

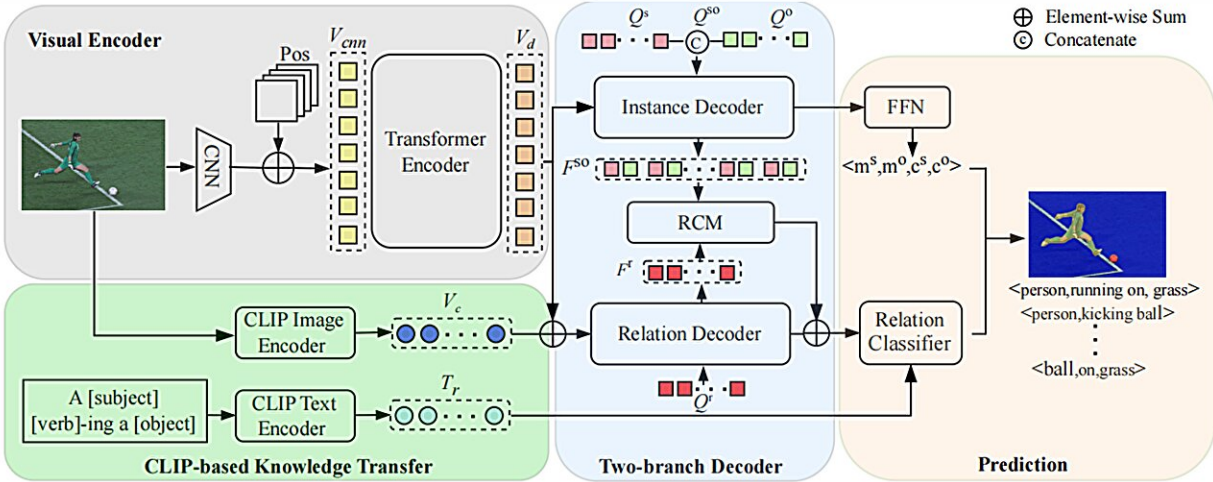
Researchers improve scene perception with innovative framework

May 13 2024, by Zhang Nannan

CKT-RCM: CLIP-Based Knowledge Transfer and Relational Context Mining for Unbiased Panoptic Scene Graph Generation

Nanhao Liang, Yong Liu, Wenfang Sun, Yingwei Xia, Fan Wang

Hefei Institutes of Physical Science, Chinese Academy of Sciences;
University of Science and Technology of China



The proposed Clip-based Knowledge Transfer and Relational Context Mining (CKT-RCM). Credit: Wang Fan

Led by Prof. Liu Yong from the Hefei Institutes of Physical Science of the Chinese Academy of Sciences, researchers have proposed a novel framework, called Clip-based Knowledge Transfer and Relational Context Mining (CKT-RCM), to address the long-tail distribution

problem in computer vision.

The [results](#) were published in *IEEE International Conference on Acoustics, Speech and Signal Processing*.

Panoptic Scene Graph (PSG) is a prominent research direction within scene graph generation, which requires comprehensive output of all relationships in an image alongside accurate segmentation for object localization. PSG aims to improve the understanding of scenes by [computer vision](#) models and to support downstream tasks such as [scene](#) description and visual inference.

In this study, the researchers explored how humans perceive object relationships, presenting two key perspectives. People anticipated the object relationships based on common sense or [prior knowledge](#). They also inferred relationships based on contextual information between subjects and objects.

These perspectives underscore the importance of leveraging prior knowledge: one involves correcting data biases using external data previously observed by humans, while the other relies on the prior distribution of conditions between objects.

"Therefore, we believe that sufficient prior knowledge and contextual information are crucial for PSG prediction," said Dr. Wang Fan, a member of the team.

They developed this network framework CKT-RCM. Based on the pre-trained vision-language model CLIP, CKT-RCM facilitates [relationship](#) inference during PSG processes. It integrates a cross-attention mechanism to extract relational context, ensuring a balance between value and quality in relational predictions.

This study contributes to the understanding and perception of scenes by robots and autonomous vehicles.

More information: Nanhao Liang et al, CKT-RCM: Clip-Based Knowledge Transfer and Relational Context Mining for Unbiased Panoptic Scene Graph Generation, *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2024). [DOI: 10.1109/ICASSP48485.2024.10446810](https://doi.org/10.1109/ICASSP48485.2024.10446810)

Provided by Chinese Academy of Sciences

Citation: Researchers improve scene perception with innovative framework (2024, May 13) retrieved 23 June 2024 from <https://techxplore.com/news/2024-05-scene-perception-framework.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.