ECML PKDD 2024

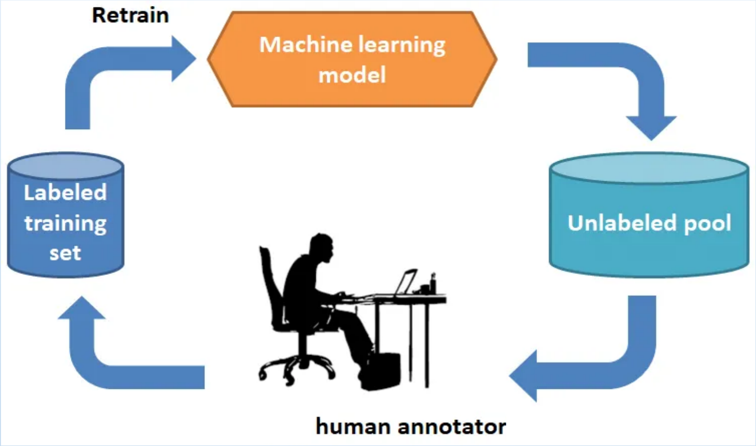
Self-SLAM: A Self-Supervised Learning Based Annotation Method to Reduce Labeling Overhead

Authors: Alfiya M. Shaikh, Hrithik Nambiar, Kshitish Ghate, Swarnali Banik, Sougata Sen, Surjya Ghosh, Vaskar Raychoudhury, Niloy Ganguly, Snehanshu Saha

Presenter: **Alfiya M. Shaikh**

Problem Statement

- Deep Neural Networks require large amounts of labeled data for accurate predictions.
- Manual annotation is time-consuming, expensive, and error-prone.
- Despite data abundance by pervasive devices, much remains unlabeled due to annotation challenges.
- Need for automated methods to handle large, unlabeled datasets.



Motivation

- **Minimizing Annotation Burden:**
  - Aim to reduce the manual data annotation workload.
- **Self-Supervised Learning:**
  - Self-SLAM (SSLAM) uses self-supervised learning that generate labels from unlabeled data.
  - This method reduces the need for labeled data by solving pretext tasks.
- **Real-World Application:**
  - SSLAM automates surface classification for wheelchair accessibility, reducing manual effort.
  - Applicable in various domains with large unlabeled datasets.

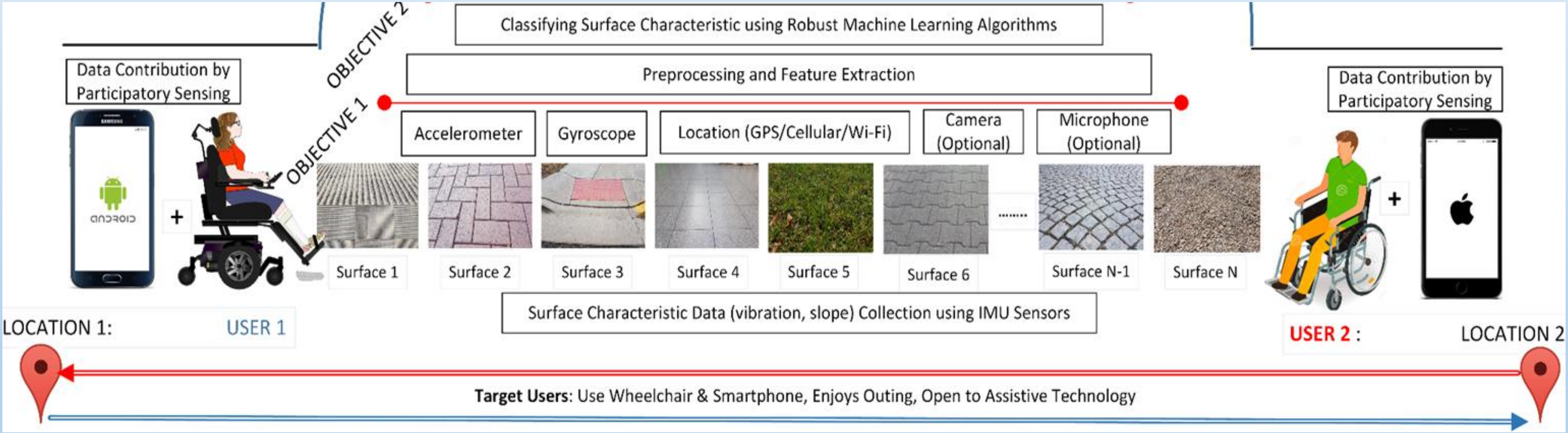
Key Contributions

- Introduction of SSLAM to reduce annotation overhead.
- Creation and sharing of a novel wheelchair-induced surface vibration dataset.
- Development of a robust log-cosh loss function.
- Demonstration of SSLAM’s effectiveness across multiple datasets.

