

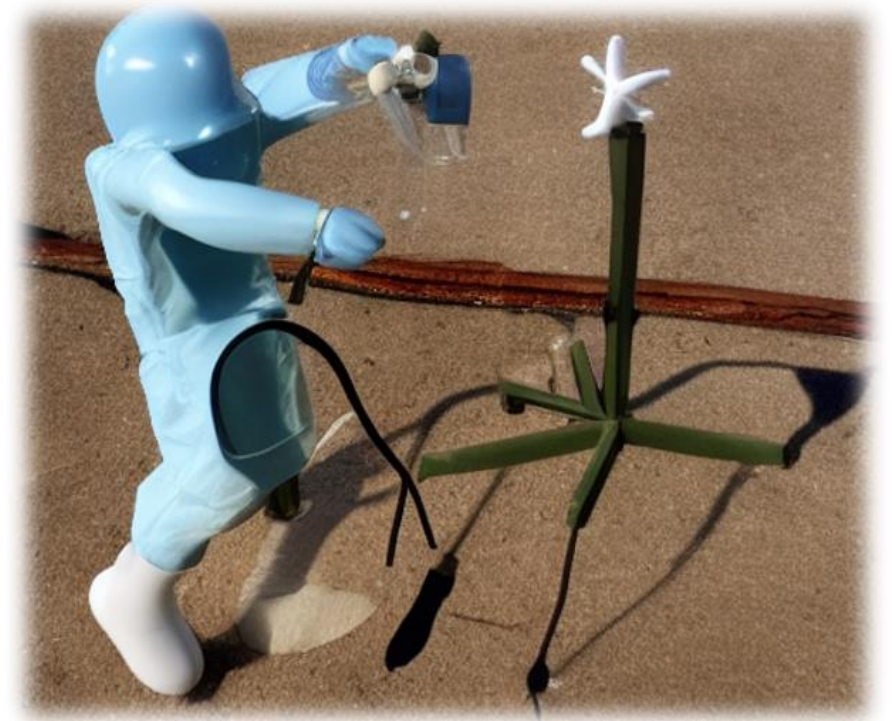


Machine Learning Seminar

Practical Course: Machine Learning in Graphics, Vision and Language
Summer Semester 2024

Why are you here? to do science and stuff

- Learn to solve a given complex problem in a group, using your programming skills
- Select a fun research topic
- Come up with a solution that you like
- Implement that solution in a group (4-5 people)
- Evaluate your solution
- Make it "nice"
- Present your work to the group
- Write a scientific report & share your code



science and stuff

Topics

- Interesting & novel research questions
- One topic can be shared by multiple groups, if they concentrate on different aspects / solutions
- If you have some other topic in mind, feel free to discuss it with us :)
- Aligned with our research interests:
 - Natural Language Processing
 - Computer Vision

Topics

- You can use whichever language and tools you want to
- Most of the papers already have implementations available which you can directly use
- We will grade you based on the novel ideas/additional experiments you do
- Just don't take credit for stuff you used from others!

Organization: ILIAS

ILIAS System (**important**):

- Important information, materials, templates, dates, ...
- Groups

Slack for group selection and communication

Registration starts on 25th April at 10am (so that you have some time to go through list of topics and decide if you would like to join)

Contact:

valay.bundele@uni-tuebingen.de

zohreh.ghaderi@uni-tuebingen.de

Organization

Next meeting in two weeks [07.05.2024]

