Cross-Modal Text-Image Retrieval Algorithm Based on Model Transfer Learning

Li, Muyao (School: Chengdu No. 7 High School)

Purpose: The application of deep learning in cross-modal retrieval can produce good results. However, the training of deep neural networks often relies on large-scale labeled dataset. It is more difficult to collect large-scale visual and text pair dataset. In this project, we propose a cross-modal text-image retrieval algorithm based on model transfer learning, which can solve the limited dataset problem to a certain degree, and the application of text-image retrieval for flowers and birds datasets is implemented. Procedure: Specially, we add two full-connection layers and a classification layer for model transferring behind the bottleneck layer of the pretrained GoogLeNet. The simple yet effective feature extractive model can extract visual features accurately, even in the case of few training samples. Additionally, we use the simple and fast CharCNN model for textual features extraction. After projecting visual and textual feature vectors into the Hilbert space, we use the cosine similarity based on the simple-triplet-loss as the objective function to train our mapping function. Results: We use the public datasets of Caltech-UCSD Birds 200-2011 and Oxford-102 to test our model. The average retrieval precision of text-based image retrieval and image-based text retrieval on these two datasets reached 92.2% and 92.3%, respectively. Conclusions: The experimental results show that the proposed method can achieve text-image retrieval on small-scale datasets, and can better implement the application of cross-modal text-image retrieval for flowers and birds.

Awards Won: