

Improving Similarity Search Using Clustered Data

Authors : Deokho Kim, Wonwoo Lee, Jaewoong Lee, Teresa Ng, Gun-Ill Lee, Jiwon Jeong

Abstract : This paper presents a method for improving object search accuracy using a deep learning model. A major limitation to provide accurate similarity with deep learning is the requirement of huge amount of data for training pairwise similarity scores (metrics), which is impractical to collect. Thus, similarity scores are usually trained with a relatively small dataset, which comes from a different domain, causing limited accuracy on measuring similarity. For this reason, this paper proposes a deep learning model that can be trained with a significantly small amount of data, a clustered data which of each cluster contains a set of visually similar images. In order to measure similarity distance with the proposed method, visual features of two images are extracted from intermediate layers of a convolutional neural network with various pooling methods, and the network is trained with pairwise similarity scores which is defined zero for images in identical cluster. The proposed method outperforms the state-of-the-art object similarity scoring techniques on evaluation for finding exact items. The proposed method achieves 86.5% of accuracy compared to the accuracy of the state-of-the-art technique, which is 59.9%. That is, an exact item can be found among four retrieved images with an accuracy of 86.5%, and the rest can possibly be similar products more than the accuracy. Therefore, the proposed method can greatly reduce the amount of training data with an order of magnitude as well as providing a reliable similarity metric.

Keywords : visual search, deep learning, convolutional neural network, machine learning

Conference Title : ICALIP 2018 : International Conference on Audio, Language and Image Processing

Conference Location : Prague, Czechia

Conference Dates : July 09-10, 2018