

# Contextual Associated Triplet Queries for Panoptic Scene Graph Generation

Jingbin Xu , Junwen Chen , Keiji Yanai

The University of Electro-Communications

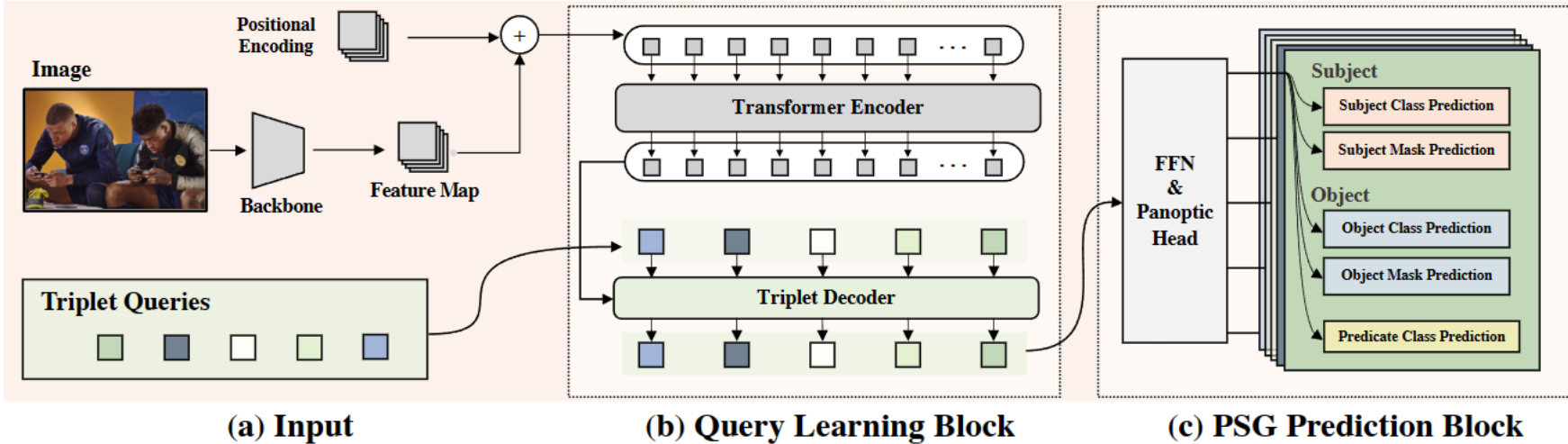


# Previous Works

method	drawback
--------	----------

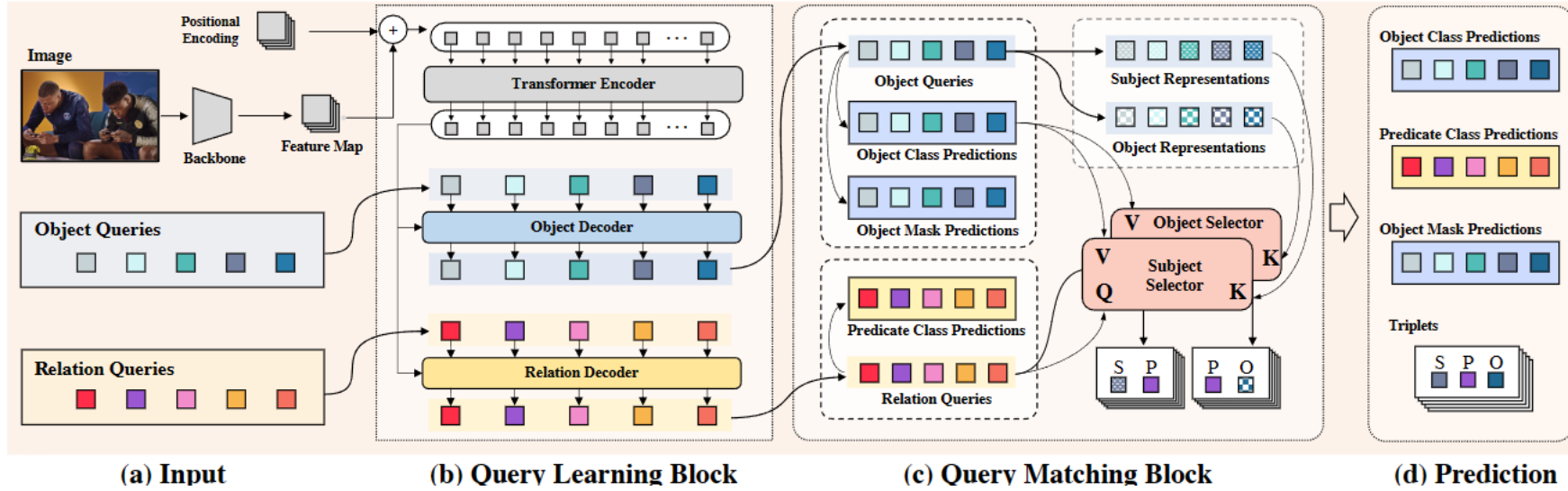
**PSGTR**  
[1]

