Abstract
Unsupervised video hashing aims to learn a nonlinear hashing function to map videos into a similarity-preserving hamming space without label supervision. In this paper, we propose Motion-Aware Graph Reasoning Hashing (MAGRH), an end-to-end framework that utilizes the motion information explicitly while keeping inference efficiency. Specifically, we design a dual-branch architecture consisting of a main and an auxiliary branch. The main (auxiliary) branch receives frame-wise (clip-wise) inputs and produces general (motion) hash codes via graph reasoning modules. We develop a combination of intra- and inter-branch objectives to simultaneously learn branch-specific hashing functions as well as transfer motion knowledge from the auxiliary branch to the main branch. In inference, the auxiliary branch is removed. Benefiting from motion guidance, our MAGRH yields superior performance on FCVID and ActivityNet, even with a small frame rate.

Graph Reasoning and Hash Layers
\[ \mathbf{x}^* = [\psi_{\text{avg}}, \Phi^*(\mathbf{x}^*) + z_p + z_m^*] \in \mathbb{R}^{(T+1) \times D_h}, \quad s \in \{g, m\} \]
\[ A = g_1(\mathbf{x})^T g_2(\mathbf{x}), \quad A_{ij} \rightarrow \sum_{t=1}^{T+1} A_{ij}^t \]
\[ \mathbf{x}^*_t = \text{ReLU}(\text{LN}(A_{\mathbf{x}_t} W_l)), \quad l \in \{1, \cdots, L\} \]
\[ \mathbf{x}^* = \frac{1}{T+1} \sum_{t=0}^{T} \mathbf{x}^*_t \in \mathbb{R}^{D_h}, \quad s \in \{g, m\} \]
\[ h^* = \Phi^*(\mathbf{x}^*) \in \mathbb{R}^{D_h}, \quad s \in \{g, m\} \]

Experiments
We compared the proposed MAGRH with: BTH, NPH, SSVH, SSTH, JTAE, MFH, ITQ, and DH. Our MAGRH outperforms all previous methods under all code lengths by a large margin in terms of MAP@K.