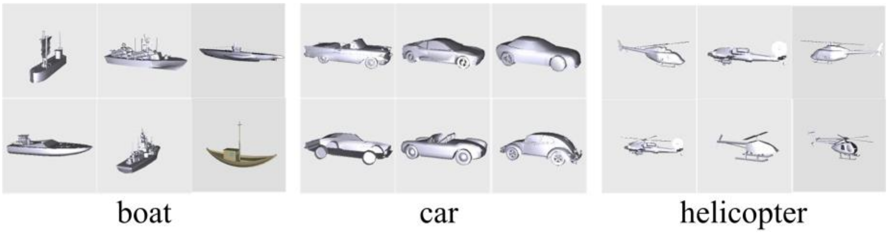
Advantages:

1. **The multi-hypergraph structure can model the high-order correlation among multi-modal data.**
2. **The cross diffusion process can combine multi-modal data effectively.**
3. **The cross diffusion process is very fast.**

Experiments



ModelNet40 (12311 objects)



NTU (2012 objects)

Two kinds of 3D features:

- Multi-View Convolutional Neural Networks (MVCNN)
- Group-View Convolutional Neural Networks (GVCNN)

State-of-the-art methods

- MVCNN [1] and GVCNN [2]
- MVCNN+HL and GVCNN+HL
- MVCNN+GVCNN+HL
- MVCNN+GVCNN+MHL [3]
- Cross Diffusion on Multi-Hypergraph (CDMH)

[1] *Su et al. Multi-View Convolutional Neural Networks for 3D Shape Recognition. CVPR'15*
[2] *Feng at al. GVCNN: Group-View Convolutional Neural Networks for 3D Shape Recognition. CVPR'18*
[3] *Gao et al. 3D Object Retrieval and Recognition with Hypergraph Analysis. TIP'12*

Experimental Results

Classification Accuracy

Method	ModelNet40	NTU
MVCNN	90.10%	79.89%
GVCNN	93.10%	82.30%
MVCNN+HL	90.68%	79.89%
GVCNN+HL	92.14%	82.84%
MVCNN+GVCNN+HL	93.23%	80.43%
MVCNN+GVCNN+MHL	96.19%	83.38%
CDMH	**96.76%**	**84.45%**

multi-modal representations

The error rate is dropped by **15%**.

Time Cost

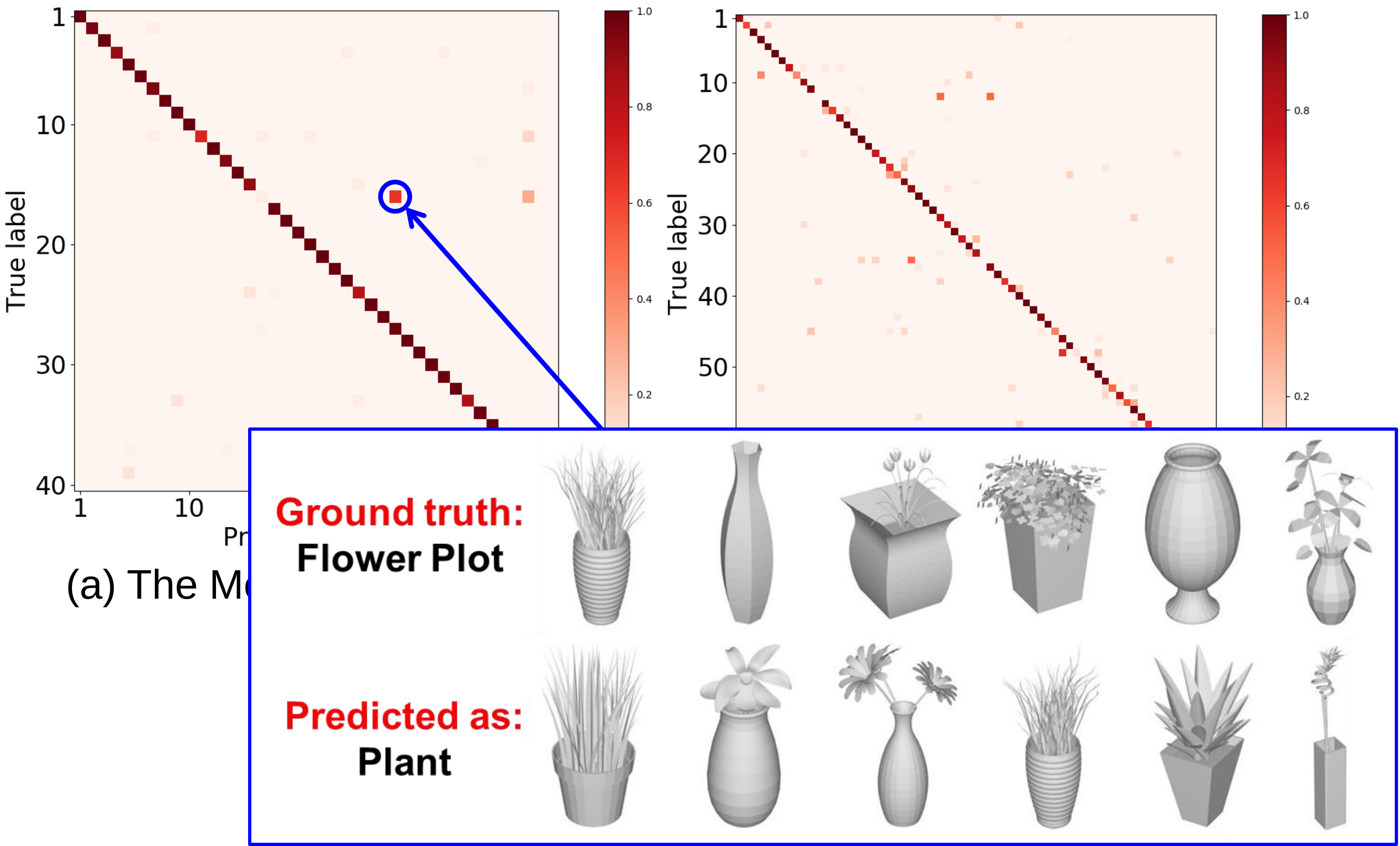
Method	ModelNet40	NTU
MVCNN+GVCNN+MHL	869.4s	6.652s
CDMH	**2.233s**	**0.332s**