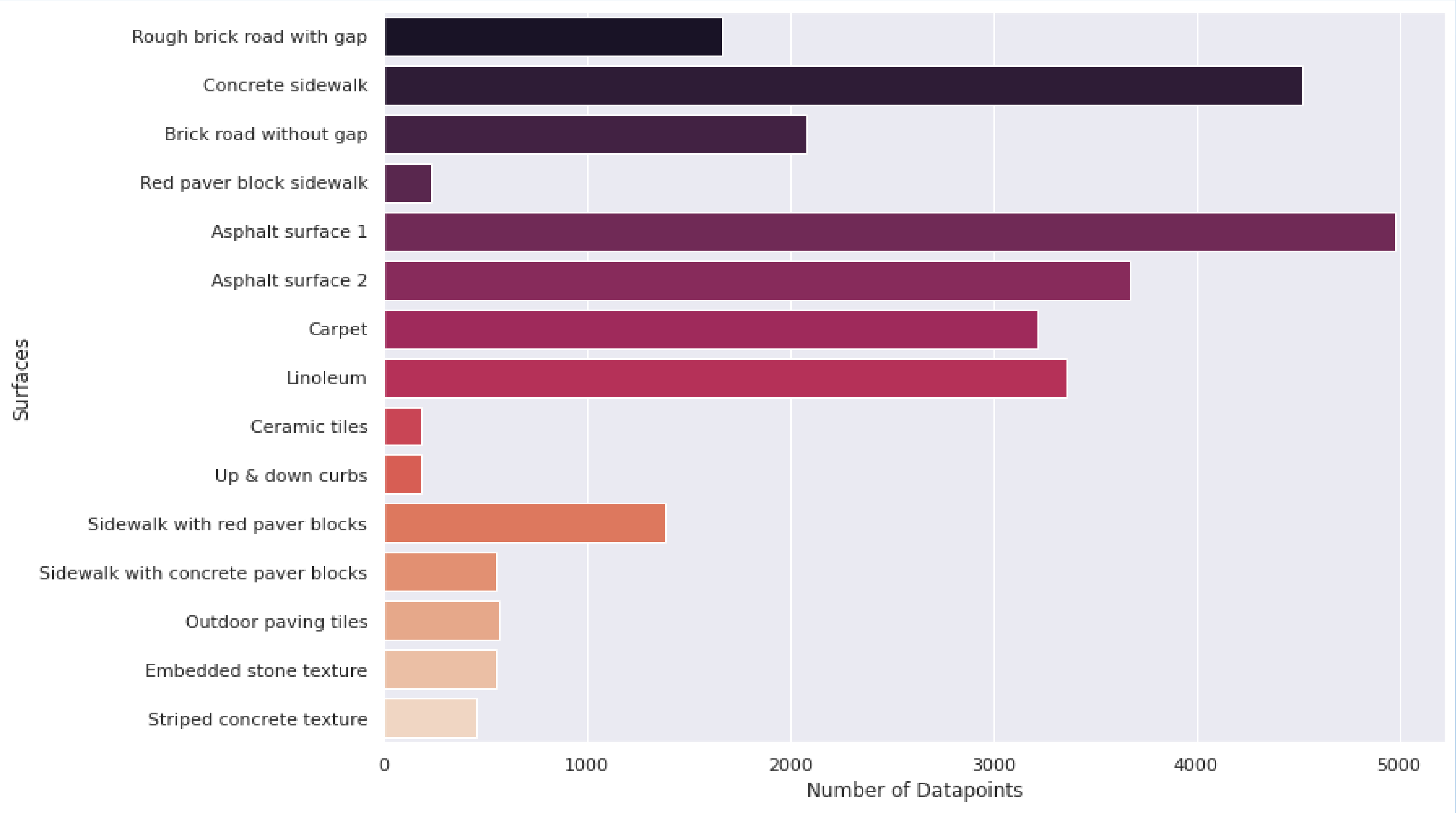
Datasets

1. Wheelchair dataset

- Data from multiple surfaces across countries

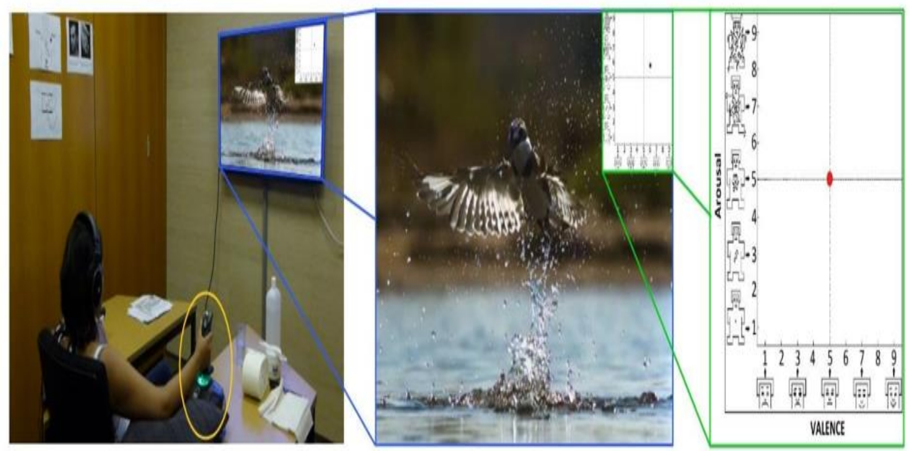


Class Distribution of the collected wheelchair dataset

