National University of Computer & Emerging Sciences
Karachi Campus

PROJECT PROPOSAL

Course Name: Artificial Neural Networks (ANN)

Course Instructor: Sir Bilal Ahsan

"Zero-Shot Image Classification for Assistive Technologies"

Group Members

Name	Roll No	Class
Mirza Abdul Wasay Uddin	22K-4087	BAI-6A
Mudasir	22K-8732	BAI-6A
Irteza	22K-8731	BAI-6A
Nihal Ali	22K-4054	BAI-6A

Introduction

- ## Background

Image classification has been transformative in many applications, especially for assistive technologies. This challenge directly addresses the needs of individuals who are blind by leveraging a dataset of images captured by blind photographers. The VizWiz Zero-Shot Image Classification Challenge offers a unique opportunity to advance computer vision methods in real-world, low-quality imaging conditions.

- ## Objective

Develop a robust image classification model that can accurately predict object categories in images captured by blind individuals, even when these images present domain shifts compared to traditional datasets.

Problem Statement

- ## Challenge Overview

  The VizWiz Challenge requires predicting the presence or absence of 200 object categories in 8,900 real-world images taken by blind photographers. The inherent issues include variable image quality, unconventional framing, and challenging lighting conditions.

  Challenge Link : https://vizwiz.org/tasks-and-datasets/image-classification/

- ## Key Difficulties

  Models must generalize from training domains (e.g., ImageNet) to the visual characteristics of the VizWiz images.

- ## Image Quality

  Blurry or low-contrast images require robust pre-processing and feature extraction techniques.

- ## Multiple Objects

  Images may contain several objects with potential overlaps or occlusions, necessitating careful label assignment.

Objectives and Expected Outcomes

- ## Primary Objectives

Develop a zero-shot image classification model that leverages transfer learning to generalize to unseen visual domains.

Address quality degradation and domain-specific challenges using innovative preprocessing and domain adaptation strategies.

Enhance model performance through strategic data augmentation, semantic embedding, and cross-modal learning techniques.

- <u>Expected Outcome</u>

A model that achieves superior classification accuracy on the VizWiz dataset.

A set of insights and methodologies for domain adaptation in low-quality imaging scenarios.

A contribution to assistive technologies that improve accessibility for blind individuals.

VizWiz Classification Dataset (CVPR 2025)

The Dataset is selected from this Website (this is our challenge website as well).

- https://vizwiz.org/tasks-and-datasets/image-classification/

VizWiz-Classification Dataset

The VizWiz-Classification dataset includes:

- 8,900 images

You may download the individual sets of components listed below.

- train, validation, and test: raw images
- annotations.json: including the list of categories and images of our dataset.

Example code is provided to demonstrate how to parse the JSON files and transform predictions to an accepted file for evaluation on the EvalAI server.

The download files are organized as follows:

- JSON annotation record has the following format:

We Downloaded the Training Dataset and Upload into Drive to further use in Project implementation.

- https://drive.google.com/drive/folders/12b4Xm9NxRNe7I-dYxqrX6NUQ9irgSwJf?usp=sharing

Methodology and Strategies

- ## Data Preprocessing and Augmentation

1. ### Preprocessing

Noise Reduction & Contrast Enhancement, Apply filters and histogram equalization to improve the quality of low-resolution images.

2. ### Normalization

Standardize images to match the distribution of pre-trained models (e.g., ImageNet normalization).

3. ### Data Augmentation

Utilize rotations, scaling, and flipping to increase dataset diversity.
Simulate real-world distortions to train the model to be robust against image quality variations.

- ## Model Architecture and Transfer Learning

1. ### Model Selection

With Pre-trained CNNs, start with architectures such as ResNet, EfficientNet, or vision transformers that have been pre-trained on large datasets.

2. ### Zero-Shot Learning Frameworks

Integrate models like CLIP that leverage joint text-image embeddings, enabling the model to interpret object categories even without explicit examples from the VizWiz dataset.

3. ### Fine-Tuning

Fine-tune the pre-trained models on a curated subset of VizWiz images (or similar external data) using transfer learning.
Explore domain adaptation layers that help bridge the gap between the source (ImageNet) and target (VizWiz) distributions.

- Semantic and Cross-Modal Embeddings

  1. Semantic Embeddings

     Use word embeddings (e.g., Word2Vec, GloVe, or BERT representations) for the 200 object categories.
     Align visual features with semantic embeddings to enhance zero-shot capabilities.

  2. Cross-Model Matching

     Incorporate a matching mechanism where the model learns to correlate image features with corresponding semantic descriptors.
     Evaluate similarity using cosine distance or learned metric spaces.

Training Strategy and Evaluation

- Loss Functions

  Experiment with multi-label classification loss functions (e.g., binary cross-entropy) alongside contrastive losses to better align cross-modal features.

- Optimization

  Use adaptive optimizers (Adam, AdamW) with scheduled learning rate decay to ensure stable convergence.

- Regularization

  Employ dropout and data augmentation to avoid overfitting on the limited VizWiz data.

- Evaluation Metrics

  1. **Accuracy and F1-Score:** Evaluate the predictions based on both overall accuracy and F1-score to balance precision and recall.
  2. **Cross-Validation:** Use k-fold cross-validation to ensure the model's generalizability across different subsets of the data.

3. **Real-World Testing:** Validate the model's performance on a hold-out set of images with varying quality and context.

Conclusion

This project will enhance our understanding of deep learning and help us apply Artificial Neural Networks (ANN) techniques to real-world classification problems. We look forward to exploring different methods and improving our skills through this challenge.