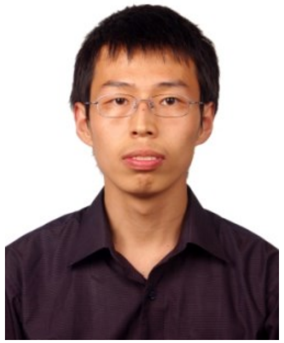
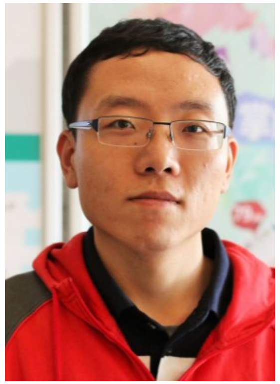
JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

VectorFloorSeg: Two-Stream Graph Attention Network for Vectorized Roughcast Floorplan Segmentation



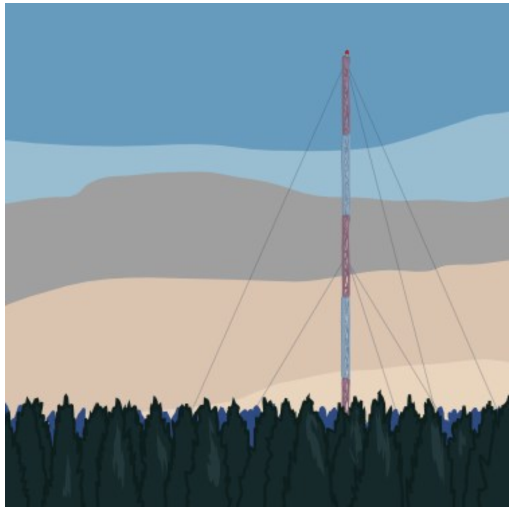


[1]Bingchen Yang [1]Haiyong Jiang [2]Hao Pan [1]Jun Xiao

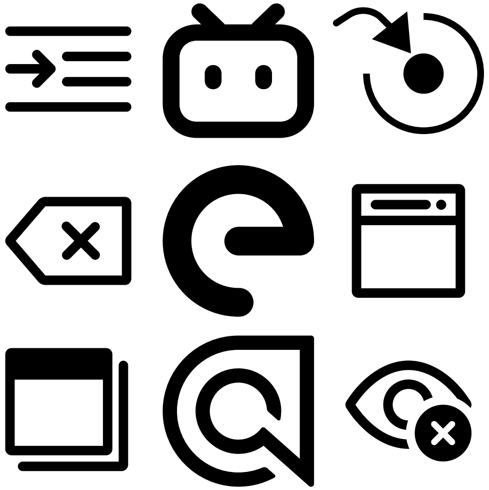
[1]University of Chinese Academy of Sciences [2]Microsoft Research Asia

Introduction

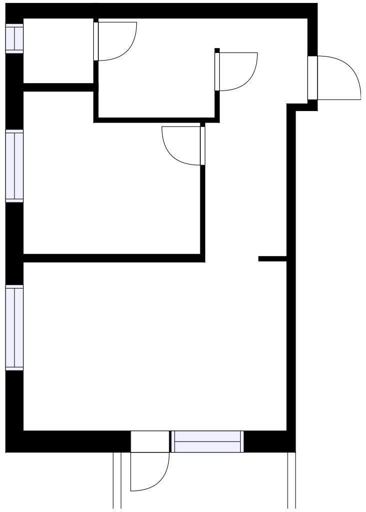
Broad application scenarios of vector graphics (VG)



