CLIPath: Fine-tune CLIP with Visual Feature Fusion for Pathology Image Analysis Towards Minimizing Data Collection Efforts

Zhengfeng Lai, Zhuoheng Li, Luca Cerny Oliveira, Joohi Chauhan, Brittany Dugger, Chen-Nee Chuah

at ICCV 2023 Workshop on Computer Vision for Automated Medical Diagnosis

ICCV23 PARIS

UCDAVIS ELECTRICAL AND COMPUTER ENGINEERING

Introduction

➢ CLIP has shown its strong transferability
  ○ It was pre-trained on 400M image-text pairs
  ○ It shows promising zero-shot ability
➢ Adaptation of CLIP in downstream tasks
  ○ Full-model fine-tuning
    ■ Time-consuming, require heavy computes
    ■ Not scalable for multiple tasks
  ○ Adapter-based fine-tuning
    ■ Freeze the whole network
    ■ Design a specific layer and only train it
    ■ Simple and scalable
   ■ Learning ability may be limited

➢ CLIP zero-shot inference
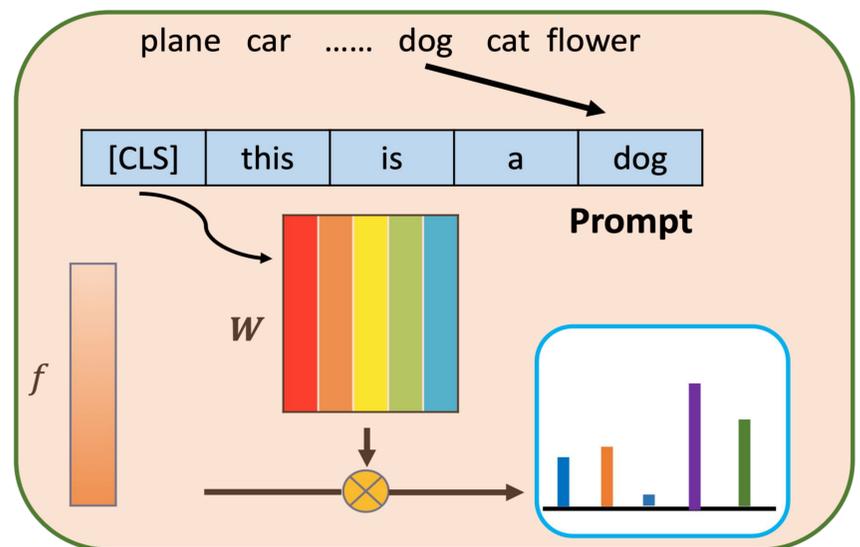
Methodology

➢ CLIP Residual Feature Connection (RFC)
  ○ Preserve pre-trained knowledge
  ○ Learn new knowledge
  ○ α: residual ratio to balance the above knowledge
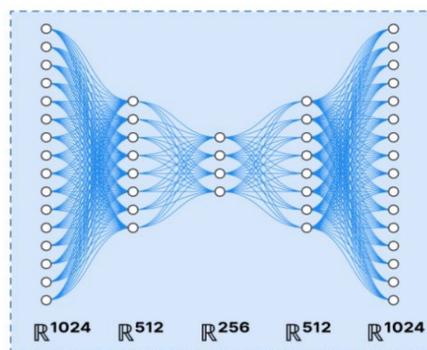
➢ Contrastive Loss
  ○ Keep contrastive loss in tuning
  ○ Preserve the pre-training properties
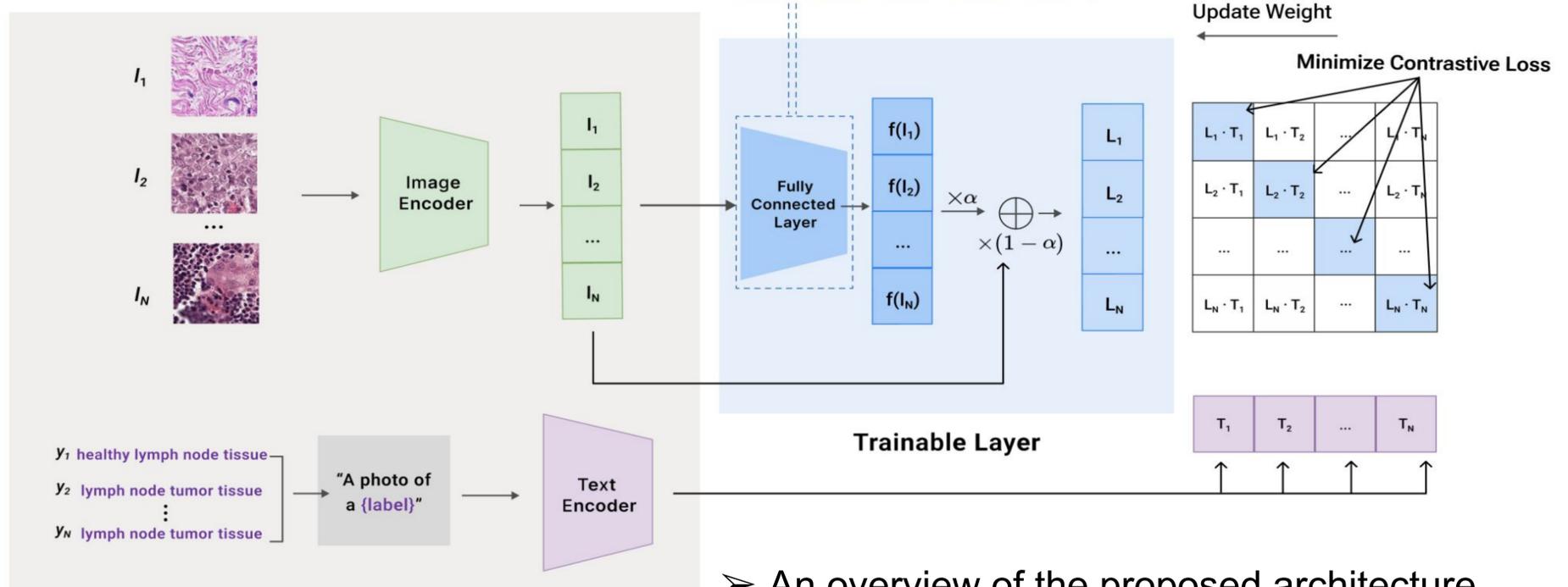
➢ An overview of the proposed architecture.

Main Experimental Results on Two Public Pathology Datasets: PCam and MHIST