A single type query introduces instability in the learning process.



**PSGFormer**  
[1]

Easy matching make hindering the identification of the best match



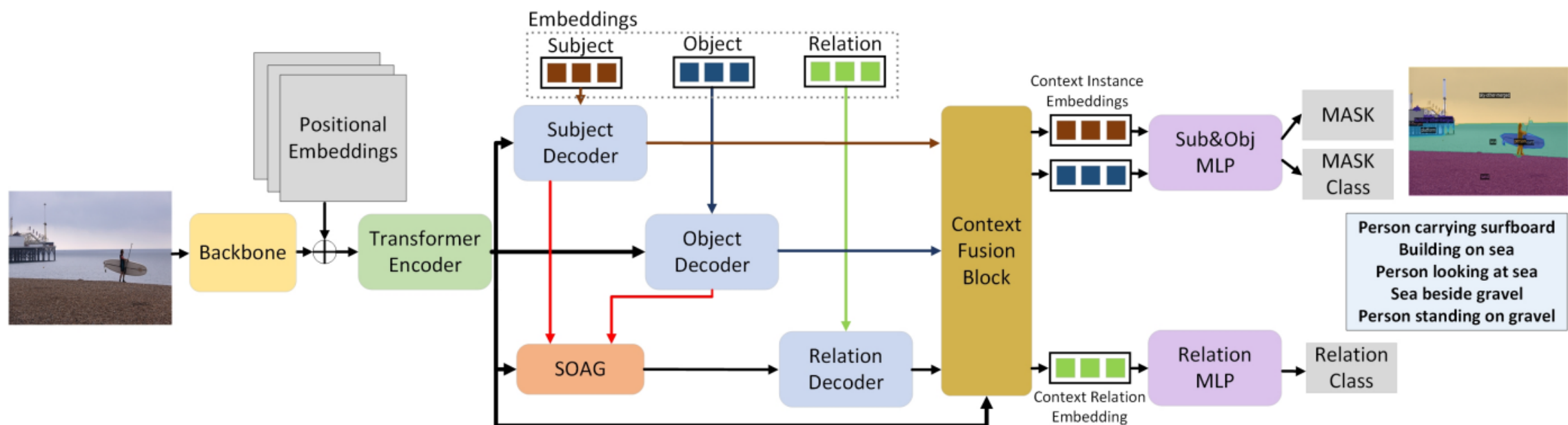
[1] Jingkan Yang, Yi Zhe Ang, Zujin Guo, Kaiyang Zhou, Wayne Zhang, and Ziwei Liu. Panoptic scene graph generation. In European Conference on Computer Vision, pages 178–196. Springer, 2022

