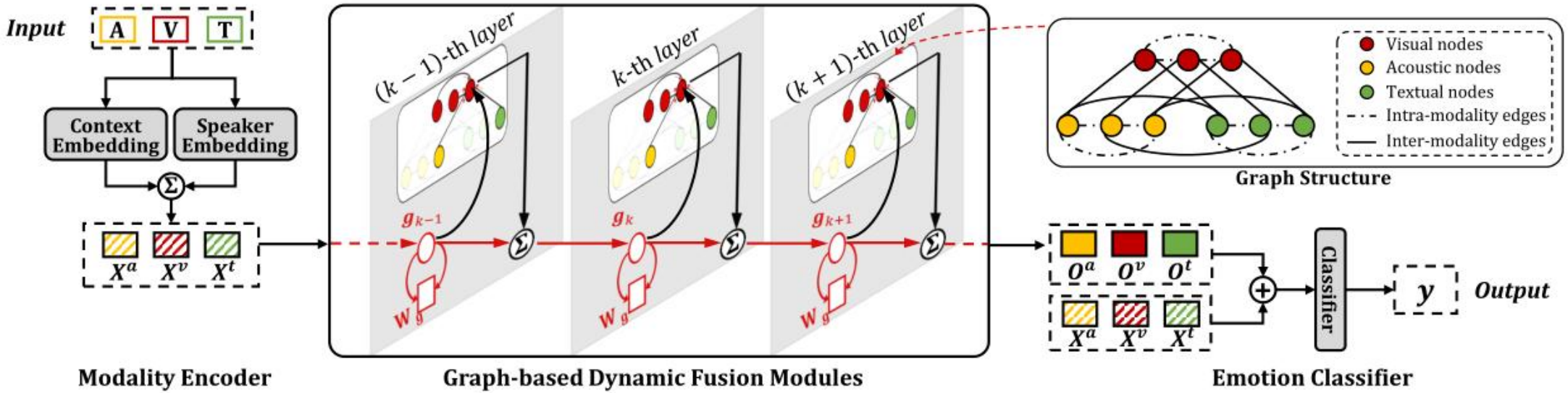
Multimodal Emotion Recognition

Made by JiaWei Cheng

MM-DFN: MULTIMODAL DYNAMIC FUSION NETWORK FOR EMOTION RECOGNITIONINCONVERSATIONS(ICASSP2022)



Implementation Details. Following [13], raw utterance-level features of acoustic, visual, and textual modality are extracted by TextCNN [21], *OpenSmile* [22], and DenseNet [23], respectively. We use focal loss [24] for training due to the class imbalance. The number of layers K are 16 and 32 for IEMOCAP and MELD. α is set to 0.2 and ρ is set to 0.5.

Improving Multimodal Fusion with Main Modal Transformer for Emotion Recognition in Conversation(KBS2022)

