UNIVERSITY OF HELSINKI
FACULTY OF MEDICINE

LANISTR: Multimodal Learning from Structured and Unstructured Data

Huaiwu ZHANG, PhD

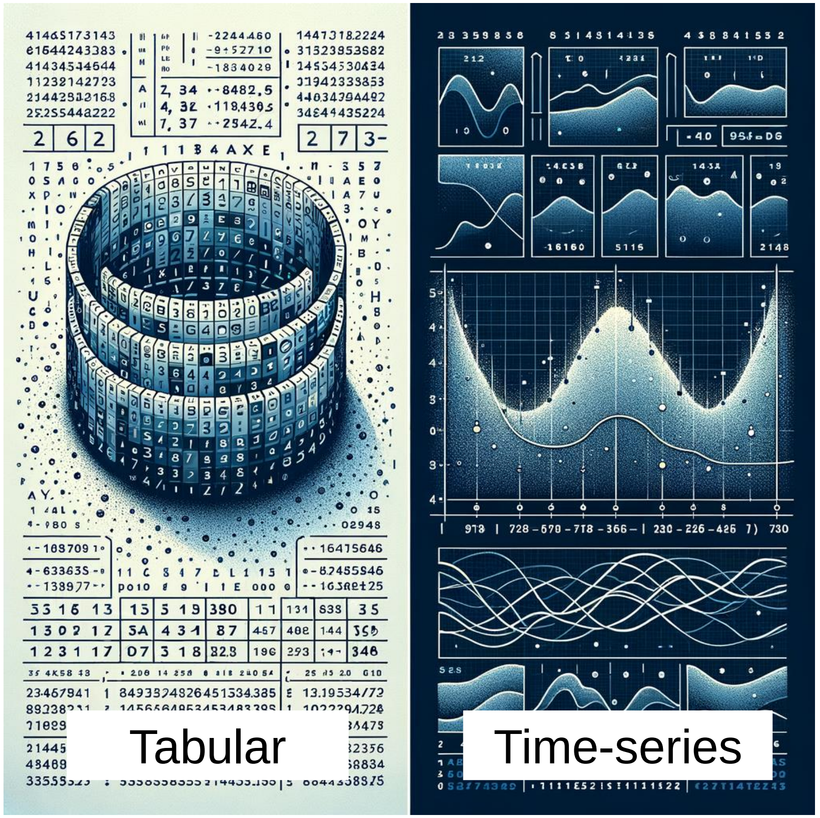
Network pharmacology for precision medicine
Faculty of Medicine, University of Helsinki

Content

- What is structured and unstructured data.
- Motivation of integrating structured and unstructured data.
- Challenges of integrating structured and unstructured data.
- Motivation of **LANISTR**.
- How **LANISTR** overcome these challenges.
- Performance of **LANISTR**.

What is structured and unstructured data.

Unstructured data

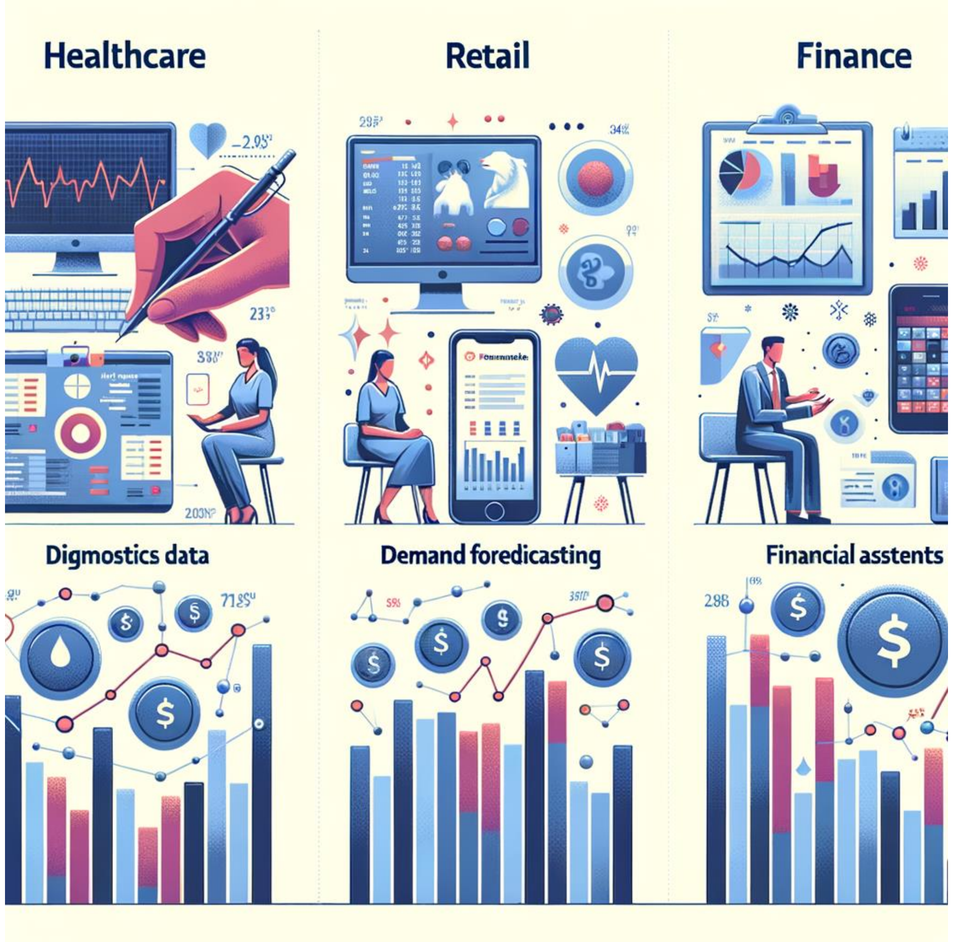


Structured data

Figures are generated by DALLE-3.

Motivation of integrate structured and unstructured data.

UNIVERSITY OF HELSINKI
FACULTY OF MEDICINE



There are more and more unstructured data in our life…

Example:

- healthcare diagnosis prediction
- financial asset price prediction

Figures are generated by DALLE-3.

Challenges of integrate structured and unstructured data.

