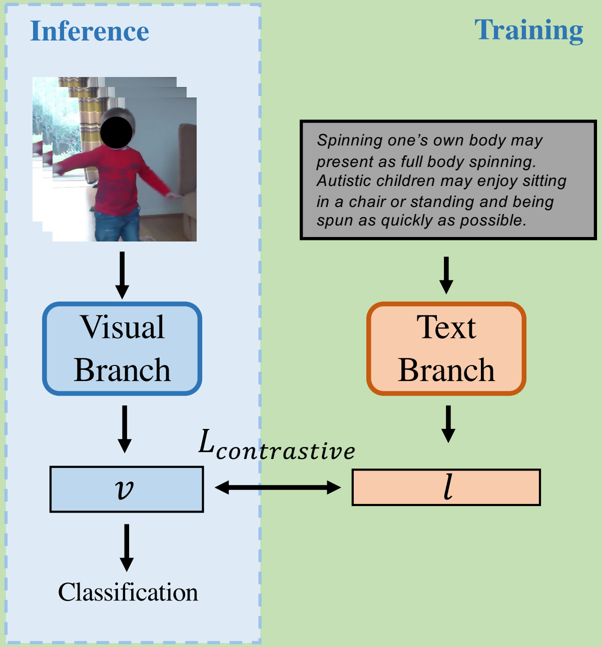
Language-Assisted Deep Learning for Autistic Behaviors Recognition

Framework



Spinning one's own body may present as full body spinning. Autistic children may enjoy sitting in a chair or standing and being spun as quickly as possible.

Language as free supervision during training stage

Improvement

Compared with previous networks in autism behaviors recognition,VideoSwinTransformer brings

$\geq$ **10%** performance gain

After incorporating language supervision, further improvements are obtained:

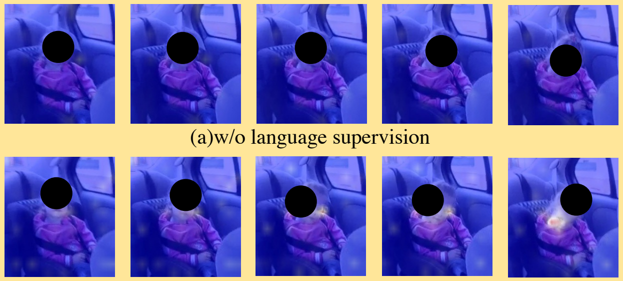
+3.49% on ESDB

+1.46% on SSDB

However,

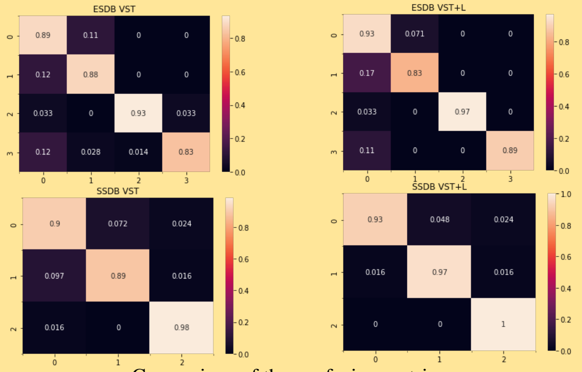
- No additional annotation efforts
- Text branch is **only used in training**

Visualization


(a)w/o language supervision

(b)w/ language supervision

Language supervision brings higher attention scores in action-related region



Comparison of the confusion matrices