# Contributions

## ◆ Triple Decoder

## ◆ Subject-Object Attention Guide (SOAG)

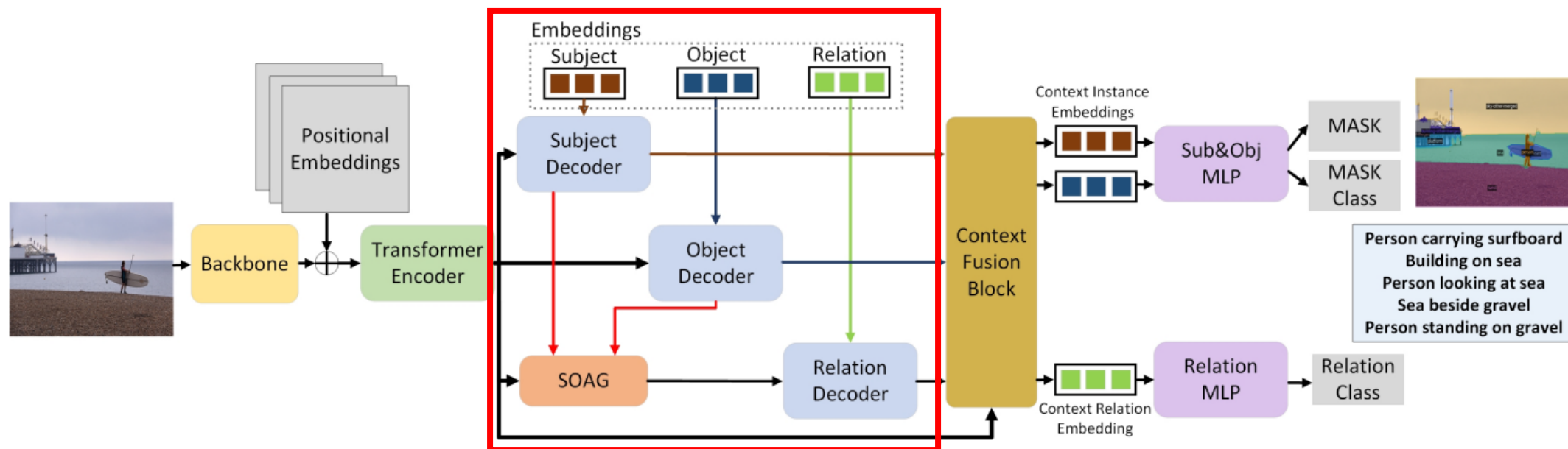
## ◆ Context Fusion Block(CFB)



