Clustering Millions of Faces By Identity

Nisim Hurst

Friday 3 May 2019

Clustering Millions of Faces by Identity

The article was written by (**otto2018**). It was was cited 44 times according to Google Scholar. The task performed was face clustering. They used the F-measure metric over clusters with distractor images.

Hypothesis

Deep features clustered using only the top-k nearest neighbors in rank-order clustering will produce a more scalable and a more accurate face clustering algorithm.

The network architecture to produce a 320D feature vector was VGG16 proposed by (**Simonyan2014**).

Evidence and Results

Evidence is presented first over a small dataset and the over an augmented version of the datasets with million of distractor images.

Dataset

Results

Contribution

A first contribution of this paper stems from an improvement of the clustering algorithm. The Rank-Order cluster proposed by (**zhu2011**) has the disadvantage that it requires $O(n^2)$. The authors propose to use the FLANN library implementation of the randomized k-d tree algorithm to compute the list of top-k nearest neighbors. Just one iteration is used.

Weaknesses

Future Work