Multimodal Learning for Medicine & Healthcare: The Challenges and Opportunities (IN2107)

Marta Hasny & Laura Daza
Chair for Computational Imaging and AI in Medicine
SS2026

Who are we?

Chair for Computational Imaging and AI in Medicine
Prof. Dr. Julia Schnabel

Who are we?

Dr. Laura Daza

- Medical image segmentation
- Foundation models
- Multimodal learning

laura.daza@helmholtz-munich.de

Marta Hasny

- Foundation models
- Multimodal learning
- Cardiology

marta.hasny@helmholtz-munich.de

1) Organization

Preliminary schedule

Week	Session
1	Introduction to the Seminar
2	How to read a paper and do poster presentations
3	Introduction to Multimodal Learning
4	Invited Talk
5	Student Presentations
6	Free
7	Student Presentations
8	Challenges of Multimodal Learning in Medicine
9	Student Presentations
10	Invited Talk
11	Student Presentations
12	Multimodal Medical Foundation Models
13	Student Presentations
14	**Poster Session**

Subject to change!

Deliverables & grading

Oral presentations
- 30 minutes presentation + 10 minutes questions
- Presentation date depends on the topic

Poster presentations
- Scientific poster
  - About the paper from the oral presentation + small review of the topic
- 5 minutes presentation

Paper selection
- We will provide a list of papers
- We will assign the papers based on your submitted selections
- Papers on both multimodal methodologies and medical applications

Grading
Oral + poster presentations + attendance + participation

Poster session at end of semester

Goals of the seminar

Theoretical skills

- Learn about state-of-the-art methods for multimodal learning
- Understand how multimodal learning can be used in clinical setting
- Understand practical challenges of deploying multimodal learning in clinical settings and learn strategies to overcome them

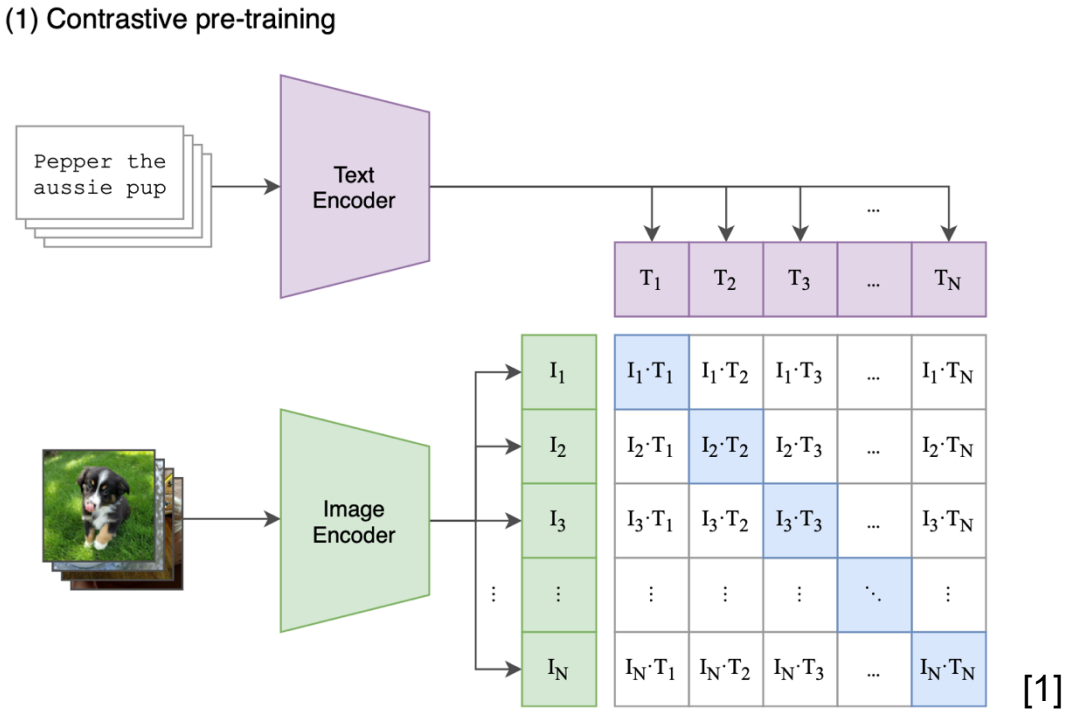
Research skills

- How to critically read, analyze, and present a research paper
- How to design and present a scientific poster

2) Multimodal learning
What will the seminar be about?