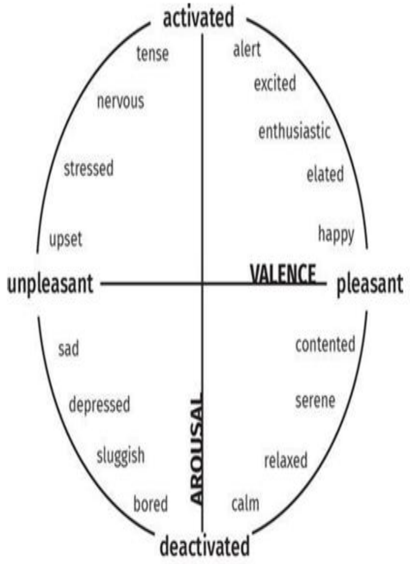


Datasets

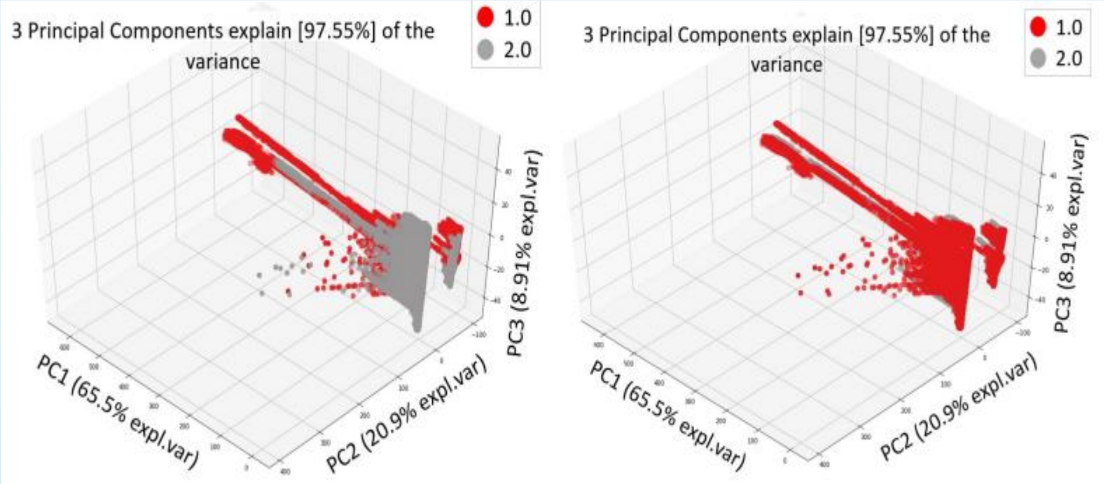
2. CASE Dataset: Continuous emotion annotation using physiological signals (ECG, BVP, GSR).