UNIVERSITY OF HELSINKI
FACULTY OF MEDICINE

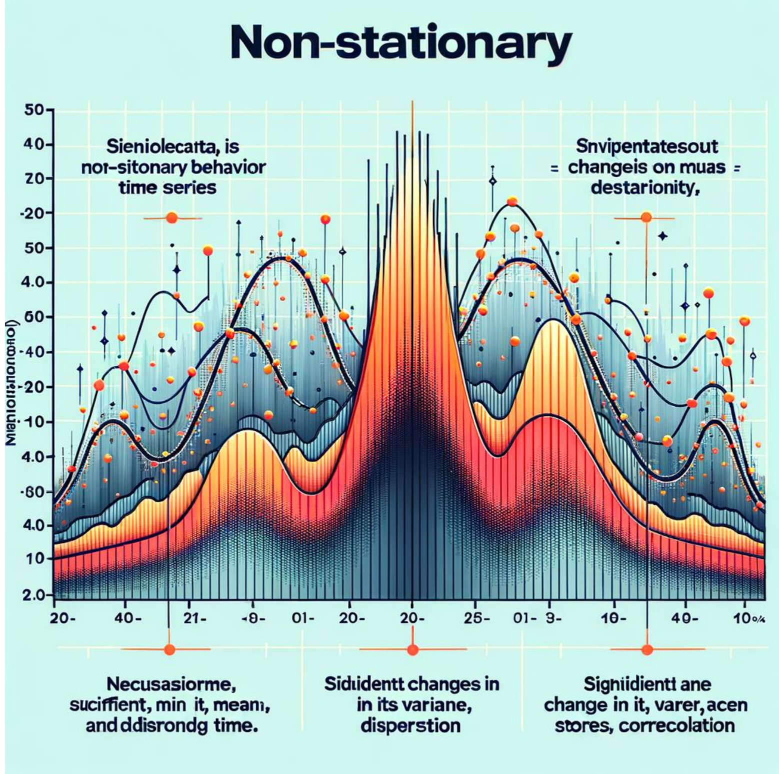
Two main challenges

- Deep neural networks can become susceptible to **overfitting** and suboptimal generalization.
- Modality **missingness** becomes a more prominent issue when dealing with multimodal data beyond two modalities.

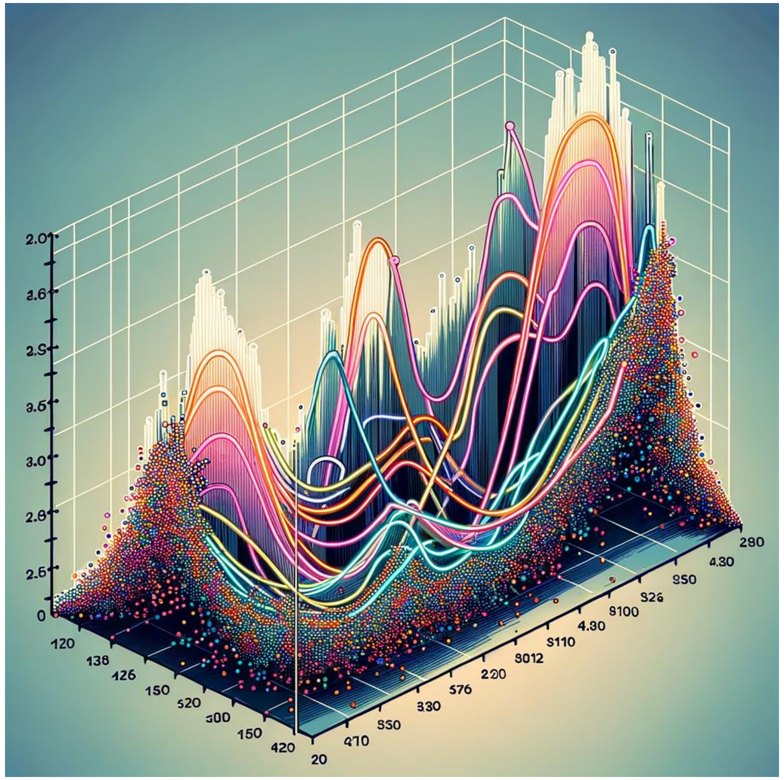
Challenges of integrate structured and unstructured data.

UNIVERSITY OF HELSINKI
FACULTY OF MEDICINE

- Deep neural networks can become susceptible to overfitting and suboptimal generalization.



Time-series



Tabular

Figures are generated by DALLE-3.

Challenges of integrate structured and unstructured data.

- Modality missingness becomes a more prominent issue when dealing with multimodal data beyond two modalities.