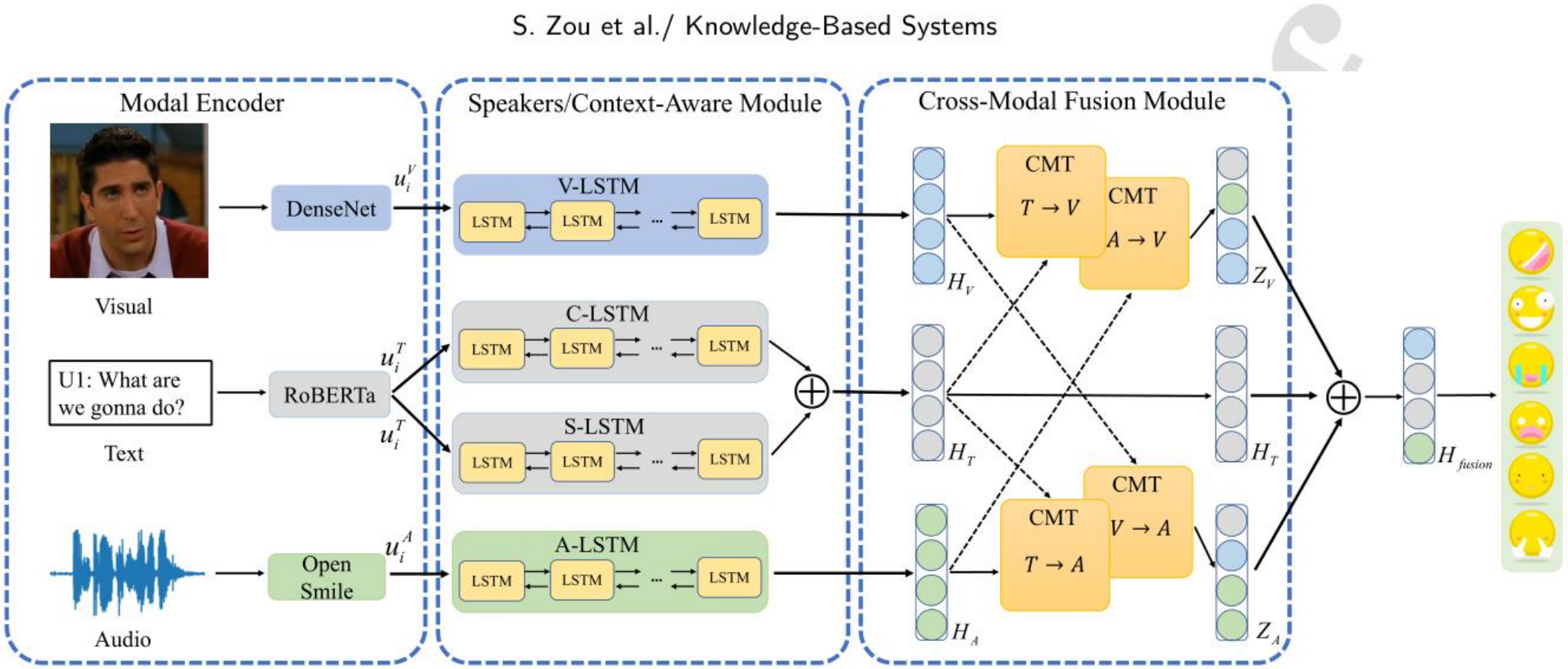
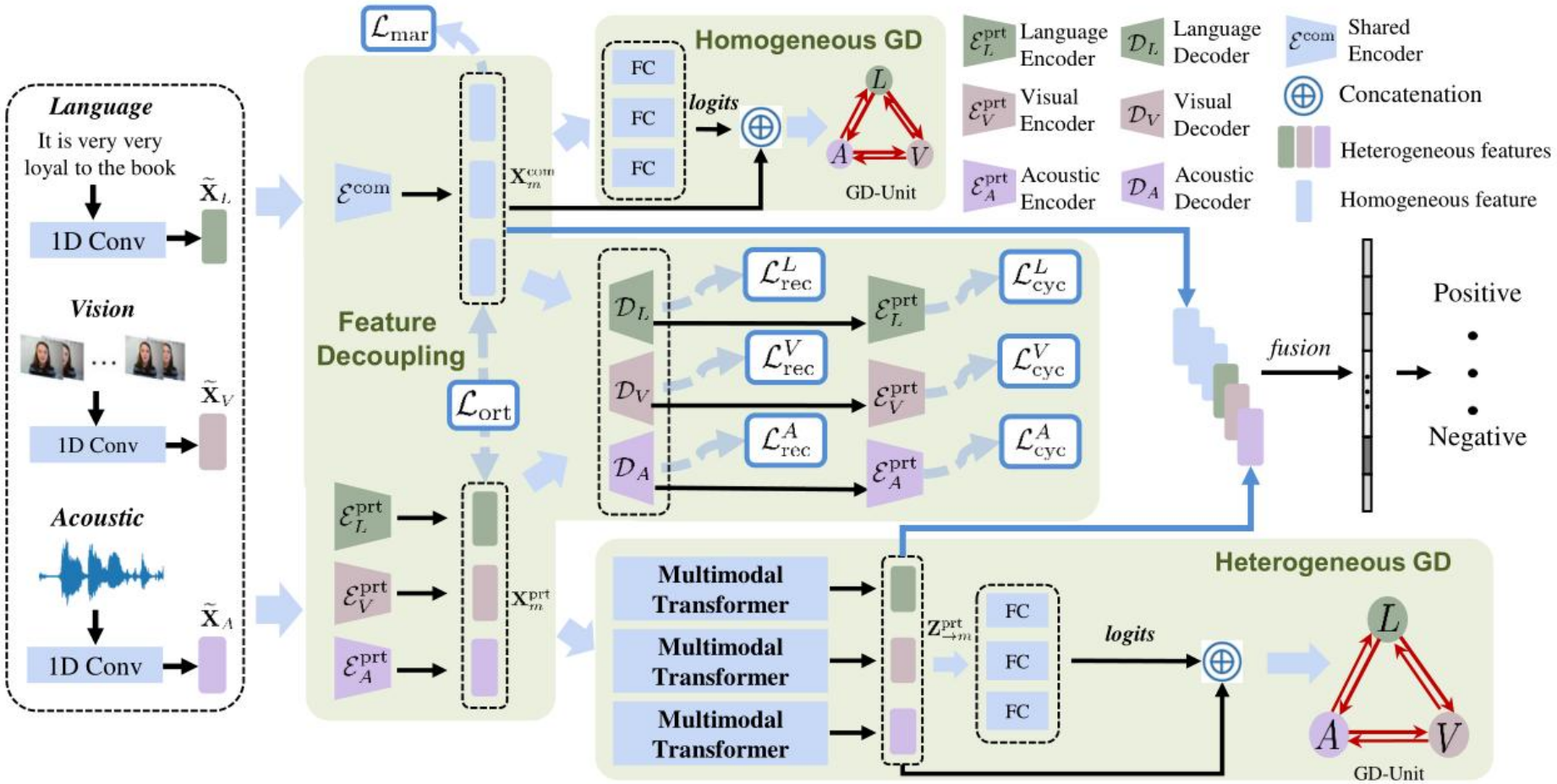


Figure 2: Overall architecture of the proposed MMTr.

Decoupled Multimodal Distilling for Emotion Recognition(CVPR2023)



文中没有提及三个模态初始特征的提取方法