A semi-supervised Teacher-Student framework for surgical tool detection and localization

Carlos Aparicio-Viveros[1], Mansoor Ali Teevno[1], Gilberto Ochoa Ruiz[1], Sharib Ali[2]

[1]Tecnologico de Monterrey, Escuela de Ingeniería y Ciencias, 38115, México

[2]School of Computing, University of Leeds, Leeds, UK

Tecnológico de Monterrey · LXAI · UNIVERSITY OF LEEDS · CVPR JUNE 17-21, 2024 SEATTLE, WA

Introduction

Surgical tool detection in minimally invasive surgery is an essential part of computer-assisted interventions. Current approaches are mostly based on supervised methods which require large fully labeled data to train supervised models and suffer from pseudo label bias because of class imbalance issues. In this work, we propose an end-to-end Teacher-Student network for the semi-supervised detection and localization of surgical tools (Fig. 1)

Dataset

We used the extended version of m2cai16-tool dataset called m2cai16-tool-locations. It consists of 2812 images with bounding box annotations for 7 types of surgical tools. The dataset contains the classes depicted in the figure below (Fig. 2)

Fig. 2. Examples of m2cai16-tool dataset classes

Ablation Studies

We measured the impact of EMA rates and confidence threshold for semi-supervised learning

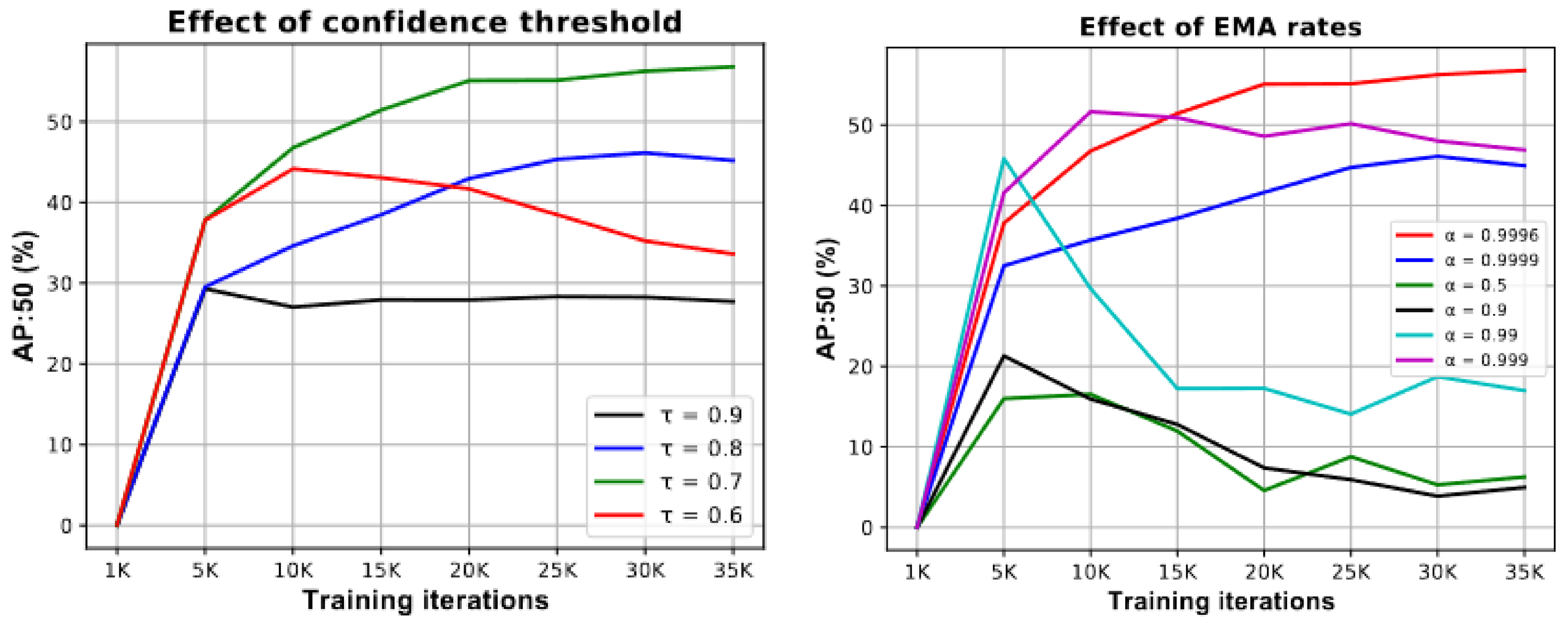

Fig. 3 Ablation studies with different confidence thresholds and EMA rates

Model

The Teacher-Student joint learning (Fig. 1) begins with an initialization stage using labeled data to set the weights for both the Teacher and Student models. We apply weak and strong augmentations to enhance learning from unlabeled data, where the Teacher generates pseudo-labels for the Student. The Student is trained on these pseudo-labels and transfers learned weights to the Teacher via Exponential Moving Average (EMA)

