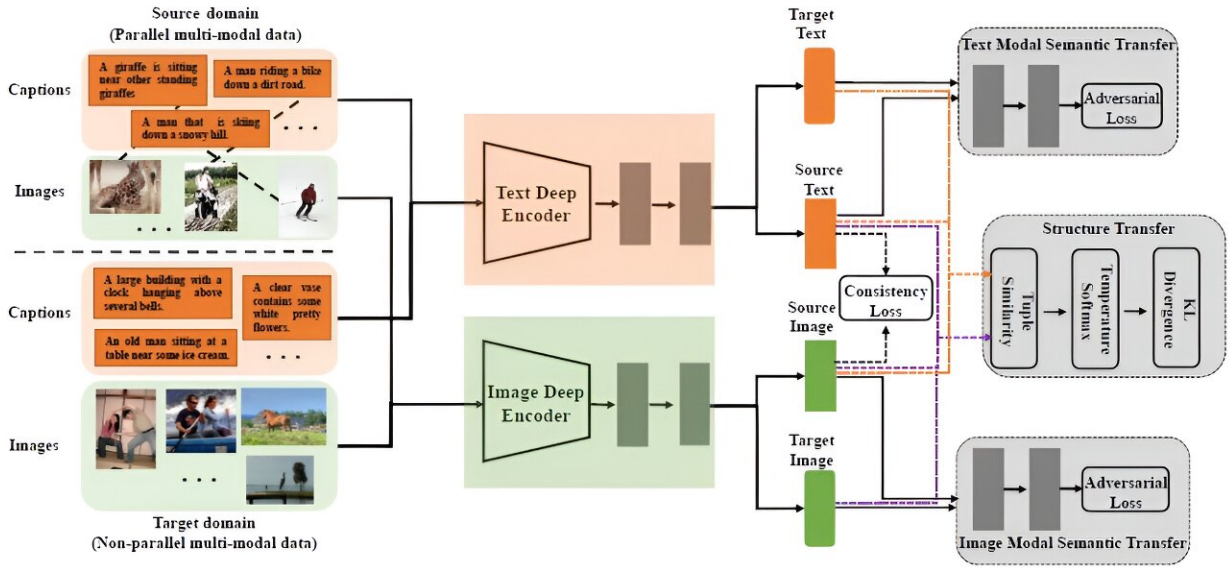
# Alignment efficient image-sentence retrieval considering transferable cross-modal representation learning

February 26 2024




The processing flow of AEIR. Credit: Yang Yang, Jinyi Guo, Guangyu Li, Lanyu Li, Wenjie Li, Jian Yang

Image-sentence retrieval task aims to search images for given sentences and retrieve sentences from image queries. The current retrieval methods are all supervised methods that require a large number of annotations for training. However, considering the labor cost, it is difficult to re-align large amounts of multimodal data in many

applications (e.g., medical retrieval), which results in unsupervised multimodal data.

A research team led by Yang Yang published their new research in *Frontiers of Computer Science*.

To solve the problem the team strive to take a step towards non-parallel image-sentence retrieval by designing the alignment transfer, and propose a novel Alignment Efficient Image-Sentence Retrieval method (AEIR).

In the research, AEIR use other auxiliary parallel data with multimodal consistency as the source domain and non-parallel data with missing consistency as the target domain. Unlike unimodal transfer learning, AEIR transfers semantic representations and modal consistency relations together from the source domain to the target domain.

Firstly, AEIR learns cross-modal consistency representations using cross-modal parallel data in the source domain. Then AEIR jointly optimizes adversarial learning-based semantic transfer constraints and metric learning-based structural transfer constraints to learn cross-domain cross-modal consistency representations to achieve transfer of consistency knowledge from the source domain to the target domain.

A large number of experimental experiments conducted in different transfer scenarios show that semantic transfer and structural transfer can effectively learn invariant features across modalities across domains. The proposed efficient alignment-based image-sentence retrieval network verifies that AEIR is more advantageous than current cross-modal retrieval methods, semi-supervised cross-modal retrieval methods and cross-modal transfer methods.

Future work can focus on the conduction of positive cross-modal

transfer considering the domain discrepancy.

**More information:** Yang Yang et al, Alignment efficient image-sentence retrieval considering transferable cross-modal representation learning, *Frontiers of Computer Science* (2023). DOI: 10.1007/s11704-023-3186-6