Art



UI

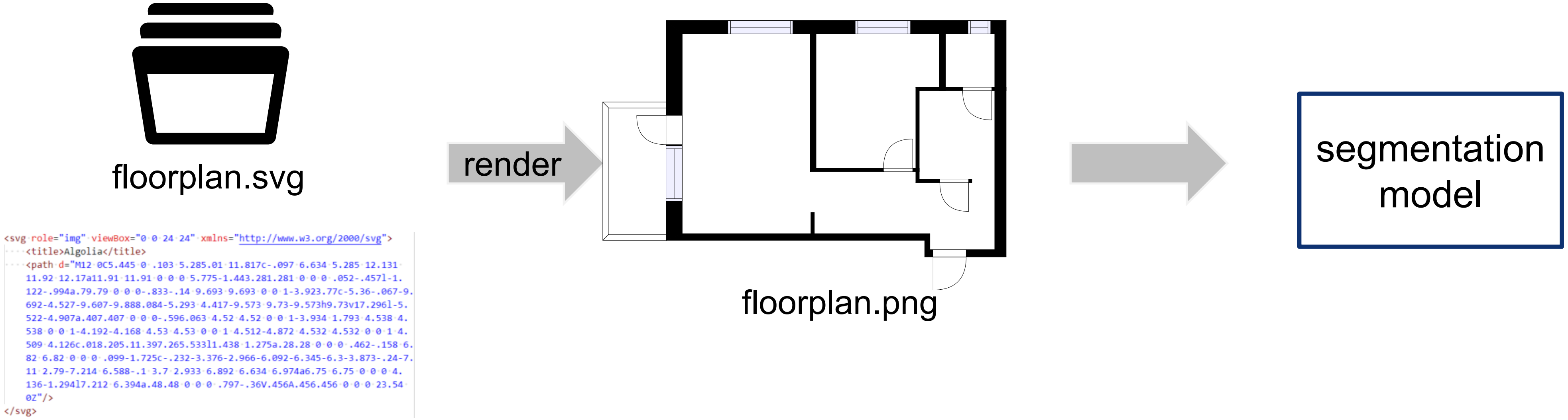


Interior design

- editable
- arbitrary scaling, no loss of details, no cause of sawtooth
- Autonomous analysis of vector graphics is necessary for downstream applications

Introduction

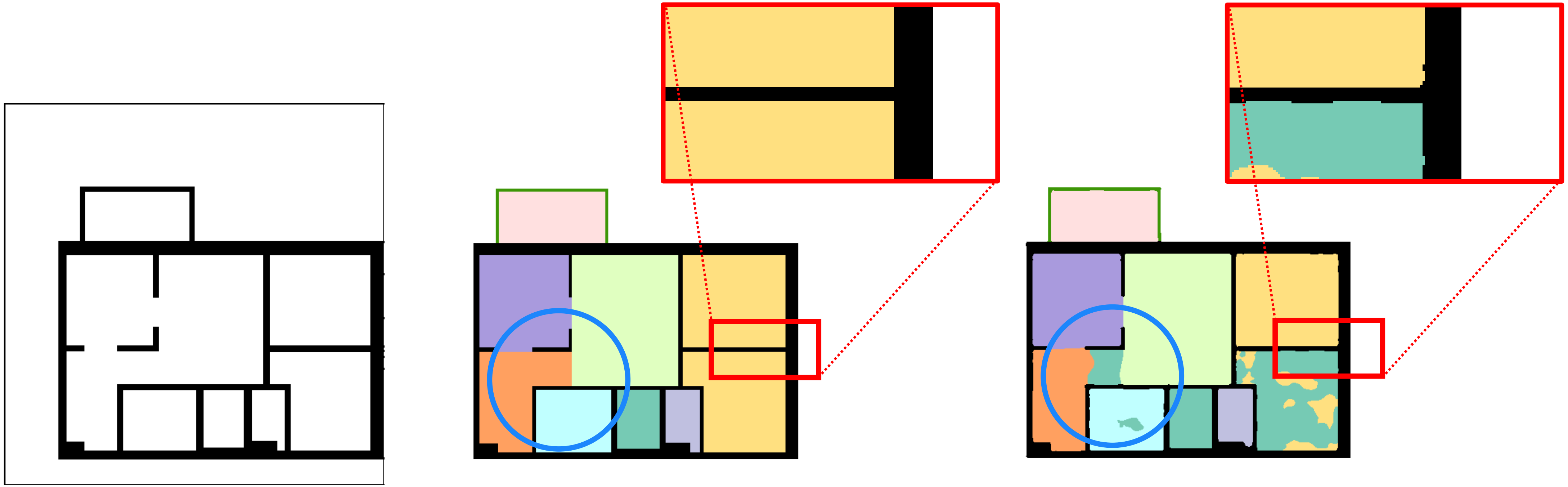
Challenges of vectorized roughcast floorplan segmentation



- Irregularly structured data pose challenge on ***directly*** applying image backbone

Introduction

Disadvantages of rasterized floorplan segmentation

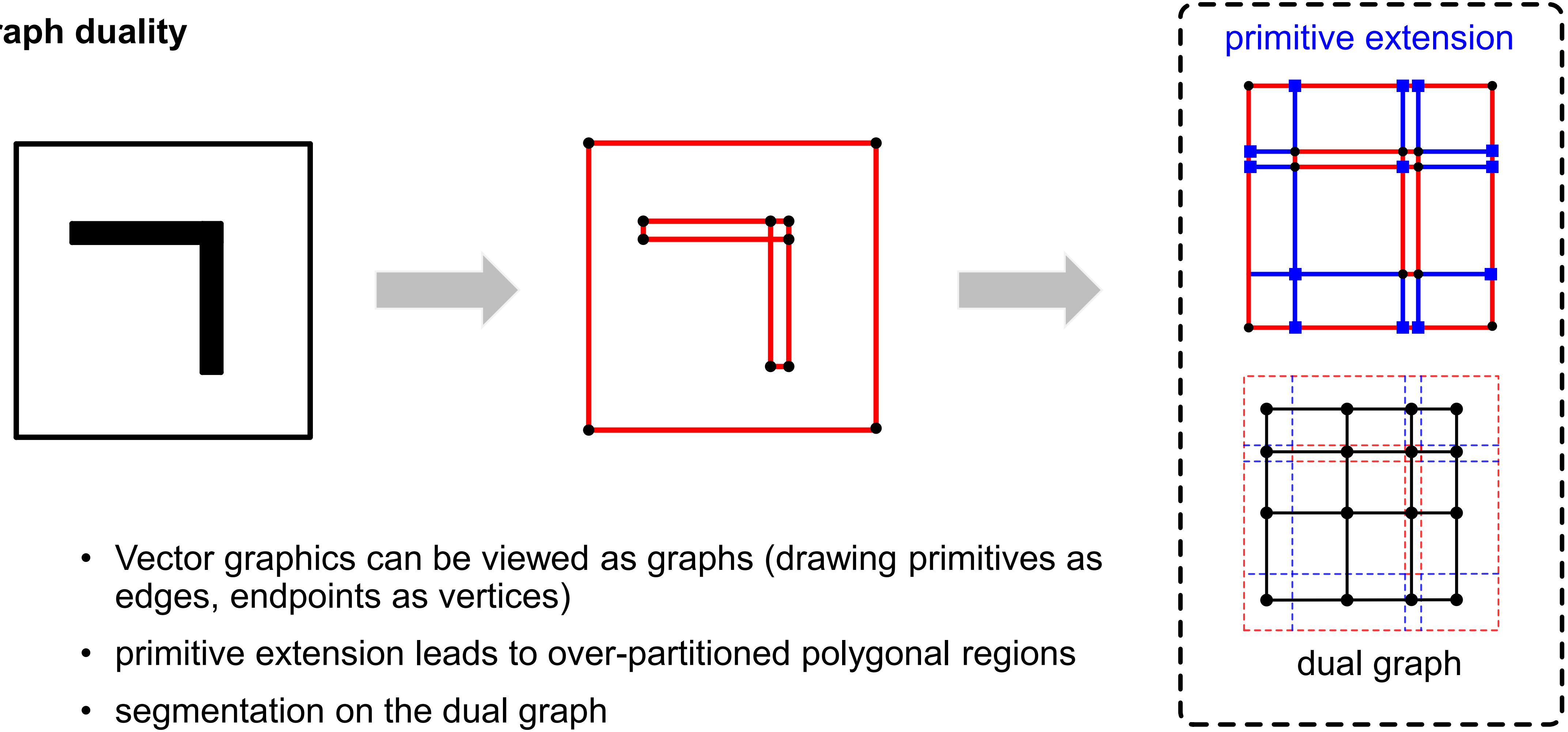


input groundtruth prediction

- jigsaw boundaries (red square) and fragmented semantic regions (blue circle)

How to segment on VG floorplans?

