Self-supervised representation learning from random data projectors

Yi Sui, Tongzi Wu, Jesse C. Cresswell, Ga Wu,
George Stein, Xiao Shi Huang, Xiaochen Zhang, Maksims Volkovs

Spotlight talk TRL@NeurIPS 2023
arXiv: 2310.07756
Code: github.com/layer6ai-labs/LFR

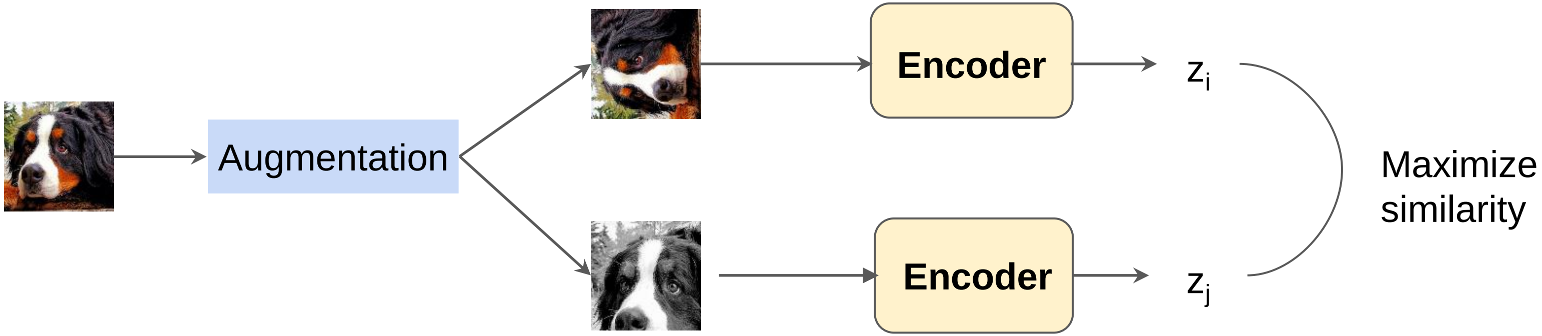
NEURAL INFORMATION PROCESSING SYSTEMS

DALHOUSIE UNIVERSITY
FACULTY OF COMPUTER SCIENCE

layer 6

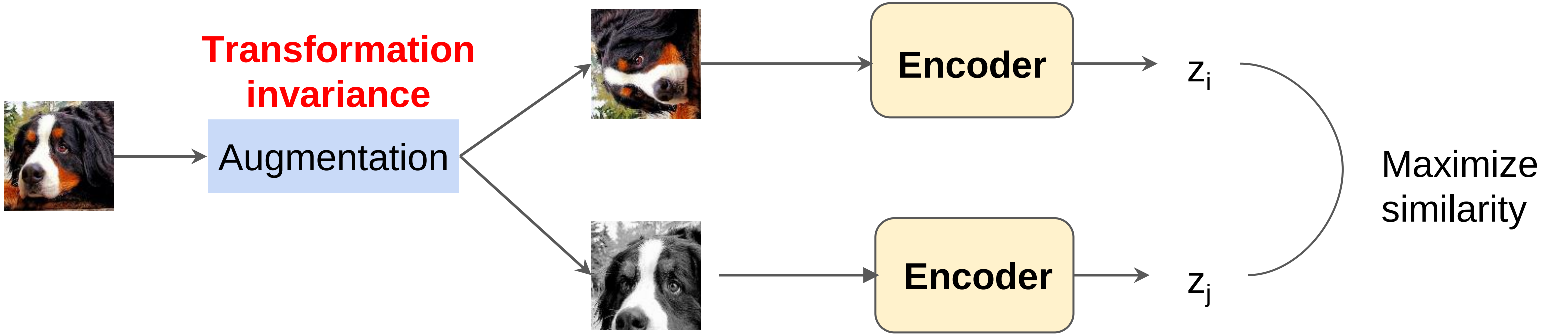
Self-supervised representation learning

Learn representations from unlabelled data



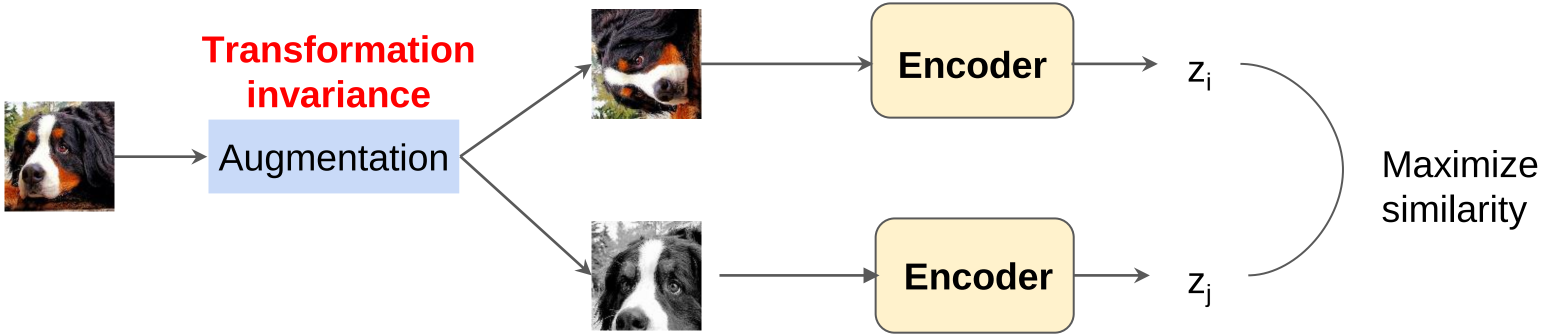
Self-supervised representation learning

Learn representations from unlabelled data

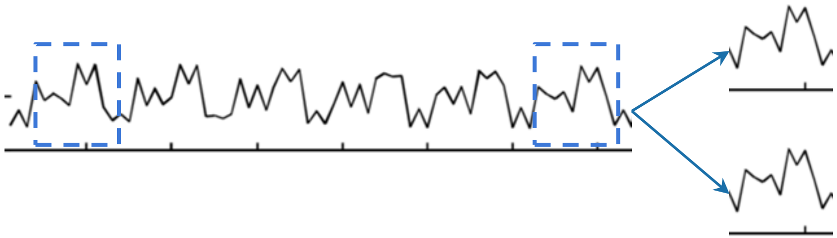


Self-supervised representation learning

Learn representations from unlabelled data