Sample	Image	Language	…
1	✓	✓	
2	✓	✗	
3	✗	✓	
…			

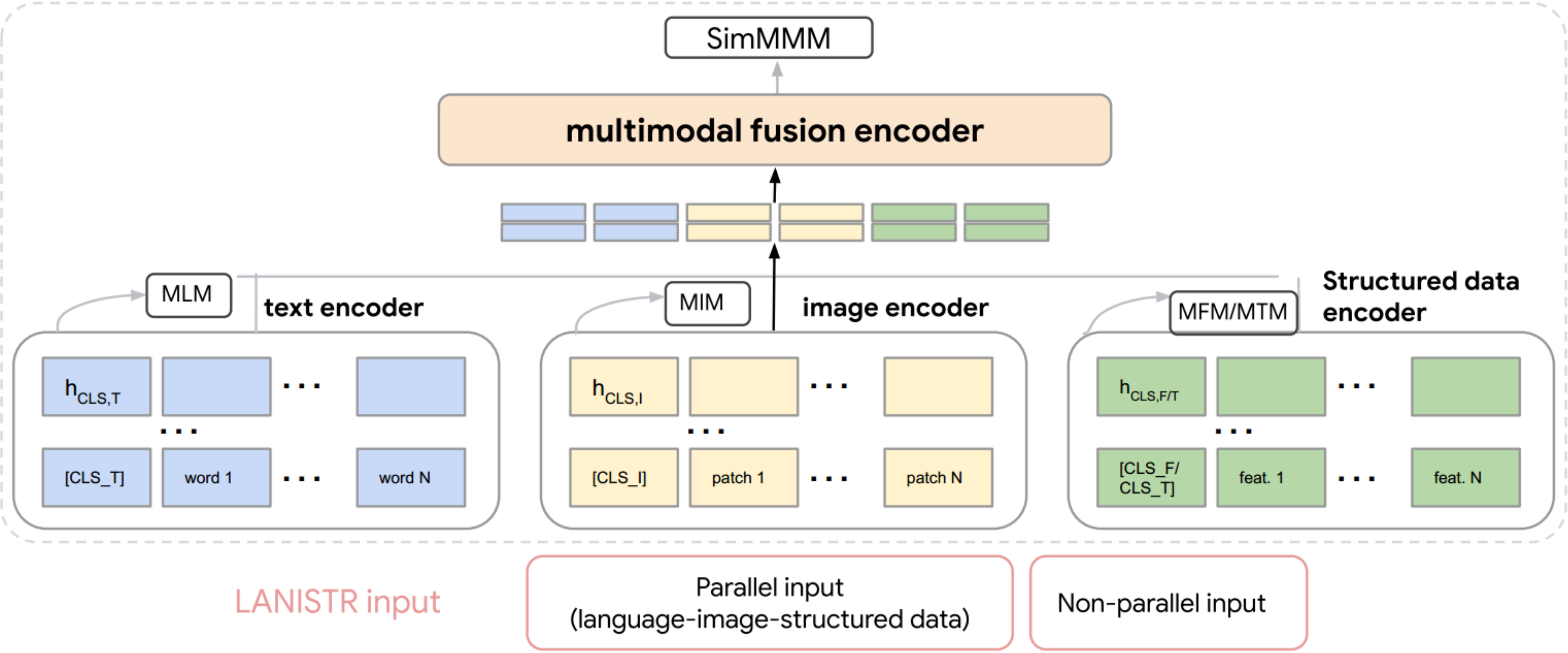
Modality missingness

Motivation of LANISTR

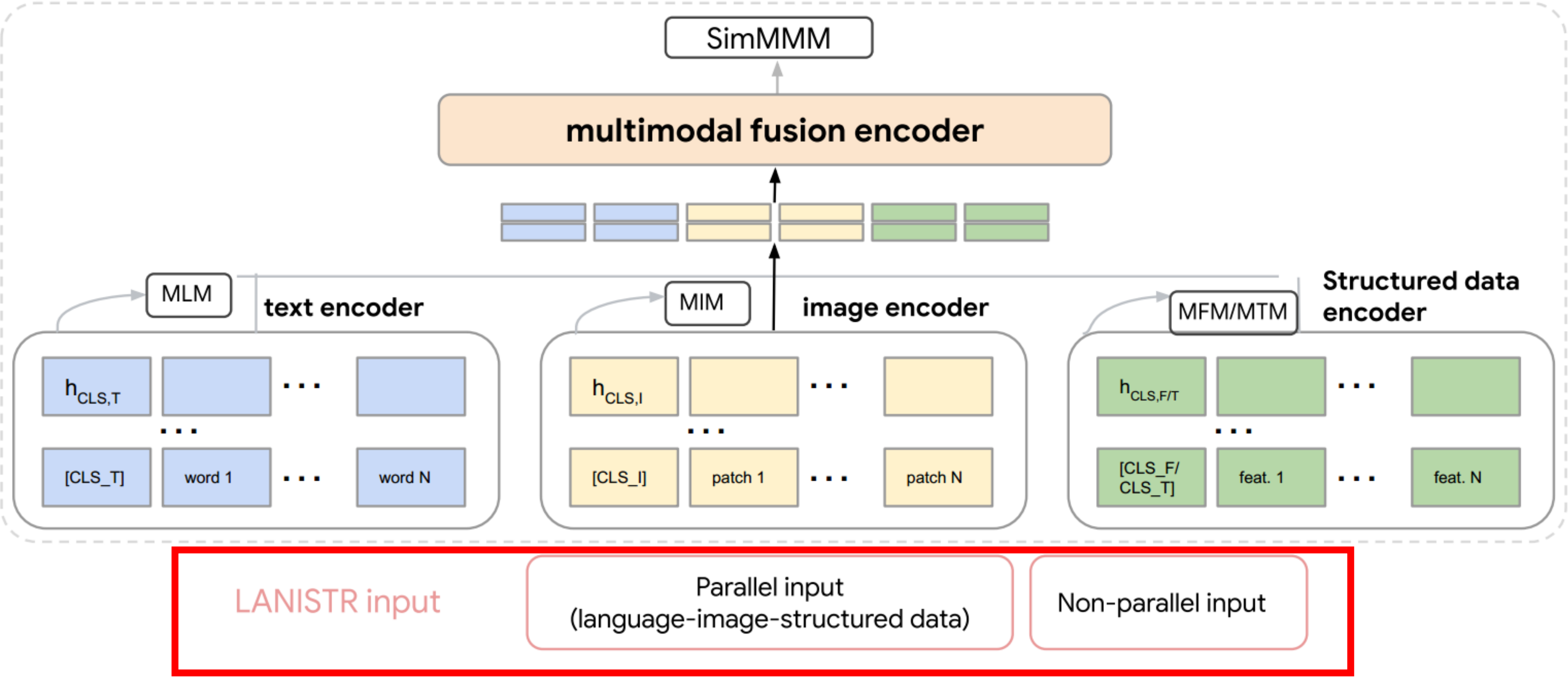
UNIVERSITY OF HELSINKI
FACULTY OF MEDICINE

- Empower the overall representation when we learn structured and unstructured data together.
- Design a unified architecture and unique pretraining strategies for two seemingly very different data types.