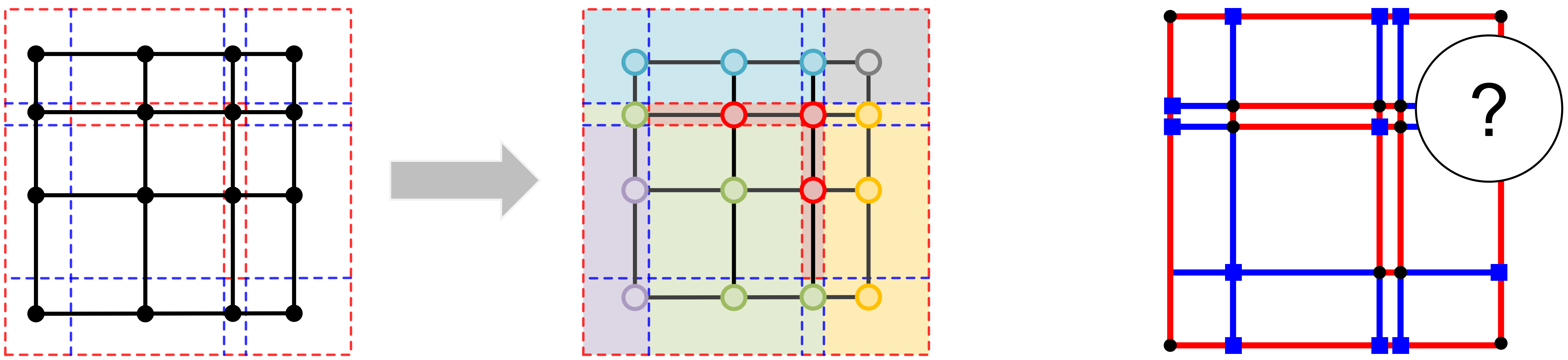
Graph duality



- Vector graphics can be viewed as graphs (drawing primitives as edges, endpoints as vertices)
- primitive extension leads to over-partitioned polygonal regions
- segmentation on the dual graph

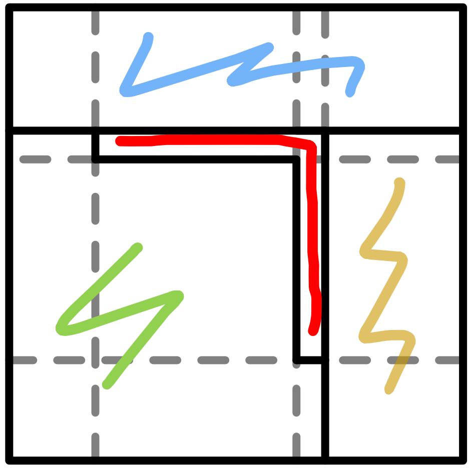
How to leverage both graphs?

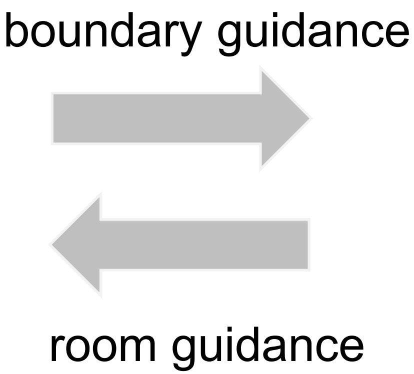
Unsatisfactory results solely based on dual graph

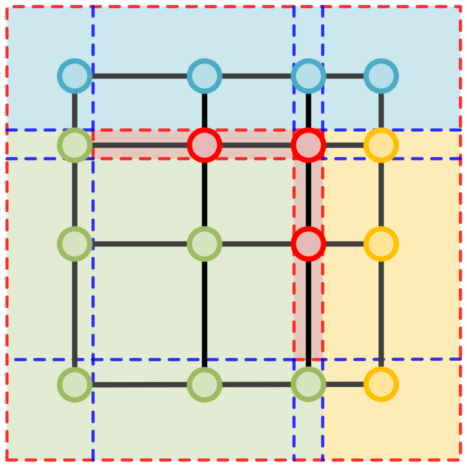


- fragmented segmentation still exists
- Remainder to use information in the primal graph

How to leverage both graphs?