The speed is increased by **400** times.

Our proposed method can achieve better performance and faster speed than state-of-the-art methods.

Experimental Results

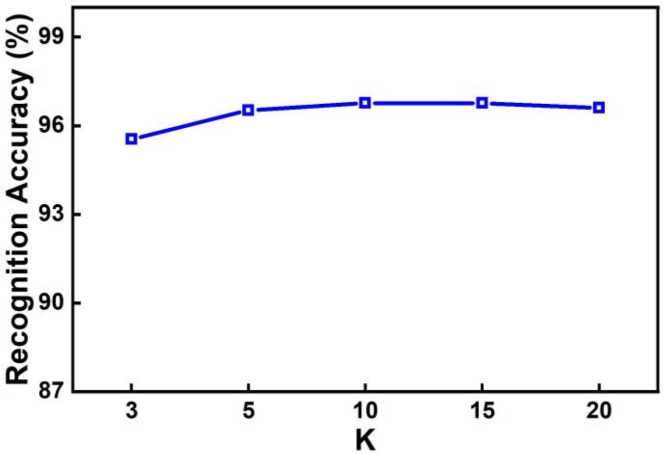
Confusion Matrix



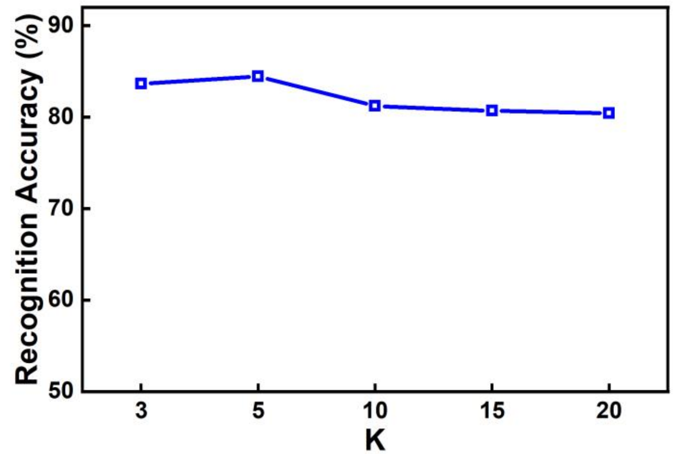
(a) The M

On Hypergraph Construction and Diffusion Process

On Hypergraph Construction

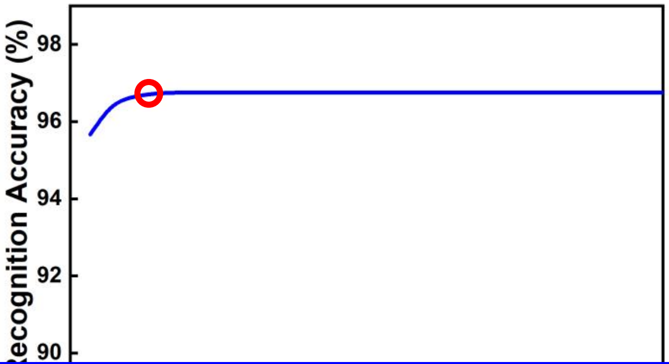


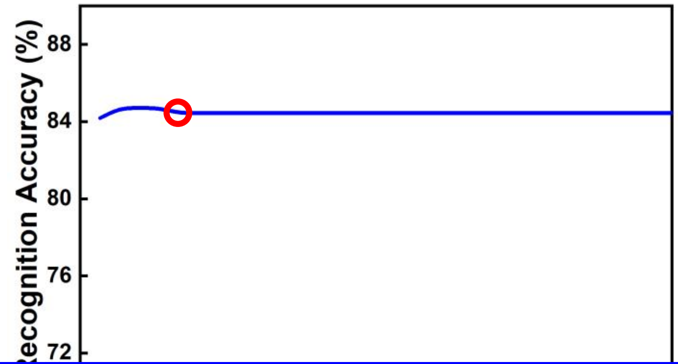
(a) ModelNet40



(b) NTU

On Diffusion Process

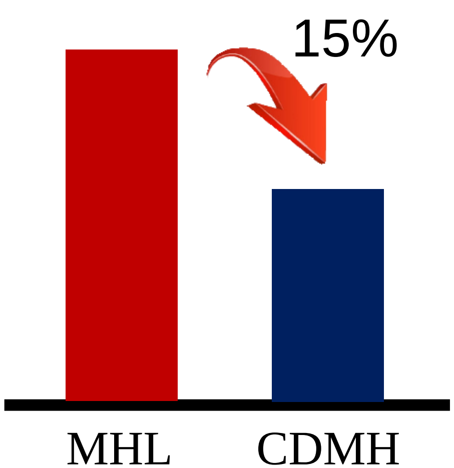




Our proposed method can achieve stable performance with different parameters and converge fast.

Conclusion

- We propose a cross diffusion method on multi-hypergraph for multi-modal 3D object recognition.
- The proposed method is more **effective and efficient** than the state-ot-the-art methods.
- The proposed method is **a general framework** which can be used in other applications with multi-modal data.



Error Rate

Speed

Medical Image Analysis

Social Media Analysis

References

[1] Chen, D.Y., Tian, X.P., Shen, Y.T., Ouhyoung, M.: On Visual Similarity Based 3D Model Retrieval. Computer Graphics Forum 22(3), 223-232 (2003)

[2] Chen, X., Ma, H., Wan, J., Li, B., Xia, T.: Multi-View 3D Object Detection Network for Autonomous Driving. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. vol. 1 (2017)

[3] Feng, Y., Zhang, Z., Zhao, X., Ji, R., Gao, Y.: GVCNN: Group-View Convolutional Neural Networks for 3D Shape Recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2018)