- 30 users
- 8 videos in randomised order
- 2D plane - Joystick Input
- Emotional data collected -



Circumplex Model of emotions

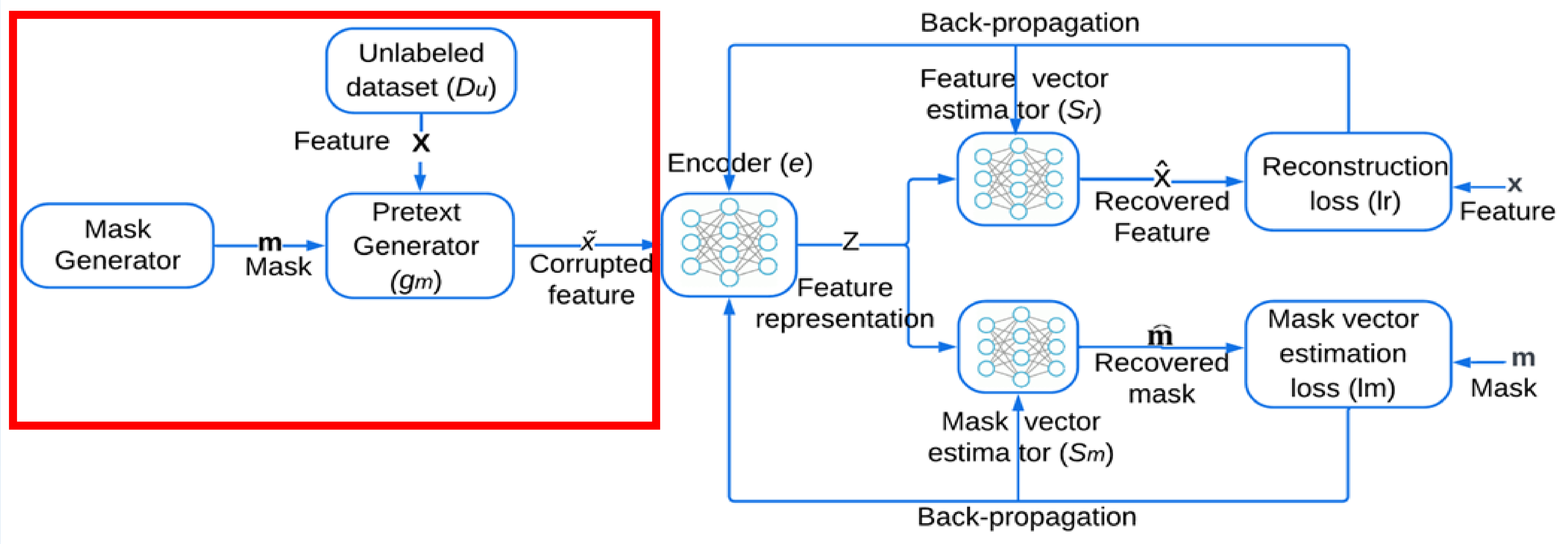


(a)(Classes 1 and 2) Low and high arousal levels. **(b)**(Classes 1 and 2) Low and high valence levels.

Fig. : CASE dataset distribution: Classes 1 and 2 correspond to low and high levels of (a) arousal and (b) valence respectively.