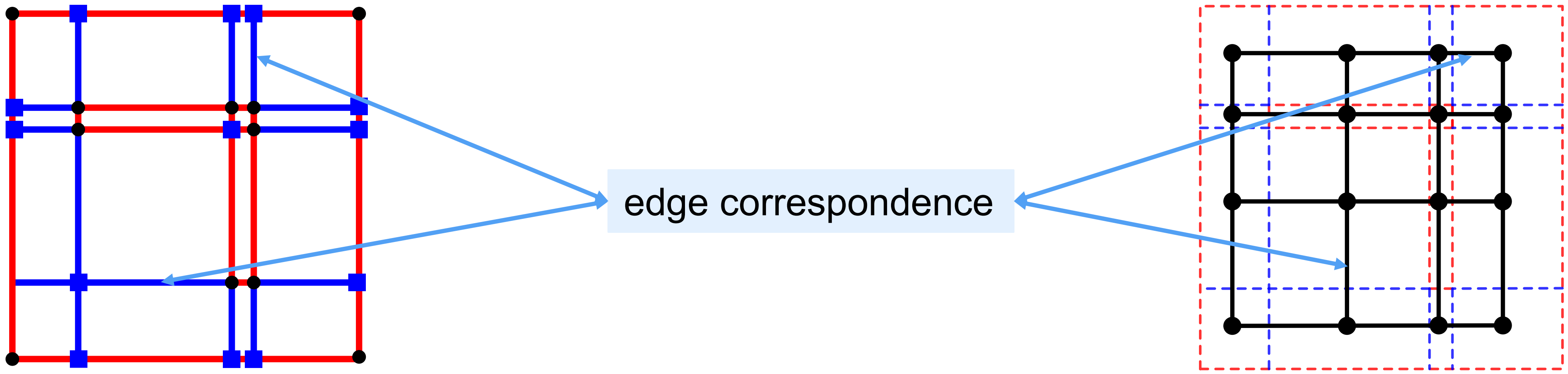




- boundary lines pose room separation guidance
- lines (including extended lines) between different rooms are potential boundaries

How to leverage both graphs?

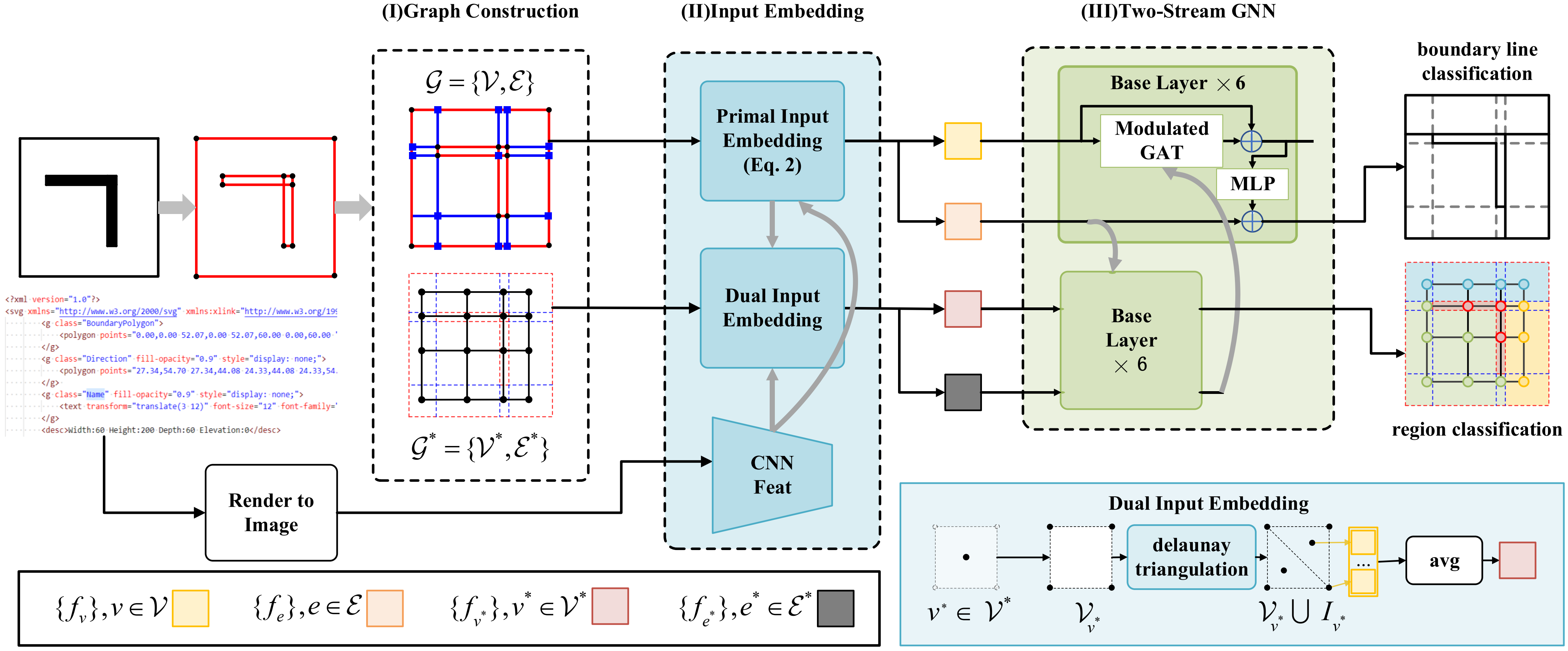
Edge correspondence as a bridge between two graphs



- graph neural network incorporates graph information into edges
- Information exchange can be established by edge correspondence