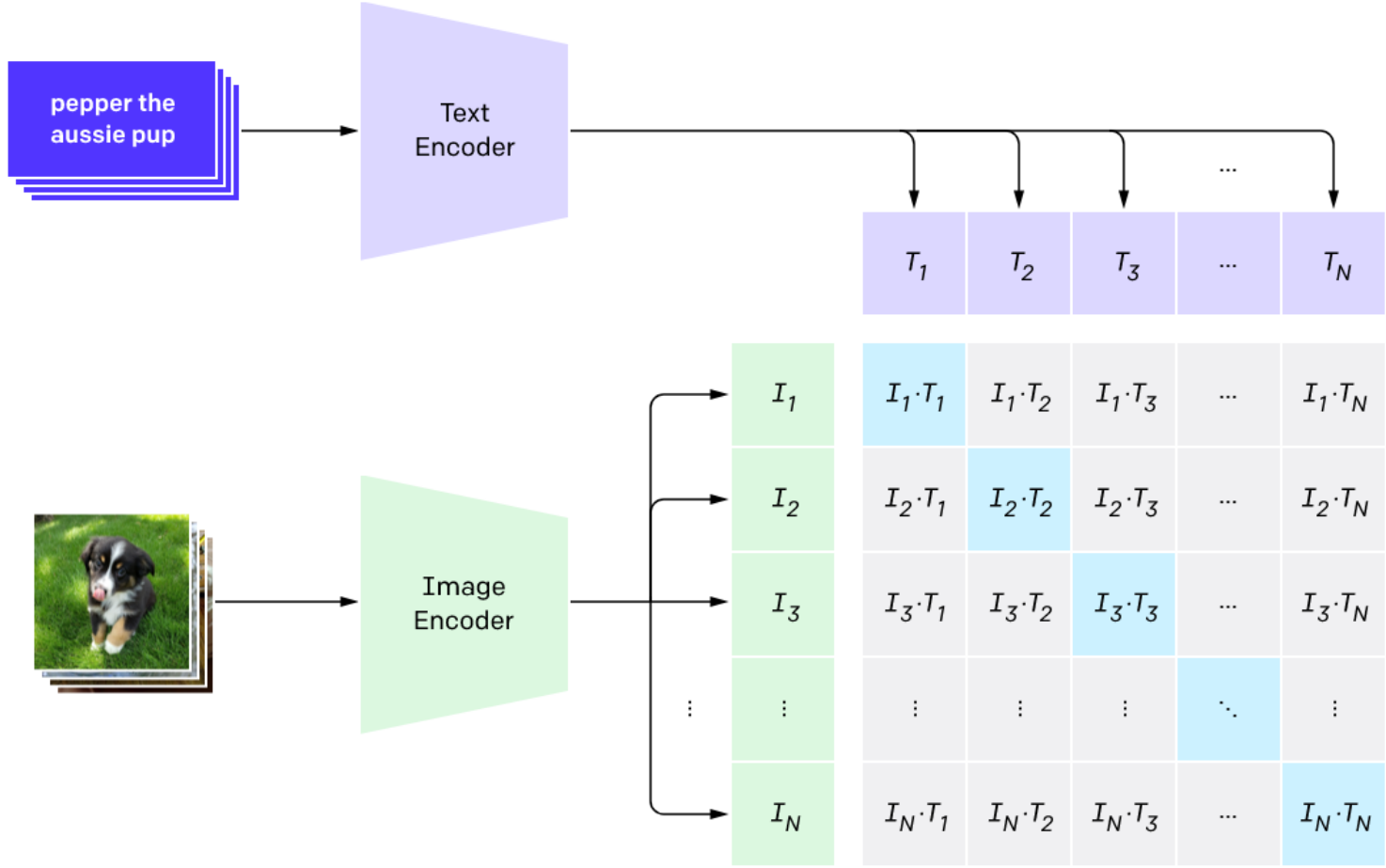
- Fix groups and topics
 - Fill out the form for WSI account
 - So you could use our machines with GPUs
 - Present short brainstorm of ideas and a rough plan for the semester
-
- Regular individual group meetings
 - Progress discussion
 - Final Presentation [16.07.2024]
 - 15 mins presentation of the project & submission of the slides
 - Written Report(paper?) until [15.09.2024]



Topics


CLIP (a vision-language model)

1. Contrastive pre-training




1. Zero-shot Image Classification with vision-language models

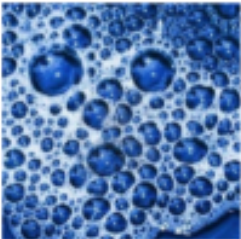
- Zero-shot classification performance depends a lot on the input prompt
- CoOp makes the input prompt learnable; trains the prompt vectors on a small subset of dataset

Caltech101	Prompt	Accuracy
	a [CLASS].	82.68
	a photo of [CLASS].	80.81
	a photo of a [CLASS].	86.29
	[V]₁ [V]₂ ... [V]_M [CLASS].	91.83


(a)

