GRE: Evaluating Computer Vision Models on Generalizability Robustness and Extensibility

AARYA UPADHYA (PES1UG23AM006)

AARAV ADARSH(PES1UG23AM003)

MOTIVATION

- Traditional Computer Vision models perform well only on datasets similar to training data -> they overestimate the performance
- Our project works on unseen compositions of objects and scenes
- GRE FRAMEWORK
  - Generability: Introduce objects of the same class
  - Robustness: Introduce same object in new scenes
  - Extensibility: Introduce objects from different classes

  AIM: Asses models' ability to generalize , reason and remain coherent under new conditions

DATASET

- Created a custom dataset (since original GRE/VQA datasets are deprecated) or are way out of scope for subset training
- Simulates : Different backgrounds , object placements , lighting and distractor objects
- QA PAIR EXAMPLE
  - “what color is the cup”?
  - Used for low level GRE evaluation with recorded metadata

MODEL AND PIPELINE

- Model Used: BLIP (Bootstrapped Language Image Pretraining)
- Chosen for strong Vision language understanding
- Pipeline:
  - Vision Encoder: Extracts the spatial and semantic features
  - Text Encoder: Converts the questions into contextual embeddings
  - Cross-Attention: Aligns image and text features
  - Decoder: Generates the final Textual answer

EVALUATION AND RESULTS

- Metrics: Prediction Match% , GRE score , Performance GAP
- Datasets: Original VQA , GRE-G , GRE –R , GRE-E
- Findings:
  - Model struggles to maintain accuracy across GRE subsets
  - Performance gaps due to low realism in the synthetic data
  - Demonstrates need for more diverse testing environments