[4] Gao, Y., Wang, M., Tao, D., Ji, R., Dai, Q.: 3D Object Retrieval and Recognition with Hypergraph Analysis. IEEE Transactions on Image Processing 21(9), 4290-4303 (2012)

[5] Gao, Y., Wang, M., Zha, Z.J., Shen, J., Li, X., Wu, X.: Visual-textual joint relevance learning for tag-based social image search. IEEE Transactions on Image Processing 22(1), 363-376 (2013)

[6] Guo, H., Wang, J., Gao, Y., Li, J., Lu, H.: Multi-View 3D Object Retrieval with Deep Embedding Network. IEEE Transactions on Image Processing 25(12), 5526-5537 (2016)

[7] Hong, R., Hu, Z., Wang, R., Wang, M., Tao, D.: Multi-View Object Retrieval Via Multi-Scale Topic Models. IEEE Transactions on Image Processing 25(12), 5814-5827 (2016)

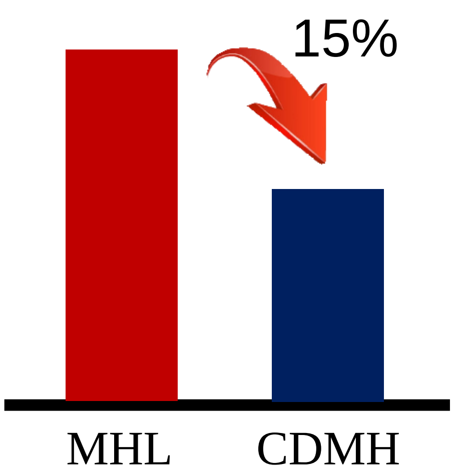
[8] Hong, R., Zhang, L., Zhang, C., Zimmermann, R.: Flickr Circles: Aesthetic Tendency Discovery by Multi-View Regularized Topic Modeling. IEEE Transactions on Multimedia 18(8), 1555-1567 (2016)

[9] Huang, Y., Liu, Q., Zhang, S., Metaxas, D.N.: Image retrieval via probabilistic hypergraph ranking. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3376-3383. IEEE (2010)

Thanks!

Conclusion

- We propose a cross diffusion method on multi-hypergraph for multimodal 3D object recognition.
- The proposed method is more **effective and efficient** than the state-ot-the-art methods.
- The proposed method is **a general framework** which can be used in other applications with multi-modal data.



Error Rate

Speed

Medical Image Analysis

Social Media Analysis