Flowers102	Prompt	Accuracy
	a photo of a [CLASS].	60.86
	a flower photo of a [CLASS].	65.81
	a photo of a [CLASS], a type of flower.	66.14
	[V]₁ [V]₂ ... [V]_M [CLASS].	94.51

(b)

Describable Textures (DTD)	Prompt	Accuracy
	a photo of a [CLASS].	39.83
	a photo of a [CLASS] texture.	40.25
	[CLASS] texture.	42.32
	[V]₁ [V]₂ ... [V]_M [CLASS].	63.58

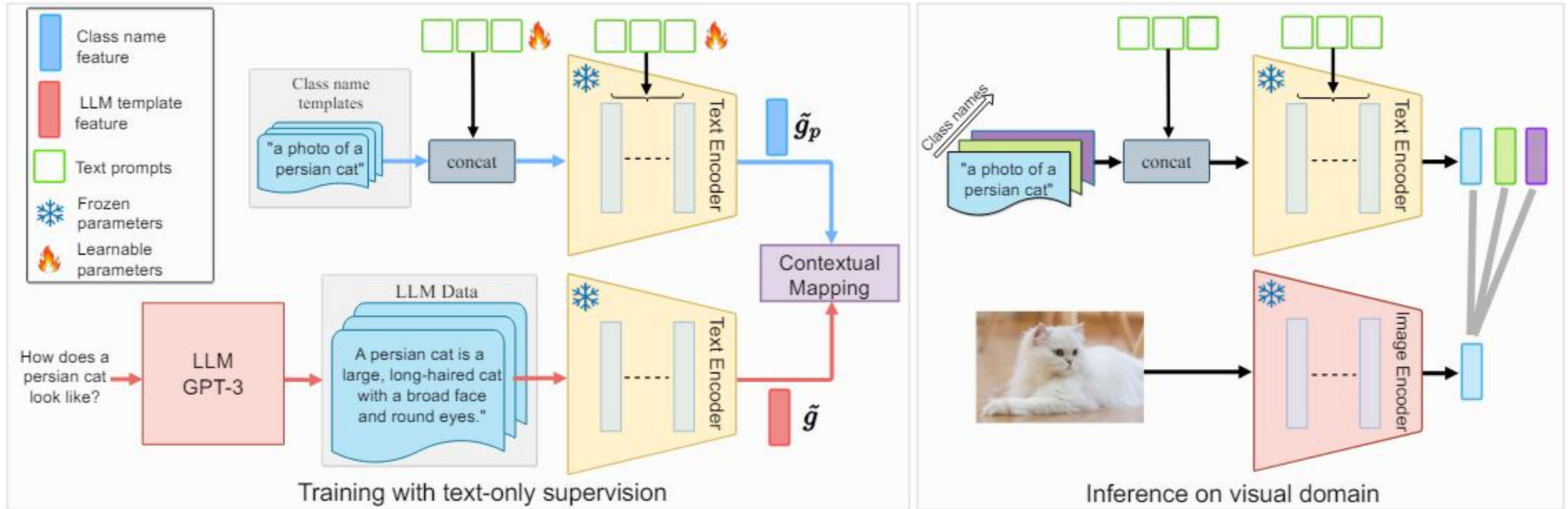
(c)

EuroSAT	Prompt	Accuracy
	a photo of a [CLASS].	24.17
	a satellite photo of [CLASS].	37.46
	a centered satellite photo of [CLASS].	37.56
	[V]₁ [V]₂ ... [V]_M [CLASS].	83.53

(d)