Multimodal Learning

Vision-Language Models



[1]

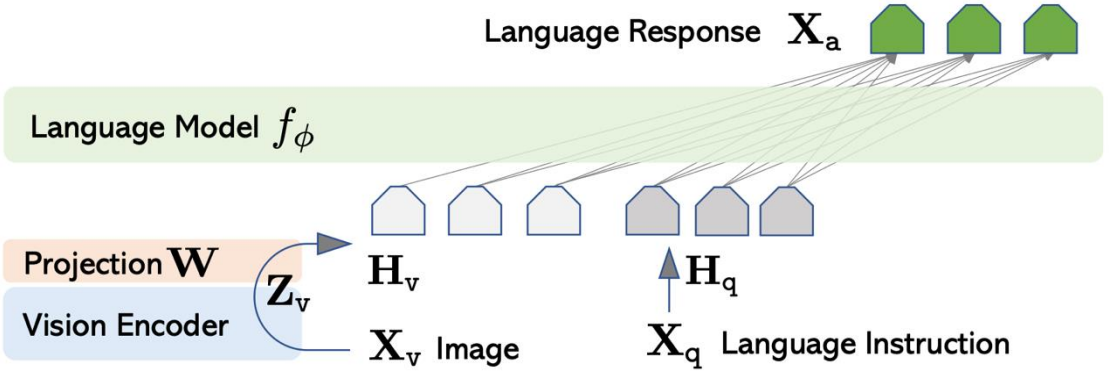
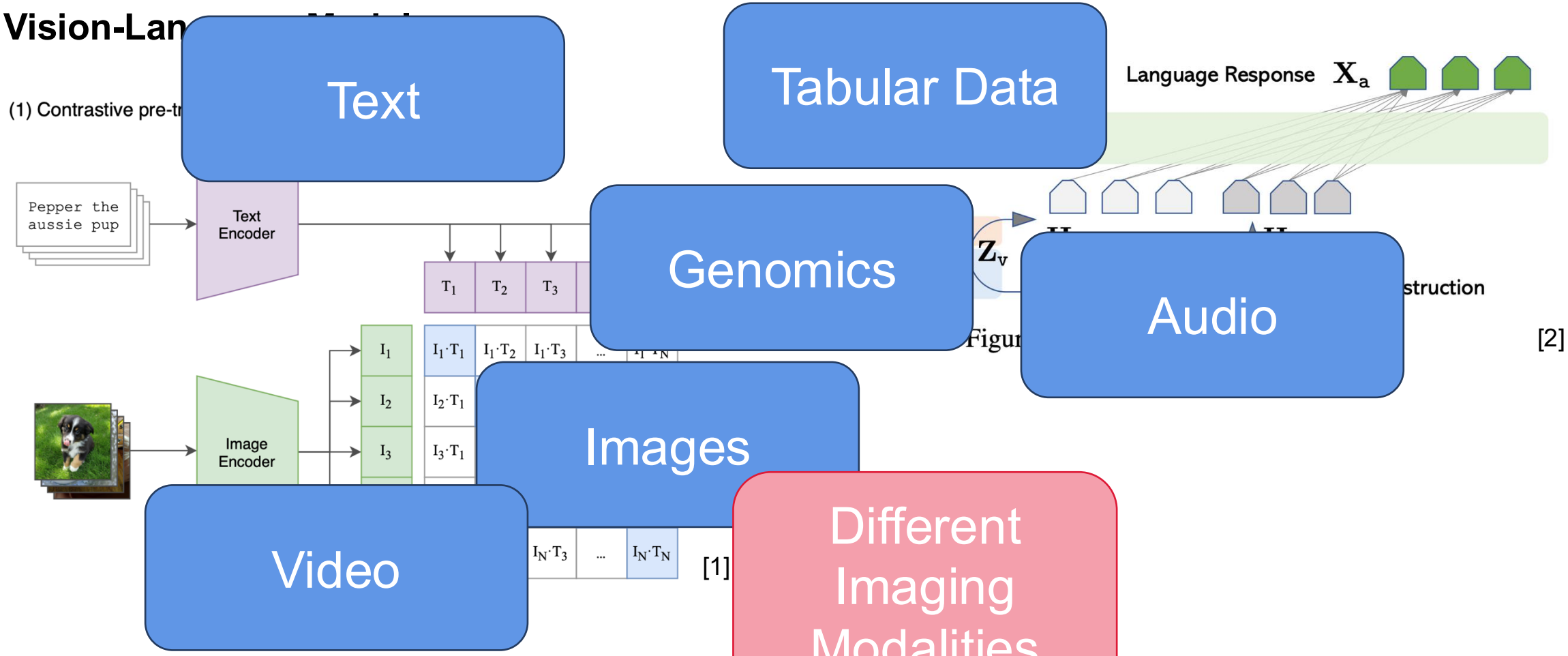


Figure 1: LLaVA network architecture. [2]

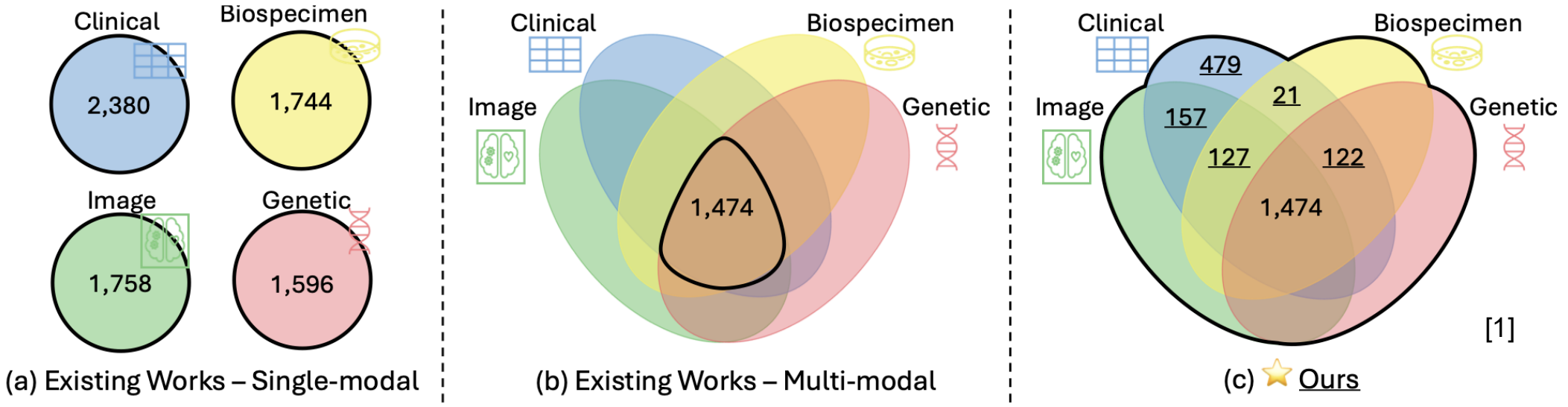
[1] Radford, Alec, et al. "Learning transferable visual models from natural language supervision." *International conference on machine learning*. PmLR, 2021.
[2] Liu, Haotian, et al. "Visual instruction tuning." *Advances in neural information processing systems* 36 (2023): 34892-34916.

Multimodal Learning



[1] Radford, Alec, et al. "Learning transferable visual models from natural language supervision." *International con*
[2] Liu, Haotian, et al. "Visual instruction tuning." *Advances in neural information processing systems* 36 (2023): 34892

Multimodal Learning



Multimodal methods often assume that all modalities are available

But in clinical practice that's not the case…

[1] Yun, Sukwon, et al. "Flex-moe: Modeling arbitrary modality combination via the flexible mixture-of-experts." Advances in Neural Information Processing Systems 37 (2024): 98782-98805.

In this seminar…

- Methods for multimodal learning
- Multimodal learning with more than two modalities
- Multimodal learning with missing modalities
- Multimodal foundation models
- And more… ☺

If you are interested, make sure to register through the matching system!

Multimodal Learning for Medicine & Healthcare: The Challenges and Opportunities (IN2107)

Contact: laura.daza@helmholtz-munich.de, marta.hasny@helmholtz.munich.de