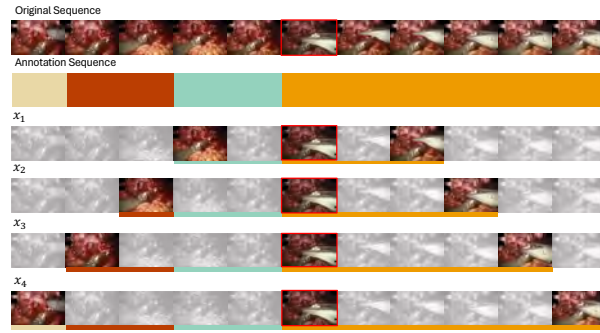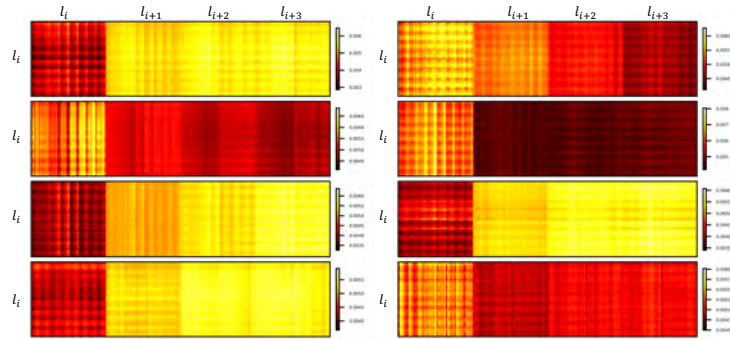# MuST: Supplementary Material

Fig. 1. Temporal Multi-Scale Pyramid Example. Illustration of a temporal multi-sequence pyramid centered on the red keyframe. Sampled frames at each level are highlighted. The top levels of the pyramid capture short-term temporal information, while the lower levels comprehend a broader temporal context.

Fig. 2. Cross-attention maps. Attention weights from the 8 attention heads of one of the layers of the cross-attention described in Section 2 (Method) for the l_i sequence of the set L. These attention maps illustrate that each head specializes in capturing different segments of the L set, featuring inter-dependencies across temporal scales.
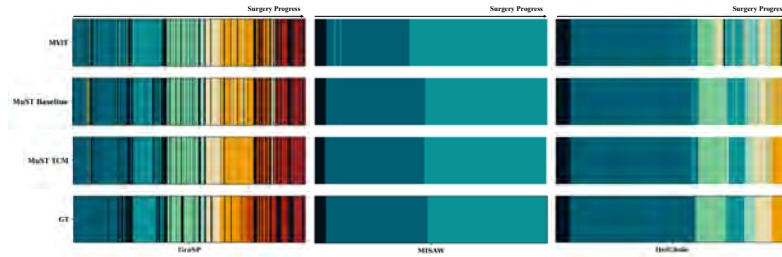
Fig. 3. MuST visualization performance in every benchmark. MuST baseline refers to our model without the Temporal Consistency Module (TCM), whereas MuST TCM represents the complete model. Overall visualizations demonstrate that MuST achieves consistent recognition of phases, including reclassification of some segments when using the Temporal Consistency Module.

Table 1. Class comparison of MuST with state-of-the-art methods. MuST baseline refers to our model without the Temporal Consistency Module (TCM), whereas MuST TCM represents the complete model. Consistent results across classes are evident across diverse temporalities within the databases. These results underscore MuST's proficiency in identifying multiple duration phases, attributed to its multi-temporal reasoning. Leveraging information from distinct temporal contexts enhances MuST's predictive capabilities for phases of different durations. The best results are highlighted in bold.

GraSP

Phase	Duration	TAPIS	TeCNO	Trans-SVNet	MuST Baseline (Ours)	MuST TCM (Ours)
0	50.60 ±107.06	85.5	77.9	74.7	86.8	**87.1**
1	285.28±318.04	86.0	95.8	94.2	95.6	**97.7**
2	221.55±256.20	76.0	87.7	84.5	91.6	**93.0**
3	143.33±159.81	83.0	83.2	85.3	91.9	**94.2**
4	84.73±55.18	83.8	67.1	76.1	**90.8**	89.5
5	149.16±149.92	85.0	78.5	82.3	**89.5**	87.8
6	222.21±264.64	73.5	81.6	79.4	79.4	**82.1**
7	108.00±56.66	8.6	**68.1**	61.1	16.6	22.6
8	74.89±76.06	71.6	70.1	66.1	71.8	**76.2**
9	107.00±77.22	59.3	**70.2**	67.9	63.8	60.4
10	70.48±80.06	76.3	67.8	70.2	78.9	**79.8**

MISAW

Phase	Duration	TAPIS	TeCNO	Trans-SVNet	MuST (Ours)	MuST TCM (Ours)
0	114.70±62.62	92.1	87.8	75.7	94.2	**95.0**
1	2318.22±931.10	97.9	99.4	96.6	98.9	**99.5**
2	3651.41±2759.67	98.9	99.6	98.8	99.4	**99.8**

HeiChole

Phase	Duration	TAPIS	TeCNO	Trans-SVNet	MuST (Ours)	MuST TCM (Ours)
0	170.38±63.18	86.4	89.3	92.3	96.2	**97.6**
1	618.67±461.98	83.0	81.9	84.2	**92.3**	91.4
2	127.85±61.96	46.4	37.8	40.3	**61.6**	58.0
3	384.68±222.82	73.8	79.2	78.1	76.6	**77.6**
4	86.33±37.20	72.1	64.1	**75.4**	67.2	68.5
5	279.48±208.76	58.5	65.6	67.7	70.0	**73.2**
6	70.12±63.01	64.7	69.3	64.9	67.9	**73.0**