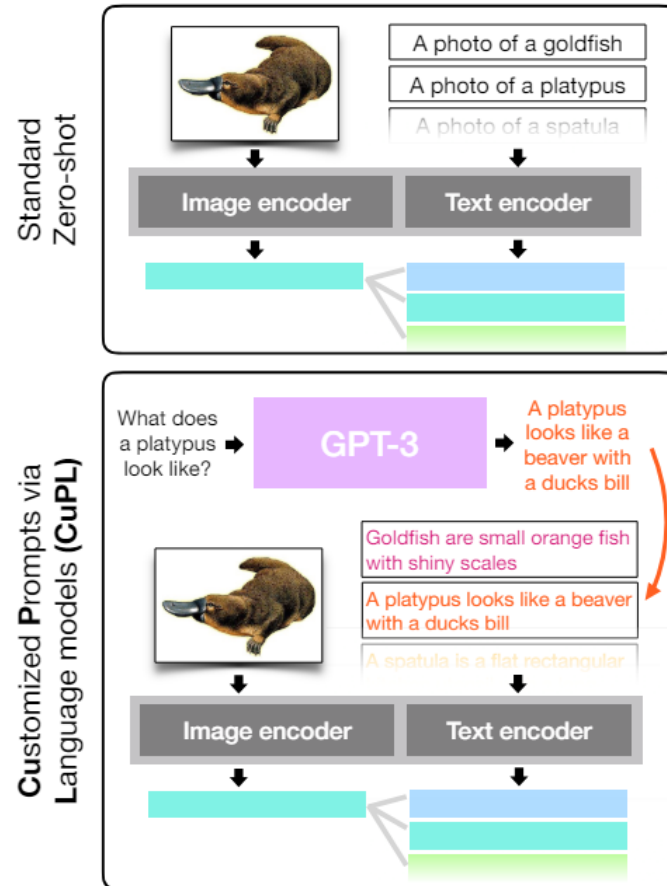
1. Zero-shot Image Classification with vision-language models

Learning to Prompt with Text Only Supervision for Vision-Language Models



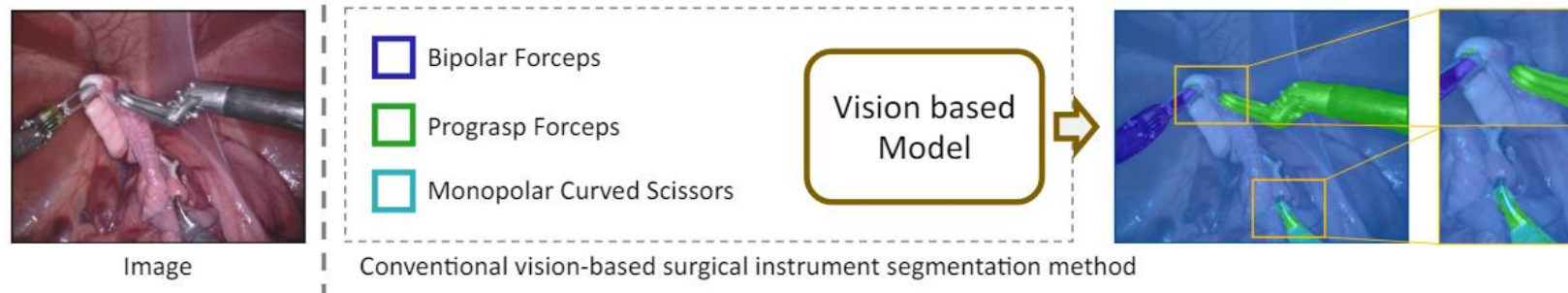
1. Zero-shot Image Classification with vision-language models

What does a platypus look like? Generating customized prompts for zero-shot image classification



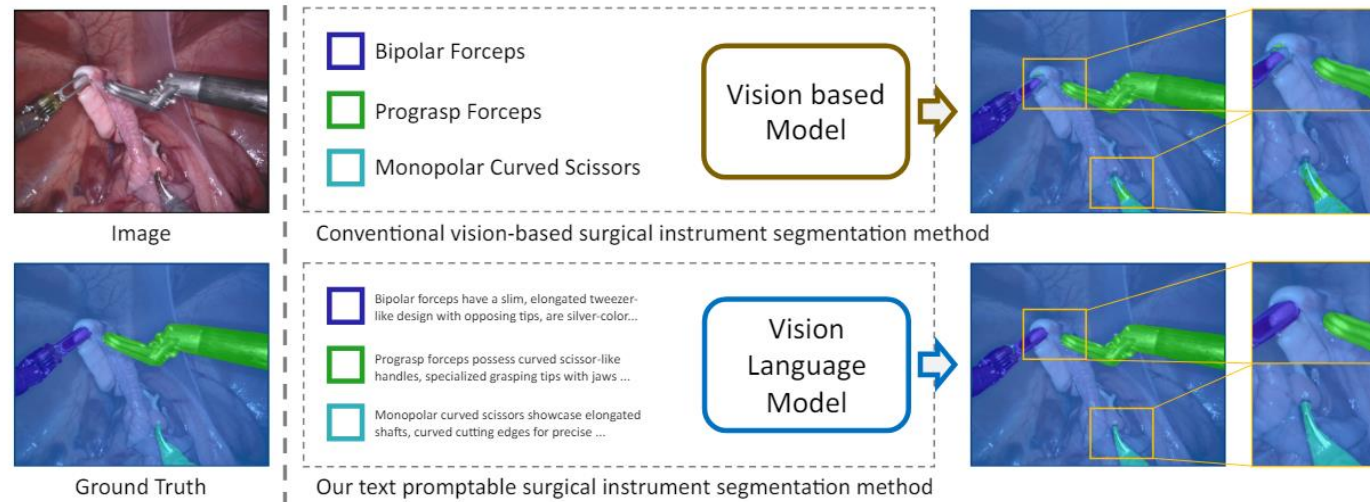
2. Generalizable Surgical Instrument Segmentation using Vision-Language models

