

CoSMo: Content-Style Modulation for Image Retrieval with Text Feedback

Seungmin Lee*, Dongwan Kim*, Bohyung Han
Computer Vision Lab, Seoul National University

Introduction

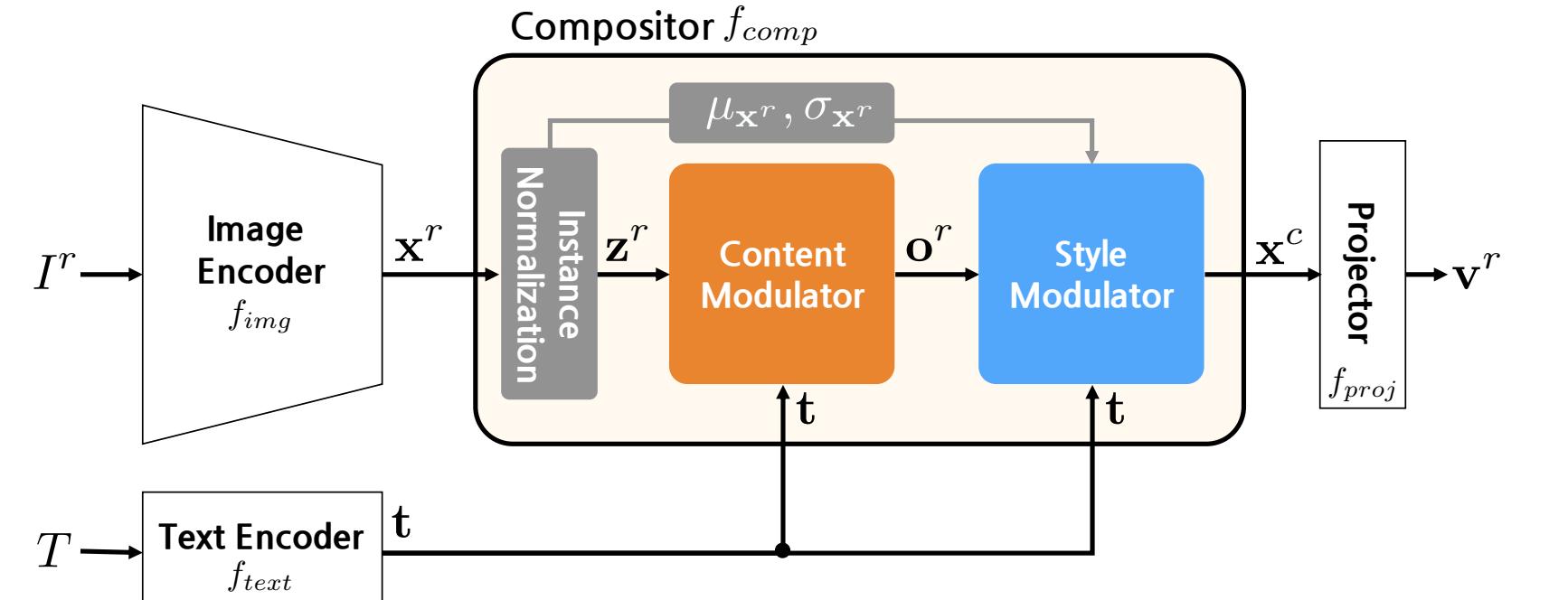
Image Retrieval with Text Feedback

We study the task of image retrieval with text feedback, where a **reference image** and **modifier text** are used jointly as the query. The modifier text describes how the reference image should be changed to obtain the **target image**.



We propose a novel image-text composer, called **Content-Style Modulator [CoSMo]**. CoSMo is able to modulate both the **contents** and the **style** of the reference image based on the modifier text.

Our Model



The image-text compositor is composed of the content modulator [CM] and style modulator [SM]. The CM takes an instance-normalized image feature as input so that it can solely focus on modifying the content information. The style modulator applies affine transformations to channels of the modified feature, allowing it to focus on style transformations .