Structure of LANISTR

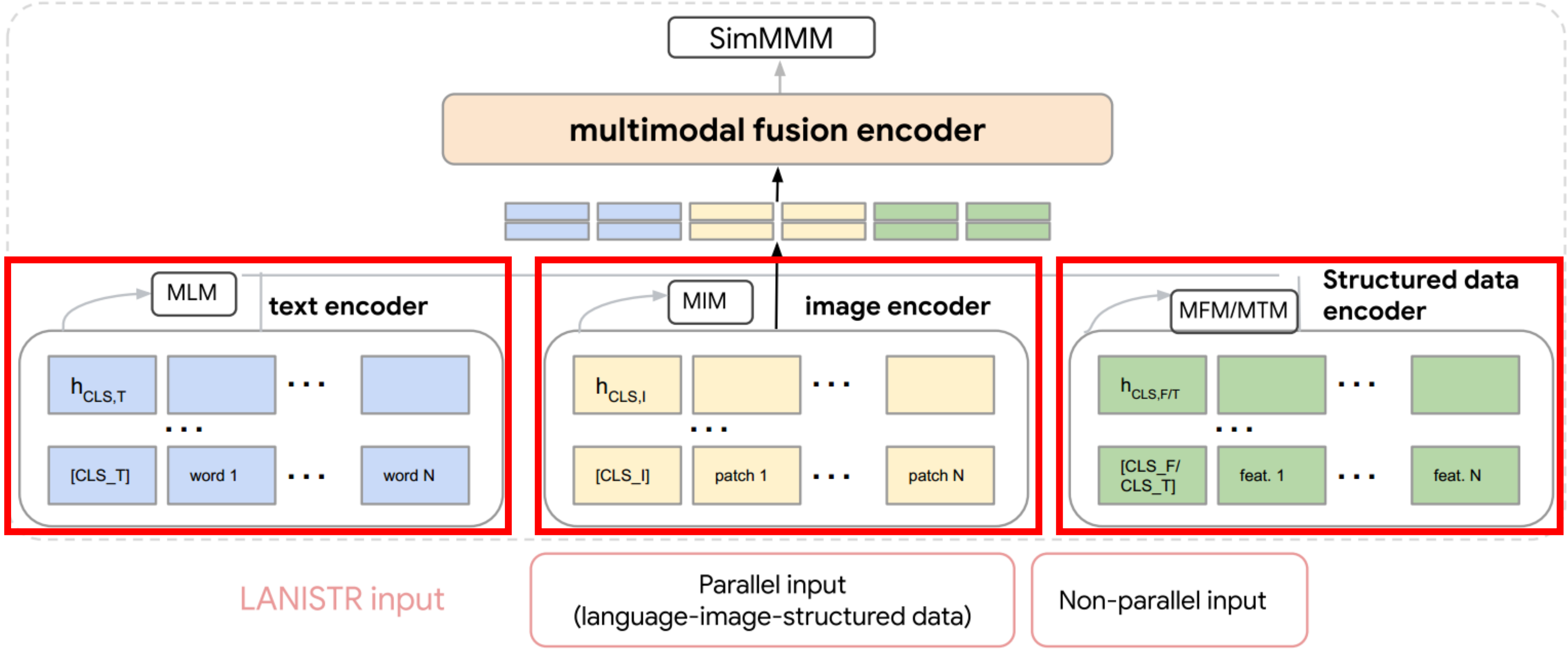


Structure of LANISTR

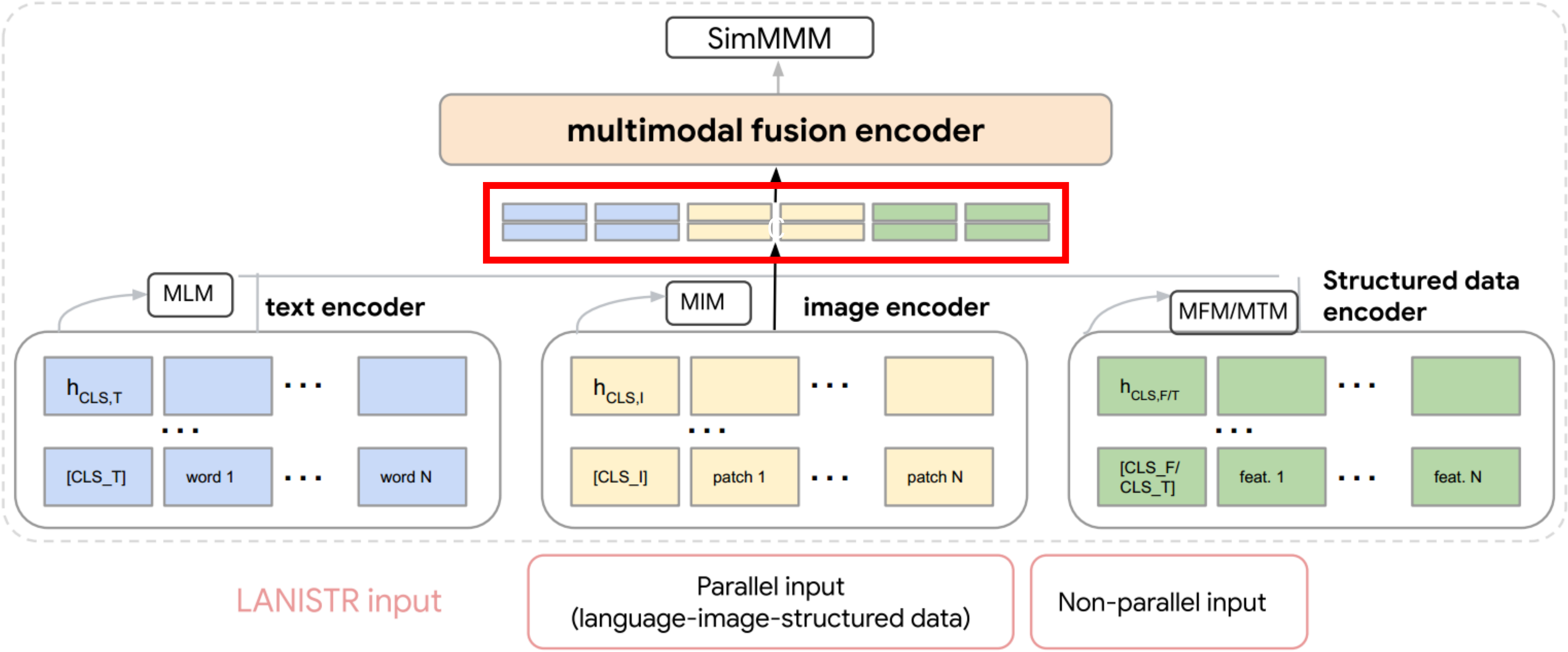


Structure of LANISTR

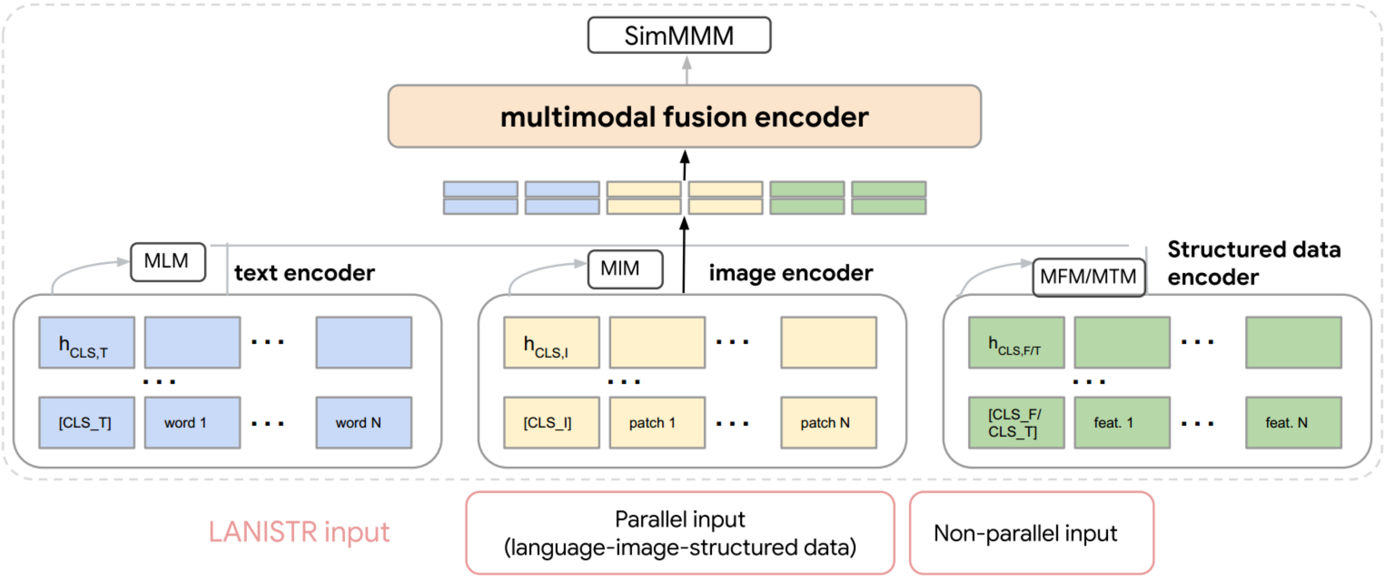
UNIVERSITY OF HELSINKI
FACULTY OF MEDICINE