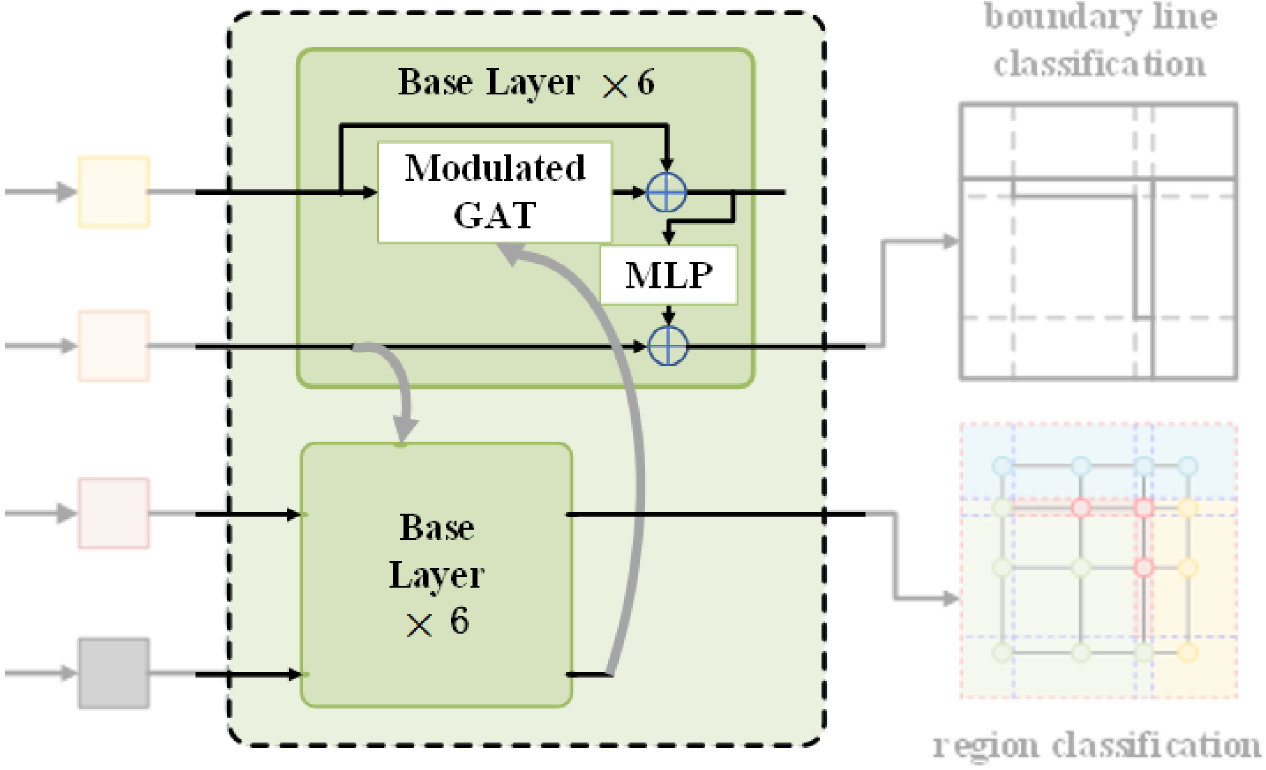
Methodology

Framework of VectorFloorSeg



Methodology

Modulated GAT



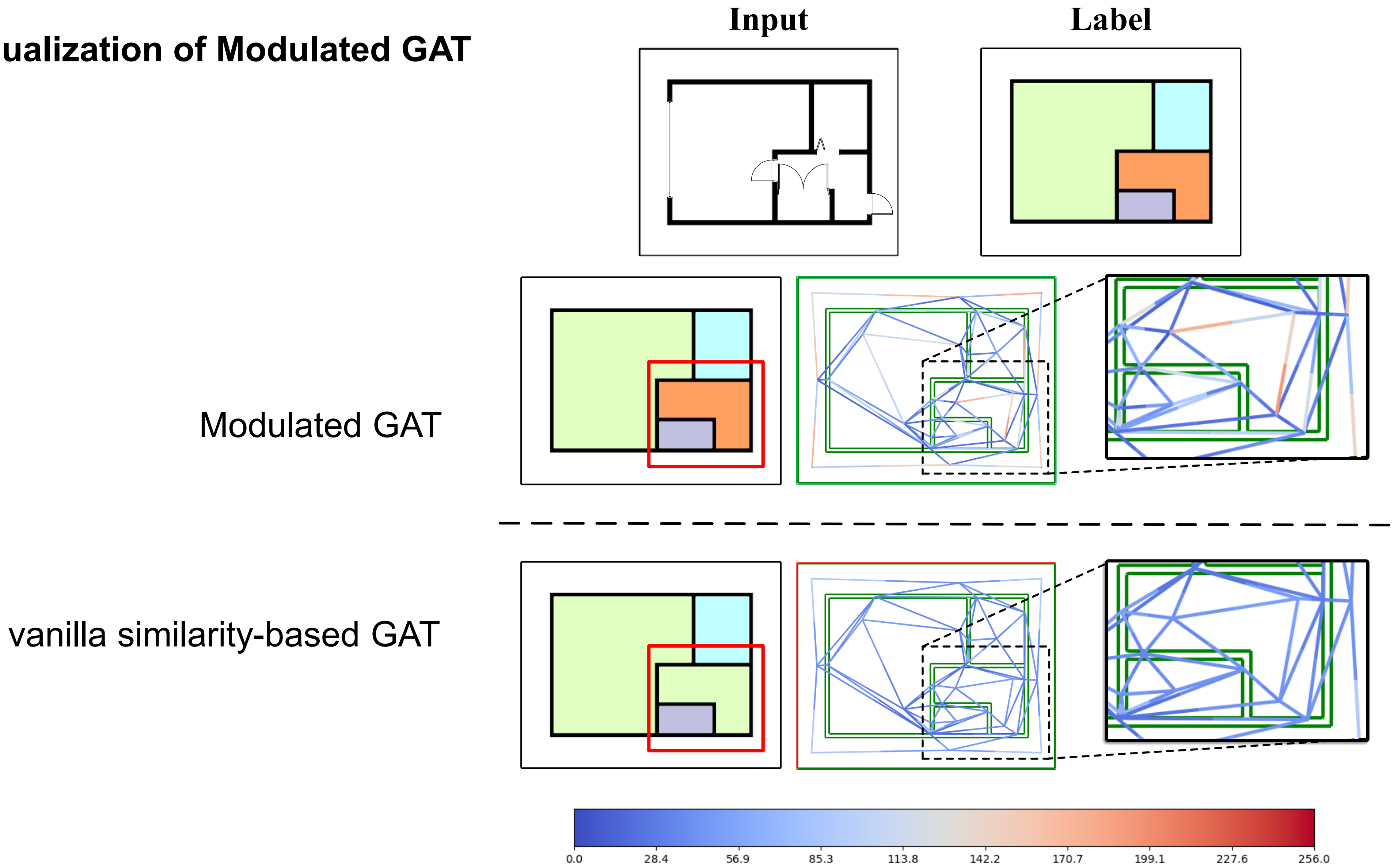
$$f_{v_i}^{l+1} = f_{v_i}^{l} + \Theta_{\gamma}\Big(\alpha_{ii} W_v f_{v_i}^{l} + \sum_{v_j \in N_{v_i}} \alpha_{ij} W_v f_{v_j}^{l}\Big)$$

$$\alpha_{ij} = \frac{\exp\left([W_q f_{v_i}^{l}]^{T} \cdot \boxed{g(f_{*e_{ij}}^{l})} \cdot [W_k f_{v_j}^{l}]\right)}{\sum_{v_m \in N_{v_i} \cup \{v_i\}} \exp\left([W_q f_{v_i}^{l}]^{T} \cdot g(f_{*e_{im}}^{l}) \cdot [W_k f_{v_m}^{l}]\right)}$$

- calculate adaptive attention weights using edge feature from another stream of graph

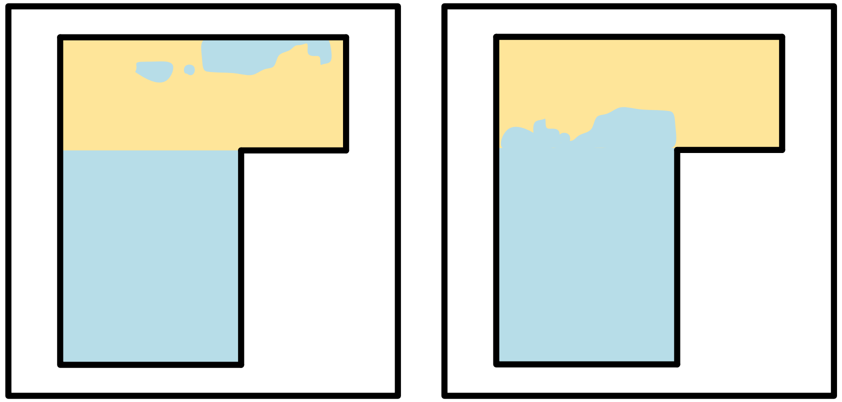
Experiments

Visualization of Modulated GAT



Experiments

Measurement: Room integrity



(a) (b)

metric	value-(a)	value-(b)
mIoU	81.67	81.67
mAcc	90.00	90.00
RI	51.43	81.67

- (a) is less tolerable than (b) for downstream applications
- Room Integrity metric is proposed to penalize fragmented segments, by measuring the bipartite matching between predicted rooms and groundtruths

Experiments

Outperforming other methods on two roughcast floorplan dataset

Methods	Backbone	Params(M)	GFLOPs	mIoU	mAcc	RI
DFPR [39]	VGG-16	28.91	223.39	69.47	81.5	50.96
Ours	–	22.57	91.63	75.56	87.32	80.48
DeepLabV3+ [6]	ResNet-50	43.59	176.76	74.59	83.46	49.99
DNL [36]	–	50.02	200.16	71.61	81.90	51.99
UperNet [35]	–	66.41	237.02	73.23	83.84	63.46
Ours	–	33.29	115.15	**79.77**	**88.41**	**84.67**
OCRNet [38]	ResNet-101	55.51	231.11	77.98	85.34	70.82
Ours	–	51.39	193.01	**81.38**	**89.86**	**86.20**

Table 1. Comparison results on R2V.

Methods	Backbone	*val*-set			*test*-set		
		mIoU	mAcc	RI	mIoU	mAcc	RI
DFPR [39]	VGG-16	49.68	60.37	38.44	47.73	58.68	38.57
Ours	–	60.27	72.32	66.89	57.48	69.89	64.09
DeepLabV3+ [6]	ResNet-50	60.46	73.18	38.41	58.18	71.75	35.16
DNL [36]	–	59.61	72.15	42.38	55.29	68.36	40.49
UperNet [35]	–	59.31	72.22	45.90	57.04	70.50	44.71
Ours	–	**63.09**	**75.48**	**69.74**	**61.35**	**74.45**	**67.99**
OCRNet [38]	ResNet-101	60.44	72.94	43.99	57.13	70.62	41.89
Ours	–	**64.36**	**76.98**	**69.55**	**62.49**	**75.48**	**67.51**

Table 2. Comparison results on CubiCasa-5k.

Experiments

Ablation studies on network setting

Vertex Embedding			Architecture		mIoU	mAcc	RI
pos.	vertex samp.	backbone	p-stream	GAT Ours			
					45.92	60.30	53.45
✓					59.01	73.29	65.12
✓	✓				61.32	75.03	67.35
✓	✓	✓			75.60	85.34	81.03
✓	✓	✓	✓		77.08	85.77	83.04
✓	✓	✓	✓	✓	79.77	88.41	84.67

Table 3. Ablation studies on vertex embedding and the network architecture. *pos.*: using sine positional encodings in Eq. 2. *vertex samp.*: using interior sampling (Eq. 3) to compute the dual vertex embedding. *backbone*: adding image features from rasterized floorplans. *p-stream*: with the primal stream that predicts boundary lines. *GAT Ours*: using the modulated GAT layer that enables feature interaction between two streams; if not marked, the GAT layer degrades to vanilla similarity-based attention [31] by replacing the learnt weight matrix $g(f^l_{*e_{ij}})$ in Eq. 7 with an identity matrix. If none of the above is marked, pe_v of Eq. 2 is set to zero, and x_v is a learnable embedding indexed by vertex category (i.e. primal vertex or sampled vertex). $W_e(\cdot)$ and $W_{e^*}(\cdot)$ of Eqs. (4,5) are also set to zero.

dual map	edge ini.	mIoU	mAcc	RI
		77.38	86.37	83.95
✓		78.22	87.67	83.06
✓	✓	79.77	88.41	84.67

Table 4. Ablation studies on edge features. *dual map*: using dual edge features from the other stream (cf. Fig. 2(III)); if not marked, the modulated GAT in each stream uses edge features from its own stream instead. *edge ini.*: using the edge embeddings of Eq. 4 and 5; if not marked, the edge embeddings are set to zero.

中国科学院大学
University of Chinese Academy of Sciences

Thanks for listening!

https://github.com/DrZiji/VecFloorSeg