Supplementary Material

Kun Yuan[1,3], Vinkle Srivastav[1,2], Nassir Navab[3], and Nicolas Padoy[1,2]

[1] University of Strasbourg, CNRS, INSERM, ICube, UMR7357, Strasbourg, France
[2] IHU Strasbourg, Strasbourg, France
[3] CAMP, Technische Universität München, Munich, Germany

$$\mathcal{L}_{clip}^{aws} = -\frac{1}{B}\sum_{i=1}^{B}\log\left(\frac{\exp(\mathcal{F}_v(v_i)^T\cdot\mathcal{F}_t(a_i)/\tau)}{\sum_{j=1}^{B}\exp(\mathcal{F}_v(v_j)^T\cdot\mathcal{F}_t(a_j)/\tau)}\right)$$

$$\mathcal{L}_{clip}^{whisper} = -\frac{1}{B}\sum_{i=1}^{B}\log\left(\frac{\sum_{m=1}^{M}\exp(\mathcal{F}_v(v_i)^T\cdot\mathcal{F}_t(w_i^m)/\tau)}{\sum_{j=1}^{B}\sum_{m=1}^{M}\exp(\mathcal{F}_v(v_j)^T\cdot\mathcal{F}_t(w_j^m)/\tau)}\right)$$

$$L_{clip} = L_{clip}^{aws} + L_{clip}^{whisper}$$

Fig. 1: Loss function of clip-level video-text pairs in Sec. 2.3. a: narration text (n_{ij} in the main paper) transcribed from AWS Medical Transcribe [1]. w: narration text transcribed from Whisper [2] system. v: short-term video clips, which correspond to 1 aws text and M whisper texts.

Table 1: Manually designed prompts for the class names to recognize the surgical phase in Cholec80 dataset.

Phase Labels	**Prompts**
Preparation	In preparation phase I insert trocars to patient abdomen cavity
CalotTriangleDissection	In calot triangle dissection phase I use grasper to hold gallbladder and use hook to expose the hepatic triangle area and cystic duct and cystic artery
ClippingCutting	In clip and cut phase I use clipper to clip the cystic duct and artery then use scissor to cut them
GallbladderDissection	In dissection phase I use the hook to dissect the connective tissue between gallbladder and liver
GallbladderPacking	In packaging phase I put the gallbladder into the specimen bag
CleaningCoagulation	In clean and coagulation phase I use suction and irrigation to clear the surgical field and coagulate bleeding vessels
GallbladderRetraction	In retraction phase I grasp the specimen bag and remove it from trocar

Table 2: Manually designed prompts for the class names to recognize the surgical phase in AutoLaparo dataset.

Phase Labels	**Prompts**
Preparation	I use grasper to grasp and explore the field
Dividing Ligament and Peritoneum	I divide ligament and peritoneum
Dividing Uterine Vessels and Ligament	I divide uterine vessels and ligament
Transecting the Vagina	I use the dissecting hook to transect the vagina
Specimen Removal	I remove the specimen bag and uterus
Suturing	I suture the tissue
Washing	Washing

Table 3: Manually designed prompts for the class names to recognize the surgical phase in gastric bypass dataset. We use the same prompts for both StrasBypass70 and BernBypass70.

Phase Labels	**Prompts**
Preparation	In preparation phase I insert trocars to the abdominal cavity and expose of the operating field
Gastric pouch creation	I cut the fat tissue and open retrogastric window at stomach
Omentum division	I grasp and lift the omentum and divide it
Gastrojejunal anastomosis	I see the proximal jejunum and determine the length of the biliary limb. I open the distal jejunum and create the gastrojejunostomy using a stapler. I reinforcement of the gastrojejunostomy with an additional suture.
Anastomosis test	I place the retractor and move the gastric tube and detect any leakage of the gastrojejunostomy
Jejunal separation	I open the mesentery to facilitate the introduction of the stapler and transect the jejunum proximal
Petersen space closure	I expose between the alimentary limb and the transverse colon and close it with sutures
Jejunojejunal anastomosis	I expose between the alimentary limb and the transverse colon and close it with sutures
Mesenteric defect closure	I expose the mesenteric defect and then close it by stitches
Cleaning and coagulation	In clean and coagulation phase I use suction and irrigation to clear the surgical field and coagulate bleeding vessels
Disassembling	I remove the instruments, retractor, ports, and camera
Other	Other

References

1. AWS: Amazon transcribe medical (2023), https://aws.amazon.com/transcribe/medical/
2. Radford, A., Kim, J.W., Xu, T., Brockman, G., McLeavey, C., Sutskever, I.: Robust speech recognition via large-scale weak supervision. In: International Conference on Machine Learning. pp. 28492–28518. PMLR (2023)