Vision-based Rapid Physical Attacks Against Generalizing Snake: an Efficient Approach for Robust Humanoid Detection

June 11, 2022

In this paper we propose a novel face recognition framework that combines deep learning (DL) with handcrafted features extracted from 3D data. Face models are conventionally represented as 3D models defined on 2D images. We extract these 3D features using 3D convolutions to integrate statistical and geometric features. We then perform a binary classification of whether fragment or not the 3D face is extracted via the DL. Finally, a majority voting is performed on the extracted 3D features resulting in a robust decision model. The proposed framework was evaluated on a standard publicly available dataset consisting of more than 10,000 facial images, with a test error of 4.05 comparison with state-of-the-art methods shows that our method yields a gradient-free method that is less sensitive to the error range and provides a generalizable face representation. Furthermore, the performance of the proposed method is demonstrated on a number of face datasets consisting of different facial expressions and different types of poses. A comparison with existing methods shows that the proposed method has superior recognition performance.

Recent works have made great progress in learning video representations from large-scale data. In contrast, there are two remaining challenges: 1) how to take advantage of intermediate feature representations and 2) how to further improve the quality of the spatial representations through machine-learning approach. In this paper, we address both challenges. First, we build a temporal hierarchical architecture over the video features by using the sparse coding ones. We develop a convolutional neural network for learning to stream from the features, followed by a temporal hierarchical network for learning to aggregate the temporal dynamics. Second, to address the large appearance variations caused by the ego-motion, we propose a novel appearance variation-based aggregation method for modeling appearance change of an object. We learn to update the neural network via the mean field algorithm to obtain a multidimensional model parameterization. The optimization is performed based on the multidimensional model parameterization and the low rank matrix estimation. We provide theoretical analysis to interpret the proposed method. Our method achieves state-of-the-art results on three video representation benchmarks: video ads database (VidFuse), video face database (YouTubeFace), and video object database (LSUN). In particular, we achieve state-of-the-art performance on YouTubeFace, and previous best results on Imagenet. Our approach yields the best performance on video face dataset (48 labeled). We also observe that our model is competitive on some classical few-shot video representation benchmarks.

Existing studies in video prediction focus on the semantic prediction of intermediate video frame, such as the one before the semantic prediction of generic object. However, few efforts paid attention to the task of video prediction with fine-grained video frame, such as recognizing the action of diverse human actions towards recognizing the humans that are performing the activities towards some specific human body parts (e.g. kicking, punching disaster). In this paper, we propose a novel framework for video prediction with fine-grained video frame, i.e. human action recognition and action recognition with a gated recurrent network. An action classifier is employed to predict the action class for each video frame. A gated fusion module is presented to integrate the predicted human action sequence and action sequence with the output of the backbone network. The proposed model is trained end-to-end in a self-supervised manner, in which fewer action classes are required. The proposed system is evaluated on two large benchmarks. Results on video prediction datasets including UCF-101 and HMDB-51. We also conduct extensive experiments on action recognition in videos with action classes considered as one category.

References