Referring Expression Counting

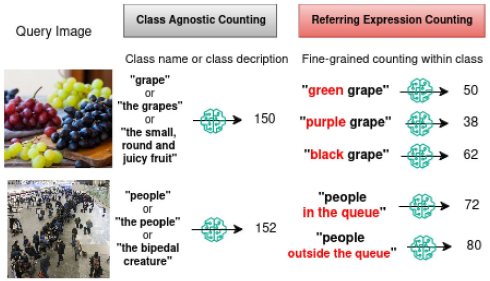
Siyang Dai1 , Jun Liu1*, Ngai-Man Cheung1*

Published on 21 June 2024

What is the problem?

Existing counting tasks (class-specific and class-agnostic) are limited to the class level, which don't account for fine-grained details within the class. In real applications, it often requires in-context or referring human input for counting target objects.

- traffic monitoring and crowd analysis
- customer demographic analysis
- inventory count
- livestock management
- wildlife monitoring



What has been done earlier?

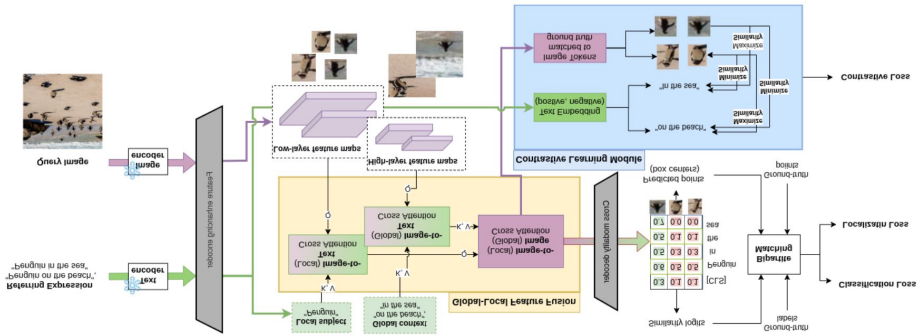
1. **Category-specific counting**: Models were trained to count objects within a specific category (e.g., people, vehicles) but performed poorly when encountering unseen categories.
2. **Class-agnostic counting**: This evolved to allow models to count objects from both seen and unseen categories, treating all objects within a class similarly without considering their specific attributes. Existing methods focused on counting at the class level and did not differentiate objects based on finer details like attributes (e.g., color, size, or location within the same class).

Tusar Kumar Nayak,B421065

Referring Expression Counting

Remaining Challenges:

Lack of Fine-Grained Counting: Existing counting methods cannot differentiate between objects of the same class with different attributes (e.g., people with different clothing or in different locations).
Contextual Understanding: Current models struggle to handle in-context human inputs that specify detailed attributes, limiting their use in real-world scenarios like traffic monitoring, crowd analysis, or retail applications.
Ineffective Object Attribute Recognition: Previous methods are limited to class-specific or class-agnostic counting but do not account for objects' varying attributes within the same class.



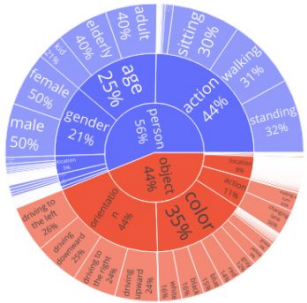
Novel Solution Proposed:

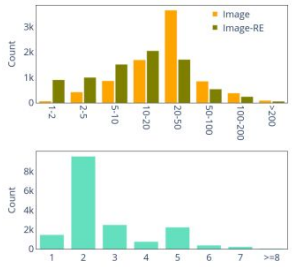
• We introduce Referring Expression Counting which takes referring expressions and a query image as inputs and outputs the target object count as well as the location.

The REC-8K Dataset

• We create REC-8K, a novel benchmark for evaluation of the REC task. Images are sourced from the domains of crowd analysis, traffic surveillance, retail and warehousing, etc.

• REC-8K contains 8011 images with 17122 referring expressions and a total of 286621 point annotations.





Tusar Kumar Nayak,B421065