

Input module

Feature extraction module

Profile generation

Proposed attention based fusion module

Image 1



Image 2

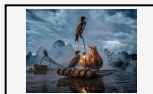


Image 3



Image pre-processing and normalization

Mel-spectrogram generation

ResNet-50

ResNet-50

Deep image features

Deep audio features

Residual feature propagation

Residual feature propagation

Residual feature propagation

