

DAVD-Net: Deep Audio-aided Video Decompression of Talking Heads

Abstract: Close-up talking heads are among the most common and salient object in video contents, such as face-to-face conversations in social media, teleconferences, news broadcasting, talk shows, etc. Due to the high sensitivity of human visual system to faces, compression distortions in talking head videos are highly visible and annoying. To address this problem, we present a novel deep convolutional neural network method for very low bit rate video reconstruction of talking heads. The key innovation is a new DCNN architecture that can exploit the audio-video correlations to repair compression defects in the face region. We further improve reconstruction quality by embedding into our DCNN the encoder information of the video compression standards and introducing a constraining projection module in the network. Extensive experiments demonstrate that the proposed DCNN method outperforms the existing state-of-the-art methods on videos of talking heads.