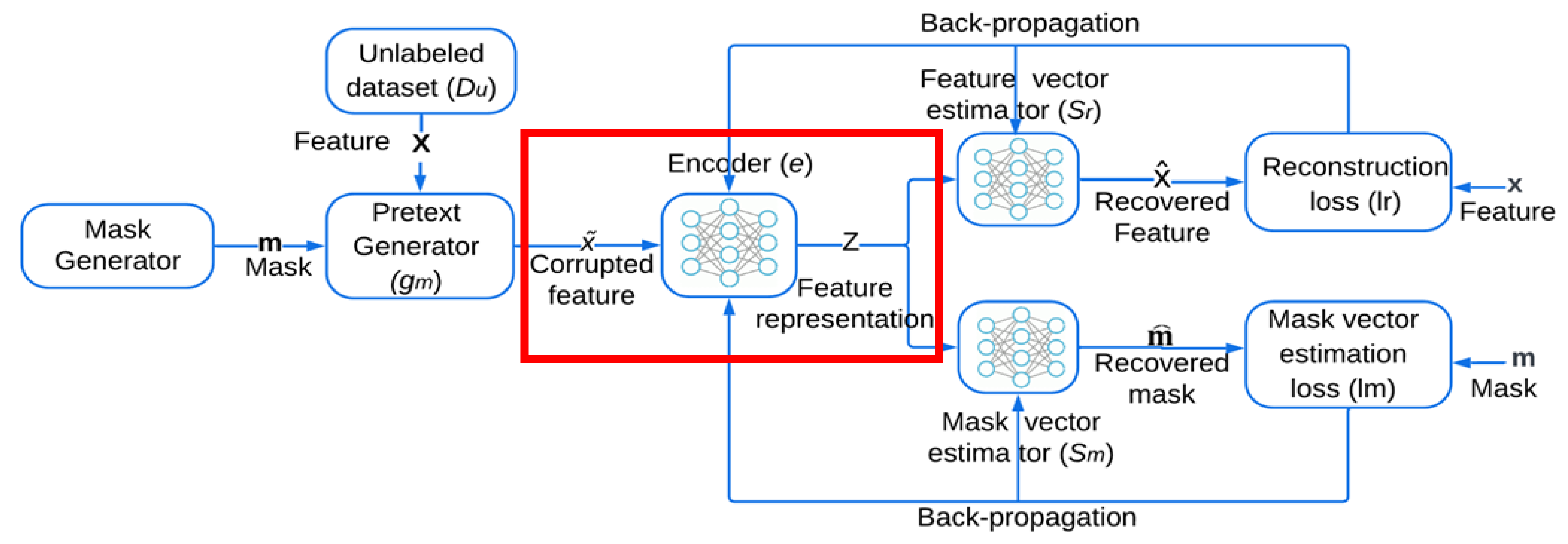
Proposed Method: Self-SLAM Framework

- **Data Corruption:**
- A portion of the training data is intentionally corrupted using a random binary mask to simulate missing or noisy features.



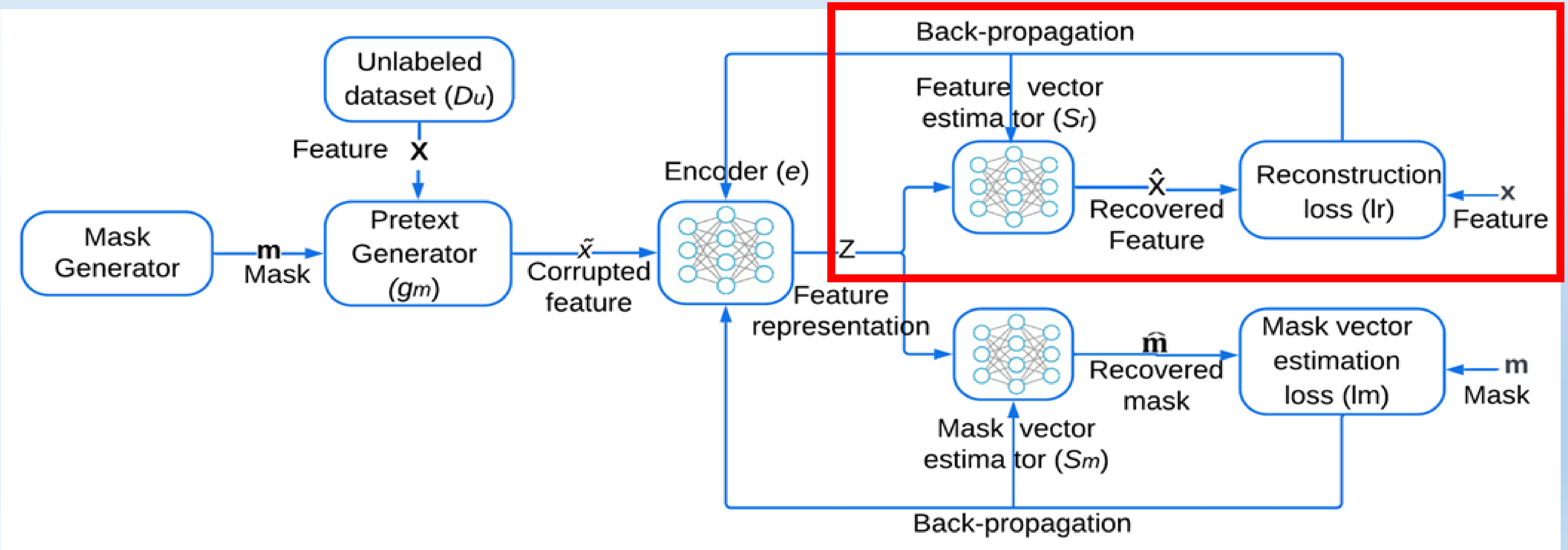
Proposed Method: Self-SLAM Framework

- **Encoder Processing:**
- Corrupted data fed into an encoder to learn low-dimensional representations.