Structure of LANISTR



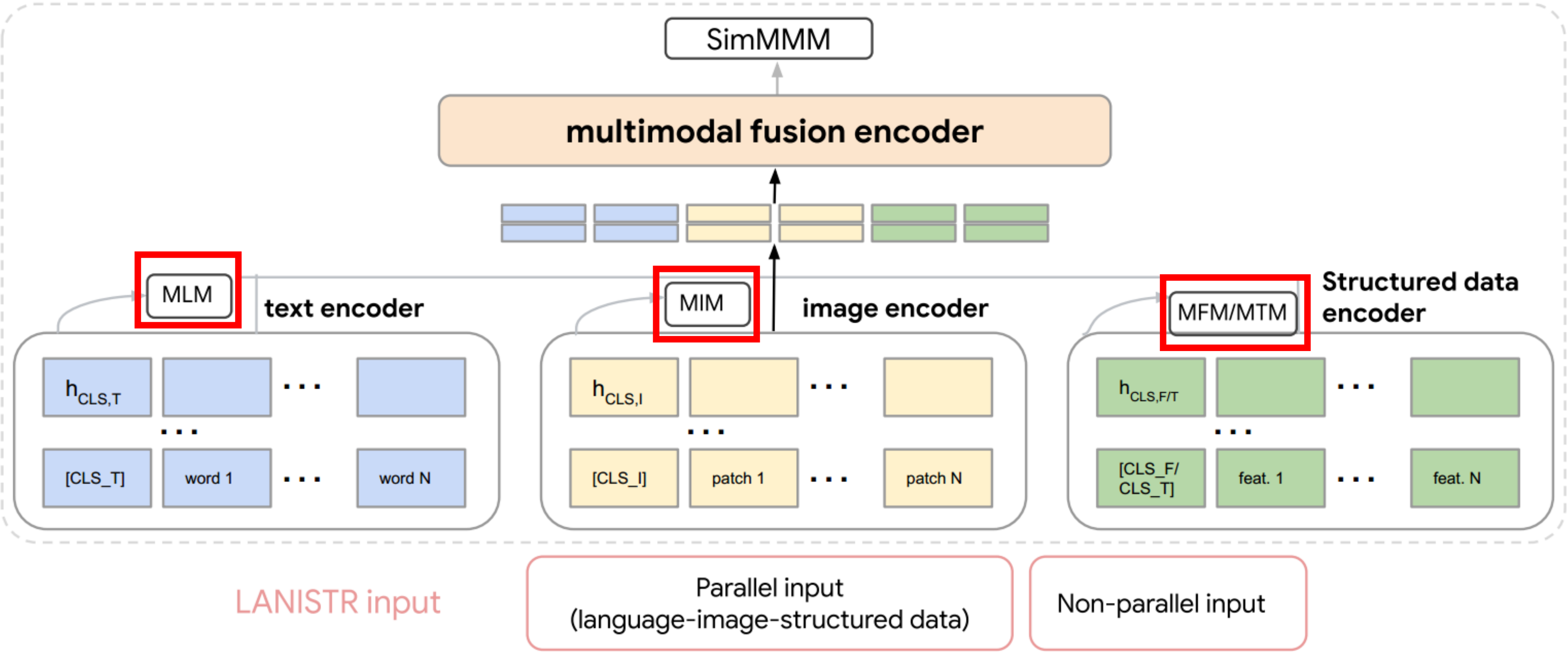
Pre-training of LANISTR



- *Unimodal* masking losses
- Similarity-based *multimodal* masking loss

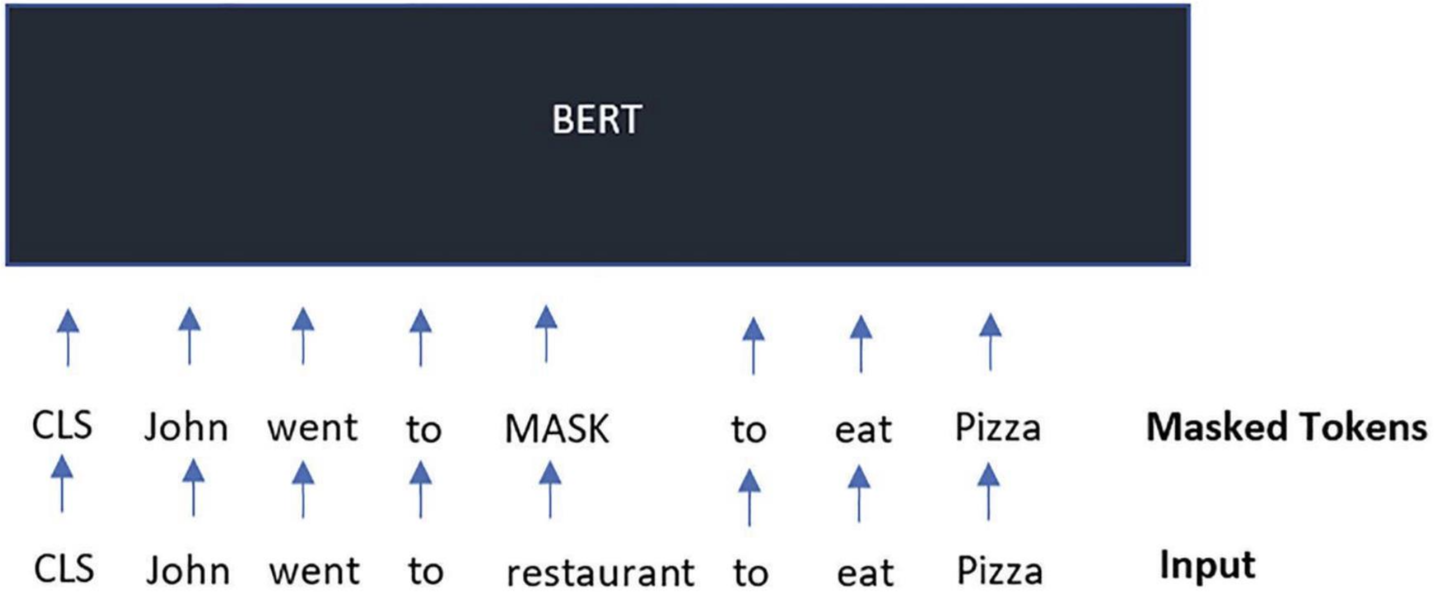
Pre-training of LANISTR

- Unimodal masking losses



Pre-training of LANISTR

- Unimodal masking losses

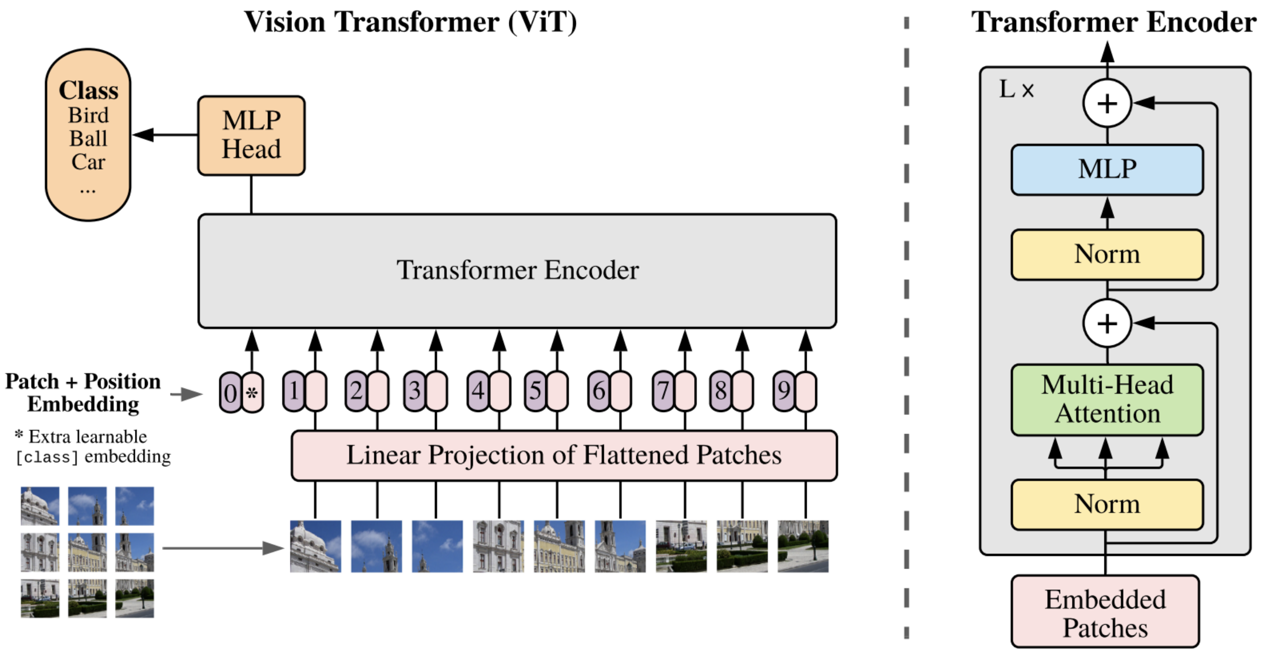


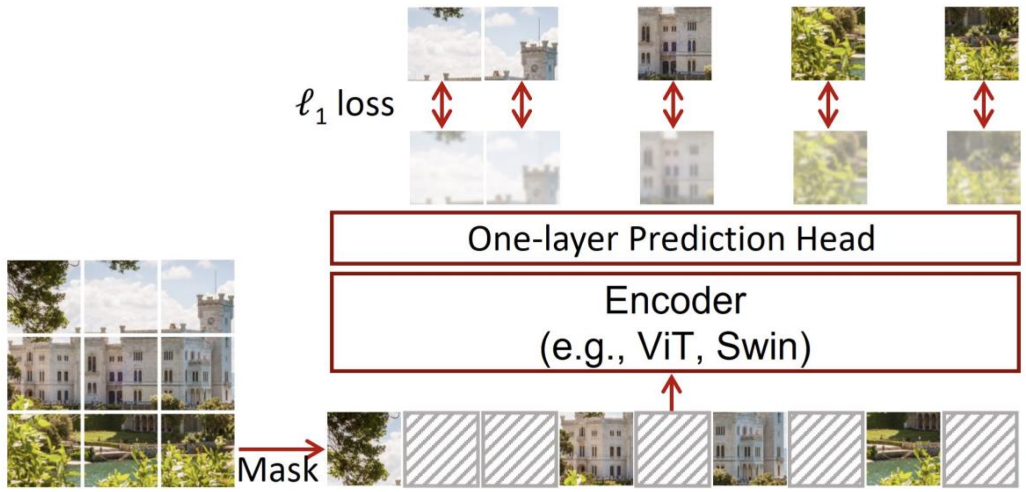
MLM: Masked Language Modeling

Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)

Pre-training of LANISTR

- Unimodal masking losses





MIM: Masked Image Modeling

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (2021)

Xie, Z., Zhang, Z., Cao, Y., Lin, Y., Bao, J., Yao, Z., Dai, Q., Hu, H.: Simmim: A simple framework for masked image modeling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9653–9663 (2022)

Pre-training of LANISTR

