PhD

Learning Features by Watching Objects Move ax170412 cvpr17

2018/03/24 8:40 AM

The main idea of this paper seems to be to use motion segmentation as an alternative to pretext tasks as a means of performing unsupervised training.

The intuition is based on the fact that humans and other animals are able to segment objects based on their movement before they are able to do it based on their appearance;
for instance the same object may have different parts of different colours but these will all move together when the object moves so it becomes a more reliable cue for detecting that object.

The most commonly used pretext tasks in literature include the reconstruction tasks where auto encoders and noisy auto-encoders are used, generative tasks where GANs are used, and the rearrangement of shuffled patches but in an image and prediction of colour information from humans as well as visual predictions from Sound apparently.

For videos, the tasks include pixel level prediction in future frames, temporal reordering of shuffled frames, distinguishing between pairs of tracked patches in the same as well as different frames [not quite clear what this is].

Experiments have shown that using segmentations from the coco data set without class labels were enough to almost match the state of the art classification performance on imagenet and also showed that it is very resilient to noisy segments as well as to training only part of the network.
As compared to pretext tasks which degrade rapidly as more and more of the network is kept frozen, using only supervised segments results in a network whose performance remains relatively unaffected.