

In this paper, we present a unified, end-to-end trainable spatiotemporal CNN model for VOS, which consists of two branches, i.e., the temporal coherence branch