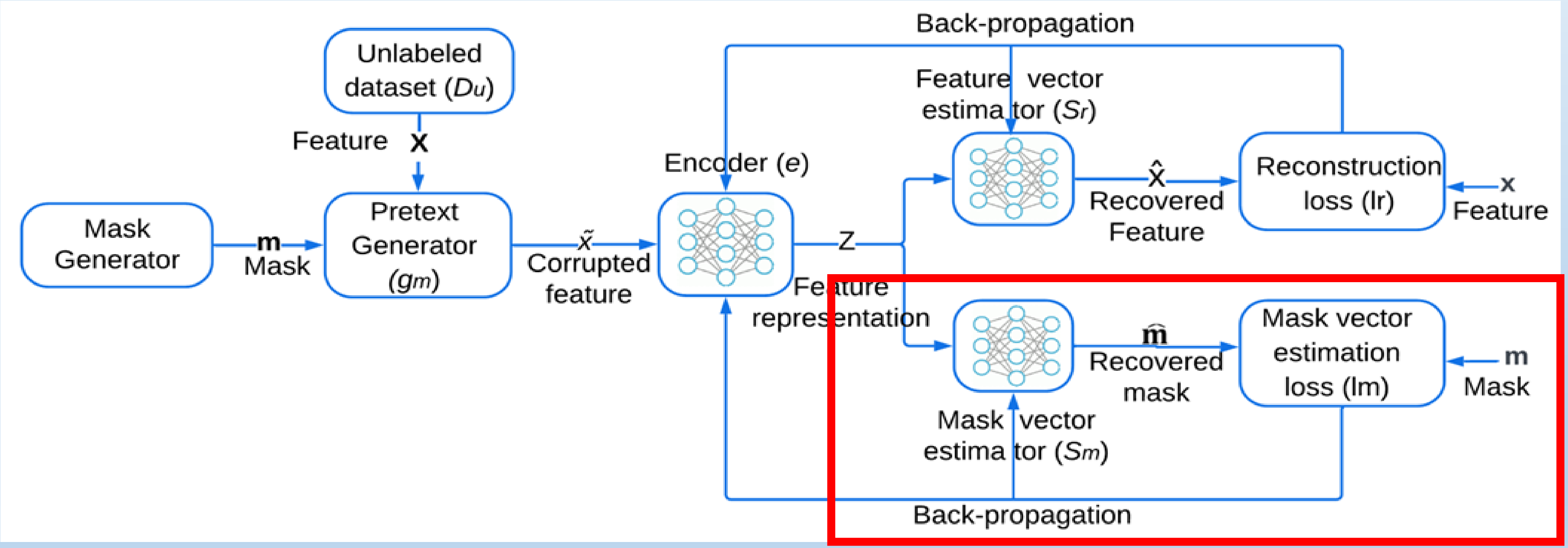
Proposed Method: Self-SLAM Framework

- **Mask and Feature Vector Estimation:**
- Feature Vector Estimation: Reconstructs original data from corrupted version.



