**Time Understanding**

Experiments:

* Reverse (randomly reverse the order of the clip 🡪 classification)
* Blackout (randomly set one out of n chunks to zero 🡪 predict which (classification))
* Freeze (randomly freeze one out of n chunks 🡪 predict which (classification))
* Permutation (randomly permute n chunks 🡪 predict perm. order (classification)

Models:

* Slow (3DResnet based, supervised)
* SlowFast (3DResnet based, supervised)
* X3D (3DResnet based, supervised)
* MViT (Transformer based, supervised)
* VIMPAC (Transformer based, self-supervised)

Results:

**Reverse Blackout**

Chart, line chart

Description automatically generated Chart, line chart

Description automatically generated

**Freeze Permutation**

Chart, line chart

Description automatically generated Chart, line chart

Description automatically generated

Inference for Action Classification is average over 10 randomly sampled sub-clips. Accuracy for 10 randomly **frozen** sub-clips is quite high (Slow: 58 % vs 72 % Kinetics400).