UNIVERSITY OF HELSINKI
FACULTY OF MEDICINE

- Unimodal masking losses

Unsupervised pre-training

Age	Cap. gain	Education	Occupation	Gender	Relationship
53	200000	?	Exec-managerial	F	Wife
19	0	?	Farming-fishing	M	?
?	5000	Doctorate	Prof-specialty	M	Husband
25	?	?	Handlers-cleaners	F	Wife
59	300000	Bachelors	?	?	Husband
33	0	Bachelors	?	F	?
?	0	High-school	Armed-Forces	?	Husband

TabNet encoder

TabNet decoder

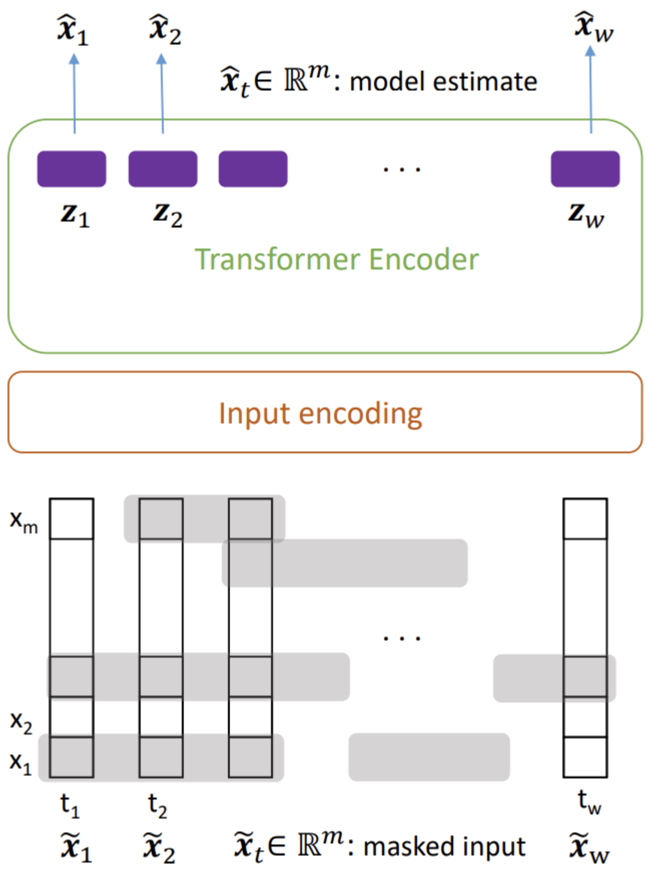
Age	Cap. gain	Education	Occupation	Gender	Relationship
		Masters			
		High-school			Unmarried
43					
	0	High-school		F	
			Exec-managerial	M	
			Adm-clerical		Wife
39				M	

MFM: Masked Feature Modeling

Arik, S.Ö., Pfister, T.: Tabnet: Attentive interpretable tabular learning. In: Proceedings of the AAAI conference on artificial intelligence. vol. 35, pp. 6679–6687 (2021)