Effective augmentations are not always available



Mass	Velocity	Momentum
2	8	16
…	…	…

LFR (Learning From Randomness)

Domain agnostic representation learning ***without augmentations***

- A good representation should capture useful information that supports various downstream predictive tasks
- Use **random data projections** to simulate arbitrary downstream tasks

LFR (Learning From Randomness)

Domain agnostic representation learning ***without augmentations***

- A good representation should capture useful information that supports various downstream predictive tasks
- Use **random data projections** to simulate arbitrary downstream tasks

LFR (Learning From Randomness)

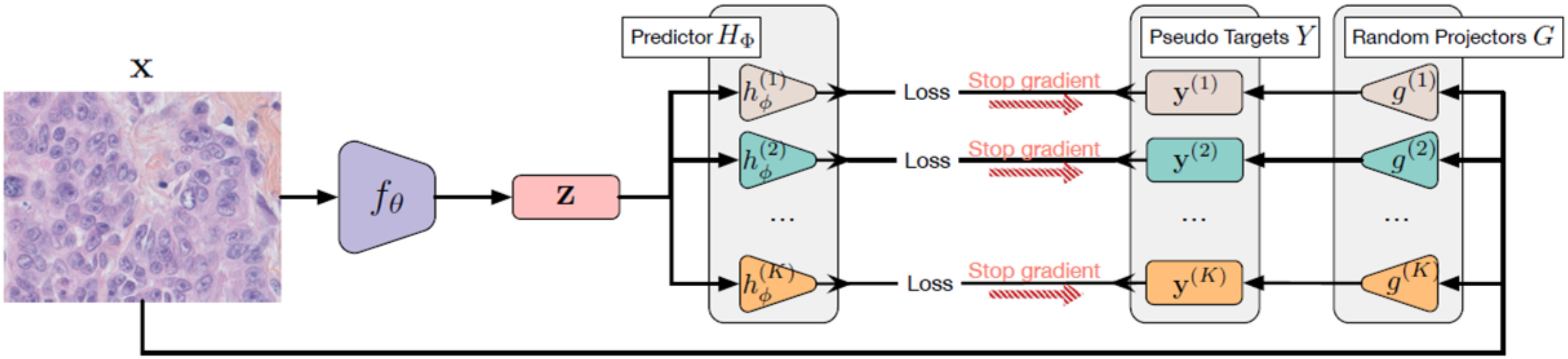
Domain agnostic representation learning ***without augmentations***