- To segment surgical instruments from videos; helpful for minimally invasive surgeries
- Wide variety of surgical instruments; lack of an annotated dataset for all of them
- Most SIS models trained only on pre-defined categories; don't generalize to other instruments



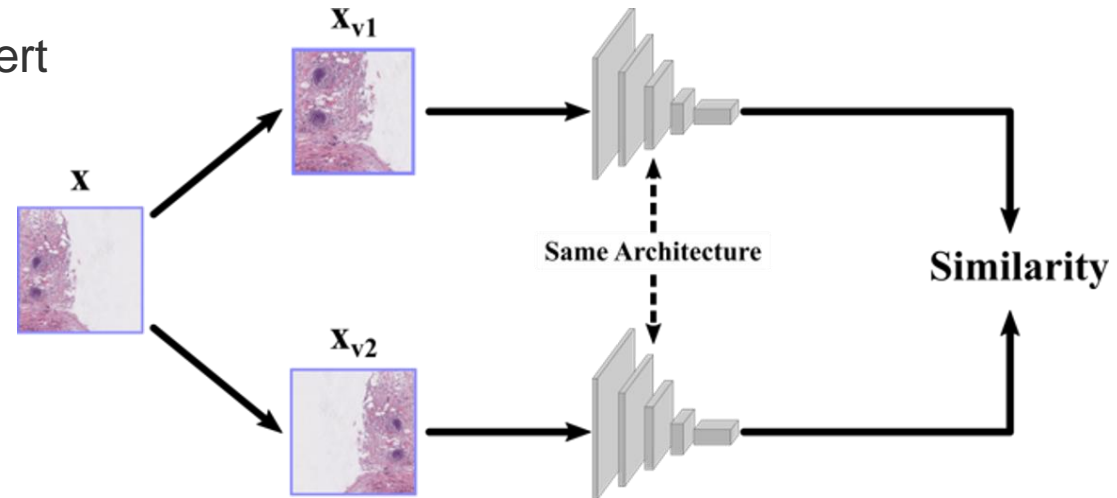
2. Generalizable Surgical Instrument Segmentation using Vision-Language models

- Make the task text-promptable and use vision-language models to enhance generalization
- What can you do?
 - Experiment with the architecture and the losses
 - Check if medical domain specific vision-language model can be used for better performance
 - Maybe try to use temporal information from videos to improve segmentation results



3. Self-Supervised Learning for Medical Image Analysis

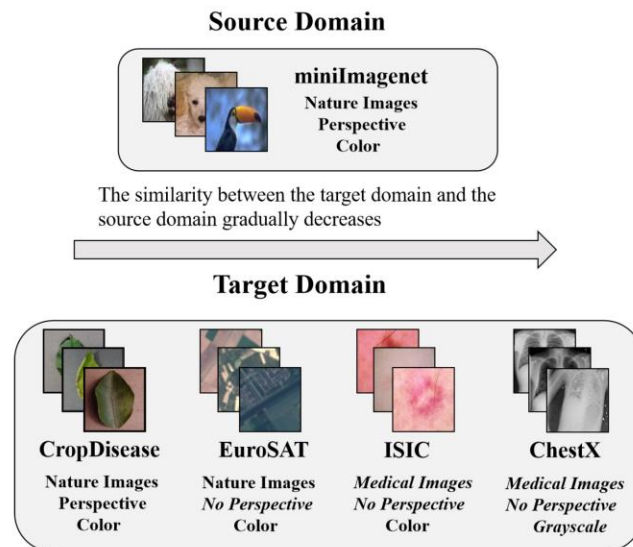
- Labeling data is time-consuming and often requires expert knowledge, e.g. in medical image analysis
- Self-supervised learning, harness the data itself as a supervisory signal