# Introduction

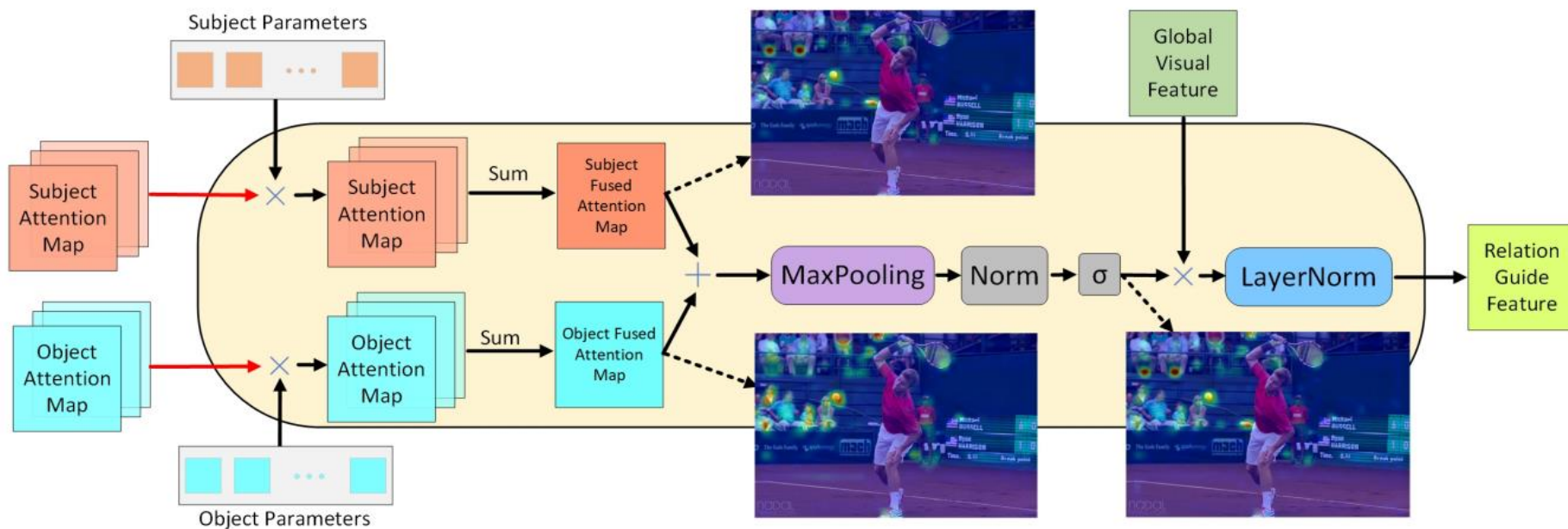
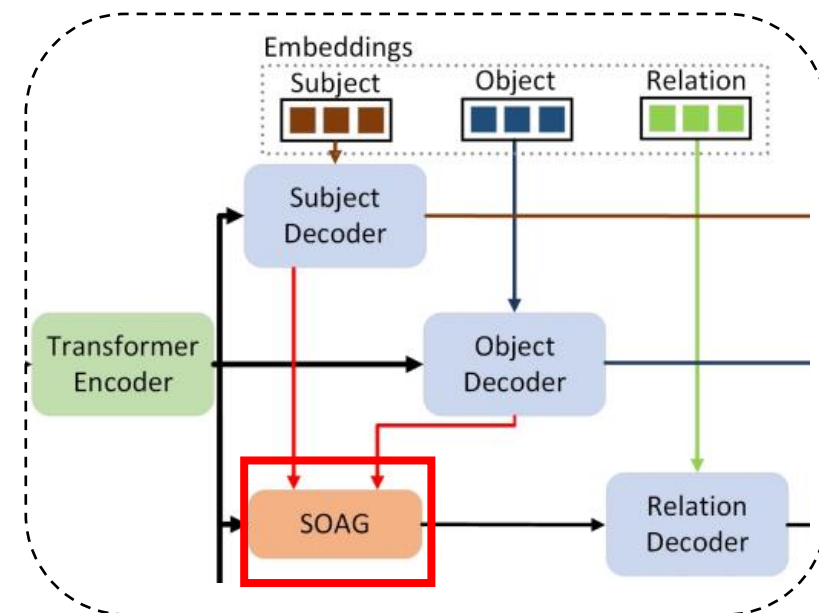
## ➤ Triple Decoder

