

Pixel Neighborhood

Frame Level

spatial feature  
extraction  
(VIF, DLM)

temporal feature  
extraction (TI)

training with  
subjective data

trained  
model

within-frame spatial  
pooling

SVM prediction

"Fusion"

temporal pooling

per-frame score