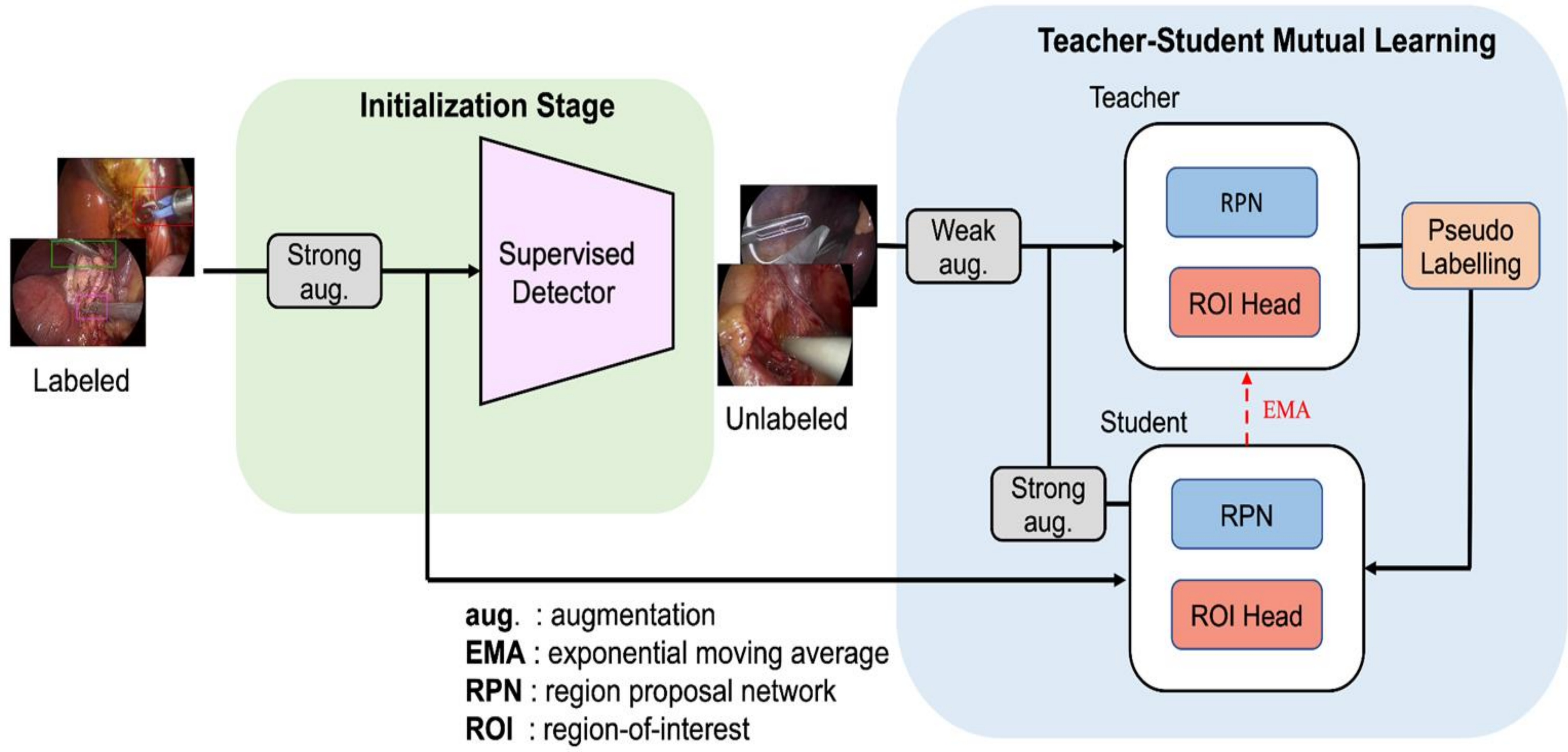

Fig. 1 Overview of the proposed surgical tool detection model. It consists of two modules: 1) An initialization module, 2) A Teacher-Student mutual learning module.

Logistic loss with added margin and distance penalization:

To address the class imbalance problem, we target the foreground and background class imbalance problem by introducing a multi-class loss function based on a margin, which tries to maximize foreground-background distance:

$$\mathcal{L}_{cls}^{roi} = \sum_{n} w_l \log\left(1 + \frac{e^{s \cdot (\beta - \rho + \sigma)}}{s}\right)$$

where wl is the loss weight, $\boldsymbol{\beta}$, $\boldsymbol{\rho}$ represent softmax of the probabilities for foreground and background logits, $\boldsymbol{\sigma}$ denotes margin, s and n are the smoothness parameter and mini-batch size respectively