- Decode triples <Subject, Relation, Object> separately



# Introduction

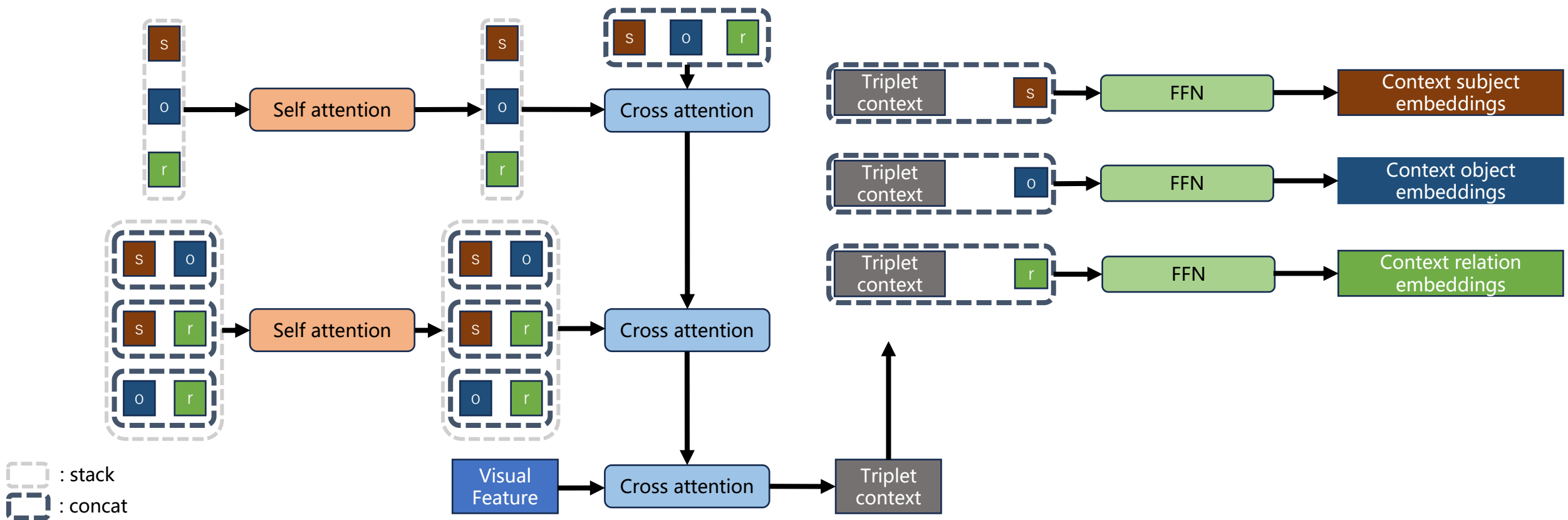
- Subject-Object Attention Guide (SOAG)
  - Refined source input for relationship branch.



# Introduction

## ➤ Context Fusion Block(CFB)

- Add contextual information to each embedding



# Experimental Results

➤ Compare CATQ with PSG baseline methods on the OpenPSG dataset.

Method	Backbone	<i>Recall@20</i>	<i>Recall@50</i>	<i>Recall@100</i>	<i>mRecall@20</i>	<i>mRecall@50</i>	<i>mRecall@100</i>	<i>PQ</i>
<b>Two-stage</b>								
IMP [15]	ResNet-50	16.5	18.2	18.6	6.5	7.1	7.2	40.2
MOTIFS [17]	ResNet-50	20.0	21.7	22.0	9.1	9.6	9.7	40.2
VCTree [13]	ResNet-50	20.6	22.1	22.5	9.7	10.2	10.2	40.2
GPSNet [10]	ResNet-50	17.8	19.6	20.1	7.0	7.5	7.7	40.2
<b>One-stage</b>								
PSGTR <sup>†</sup> [16]	ResNet-50	28.4	34.4	36.3	16.6	20.8	22.1	13.9
PSGFormer <sup>†</sup> [16]	ResNet-50	18.0	19.6	20.1	14.8	17.0	17.6	36.8
<b>CATQ</b>	ResNet-50	<b>34.8</b>	<b>39.7</b>	<b>40.3</b>	<b>20.9</b>	<b>24.9</b>	<b>25.2</b>	<b>35.9</b>

# Experimental results

## ➤ Ablation for SOAG

SOAG	<i>Recall@20</i>	<i>Recall@100</i>	<i>mRecall@20</i>	<i>mRecall@100</i>
✓	<b>34.75</b>	<b>40.26</b>	<b>20.87</b>	<b>25.19</b>
-	33.44	38.89	20.23	24.20



# Conclusion

- Proposed a Panoptic Scene Graph Generation method CATQ
  - ✓ Triple Decoder
  - ✓ Subject-Object Attention Guide (SOAG)
  - ✓ Context Fusion Block(CFB)
- SOTA performance with less training session.

**Thanks for your listening**