Pre-training of LANISTR

- Unimodal masking losses



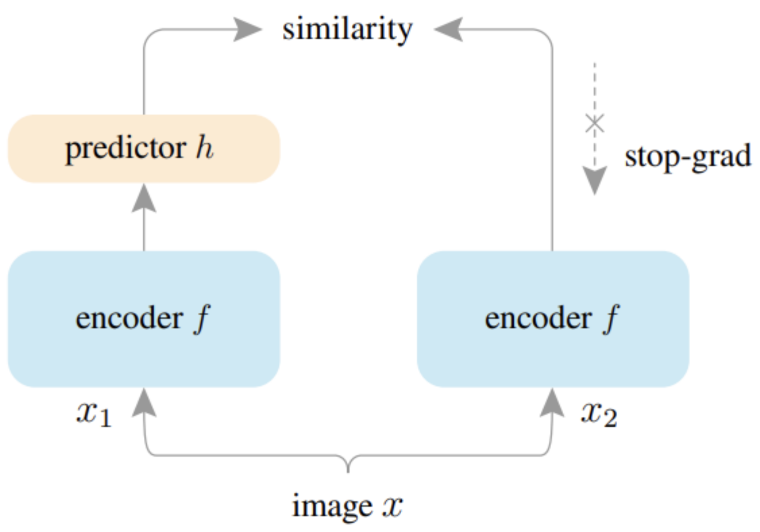
MTM: Masked Time-series Modeling

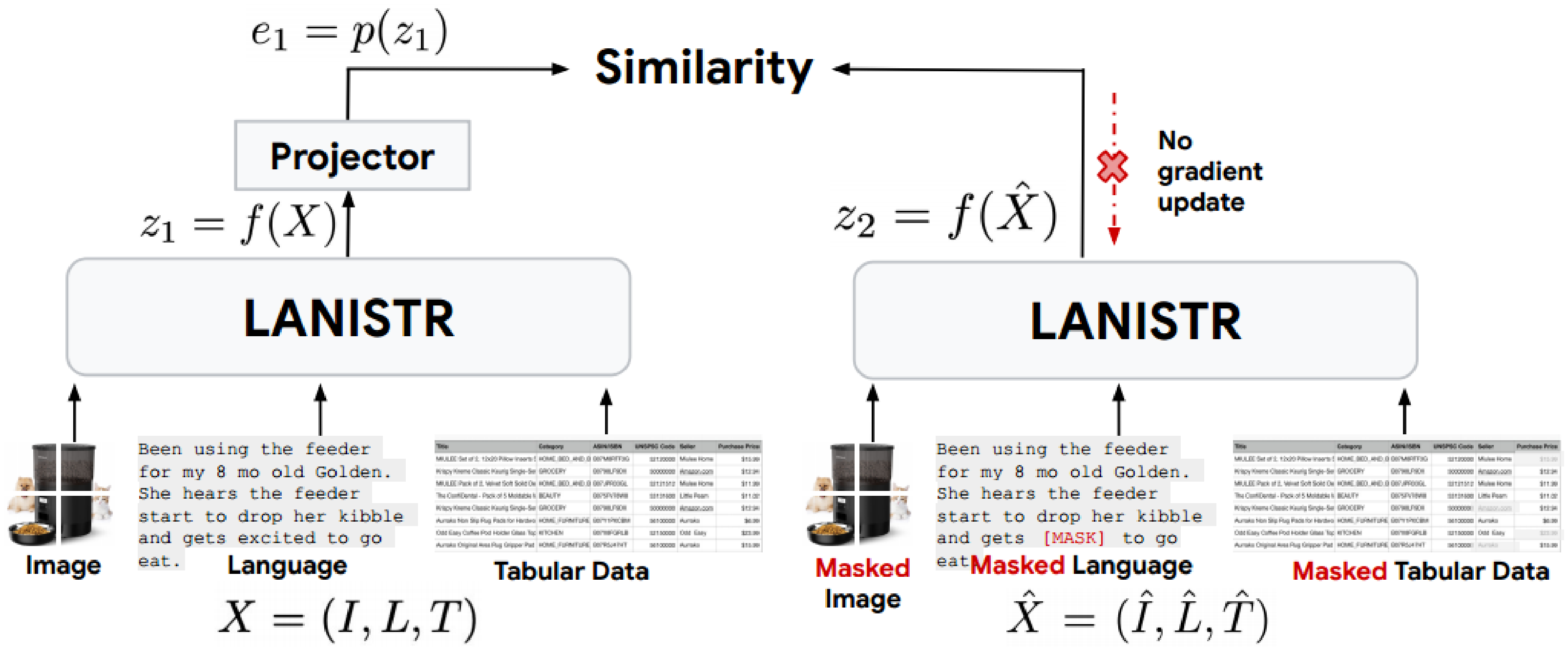
Zerveas, G., Jayaraman, S., Patel, D., Bhamidipaty, A., Eickhoff, C.: A transformerbased framework for multivariate time series representation learning. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. pp. 2114–2124 (2021)

Pre-training of LANISTR

UNIVERSITY OF HELSINKI
FACULTY OF MEDICINE

- Similarity-based multimodal masking loss

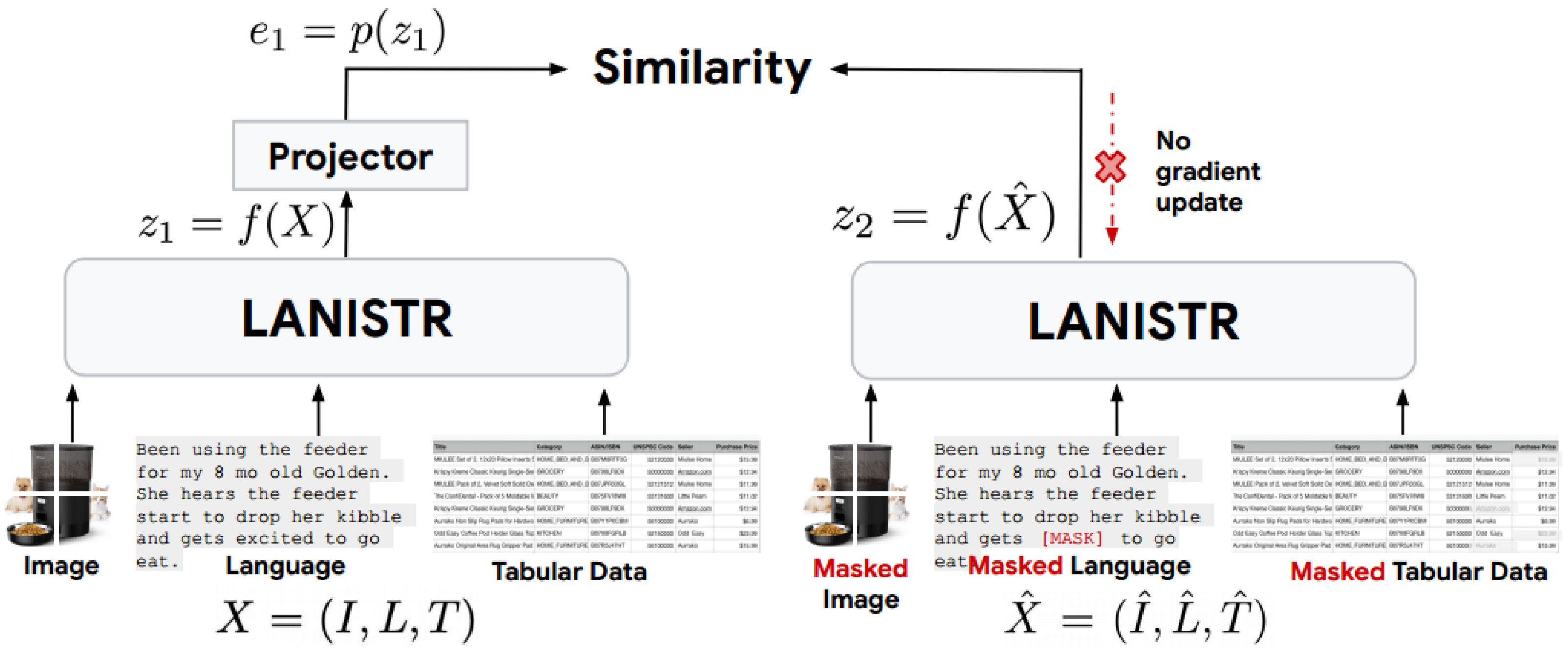




Chen, X., He, K.: Exploring simple siamese representation learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 15750–15758 (2021)