➢ Quantitative results on the hold-out test set of PCam.[1]

Algorithm	Data Usage	Accuracy	Recall	Precision	F1-score	AUC
CLIP	Zero-shot	56.5	50.3	57.4	53.7	0.600
CLIP + RFC	0.1%	76.4	90.0	70.7	79.2	0.849
	0.5%	81.5	85.0	79.4	82.1	0.894
	1%	81.9	82.9	81.3	82.1	0.900
	5%	82.9	77.1	87.2	81.8	0.918
	10%	82.8	79.2	85.4	82.1	0.914
	50%	81.4	71.0	89.6	79.3	0.918

➢ Quantitative results on the hold-out test set of MHIST.[2]

Algorithm	Data Usage	Accuracy	Recall	Precision	F1-score	AUC
CLIP	Zero-shot	36.9	100.0	36.9	53.9	0.501
CLIP + RFC	1%	63.9	7.5	57.5	13.3	0.643
	5%	66.8	42.8	56.6	48.7	0.732
	10%	70.5	79.7	57.1	66.6	0.784
	20%	70.7	86.1	56.8	68.4	0.788
	50%	74.8	75.6	63.3	68.9	0.838

➢ Performance and Complexity comparison on PCam.[1]

Data Usage	Algorithm	Accuracy	Training Time
Zero-shot	CLIP	56.5	-
0.1%	CLIP + CoOp	64.3	**7 min 6 sec**
	CLIP + RFC	**76.4**	10 min 29 sec
1%	CLIP + CoOp	61.9	53 min 21 sec
	CLIP + RFC	**81.9**	**11 min 56 sec**
10%	CLIP + CoOp	59.9	2 h 23 min 45 sec
	CLIP + RFC	**82.8**	**27 min 18 sec**

[1] Veeling, Bastiaan S., et al. "Rotation equivariant CNNs for digital pathology." Medical Image Computing and Computer Assisted Intervention–MICCAI.
[2] Wei, Jerry, et al. "A petri dish for histopathology image analysis." Artificial Intelligence in Medicine: 19th International Conference on Artificial Intelligence in Medicine, AIME 2021.