- Tasks:
 - Curate a multi-modal 2D image dataset (Histopathological, Endoscopic, Fundus Images, ...)
 - Train CNN- and Transformer-based architectures
 - Compare different cutting-edge (contrastive) self-supervised strategies
 - Evaluate their performance on different downstream tasks (Image classification, instance retrieval, ...)

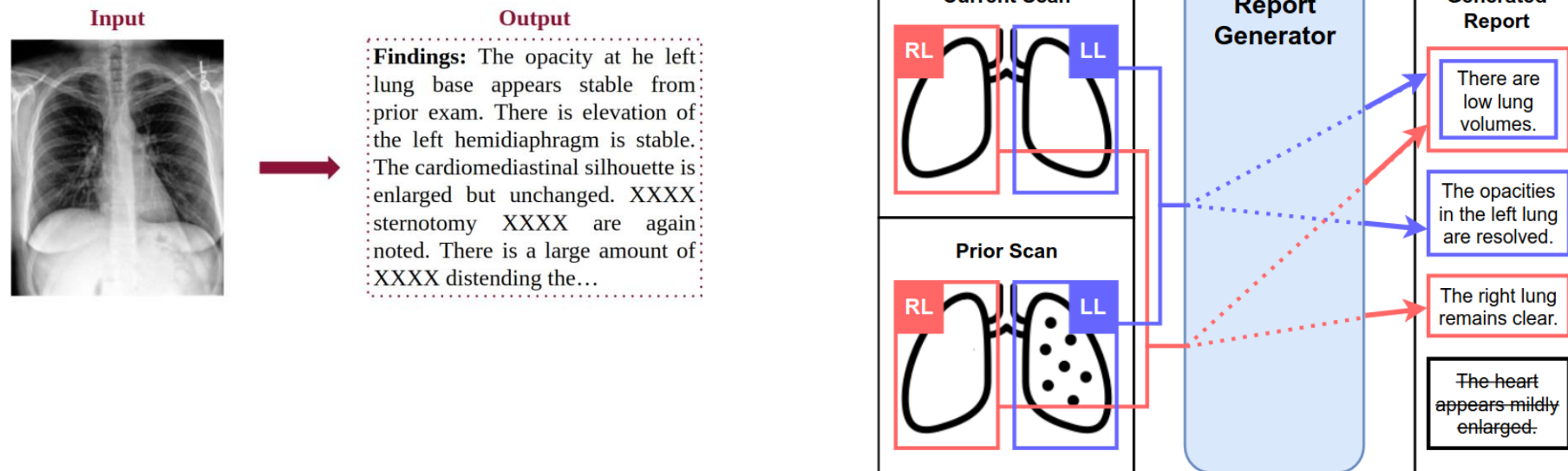
4. Cross-Domain Few-shot Medical Image Classification

- Not enough annotated data available in every domain
- Similarities exist between different medical image analysis datasets
- Can we leverage shared knowledge to learn previously unseen tasks more efficiently?
- What's the task?
 - Given datasets from different domains (Breast Ultrasound, Chest X-Ray, Colorectal cancer...)
 - Learn a model which could generalize to new domains (Dermatoscopy, Fundus Multi-disease) with just a few samples



5. Utilizing Longitudinal Information for Chest X-Ray Report Generation

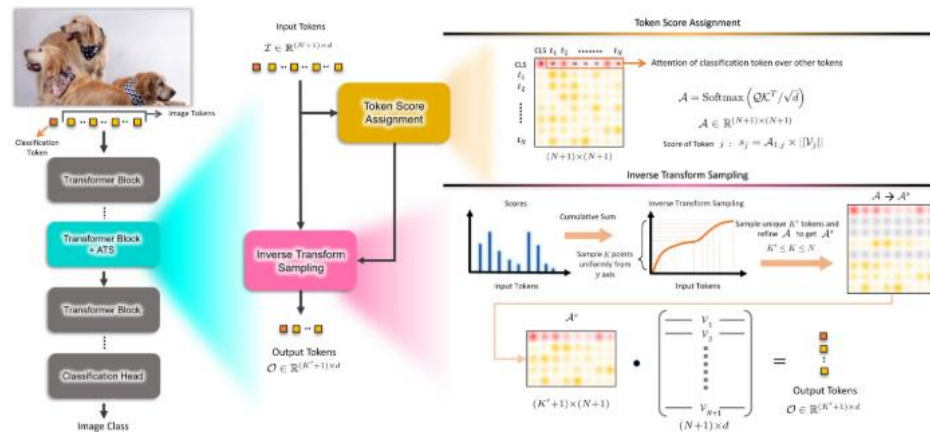
- Automatically generating a radiology report from a given patient's CXR
- What you can do?
 - Use a vision encoder like MedClip/RadDino along with gpt-2m to generate reports
 - Experiment with also using reports from previous visits of the patient
 - Use pretrained models and finetune them for the task



6. Adaptive Token Sampling for Efficient Vision Transformers

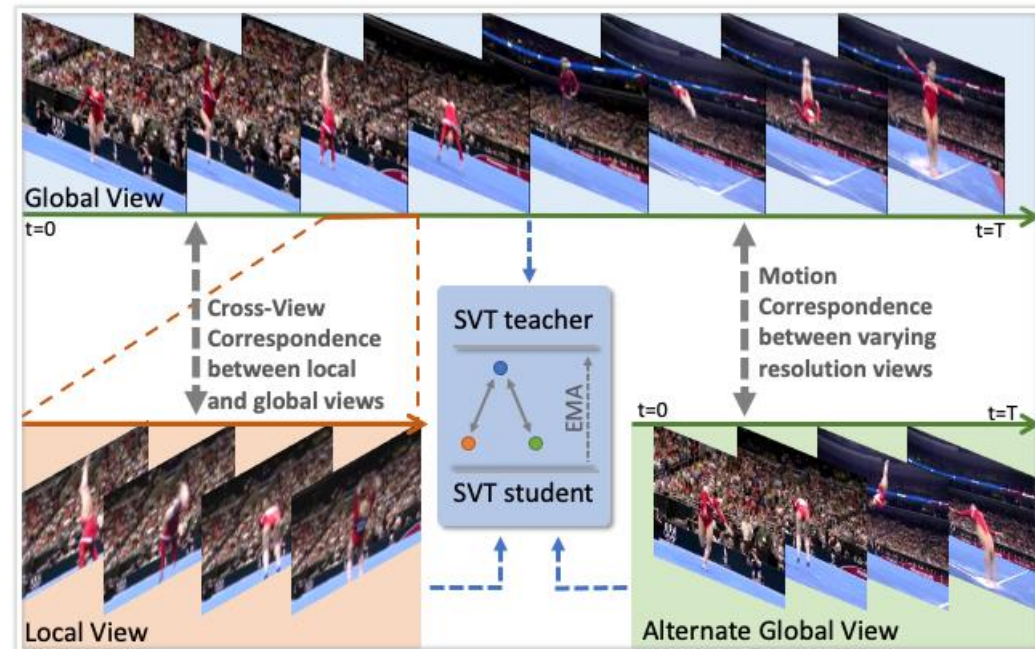
- Vision Transformers are very computationally expensive
- Authors propose a method to adaptively sample significant tokens to make ViTs efficient
- What can you do?
 - Train the proposed model on a small dataset (maybe TinyImageNet) for image classification
 - Try to improve the approach to further reduce the computational cost on the same dataset

