MAY 2025

TEXAS
The University of Texas at Austin

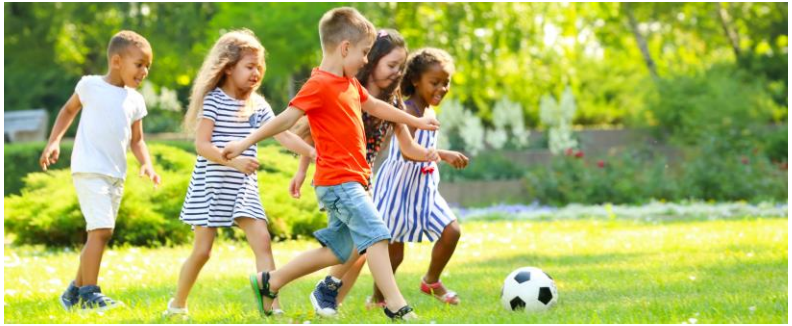
PATH-AGENT: MIMICKING A CLINICAL DIAGNOSTIC WORKFLOW FOR OPEN-ENDED PATHOLOGY VISUAL QUESTION ANSWERING

AWAIS NAEEM[1], YING DING[1], AMY COFFEE[2], CHANDRA KRISHNAN[2]
School of Information[1], Dell Medical School[2] - The University of Texas at Austin

PRESENTER: AWAIS NAEEM
Graduate Student, The University of Texas at Austin

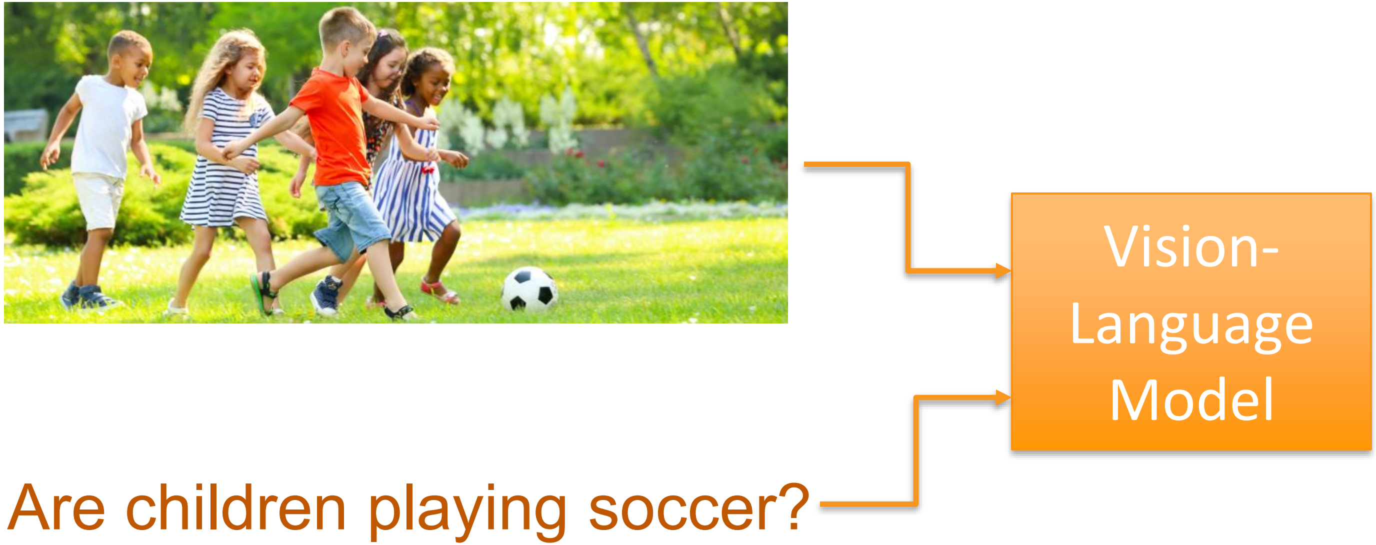
Visual Question Answering (VQA)

Visual Question Answering (VQA)



Are children playing soccer?

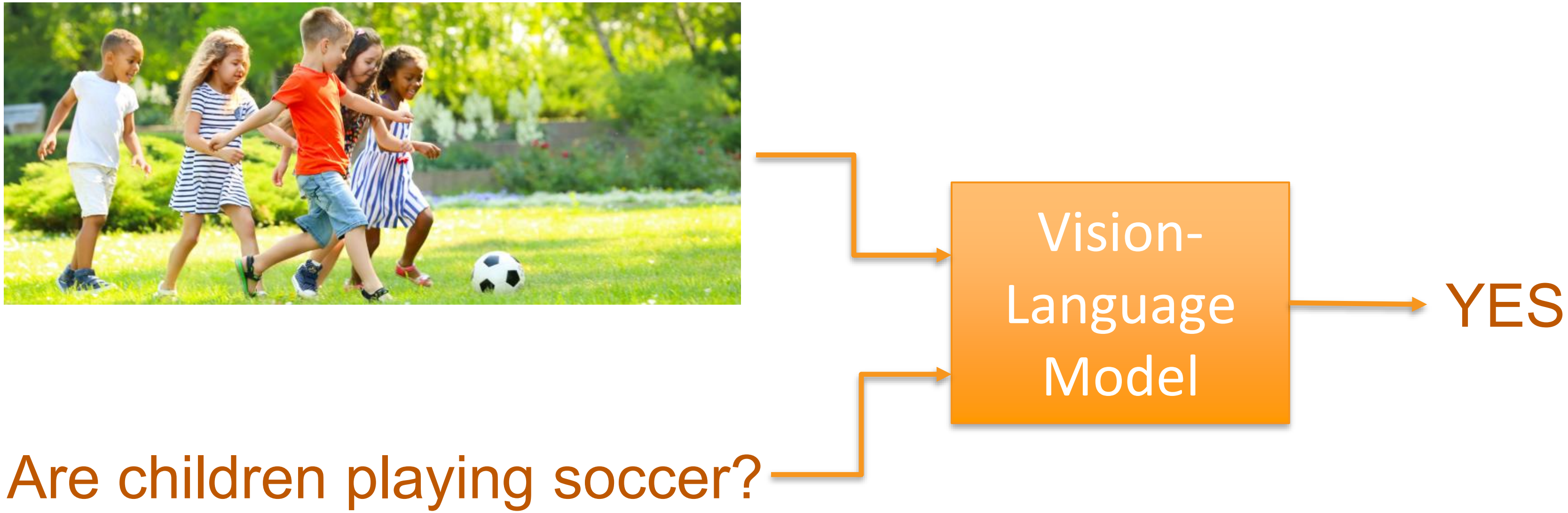
Visual Question Answering (VQA)



Vision-Language Model

Are children playing soccer?

Visual Question Answering (VQA)



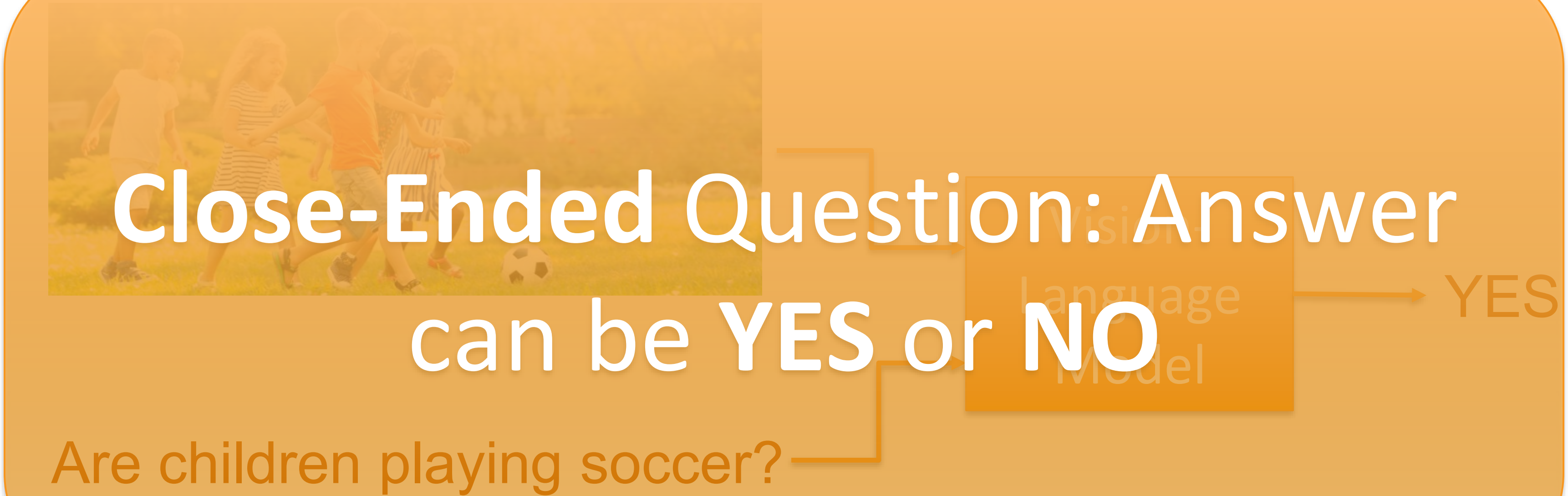
Vision-Language Model

YES

Are children playing soccer?

Visual Question Answering (VQA)

Close-Ended Question: Answer can be **YES** or **NO**

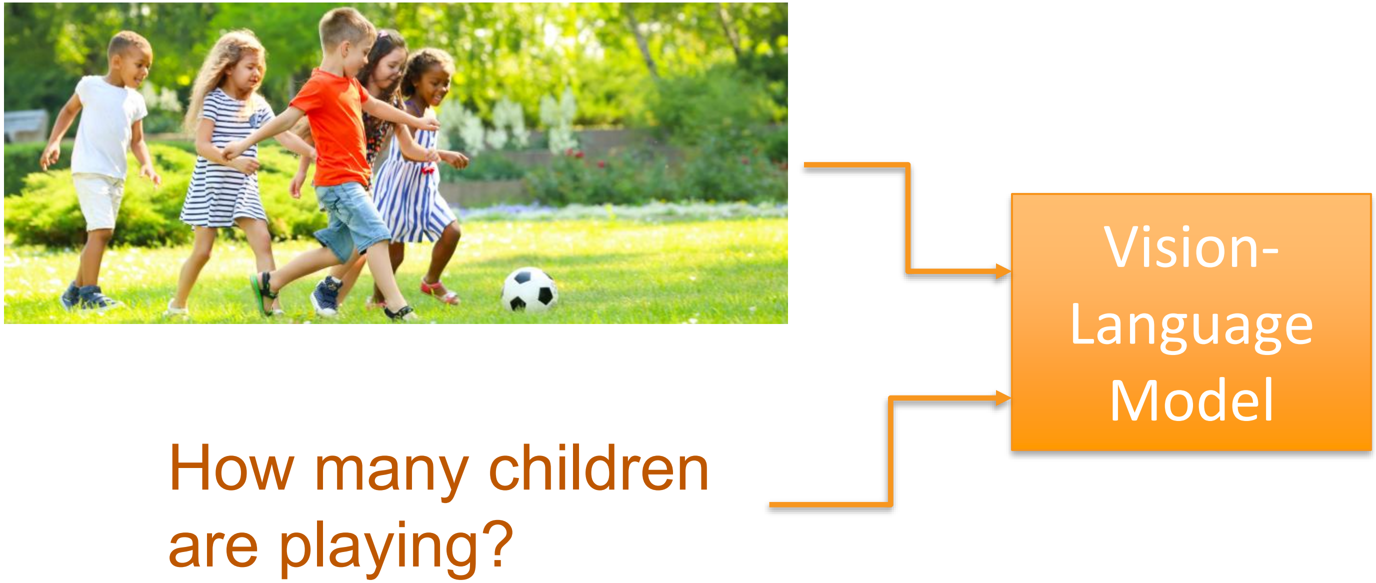


Vision-Language Model

YES

Are children playing soccer?

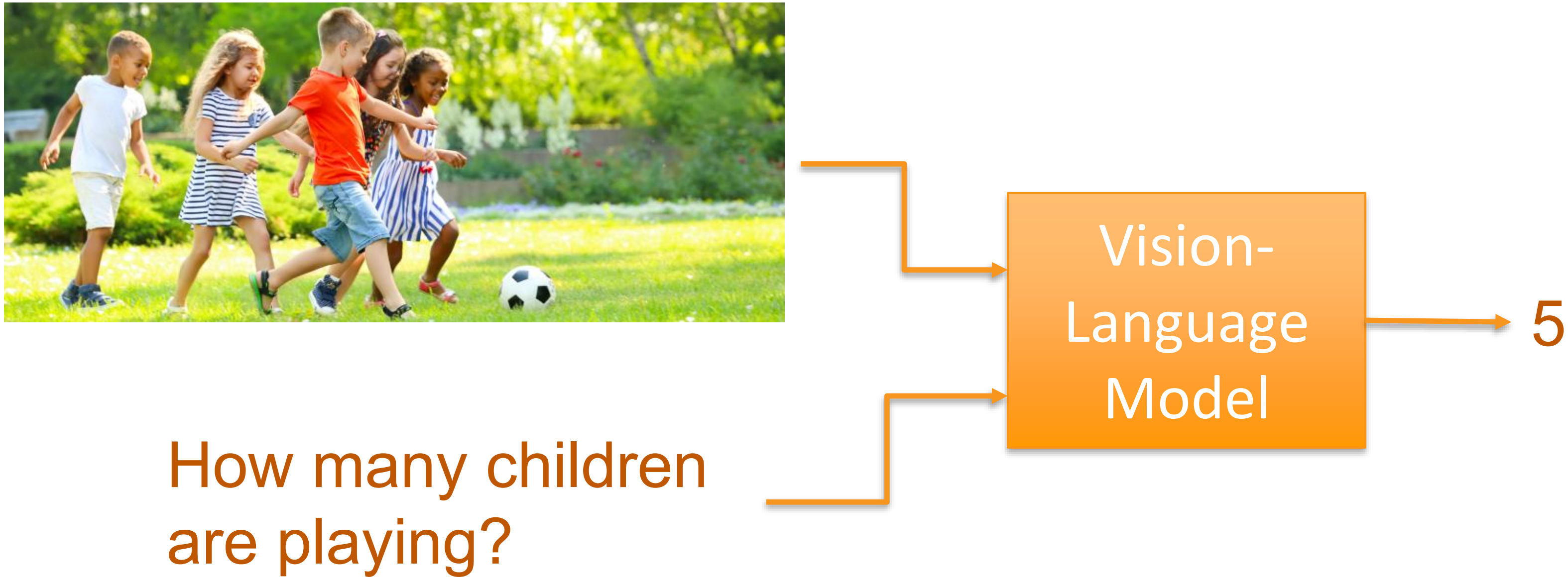
Visual Question Answering (VQA)



How many children are playing?

Vision-Language Model

Visual Question Answering (VQA)

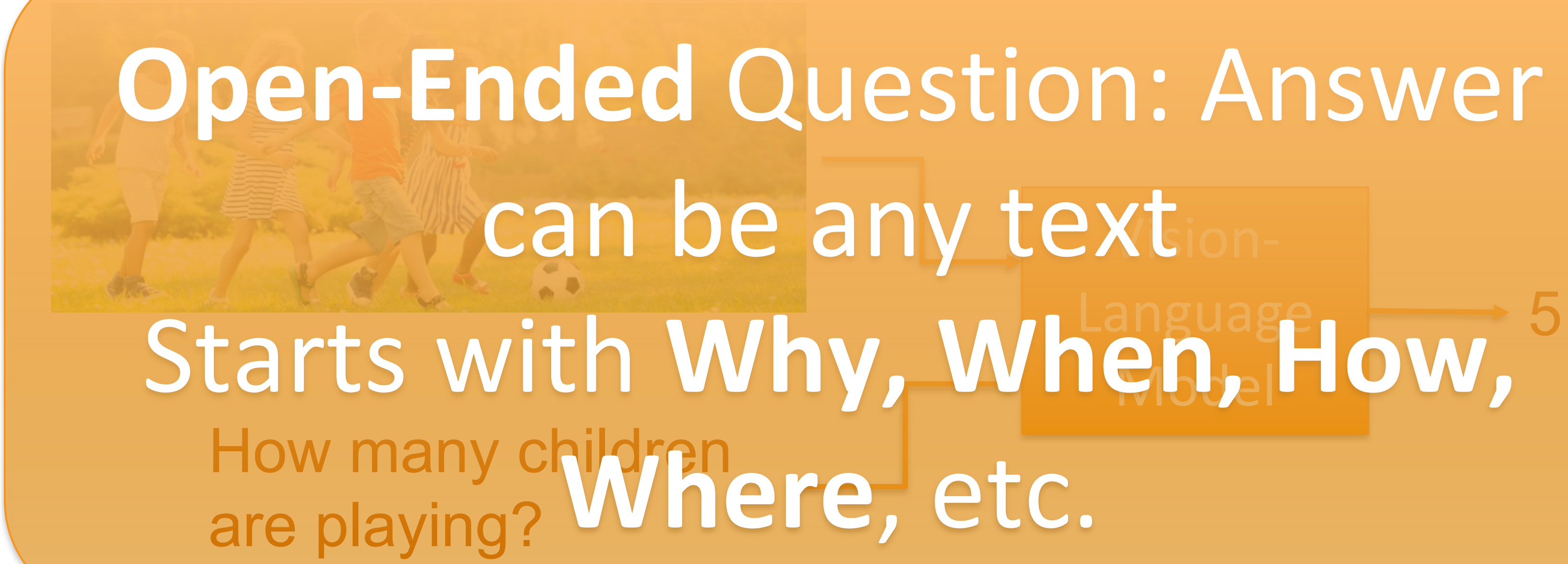


How many children are playing?

Vision-Language Model

5

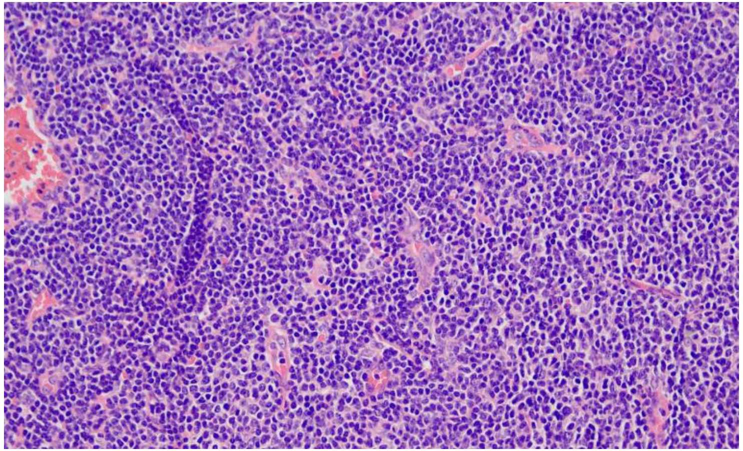
Visual Question Answering (VQA)



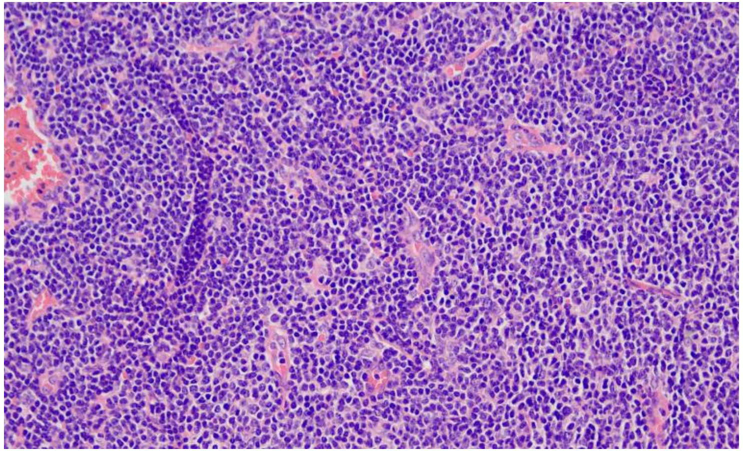
Open-Ended Question: Answer can be any text

Starts with **Why, When, How, Where**, etc.

VQA: From normal to Pathology

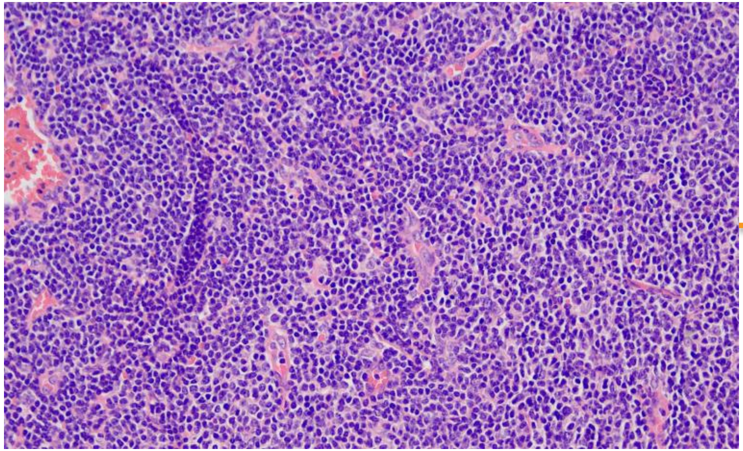


VQA: From normal to Pathology



Is the nucleus to cytoplasmic ratio of these lymphocytes high?

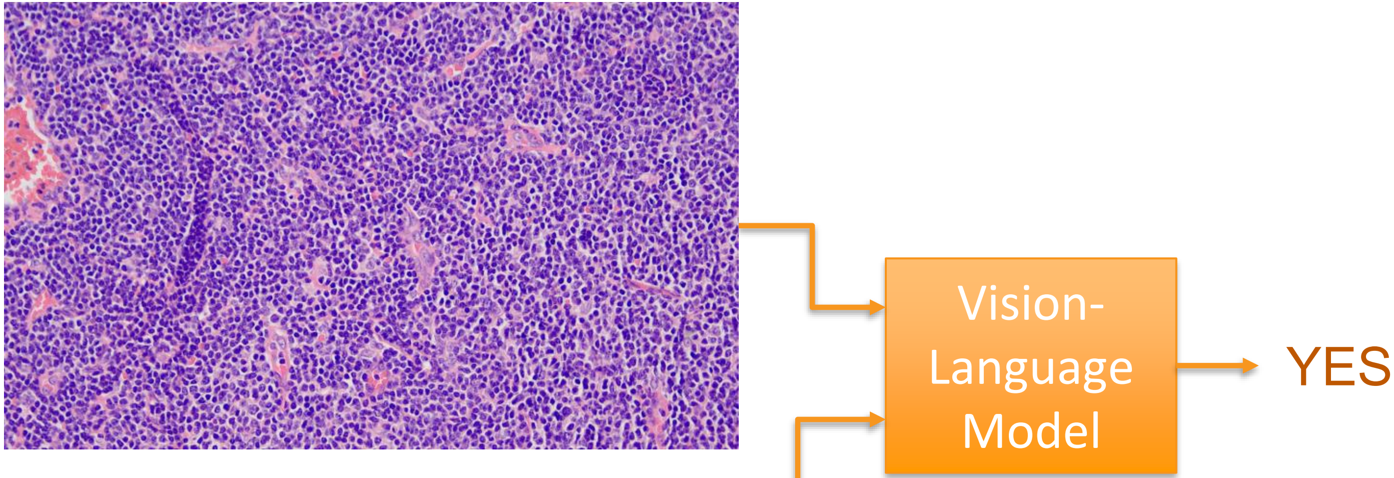
VQA: From normal to Pathology



Is the nucleus to cytoplasmic ratio of these lymphocytes high?

Vision-Language Model

Pathology VQA: Close-Ended



Is the nucleus to cytoplasmic ratio of these lymphocytes high?

Pathology VQA: Close-Ended

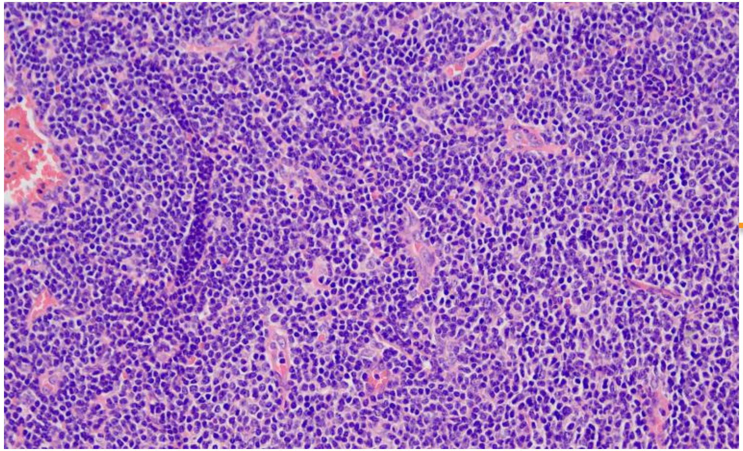
Is the nucleus to cytoplasmic ratio of these lymphocytes high?

Vision Language Model

YES

SOTA VLMs (LLaVA-MED, LLaVA-Med++) ~ **91% (Close-Ended)**

Pathology VQA: Open-Ended



What is the predominant cell type seen here?

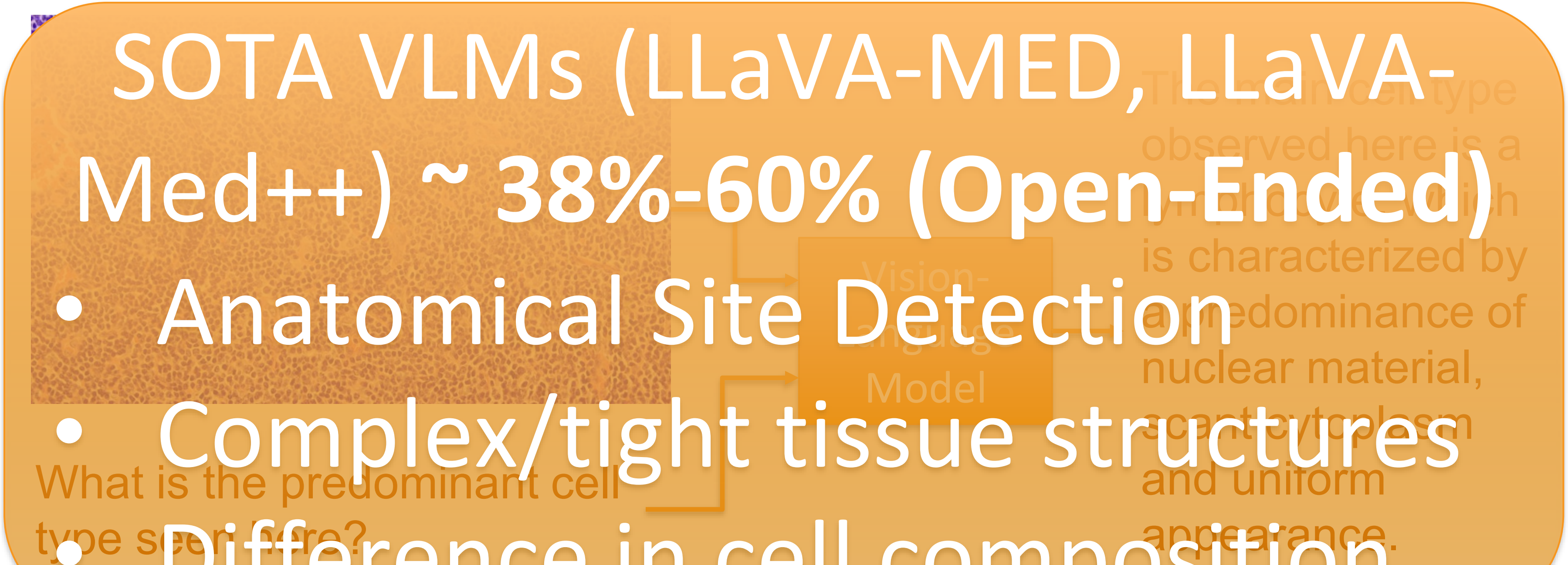
Vision-Language Model

The main cell type observed here is a lymphocyte, which is characterized by a predominance of nuclear material, scant cytoplasm and uniform appearance.

Pathology VQA: Open-Ended

SOTA VLMs (LLaVA-MED, LLaVA-Med++) ~ **38%-60% (Open-Ended)**

- Anatomical Site Detection
- Complex/tight tissue structures
- Difference in cell composition



What is the predominant cell type seen here?

The main cell type observed here is a lymphocyte, which is characterized by a predominance of nuclear material, scant cytoplasm and uniform appearance.

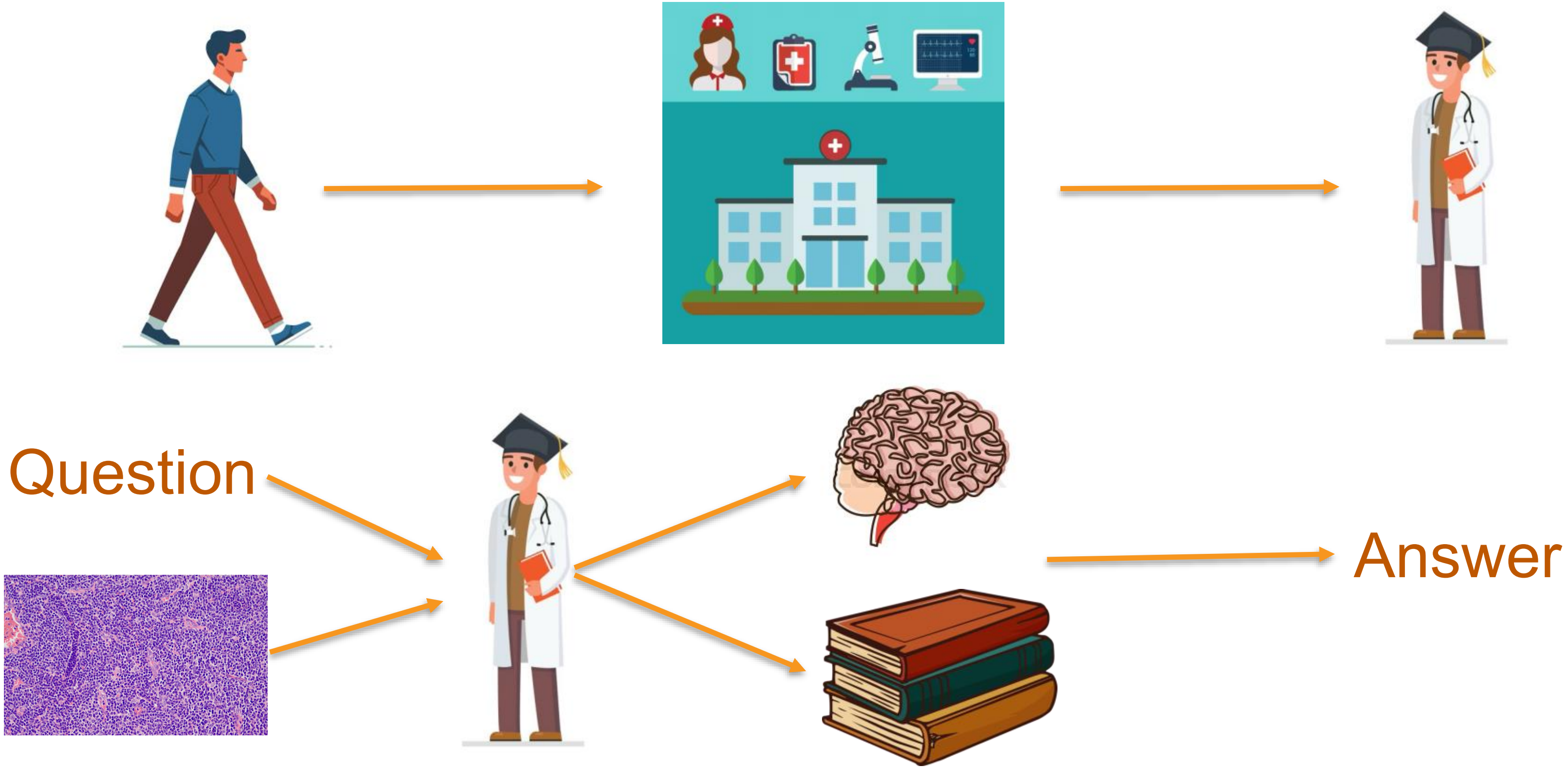
Path-Agent: Mimicking the Pathologist

A multi-agent framework mimicking the diagnostic workflow of a human pathologist

- Knowledgebase Agent
- Magnifier Agent
- ROI Agent
- Patch Agent
- Critique Agent
- Response Agent

Pathologist Knowledgebase

Pathologist Knowledgebase

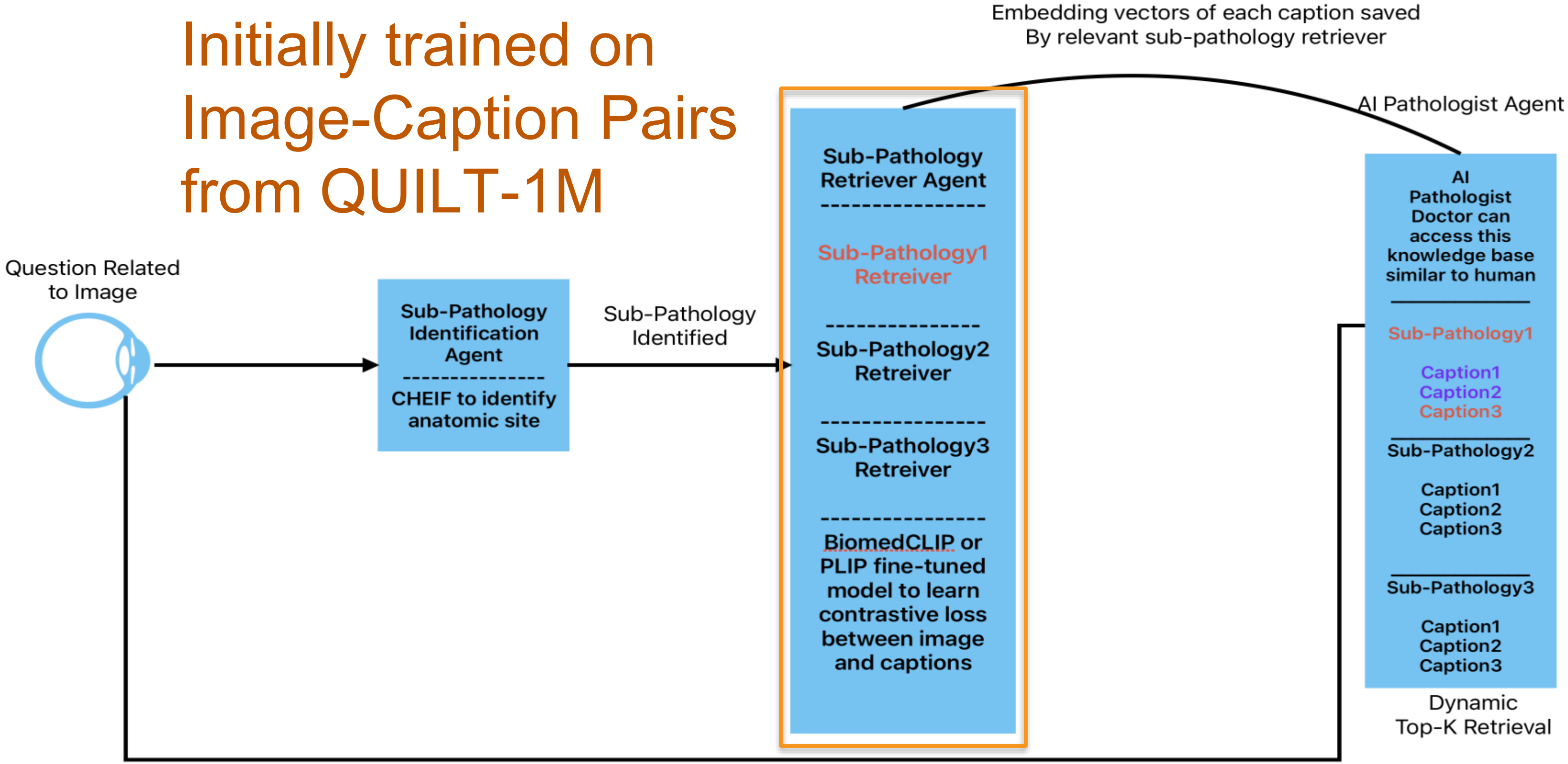


Question

Answer

Knowledgebase Agent - Training

Initially trained on Image-Caption Pairs from QUILT-1M



Embedding vectors of each caption saved
By relevant sub-pathology retriever

AI Pathologist Agent

Question Related to Image

Sub-Pathology Identification Agent

CHEIF to identify anatomic site

Sub-Pathology Identified

Sub-Pathology Retriever Agent

Sub-Pathology1 Retreiver

Sub-Pathology2 Retreiver

Sub-Pathology3 Retreiver

BiomedCLIP or PLIP fine-tuned model to learn contrastive loss between image and captions

AI Pathologist Doctor can access this knowledge base similar to human

Sub-Pathology1
Caption1
Caption2
Caption3

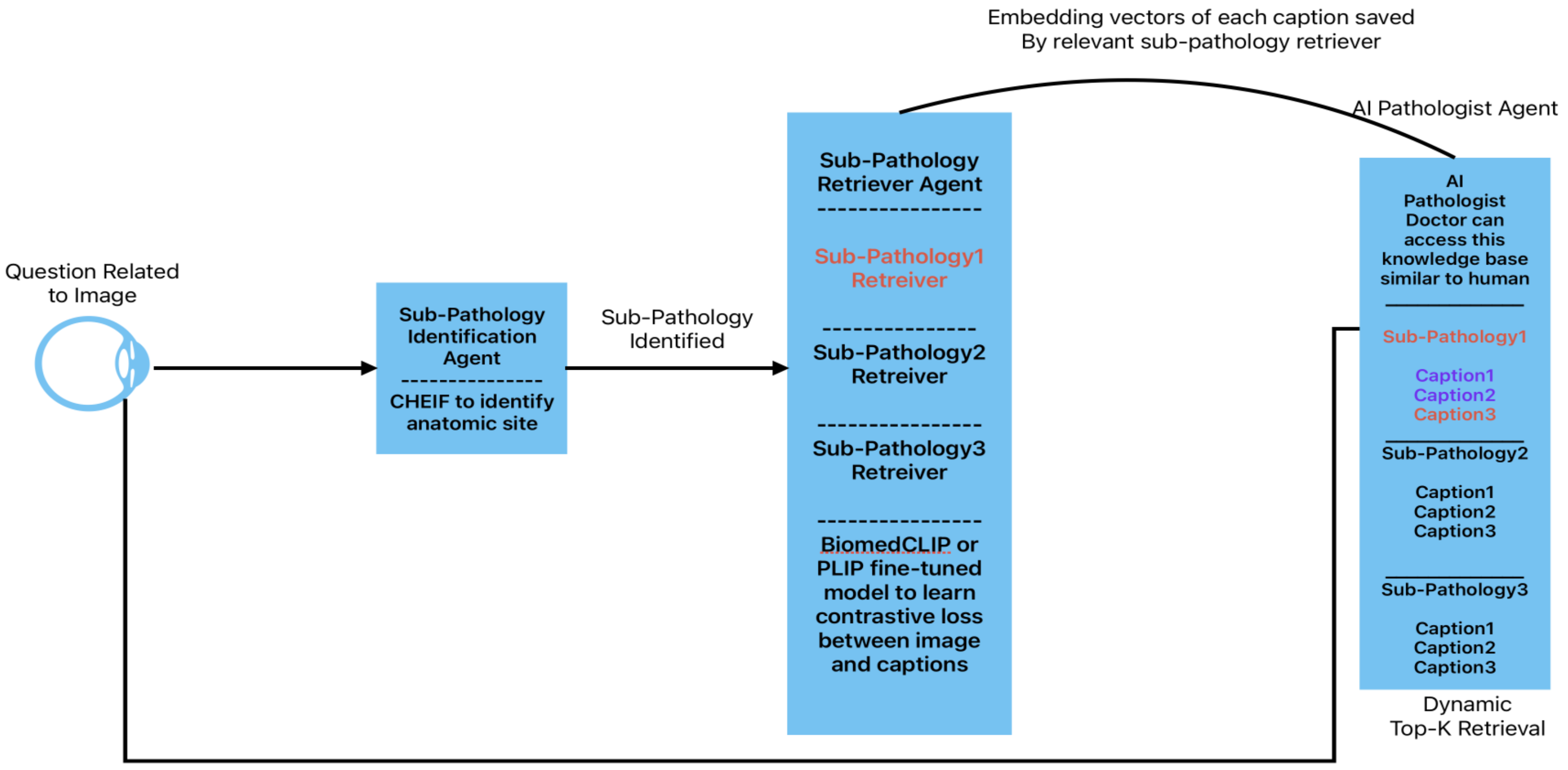
Sub-Pathology2
Caption1
Caption2
Caption3

Sub-Pathology3
Caption1
Caption2
Caption3

Dynamic Top-K Retrieval

QUILT-1M: https://quilt1m.github.io, CHIEF: https://www.nature.com/articles/s41586-024-07894-z

Knowledgebase Agent - Inference



Embedding vectors of each caption saved
By relevant sub-pathology retriever

AI Pathologist Agent

Question Related to Image

Sub-Pathology Identification Agent

CHEIF to identify anatomic site

Sub-Pathology Identified

Sub-Pathology Retriever Agent

Sub-Pathology1 Retreiver

Sub-Pathology2 Retreiver

Sub-Pathology3 Retreiver

BiomedCLIP or PLIP fine-tuned model to learn contrastive loss between image and captions

AI Pathologist Doctor can access this knowledge base similar to human

Sub-Pathology1

Caption1
Caption2
Caption3

Sub-Pathology2

Caption1
Caption2
Caption3

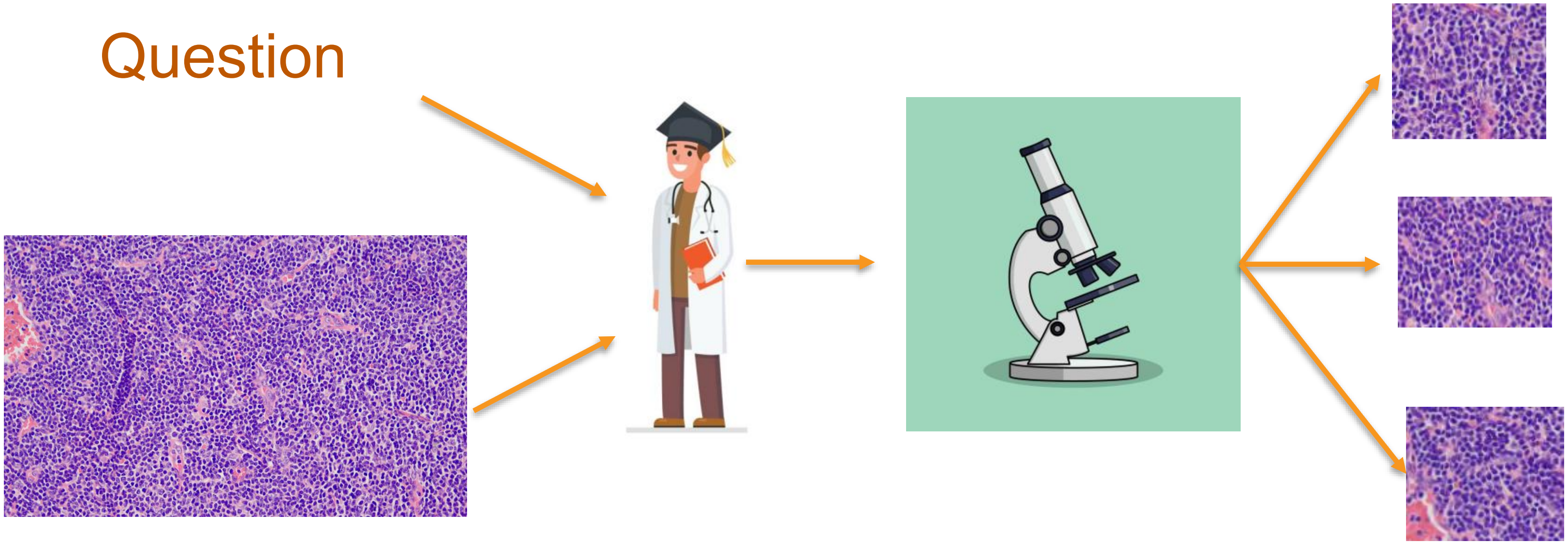
Sub-Pathology3

Caption1
Caption2
Caption3

Dynamic Top-K Retrieval

QUILT-1M: https://quilt1m.github.io, CHIEF: https://www.nature.com/articles/s41586-024-07894-z

Pathologist Focus on Magnification



Question

Magnifier Agent

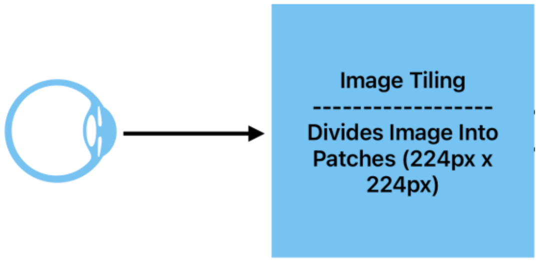


Image Tiling

Divides Image Into Patches (224px x 224px)

HistoCartography: https://arxiv.org/pdf/2107.10073, CHIEF: https://www.nature.com/articles/s41586-024-07894-z

Magnifier Agent

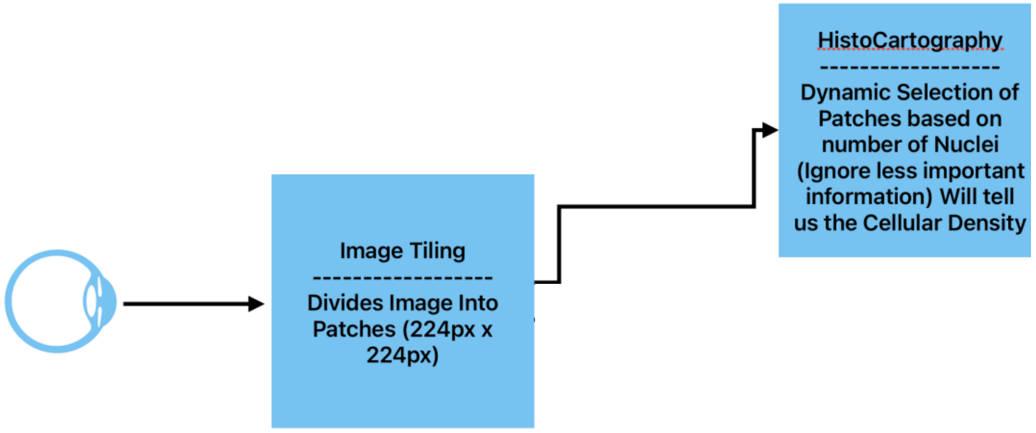


Image Tiling

Divides Image Into Patches (224px x 224px)

HistoCartography

Dynamic Selection of Patches based on number of Nuclei (Ignore less important information) Will tell us the Cellular Density

HistoCartography: https://arxiv.org/pdf/2107.10073, CHIEF: https://www.nature.com/articles/s41586-024-07894-z

Magnifier Agent

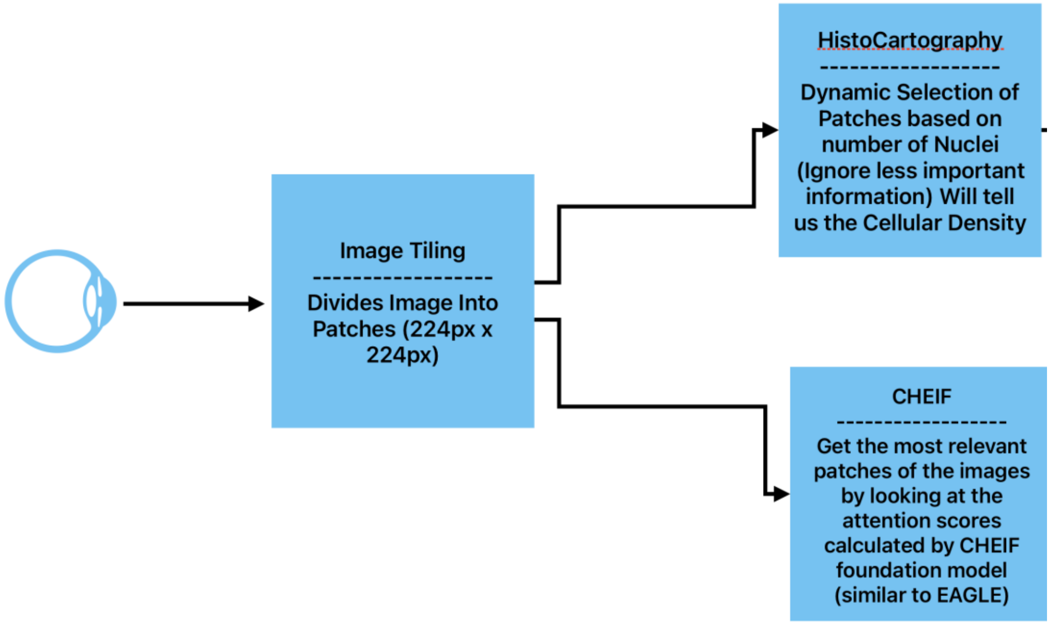


Image Tiling

Divides Image Into Patches (224px x 224px)

HistoCartography

Dynamic Selection of Patches based on number of Nuclei (Ignore less important information) Will tell us the Cellular Density

CHEIF

Get the most relevant patches of the images by looking at the attention scores calculated by CHEIF foundation model (similar to EAGLE)

HistoCartography: https://arxiv.org/pdf/2107.10073, CHIEF: https://www.nature.com/articles/s41586-024-07894-z

Magnifier Agent

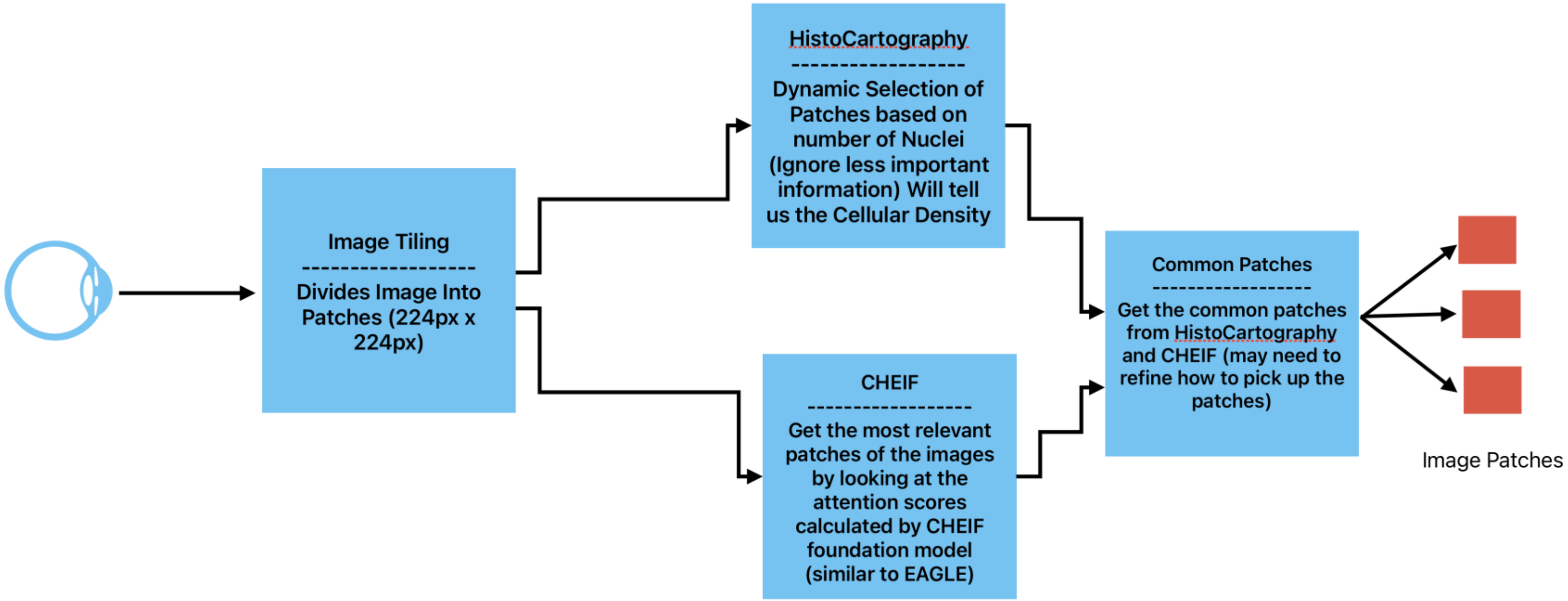


Image Tiling

Divides Image Into Patches (224px x 224px)

HistoCartography

Dynamic Selection of Patches based on number of Nuclei (Ignore less important information) Will tell us the Cellular Density

CHEIF

Get the most relevant patches of the images by looking at the attention scores calculated by CHEIF foundation model (similar to EAGLE)

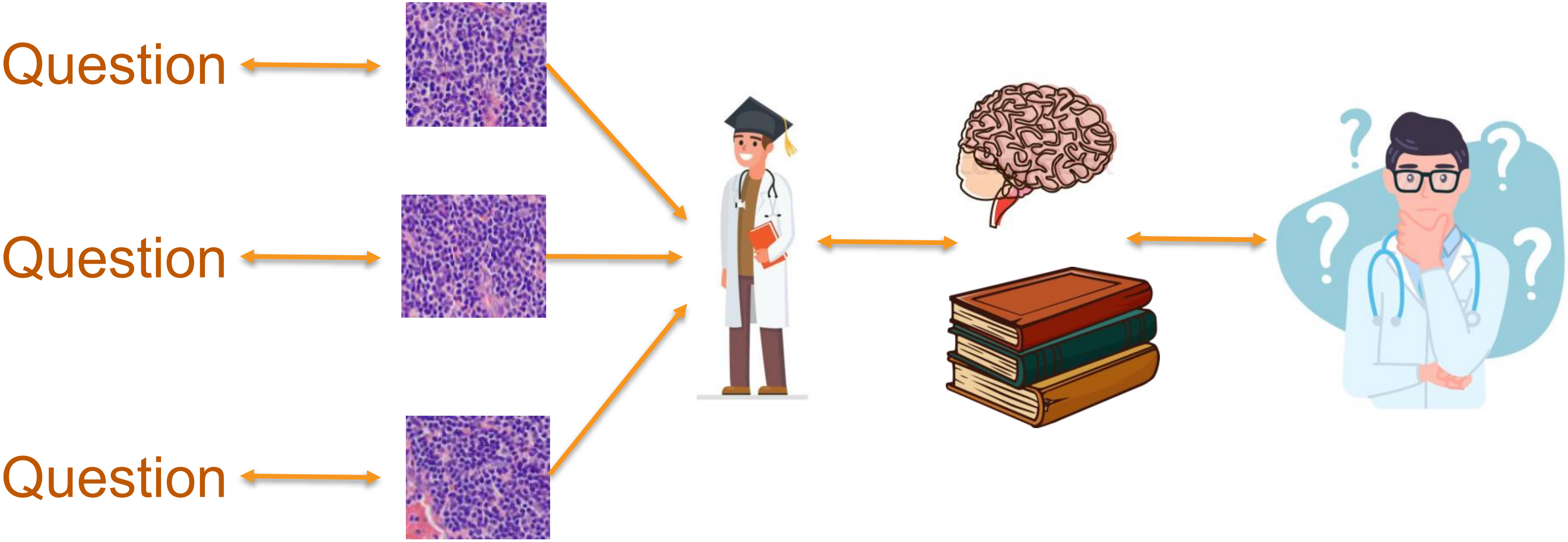
Common Patches

Get the common patches from HistoCartography and CHEIF (may need to refine how to pick up the patches)

Image Patches

HistoCartography: https://arxiv.org/pdf/2107.10073, CHIEF: https://www.nature.com/articles/s41586-024-07894-z

Pathologist Connecting Dots

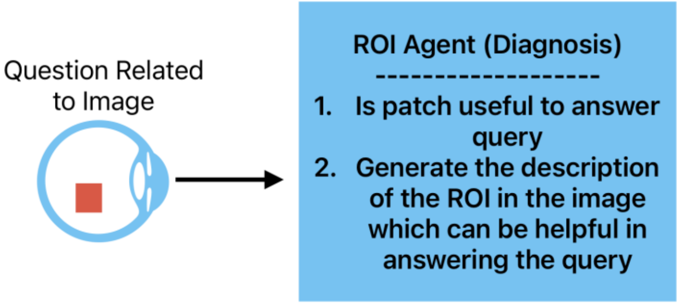


Question

Question

Question

ROI Agent – Each Selected Patch



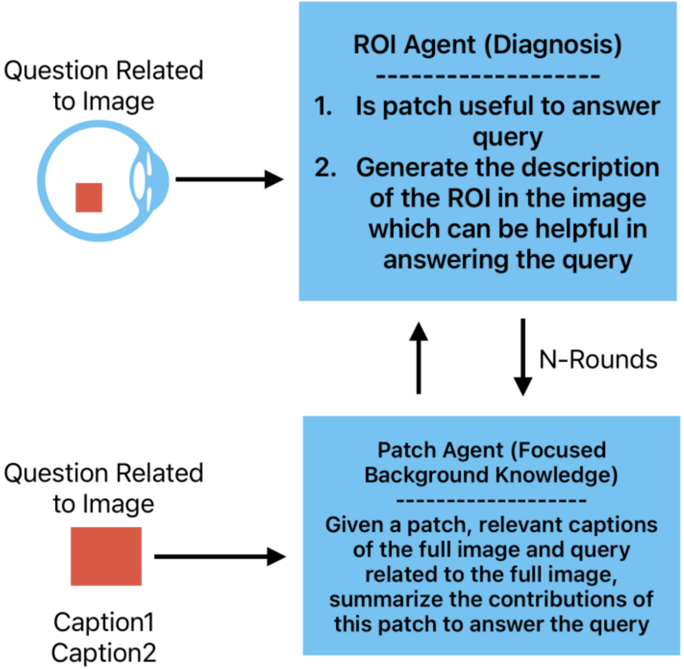
ROI Agent (Diagnosis)

----

1. Is patch useful to answer query
2. Generate the description of the ROI in the image which can be helpful in answering the query

LLaVA-MED: https://arxiv.org/abs/2306.00890, Qwen-VL:

Patch Agent – Each Selected Patch



Question Related to Image

ROI Agent (Diagnosis)
----
1. Is patch useful to answer query
2. Generate the description of the ROI in the image which can be helpful in answering the query

N-Rounds

Question Related to Image

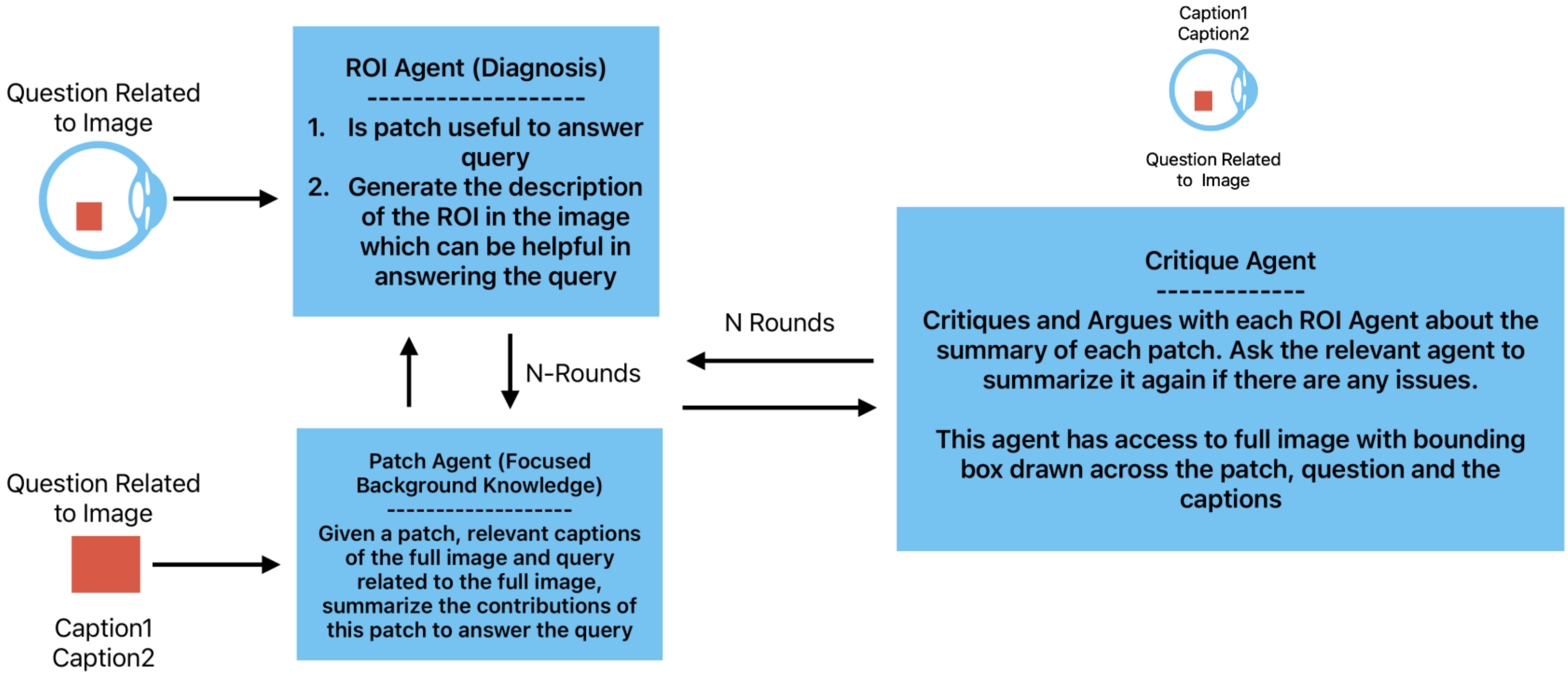
Caption1
Caption2

Patch Agent (Focused Background Knowledge)

Given a patch, relevant captions of the full image and query related to the full image, summarize the contributions of this patch to answer the query

LLaVA-MED: https://arxiv.org/abs/2306.00890, Qwen-VL: https://arxiv.org/abs/2308.12966

Critique Agent – Each Selected Patch



Caption1
Caption2

Question Related to Image

Question Related to Image

ROI Agent (Diagnosis)
----
1. Is patch useful to answer query
2. Generate the description of the ROI in the image which can be helpful in answering the query

N-Rounds

N Rounds

Critique Agent

Critiques and Argues with each ROI Agent about the summary of each patch. Ask the relevant agent to summarize it again if there are any issues.

This agent has access to full image with bounding box drawn across the patch, question and the captions

Question Related to Image

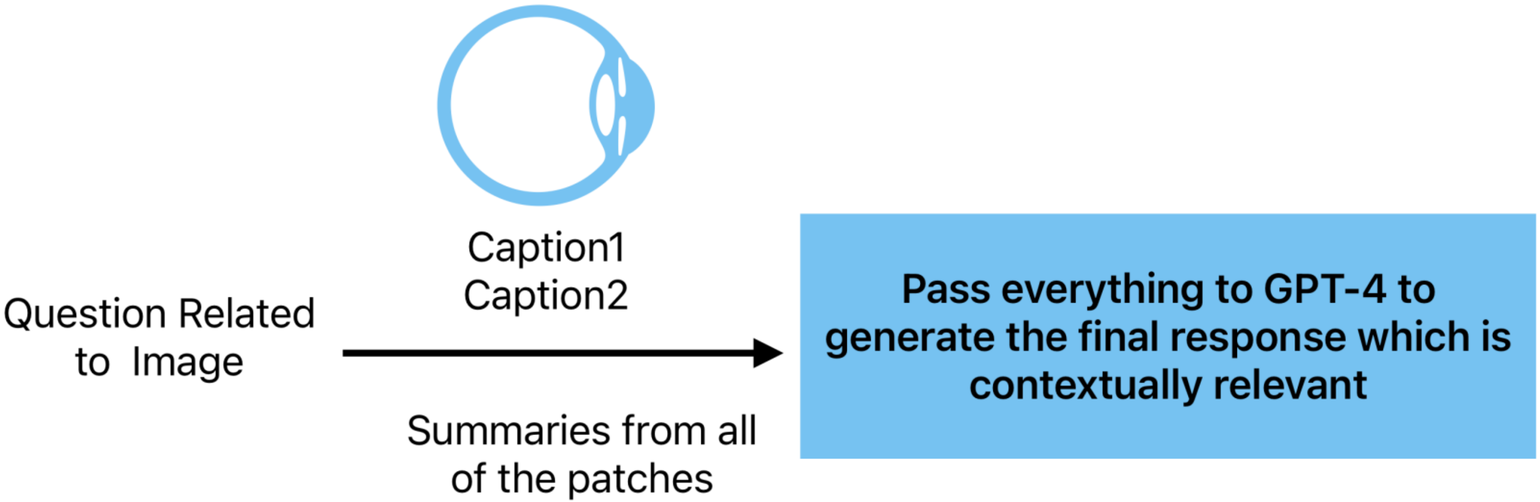
Patch Agent (Focused Background Knowledge)

Given a patch, relevant captions of the full image and query related to the full image, summarize the contributions of this patch to answer the query

Caption1
Caption2

LLaVA-MED: https://arxiv.org/abs/2306.00890, Qwen-VL: https://arxiv.org/abs/2308.12966

Final Response



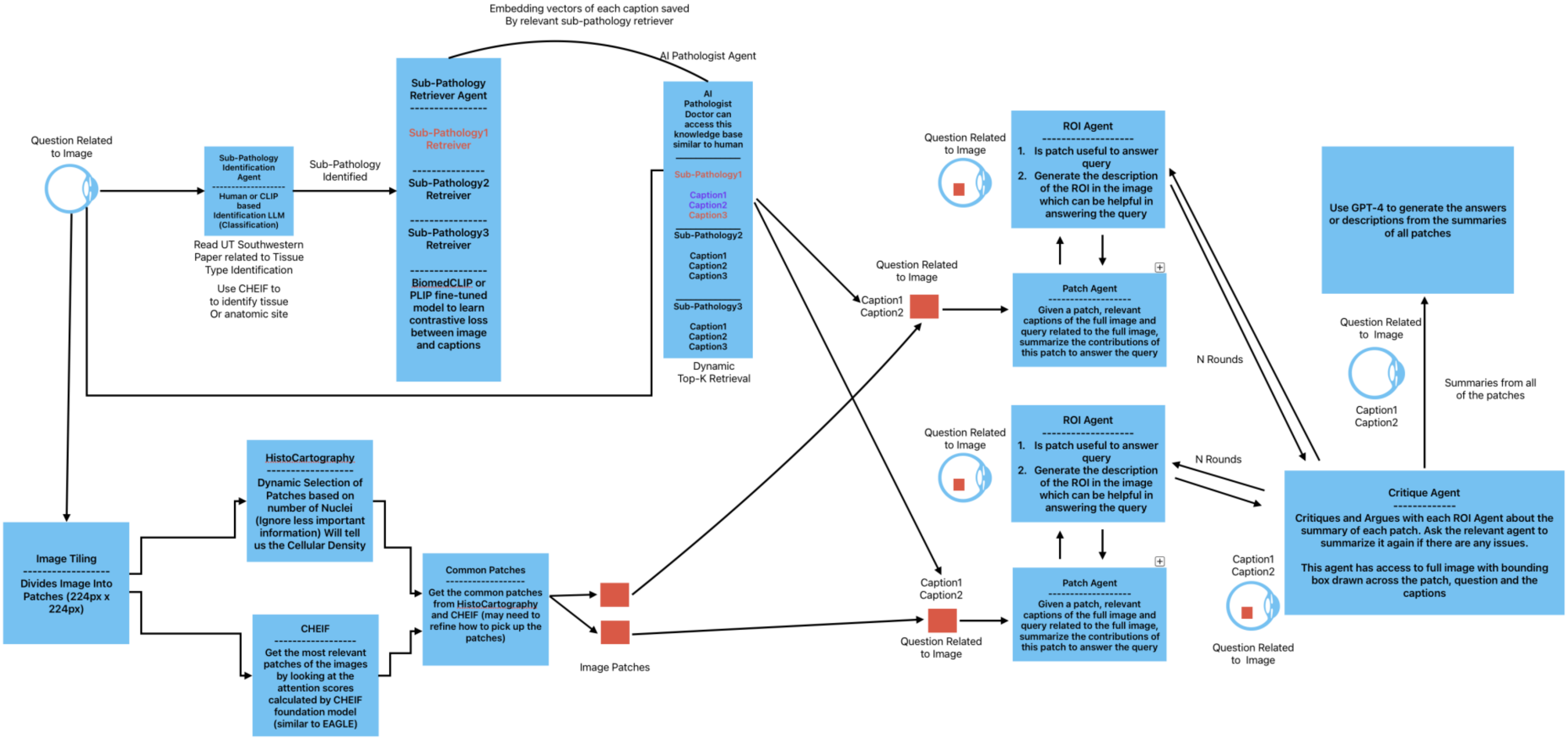
Question Related to Image

Caption1
Caption2

Summaries from all of the patches

Pass everything to GPT-4 to generate the final response which is contextually relevant

Path-Agent: Complete Architecture



Embedding vectors of each caption saved
By relevant sub-pathology retriever

Question Related to Image

Sub-Pathology Identification Agent

Human or CLIP based Identification LLM (Classification)

Read UT Southwestern Paper related to Tissue Type Identification

Use CHEIF to to identify tissue Or anatomic site

Sub-Pathology Identified

Sub-Pathology Retriever Agent

Sub-Pathology1 Retreiver

Sub-Pathology2 Retreiver

Sub-Pathology3 Retreiver

BiomedCLIP or PLIP fine-tuned model to learn contrastive loss between image and captions

AI Pathologist Agent

AI Pathologist Doctor can access this knowledge base similar to human

Sub-Pathology1
Caption1
Caption2
Caption3

Sub-Pathology2
Caption1
Caption2
Caption3

Sub-Pathology3
Caption1
Caption2
Caption3

Dynamic Top-K Retrieval

Image Tiling

Divides Image Into Patches (224px x 224px)

HistoCartography

Dynamic Selection of Patches based on number of Nuclei (Ignore less important information) Will tell us the Cellular Density

CHEIF

Get the most relevant patches of the images by looking at the attention scores calculated by CHEIF foundation model (similar to EAGLE)

Common Patches

Get the common patches from HistoCartography and CHEIF (may need to refine how to pick up the patches)

Image Patches

Question Related to Image
Caption1
Caption2

Patch Agent

Given a patch, relevant captions of the full image and query related to the full image, summarize the contributions of this patch to answer the query

ROI Agent
---
1. Is patch useful to answer query
2. Generate the description of the ROI in the image which can be helpful in answering the query

Question Related to Image

N Rounds

Caption1
Caption2
Question Related to Image

Patch Agent

Given a patch, relevant captions of the full image and query related to the full image, summarize the contributions of this patch to answer the query

ROI Agent
---
1. Is patch useful to answer query
2. Generate the description of the ROI in the image which can be helpful in answering the query

Question Related to Image

N Rounds

Critique Agent

Critiques and Argues with each ROI Agent about the summary of each patch. Ask the relevant agent to summarize it again if there are any issues.

This agent has access to full image with bounding box drawn across the patch, question and the captions

Caption1
Caption2
Question Related to Image

Question Related to Image
Caption1
Caption2

Summaries from all of the patches

Use GPT-4 to generate the answers or descriptions from the summaries of all patches

Initial Results

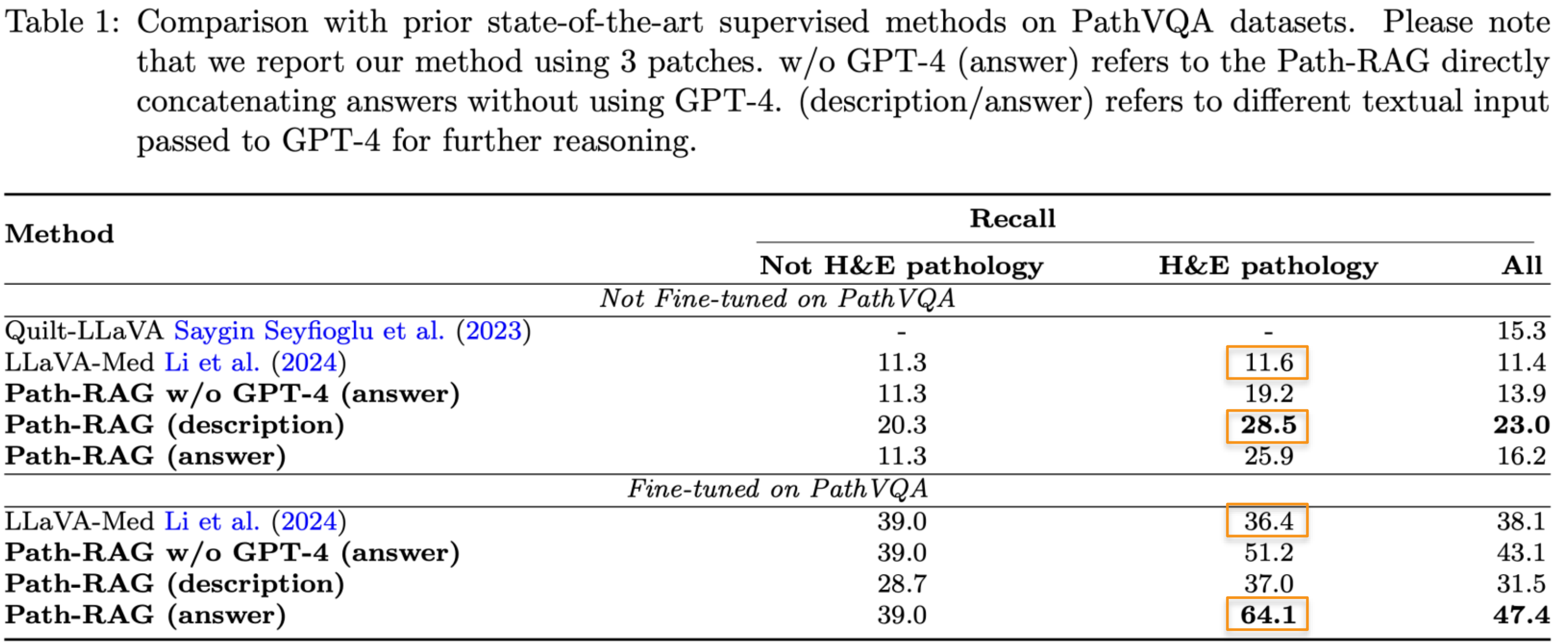


Table 1: Comparison with prior state-of-the-art supervised methods on PathVQA datasets. Please note that we report our method using 3 patches. w/o GPT-4 (answer) refers to the Path-RAG directly concatenating answers without using GPT-4. (description/answer) refers to different textual input passed to GPT-4 for further reasoning.

Method	**Recall**		
	Not H&E pathology	**H&E pathology**	**All**
Not Fine-tuned on PathVQA			
Quilt-LLaVA Saygin Seyfioglu et al. (2023)	-	-	15.3
LLaVA-Med Li et al. (2024)	11.3	11.6	11.4
Path-RAG w/o GPT-4 (answer)	11.3	19.2	13.9
Path-RAG (description)	20.3	**28.5**	**23.0**
Path-RAG (answer)	11.3	25.9	16.2
Fine-tuned on PathVQA			
LLaVA-Med Li et al. (2024)	39.0	36.4	38.1
Path-RAG w/o GPT-4 (answer)	39.0	51.2	43.1
Path-RAG (description)	28.7	37.0	31.5
Path-RAG (answer)	39.0	**64.1**	**47.4**

Path-RAG: https://arxiv.org/pdf/2411.17073

Thank You for Listening!

For Questions:

awais.naeem@utexas.edu

ying.ding@ischool.utexas.edu