

Visual-semantic embedding aims to find a shared latent space where related visual and textual instances are close to each other. Most current methods learn in