Paper : <https://arxiv.org/abs/2111.15667>



7. Self-supervised Video Transformer

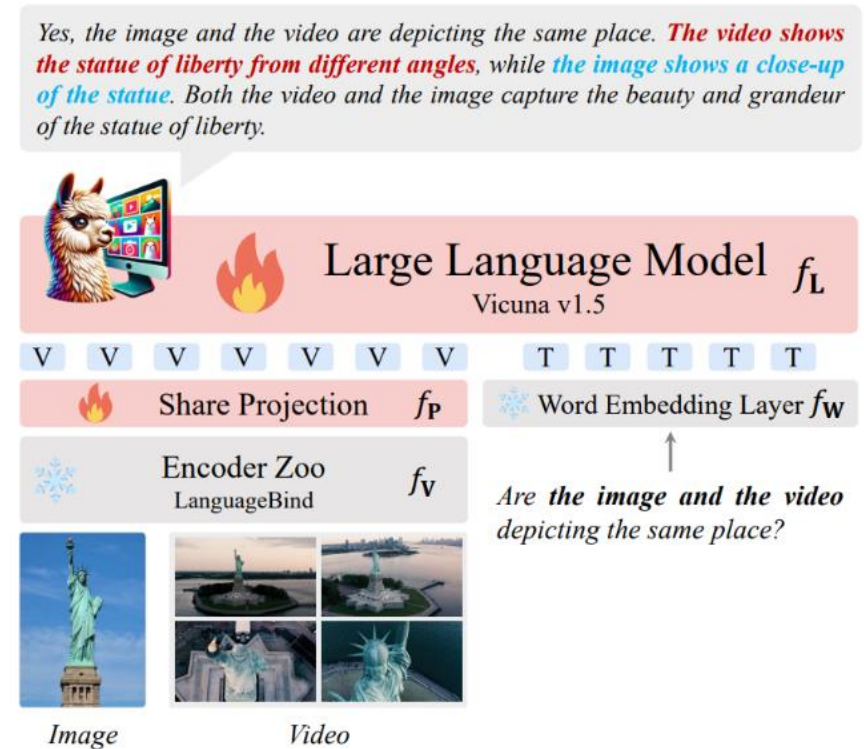
- A method for self-supervised training of video transformers using unlabeled video data
- Authors report results on action recognition benchmarks
- What can you do?
 - Adapt to some other tasks like video object recognition/video captioning
 - Use the pretrained video transformer along with something on top to perform the task
 - Keep the video transformer frozen (very expensive to train from scratch)
- Datasets:
 - Charades (object recognition)
 - MSVD (video captioning)
- Paper :
 - <https://arxiv.org/abs/2112.01514>



8. Sensitivity and Robustness of VideoLLava to Prompt Template

- Recent research on engineering relevance has surged thanks to advancements in pre-trained and large language models. However, a significant challenge has arisen: these models often struggle with sensitivity and robustness when faced with prompt templates. You can do this with Video-LLaVA:
- Prompt sensitivity - data
- Output distribution analysis
- Smoothness of video - “creativity” of the model
- Fairness, diversity, inclusion topics - comparison with other

- Papers:
 - <https://arxiv.org/abs/2305.08714>
 - <https://arxiv.org/abs/2311.10122>
 - <https://arxiv.org/abs/2310.11324>
- Dataset:
 - MSVD



(a) Illustration of Video-LLaVA

Your Next Tasks

1. Find more info about the topics: <https://docs.google.com/spreadsheets/d/1j1zbMLxR3fMb8XqaSs6ObDq-rJgXSVoCkQPveYIBAfW/edit?usp=sharing>
2. Apply for WSI user account (until 27.04)
 - Application letter is on ILIAS in the supplementary materials folder
 - Guidelines for WSI are also be available
 - Submit your filled out form using the 'WSI Application' exercise
3. Find a group (until 02.05)
 - Slack link: https://join.slack.com/t/practical-course-2024/shared_invite/zt-2hi7s3bjk-BI0IF6NzR7cOda8MaKGqkA
 - Inform us about your group members and the topic you will be working on via email
 - ⑩ One email per group
4. Have fun doing science and stuff !

Timetable

Date		
23.04.2024	Meeting Zoom	Introduction <ul style="list-style-type: none">• logistics• Topics introductions
07.05.2024	Meeting (All)	Topics selected Present your ideas (solutions) & your rough working plan
28.05.2024	Meeting (Groups)	Progress meeting (time slot for each group around 15 min)
11.06.2024	Meeting (Groups)	Progress meeting
25.06.2024	Meeting (Groups)	Progress meeting
09.07.2024	Meeting (Groups)	Progress meeting
16.07.2024	Meeting (All)	Final presentation
15.09.2024	Deadline	Final report submission