Shifted Window Based Self-Attention via Swin Transformer for Zero-Shot Learning

Authors : Yasaswi Palagummi, Sareh Rowlands

Abstract : Generalised Zero-Shot Learning, often known as GZSL, is an advanced variant of zero-shot learning in which the samples in the unseen category may be either seen or unseen. GZSL methods typically have a bias towards the seen classes because they learn a model to perform recognition for both the seen and unseen classes using data samples from the seen classes. This frequently leads to the misclassification of data from the unseen classes into the seen classes, making the task of GZSL more challenging. In this work of ours, to solve the GZSL problem, we propose an approach leveraging the Shifted Window based Self-Attention in the Swin Transformer (Swin-GZSL) to work in the inductive GSZL problem setting. We run experiments on three popular benchmark datasets: CUB, SUN, and AWA2, which are specifically used for ZSL and its other variants. The results show that our model based on Swin Transformer has achieved state-of-the-art harmonic mean for two datasets -AWA2 and SUN and near-state-of-the-art for the other dataset - CUB. More importantly, this technique has a linear computational complexity, which reduces training time significantly. We have also observed less bias than most of the existing GZSL models.

Keywords : generalised, zero-shot learning, inductive learning, shifted-window attention, Swin transformer, vision transformer

Conference Title : ICIAP 2022 : International Conference on Image Analysis and Processing

Conference Location : Paris, France

Conference Dates : October 27-28, 2022