Zero Shot Image Classification

Team AlgoUnlock

Introduction

TransZero++ is a novel zero-shot learning (ZSL) method that utilizes cross-attribute guided transformers to refine visual features and learn accurate attribute localization for key semantic knowledge representations.

TransZero++ consists of two sub-nets: an attribute→visual Transformer sub-net (AVT) and a visual→attribute Transformer sub-net (VAT). The AVT takes as input attribute descriptions and generates attribute-based visual features. The VAT takes as input visual features and generates visual-based attribute features. The two sub-nets are trained in a collaborative manner, such that they learn to refine each other's output.

Problem Statement

Currently the TransZero++ model is able to give SOTA results on benchmarks like CUB dataset but not on SUN dataset because the number of images in a single class are very less in the latter ones (approx. 16 training images per class). With this few number of training examples, the TransZero++ based ZSL does not give efficient semantic augmented visual embeddings. Therefore this heavily limits the performance of ZSL model

Proposed Solution

Since per class training examples are very limited, we can improve the model performance by data augmentation using generative models like Variational Auto Encoders (VAEs), Generative Adversarial Nerworks (GANs) or Generative Flows.

Blockers

We tried using GANs as a generative model but we were encountering many code errors

Other Contributions:

1) Currently the Attribute Regression Loss and Attribute based Cross Entropy Loss optimise the model on seen classes only. This causes the TransZero++ model to overfit on the seen classes. To mitigate this seen-unseen bias we have designed a “Debiasing Loss”.

$$\mathcal{L}_{deb} = ||\alpha_s - \alpha_u||_2^2 + ||\beta_s - \beta_u||_2^2$$

Here α_s , β_s are the mean and variance of seen prediction scores.

Here α_u , β_u are the mean and variance of unseen prediction scores.

2) We have introduced another loss function called “Correlation Loss”. The final class which we get after the zero shot prediction, should be closer to the classes which are similar to the actual class and farther from the remaining classes.

The similarity between the classes is calculated using cosine similarity.