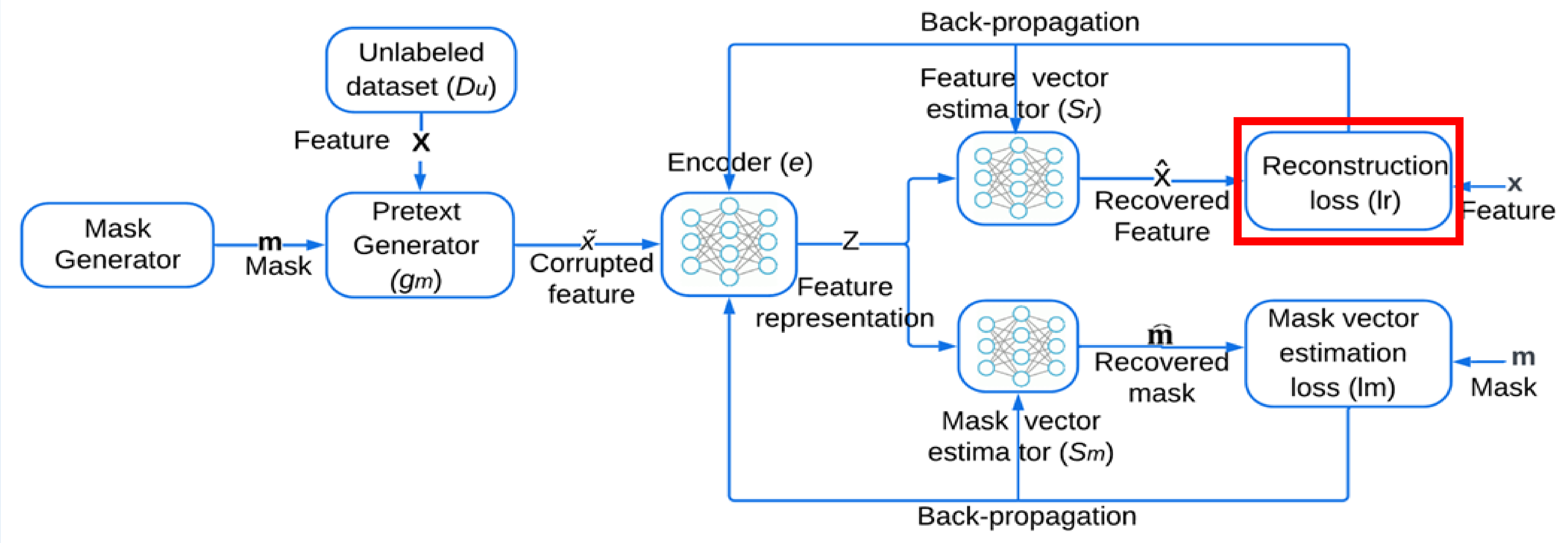
Proposed Method: Self-SLAM Framework

- **Mask Vector Estimation**: Predicts which features were corrupted.



Proposed Method: Self-SLAM Framework

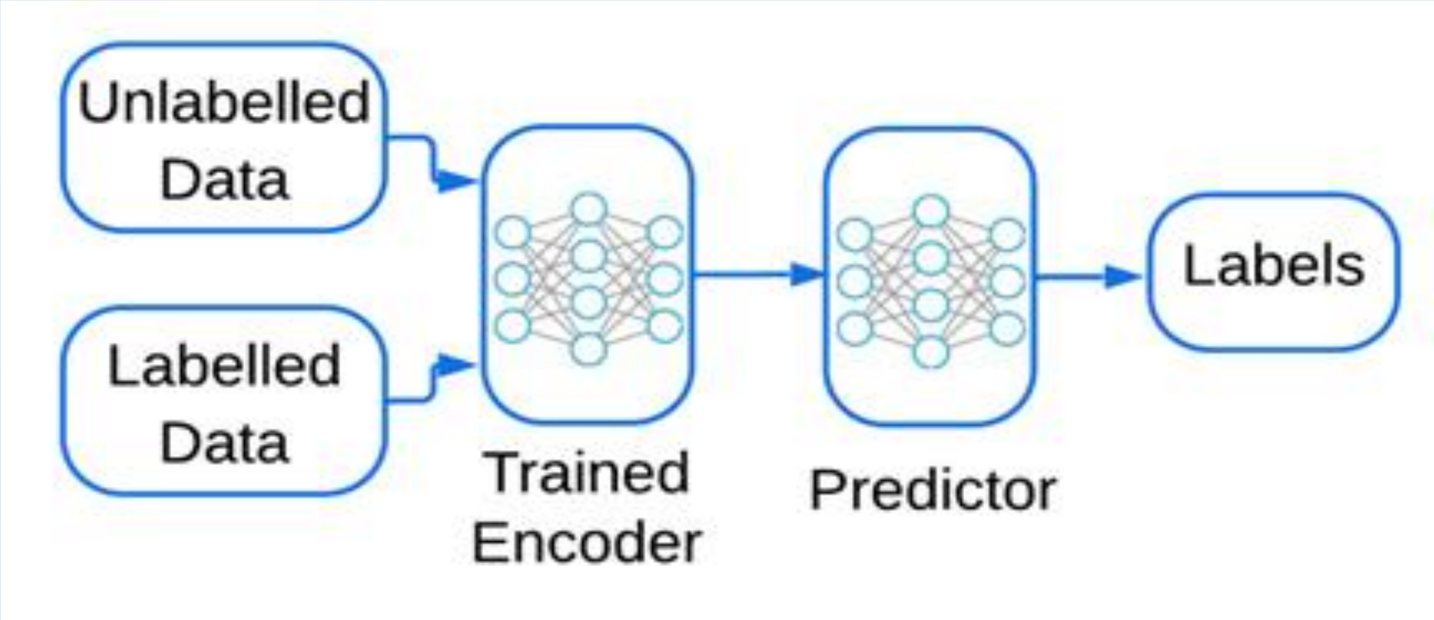
- Log-cosh[1] loss function used for balancing sensitivity and robustness.



