

# Video Transformer Network – Supplementary Material –

Daniel Neimark   Omri Bar   Maya Zohar   Dotan Asselmann

Theator

{danieln, omri, maya, dotan}@theator.io

## Appendix

### A. Additional Qualitative Results



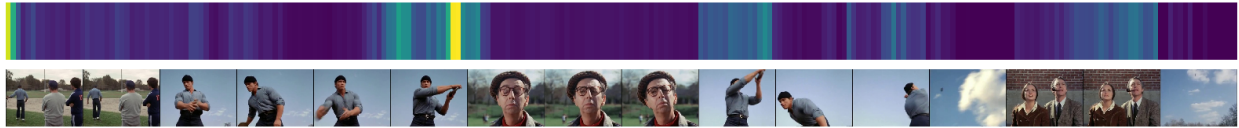
(a) Label: *Tai chi*. Prediction: *Tai chi*.



(b) Label: *Chopping wood*. Prediction: *Chopping wood*.



(c) Label: *Archery*. Prediction: *Archery*.



(d) Label: *Throwing discus*. Prediction: *Flying kite*.



(e) Label: *Surfing water*. Prediction: *Parasailing*.

Figure 6. Additional qualitative examples of why attention matters. Similar to Fig. 3, we illustrate the  $[CLS]$  token weights of the first attention layer. We show some examples for successful predictions (a, b, c) and some of the failure modes of our approach (d, e).