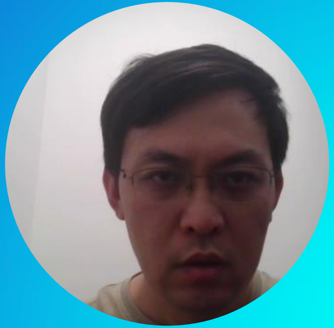
SEMANTIC-EMBEDDED KNOWLEDGE ACQUISITION AND REASONING FOR IMAGE SEGMENTATION

Wei Liu, Huigang Zhang, Xiaojie Xia, Liuan Wang and Jun Sun

Fujitsu R&D Center, Co., LTD



Ideas

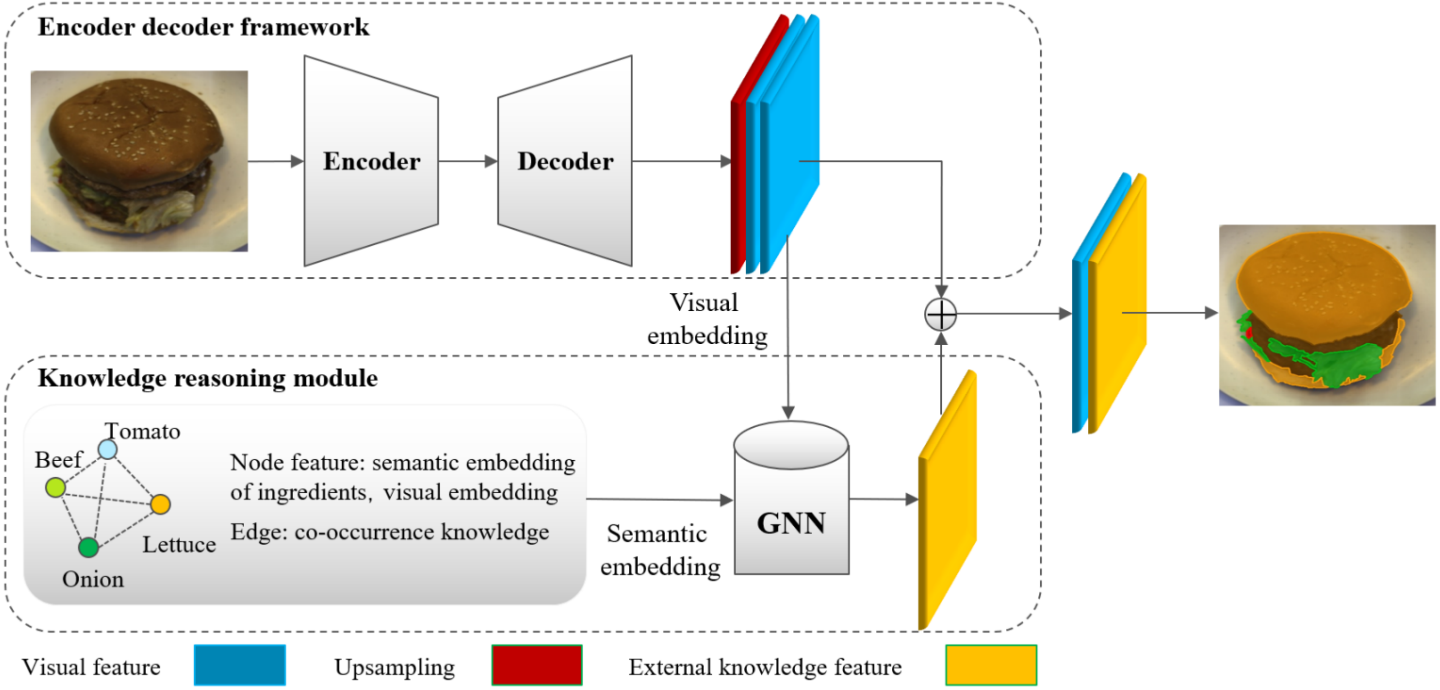
- Previous work focus on improving visual features for segmentation
- Our methods use correlation between objects
  - Introduce a knowledge reasoning module (KRM) for external knowledge aggregation
  - Use GNN to aggregate the knowledge feature
  - Enhance the visual feature with external knowledge feature for final segmentation