- A good representation should capture useful information that supports various downstream predictive tasks
- Use **random data projections** to simulate arbitrary downstream tasks

LFR (Learning From Randomness)

Domain agnostic representation learning ***without augmentations***

- A good representation should capture useful information that supports various downstream predictive tasks
- Use **random data projections** to simulate arbitrary downstream tasks



Random data projectors

- **Initialization** | Randomly initialized neural networks

Random data projectors

- **Initialization** | Randomly initialized neural networks
- **Training** | Projectors are fixed during training
    - LFR can train faster than standard methods like SimCLR with only *one* pass through the encoder and no CPU-intensive augmentations

Random data projectors

- **Initialization** | Randomly initialized neural networks
- **Training** | Projectors are fixed during training
  - LFR can train faster than standard methods like SimCLR with only *one* pass through the encoder and no CPU-intensive augmentations
- **Diversity** | Diverse projections benefit the learned representations
  - Use determinantal point process to select diverse projectors from larger candidate pool

Random data projectors

- **Initialization** | Randomly initialized neural networks
- **Training** | Projectors are fixed during training
  - LFR can train faster than standard methods like SimCLR with only *one* pass through the encoder and no CPU-intensive augmentations
- **Diversity** | Diverse projections benefit the learned representations
  - Use determinantal point process to select diverse projectors from larger candidate pool

Diverse projectors focus on different aspects

Similar projectors use redundant features

Linear evaluation performance on downstream tasks

		Time series			Tabular		
		HAR	Epilepsy	MIMIC-III	Income	Theorem	HEPMASS
	Log Reg	57.5 ± N/A	80.9 ± N/A	47.8 ± N/A	84.8 ± N/A	45.3 ± N/A	90.7 ± N/A
	Supervised	96.0 ± 0.6	98.3 ± 0.1	48.8 ± 0.0	81.5 ± 0.2	53.8 ± 0.5	91.5 ± 0.0
	Random Init	80.7 ± 2.3	89.1 ± 0.1	42.4 ± 1.1	83.1 ± 0.2	44.9 ± 0.8	84.3 ± 1.3
Domain-agnostic baselines	Autoencoder	77.2 ± 0.7	90.8 ± 1.3	44.9 ± 0.5	85.0 ± 0.1	50.0 ± 0.4	**90.7 ± 0.0**
	DIET	88.6 ± 1.3	96.8 ± 0.3	33.8 ± 5.2	82.2 ± 0.4	47.1 ± 0.5	-
	DACL	90.7 ± 0.4	97.5 ± 1.5	40.9 ± 0.6	79.8 ± 0.7	47.6 ± 1.0	88.7 ± 0.8
Domain-specific baselines	SimSiam	65.1 ± 0.8	97.4 ± 0.0	41.0 ± 1.9	79.2 ± 1.9	40.9 ± 0.9	85.3 ± 3.1
	SimCLR	87.8 ± 0.4	97.4 ± 0.2	44.1 ± 0.1	-	-	-
	SCARF	-	-	-	84.2 ± 0.1	48.5 ± 1.0	90.1 ± 0.1
	STab	-	-	-	84.2 ± 0.3	50.7 ± 0.7	83.6 ± 1.7
	TS-TCC	91.2 ± 0.8	97.6 ± 0.2	38.5 ± 1.3	-	-	-
	LFR (Ours)	**93.1 ± 0.5**	**97.9 ± 0.2**	**46.6 ± 0.3**	**85.2 ± 0.1**	**51.6 ± 0.7**	90.1 ± 0.2

Results for image modality also available, see preprint on arXiv: 2310.07756