1. More details about the log-cosh loss can be found in this paper

Proposed Method: Self-SLAM Framework

- **Label Generation and Downstream Tasks:** Encoder generates labels from unlabeled data for tasks like classification.
- **Iterative Improvement:** Process can be repeated with newly generated labels to enhance model performance.



Experimental Setup

- Experimental setup with labeled and unlabeled data split.
- Use of SSLAM to leverage unlabeled data for learning representations.
- Comparison with baseline models: MLP, Logistic Regression, XGBoost, and VIME.

Results: CASE Dataset

- Results on CASE dataset: SSLAM significantly outperformed baselines.
- Achieved 90.01% accuracy for valence prediction and 89.79% for arousal prediction.
- Demonstrated robustness in handling noisy annotations.

Model Type	Accuracy using 85-15% split		Accuracy using 80-20% split	
	Valence	Arousal	Valence	Arousal
MLP	0.8130 ± 0.0040	0.7993 ± 0.0013	0.8034 ± 0.0021	0.7911 ± 0.0057
Logistic Regression	0.6901 ± 0.0012	0.6806 ± 0.0010	0.6877 ± 0.0028	0.6899 ± 0.0025
XGBoost	0.7330 ± 0.0009	0.7339 ± 0.0022	0.8467 ± 0.0041	0.7343 ± 0.0015
VIME	0.6917 ± 0.0051	0.7213 ± 0.0042	0.7197 ± 0.0027	0.7093 ± 0.0027
SSLAM	**0.9001 ± 0.0024**	**0.8979 ± 0.0045**	**0.8949 ± 0.0046**	**0.8884 ± 0.0073**

Results: Wheelchair Dataset

- Results on the wheelchair dataset: SSLAM achieved 71.65% accuracy.
- Outperformed MLP and VIME despite class imbalance.
- Demonstrated resilience to noisy data through the log-cosh loss function.

Model Type	85-15% split		80-20% split	
	Accuracy	F1 score	Accuracy	F1 score
MLP	0.6740 ± 0.0211	0.6704 ± 0.0054	0.6347 ± 0.0294	0.6309 ± 0.0231
Logistic Regression	0.4463 ± 0.0020	0.4063 ± 0.0012	0.4307 ± 0.0054	0.4237 ± 0.0073
XGBoost	0.6305 ± 0.0089	0.6304 ± 0.0029	0.6216 ± 0.0019	0.6193 ± 0.0147
VIME	0.6366 ± 0.0393	0.6065 ± 0.0381	0.6283 ± 0.0317	0.6267 ± 0.0318
SSLAM	**0.7165 ± 0.0054**	**0.7120 ± 0.0067**	**0.7074 ± 0.0059**	**0.7040 ± 0.0093**

Implications and Future work

- SSLAM reduces annotation burden by leveraging unlabeled data.
- Demonstrated superior performance across datasets with noisy and imbalanced data.
- Future directions: Extending SSLAM to other datasets, exploring data augmentation, and semi-supervised learning.

Thank You for Your Attention

Acknowledgments:

- **Funding and Research Support:** APPCAIR, SERB-DST, SERB CRG-DST, Chanakya Fellowship from AI4CPS (Swarnali Banik)

References

1. **Jeendgar, A., Devale, T., Dhavala, S. S., & Saha, S.** (2023). *LogGENE: A smooth alternative to check loss for Deep Healthcare Inference Tasks*. arXiv preprint arXiv:2206.09333. Available at https://arxiv.org/abs/2206.09333.