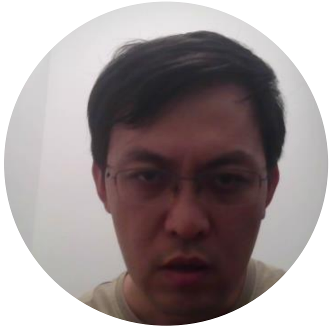
Main contribution

- A novel knowledge reasoning module is proposed to introduce the external knowledge feature. GCN is introduced to aggregate the external knowledge to enhance the features for segmentation task.
- The proposed module can be flexibly integrated with other semantic segmentation framework.
- The effectiveness of the proposed method is proved by the evaluation on the public dataset FoodSeg103 and Cityscapes

Our methods -architecture

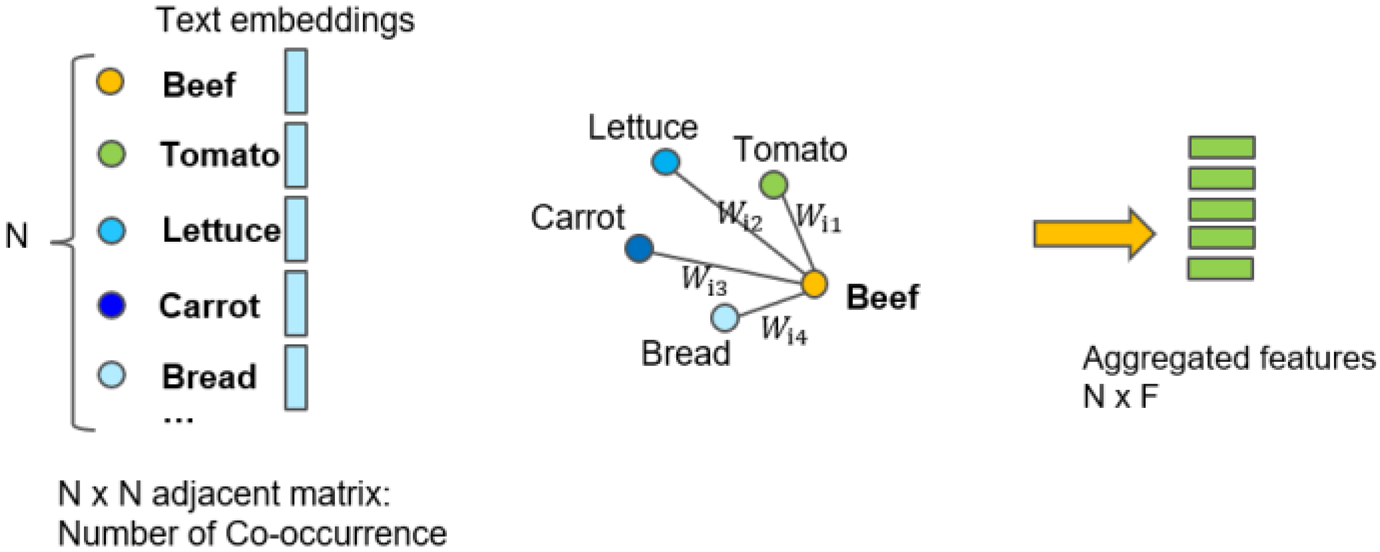
- Our method is designed on the basic encoder decoder framework and can be flexibly applied to all networks based thereon



