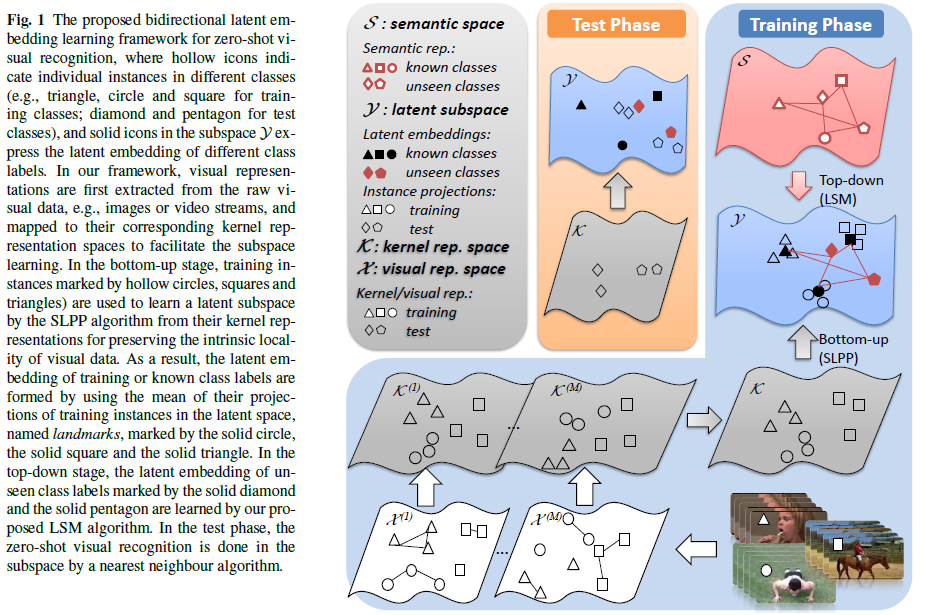
Zero-Shot Visual Recognition via Bidirectional Latent Embedding

Introduction

Zero-shot learning for visual recognition, e.g., object and action recognition, has recently attracted a lot of attention. However, it still remains challenging in bridging the semantic gap between visual features and their underlying semantics and transferring knowledge to semantic categories unseen during learning. Unlike most of the existing methods that learn either a direct mapping from visual features to their semantic representations or a common latent space by the joint use of visual features and their semantic representations, we propose a stagewise bidirectional latent embedding framework for zero-shot visual recognition. In the bottom-up stage, a latent embedding space is first created by exploring the topological and labeling information underlying training data of known classes via supervised locality preserving projection and the latent representations of training data are used to form landmarks that guide embedding semantics underlying unseen classes onto this latent space. In the top-down stage, semantic representations for unseen classes are then projected to the latent embedding space to preserve the semantic relatedness via the semi-supervised Sammon mapping with landmarks. As a result, the resultant latent embedding space allows for predicting the label of a test instance with a simple nearest neighbor algorithm. To evaluate the effectiveness of the proposed framework, we have conducted experiments on four benchmark datasets in object and action recognition, i.e., AwA, CUB-200-2011, UCF101 and HMDB51. The experimental results under comparative studies demonstrate that our proposed approach yields the state-of-the-art performance.

Framework



Datasets and splits

Experimental results

Codes and data