

# Supplementary Material

Surgformer: Surgical Transformer with Hierarchical Temporal Attention for Surgical Phase Recognition

The Hong Kong University of Science and Technology

Table 1: Performance comparisons of Surgformer training with different sequence lengths (**T**) and frame rates (**R**) on the Cholec80 dataset.

<b>T</b> × <b>R</b>	Image-level Accuracy	Unrelaxed Evaluation				Relaxed Evaluation			
		Accuracy	Precision	Recall	Jaccard	Accuracy	Precision	Recall	Jaccard
<b>Varied Frame Rates <i>R</i></b>									
16 × 1	89.4	89.6±6.6	84.5±7.1	86.6±6.1	74.5±9.6	90.5±6.6	87.5±5.5	89.3±5.0	77.8±8.6
16 × 2	90.3	90.6±5.9	87.2±4.5	86.9±6.9	76.7±8.6	91.4±6.0	89.9±3.3	89.5±5.7	80.0±7.2
16 × 4	<b>91.3</b>	<b>91.7±6.1</b>	<b>88.0±6.2</b>	<b>88.7±6.7</b>	<b>79.1±9.1</b>	<b>92.5±6.1</b>	<b>90.6±4.9</b>	<b>91.0±5.8</b>	<b>82.1±7.5</b>
16 × 6	<b>91.3</b>	91.6±5.2	87.0±7.6	88.1±7.1	78.0±10.5	92.4±5.1	89.6±6.0	90.4±5.6	81.0±8.8
16 × 8	90.5	90.7±6.9	86.1±8.6	85.4±10.0	75.4±11.1	91.7±6.9	90.0±6.0	88.7±8.4	79.5±9.1
<b>Varied Sequence Lengths <i>T</i></b>									
4 × 4	88.5	89.0±6.2	84.4±5.2	84.9±7.6	73.1±9.4	89.7±6.2	86.9±4.1	87.4±6.8	75.8±8.3
8 × 4	89.8	90.2±5.9	87.0±5.2	86.6±7.9	75.8±8.6	91.0±6.0	89.5±4.5	89.1±7.5	79.0±7.8
12 × 4	90.5	90.7±6.5	87.6±5.6	86.9±8.2	76.8±8.5	91.5±6.5	90.3±4.3	89.5±7.7	80.3±7.1
16 × 4	<b>91.3</b>	<b>91.7±6.1</b>	<b>88.0±6.2</b>	<b>88.7±6.7</b>	<b>79.1±9.1</b>	<b>92.5±6.1</b>	<b>90.6±4.9</b>	<b>91.0±5.8</b>	<b>82.1±7.5</b>
24 × 4	91.0	91.4±5.9	87.1±8.4	87.1±7.7	77.8±10.6	92.1±5.9	90.3±5.8	89.7±5.5	81.0±8.5

Table 2: Performance comparisons of Surgformer with different sequence lengths (**T**) and frame rates (**R**) in inference on the Cholec80 dataset.

<b>T</b> × <b>R</b>		Unrelaxed Evaluation				Relaxed Evaluation			
Training	Testing	Accuracy	Precision	Recall	Jaccard	Accuracy	Precision	Recall	Jaccard
16 × 4	16 × 4	91.7±6.1	88.0±6.2	88.7±6.7	79.1±9.1	92.5±6.1	90.6±4.9	91.0±5.8	82.1±7.5
16 × 4	16 × 5	92.0±6.2	86.9±7.3	87.7±9.8	77.5±12.1	<b>93.3±6.2</b>	91.6±4.6	91.9±7.1	83.1±9.3
16 × 4	16 × 6	<b>92.2±6.3</b>	<b>88.3±6.3</b>	<b>88.9±7.9</b>	<b>79.5±10.1</b>	<b>93.3±6.3</b>	<b>91.9±4.4</b>	<b>91.9±5.9</b>	<b>83.9±7.9</b>
16 × 4	16 × 7	92.1±6.1	87.3±7.1	88.4±8.8	78.5±11.1	<b>93.3±6.2</b>	<b>91.9±4.4</b>	<b>91.9±6.5</b>	83.7±8.5
16 × 4	16 × 8	92.0±6.2	86.9±7.3	87.7±9.8	77.5±12.1	<b>93.3±6.2</b>	91.6±4.6	91.9±7.1	83.1±9.3
16 × 4	16 × 4	91.7±6.1	88.0±6.2	88.7±6.7	79.1±9.1	92.5±6.1	90.6±4.9	91.0±5.8	82.1±7.5
16 × 4	20 × 4	92.1±6.3	<b>88.3±6.6</b>	89.2±7.1	<b>79.9±9.6</b>	93.0±6.3	91.5±4.9	91.7±5.5	83.6±7.7
16 × 4	24 × 4	<b>92.4±6.4</b>	87.9±6.9	<b>89.3±7.8</b>	<b>79.9±10.2</b>	<b>93.4±6.4</b>	<b>91.9±4.7</b>	<b>92.1±5.8</b>	<b>84.1±8.0</b>