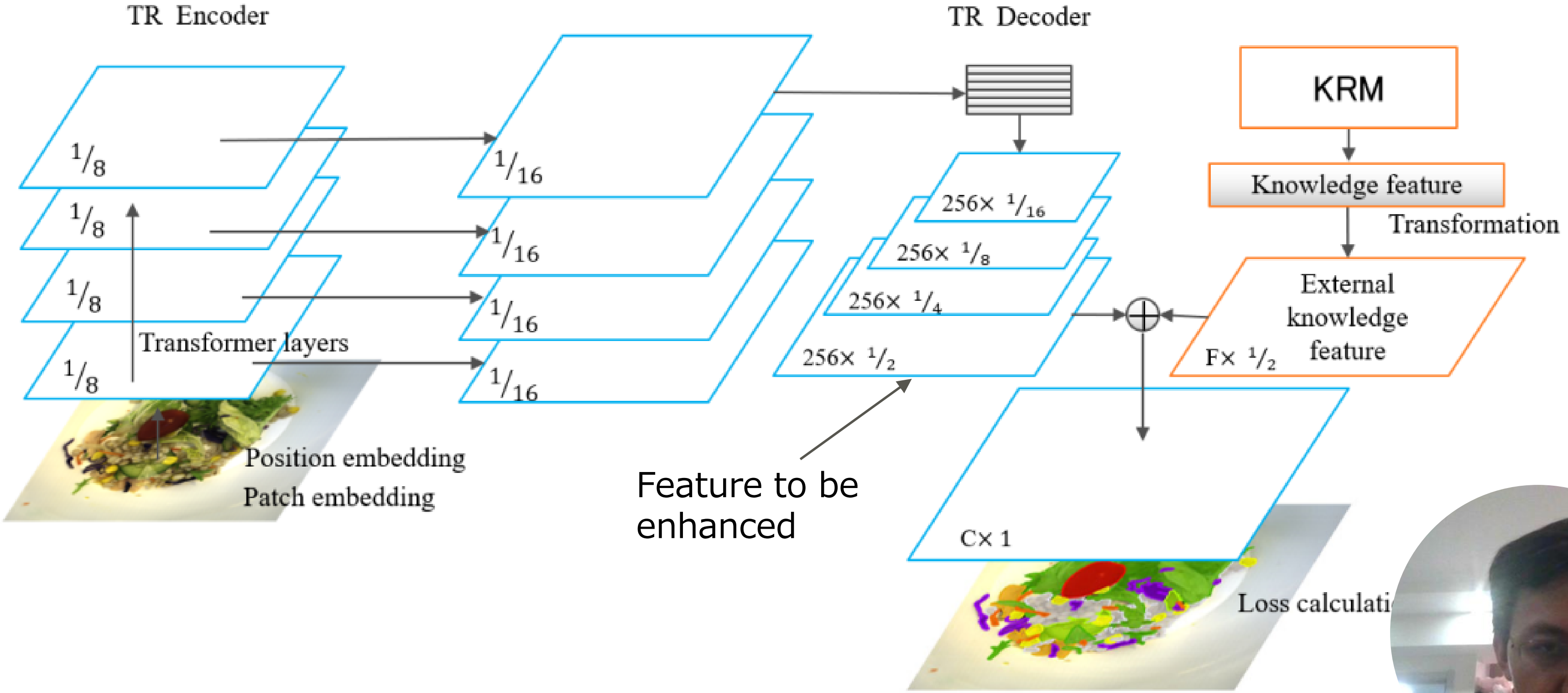


Methods: Knowledge reasoning module

- Node representation
  - comes from word embedding(by Glove or CLIP) of the object name or visual embedding of convolution transformation weights
- Edge representation
  - The edge between two nodes represent the relationship of two objects in the image



SETR plug-in: proposed knowledge reasoning module



Feature to be enhanced

Methods: Knowledge feature transformation

○ Detailed transformation

Foodseg103 dataset :

Feature to be enhanced
[8,**104**,W=768,H=768]

Nodes:
Recipe ingredients 103

cheese
……
egg

Edge：103X103
Adjacent matrix
occurrence in the foodseg103
or commonsense ingredients matches

GCN

Semantic Knowledge

[103,F = 64]
↓ Add background
[104,F]
↓ Repeat H=256
[H,104,F]

Feature transformation

Iterate each batch
[104,W,H]
↓ Transpose（0,2)
[H,W,104]
↓ bmm [H,W,104] *[H,104,F]
[H,W,F]
↓ Cat [H,W,512]
[H,W,F+512]
↓ transpose
[F+512,W,H]
↓ Batch concat
[8, F+512,W,H] → **[8,104,W,H**
1X1 conv

[8,**104,W,H**] Oringinal feature
↓ 1X1 conv
[8,**512,W,H**]

Loss manipulate

EXPERIMENTS

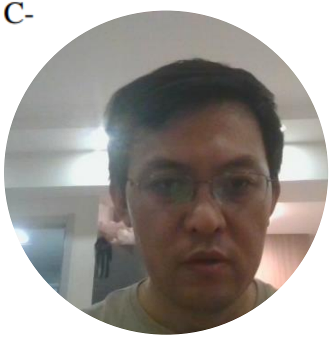
- Our experiment results outperform the baseline methods on the food dataset FoodSeg103 and Cityscapes, and demonstrate the effectiveness of our proposed method