Results

Qualitative Results

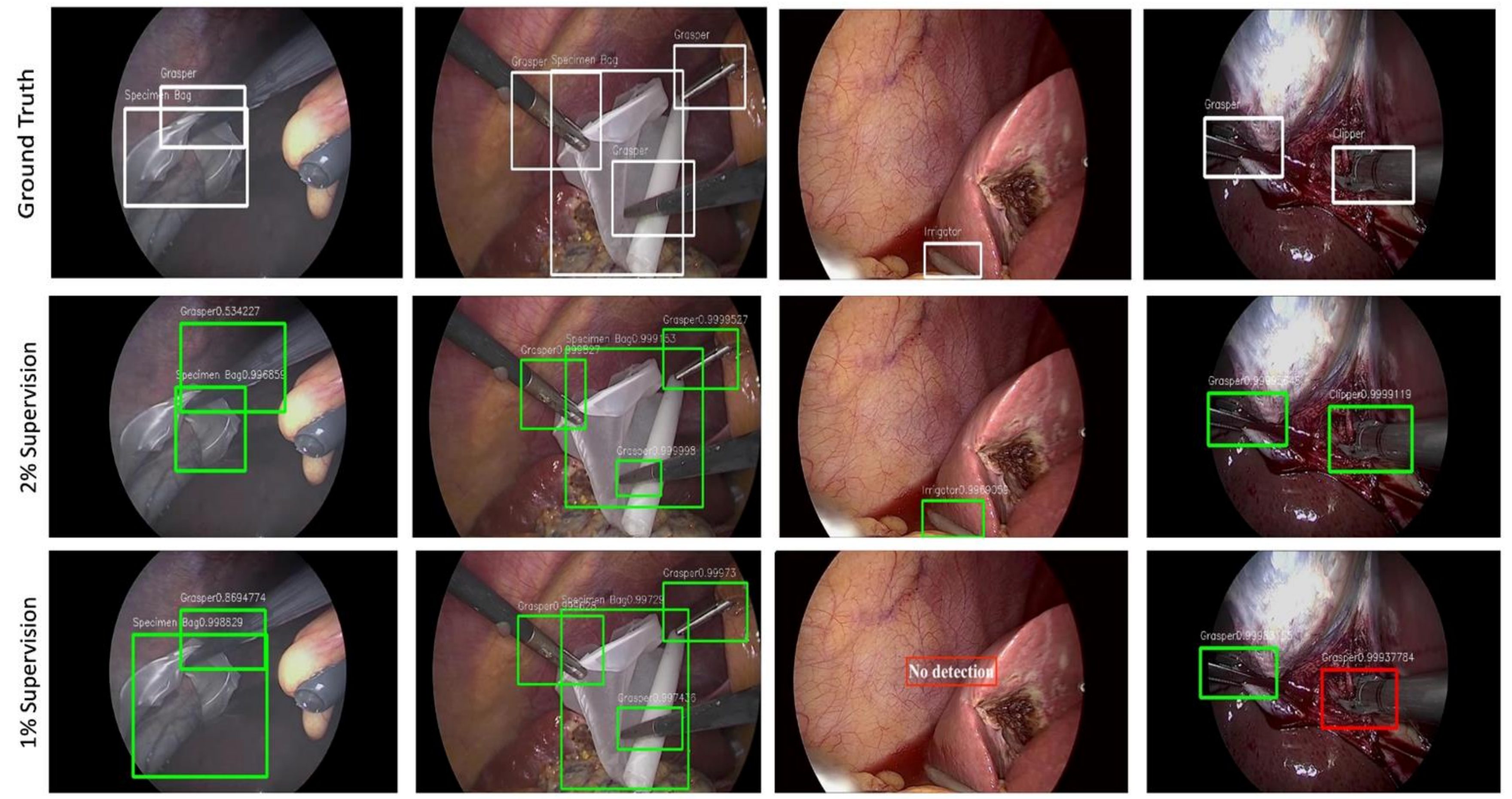

Fig. 4 Qualitative Results: First row shows images with ground truth. Second and third row presents results on 2% and 1% setting respectively

Quantitative Results

Method	mAP$_{50}$	mAP$_{50:95}$	mAP$_{75}$	mAP$_{medium}$	mAP$_{large}$	p-values
1% Labeled Data						
Supervised	23.578	7.673	2.322	6.189	9.050	5.996e-17
Unbiased Teacher$_{focal}$	34.374	14.145	7.855	10.687	15.880	5.626e02
Unbiased Teacher$_{CE}$	42.382	18.008	11.387	13.041	20.135	6.229e03
SoftTeacher	38.421	13.556	6.623	**16.756**	13.045	5.526e-02
Proposed	**50.632**	**20.094**	**12.713**	15.219	**21.774**	—
2% Labeled Data						
Supervised	47.140	18.609	9.480	24.033	18.586	2.558e-14
Unbiased Teacher$_{focal}$	71.608	31.752	20.479	27.871	32.430	3.975e-04
Unbiased Teacher$_{CE}$	**72.416**	31.490	21.446	26.767	32.666	2.010e-01
SoftTeacher	60.366	25.421	14.767	17.991	28.323	2.558e-08
Proposed	72.341	**32.311**	**21.614**	**29.780**	**33.556**	—

Table 1. Experimental results on m2cai16-tool-locations dataset with ResNet50-FPN as backbone