

# Attention-aware Deep Adversarial Hashing for Cross-Modal Retrieval

Xi Zhang<sup>1,2</sup>[0000–0002–9173–4119], Hanjiang Lai<sup>1,2</sup> [0000–0001–8057–6744], and  
Jiashi Feng<sup>3</sup>[0000–0001–6843–0064]

<sup>1</sup> School of Data and Computer Science, Sun Yat-Sen University, GuangZhou, China

<sup>2</sup> Guangdong Key Laboratory of Big Data Analysis and Processing  
zhangx368@mail2.sysu.edu.cn, lai hanj 3@mail1.sysu.edu.cn

<sup>3</sup> Department of Electrical and Computer Engineering,  
National University of Singapore, Singapore, Singapore  
elefji@nus.edu.sg

**Abstract.** Due to the rapid growth of multi-modal data, hashing methods for cross-modal retrieval have received considerable attention. However, finding content similarities between different modalities of data is still challenging due to an existing heterogeneity gap. To further address this problem, we propose an adversarial hashing network with an attention mechanism to enhance the measurement of content similarities by selectively focusing on the informative parts of multi-modal data. The proposed new deep adversarial network consists of three building blocks: 1) the feature learning module to obtain the feature representations; 2) the attention module to generate an attention mask, which is used to divide the feature representations into the attended and unattended feature representations; and 3) the hashing module to learn hash functions that preserve the similarities between different modalities. In our framework, the attention and hashing modules are trained in an adversarial way: the attention module attempts to make the hashing module unable to preserve the similarities of multi-modal data w.r.t. the unattended feature representations, while the hashing module aims to preserve the similarities of multi-modal data w.r.t. the attended and unattended feature representations. Extensive evaluations on several benchmark datasets demonstrate that the proposed method brings substantial improvements over other state-of-the-art cross-modal hashing methods.

**Keywords:** Hashing, Adversarial Learning, Attention Mechanism, Cross Modal Retrieval

## 1 Introduction

Due to the rapid development of the Internet, different types of media data are also growing rapidly, e.g., texts, images, and videos. Cross-modal retrieval, which takes one type of data as the query and returns the relevant data of

---

Corresponding author: Hanjiang Lai