Method	mIoU	Model size
FPN [3](ResNet50)	27.3	218M
FPN-KRM(CLIP+GCN)	**28.3**	227M
CCNet [4](ResNet50)	35.1	381M
CCNet-KRM(CLIP+GCN)	**36.4**	399M
SETR [6](Vit-16/B)	44.6	776M
SETR-KRM(CLIP+GCN)	**45.7**	805M

Table 1. Segmentation evaluation results of MIou on FoodSeg103 dataset

Method	mIoU	Model size
FPN [3](ResNet50)	74.5	218M
FPN-KRM(CLIP+GCN)	**76.4**	227M
CCNet [4](ResNet50)	79.3	381M
CCNet-KRM(CLIP+GCN)	**80.5**	399M
SETR [6](Vit-16/B)	78.1	776M
SETR-KRM(CLIP+GCN)	**79.7**	805M

Table 2. Segmentation evaluation results of MIou on Cityscapes(test), training schedule with 80k



Ablation study

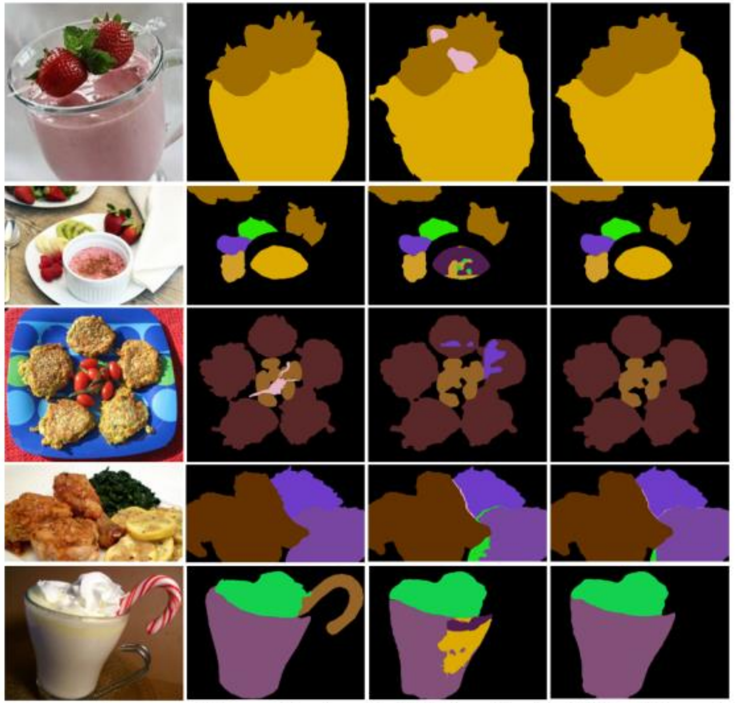
- Experimental results shows that a combination of CLIP based module and visual knowledge module can achieve the best mIoU performance

Method	mIoU
SETR[6](Vit-16/B)	44.6
SETR-KRM(GCN+CLIP)	45.7
SETR-KRM(GCN+CLIP+GLOVE)	45.3
SETR-KRM(GCN+CLIP+VISUAL)	**45.8**
SETR-KRM(GCN+CLIP+GLOVE+VISUAL)	45.2

Table 4. Results of different modules on FoodSeg103

Visualization

- Some visualization examples are demonstrated. SETR is used as the baseline method in the third column, the fourth column is proposed method
- From the visualization observations, the SETR-knowledge reasoning module (KRM) achieves better performance and more detailed results



(a).source image (b).ground truth (c). SETR baseline (d).SETR with knowledge

Fig. 4. Visualization of testing samples in FoodSeg103: SETR with knowledge achieves better performance.

Conclusion

- In this paper, a semantic segmentation framework that incorporates the KRM for image segmentation task is introduced
- GCN is used to aggregate the external knowledge feature
- Experiments show that our proposed method outperforms the baseline methods on the FoodSeg103 and Cityscapes datasets
- We hope our proposed method can contribute to the community in semantic segmentation tasks

Thank you for your watching