Pre-training of LANISTR

- Similarity-based multimodal masking loss



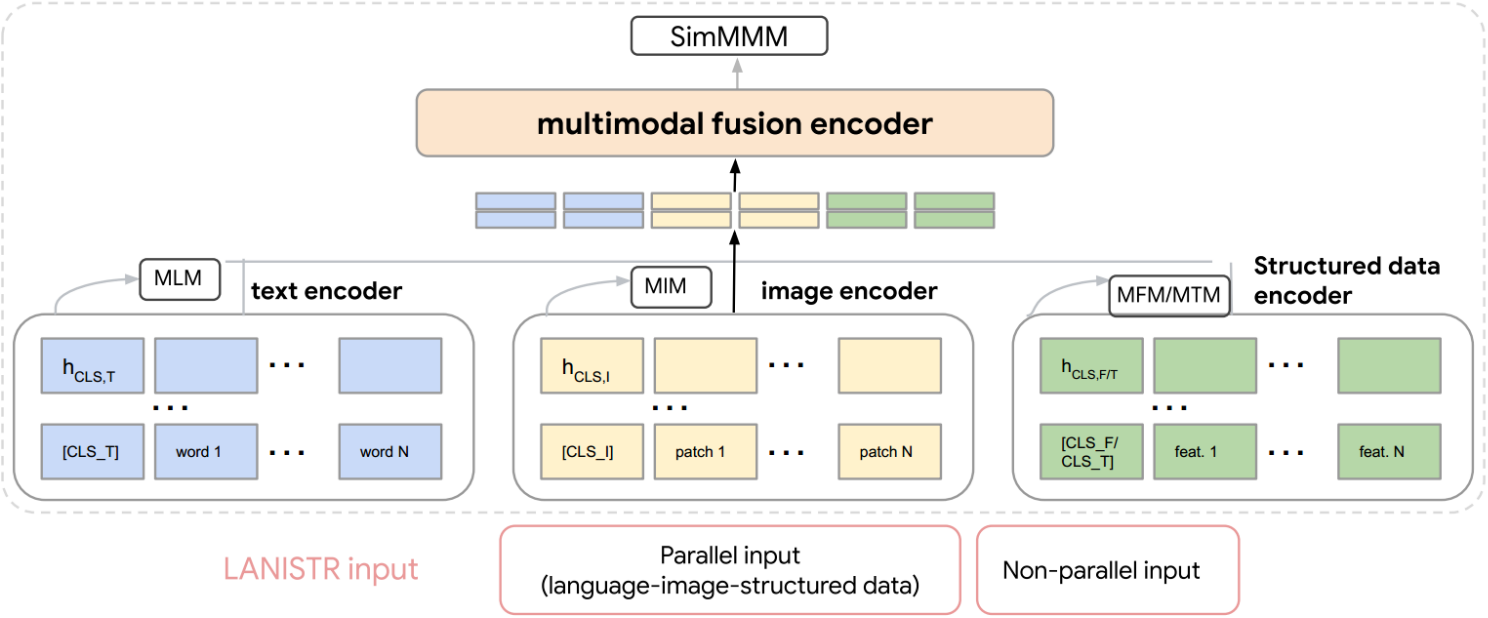
$$\mathcal{D}(e_1, z_2) = -\frac{e_1}{||e_1||_2} \cdot \frac{z_2}{||z_2||_2},$$

$$\mathcal{L}_{\text{SimMMM}} = \mathcal{D}(e_1, z_2) + \mathcal{D}(e_2, z_1).$$

Pre-training of LANISTR

UNIVERSITY OF HELSINKI
FACULTY OF MEDICINE

- Loss function



- Unimodal masking losses

$$\mathcal{L}_{\text{LANISTR}} = \boxed{\lambda_1 \mathcal{L}_{\text{MLM}} + \lambda_2 \mathcal{L}_{\text{MIM}} + \lambda_3 \mathcal{L}_{\text{MFM}} + \lambda_4 \mathcal{L}_{\text{MTM}}} + \boxed{\lambda_5 \mathcal{L}_{\text{SimMMM}}}$$

- Similarity-based multimodal masking loss

Fine-tuning of LANISTR

UNIVERSITY OF HELSINKI
FACULTY OF MEDICINE

- Unimodal masking losses

- Similarity-based multimodal masking loss
- Downstream tasks (Classification)

Performance of LANISTR

UNIVERSITY OF HELSINKI
FACULTY OF MEDICINE

- Dataset

Dataset	Language	Image	Tabular	Time-series	Missing rate	Task	Pre-training Samples	Fine-tuning Samples
MIMIC-IV (v2.2)	Clinical notes	The last chest X-ray image taken in the first 48-hour	*NA*	Clinical time series data	35.7%	Predicting in-hospital mortality after the first 48-hours of ICU	3,680,784	5923
Amazon review data (2018)	Truncated text summaries	Seller or user-provided visuals	Product ID, reviewer ID, review verification status, year, review ratings count, and timestamp	*NA*	Not mention	Predict the star rating (out of 5) a product receives	5,581,312	896

Performance of LANISTR

- Results on **MIMIC-IV**

Method/Category	AUROC
CoCa	38.45
FLAVA	77.54
MedFuse	78.12 ± 2.79
LateFusion	80.79 ± 1.12
LANISTR, no pretrain	80.87 ± 2.56
LANISTR	$\mathbf{87.37 \pm 1.28}$

Performance of LANISTR

UNIVERSITY OF HELSINKI
FACULTY OF MEDICINE

- Results on **Amazon Product Review**

Method/Category	*Beauty*	*Fashion*
AutoGluon-MLP	55.34 ± 3.55	50.39 ± 1.70
AutoGluon-TF	61.59 ± 4.50	46.10 ± 3.92
LateFusion	62.47 ± 3.32	65.83 ± 6.85
ALBEF, Tab2Txt	43.51 ± 2.91	43.23 ± 3.56
ALBEF	56.34 ± 2.09	55.78 ± 2.16
LANISTR, Tab2Txt	59.23 ± 3.76	48.21 ± 4.62
LANISTR, no pretrain	65.43 ± 7.13	52.07 ± 5.66
LANISTR	**76.27 ± 3.17**	**75.15 ± 1.20**

Performance of LANISTR

- Ablation study

Ablation	w/o time	w/o image	w/o text	w/o $\mathcal{L}_{\text{MTM}}$	w/o $\mathcal{L}_{\text{MIM}}$	w/o $\mathcal{L}_{\text{MLM}}$	w/o $\mathcal{L}_{\text{SimMIM}}$	w/o non-parallel data	LANISTR
AUROC	79.89	72.78	70.29	83.41	82.23	80.89	80.43	79.87	87.37

Ablation study for modalities and objective functions in LANISTR in the presence of different modalities in the MIMIC-IV dataset.

% Unlabeled Data	0%	25%	50%	75%	100%
AUROC (%)	80.87	81.90	83.60	85.90	87.37

Effect of pretraining dataset size on downstream task in MIMIC-IV.

Conclusion

UNIVERSITY OF HELSINKI
FACULTY OF MEDICINE

- *Structured* and *Unstructured* data.
- LANISTR, a novel framework for language, image, and structured data, utilizing unimodal and multimodal masking strategies for pretraining.
- Overcome that missing modality in large-scale unlabeled data, a prevalent issue in real-world multimodal datasets.
- Demonstrated on **real-world retail (Amazon Product Review)** and **healthcare (MIMIC-IV)** datasets, LANISTR showcases remarkable performance improvements over existing methods.

UNIVERSITY OF HELSINKI
FACULTY OF MEDICINE

Thank you