GRE: Evaluating Computer Vision Models on Generalizability Robustness and Extensibility

AARYA UPADHYA (PES1UG23AM006)

AARAV AADARSH (PES1UG23AM003)

MOTIVATION

- Evaluating computer vision models on a test dataset that is sampled from the same distribution as the train dataset leads to an overstatement of model's performance. Hence, we propose that computer vision models should be tested on unseen composition of data
- Here comes the GRE Method,
    - An image has 2 parts the foreground and the background scene, propose a method of introducing new compositions of objects and scenes that are already in the dataset to test the models called as concept learning
    - Generalizability: Introduce objects of the same class
    - Robustness: Introduce a new scene to the object
    - Extensibility: Introduce object of a different class

DATASET

- Dataset used is Synthetic dataset created during pipeline as datasets respected the paper have been deprecated or URL Extraction for images for VQA models at high levels are not possible
- Working of the Synthetic dataset
    - We try to simulate different background selections, Main object placements and distractor objects and different light variations
    - Each object is drawn symbolically, Metadata is recorded as well for each
- Question Answer Generation
    - QA pairs very simple ones are generated
    - Colour: "What is the colour of the cup" as an example
    - This ensures us to do low level GRE study

MODEL SELECTION:

- BLIP Models are specifically designed to bridge vision and language tasks making them highly suitable for synthetic GRE datasets
- BLIP Models are pretrained on large image-text pairs, learning how to visual elements mapped to descriptive language, this allows them to answer QA pairs about objects well

PERFORMANCE

- The metrics used for each dataset are (Original VQA, GRE-G, GRE-R, GRE-E)
- Prediction matching, GRE score, Performance gap are all calculated
- The stats do not resemble the same as the paper due to the low real world quality resemblance from the synthetic dataset