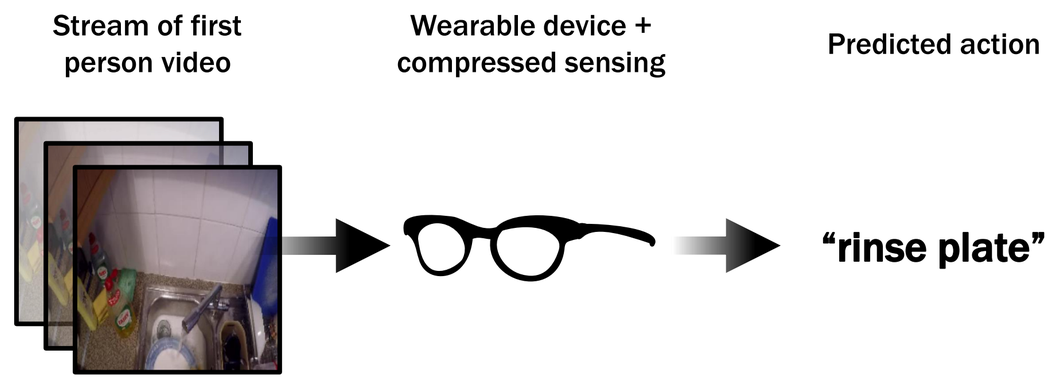
Compressed Learning for Egocentric Action Recognition

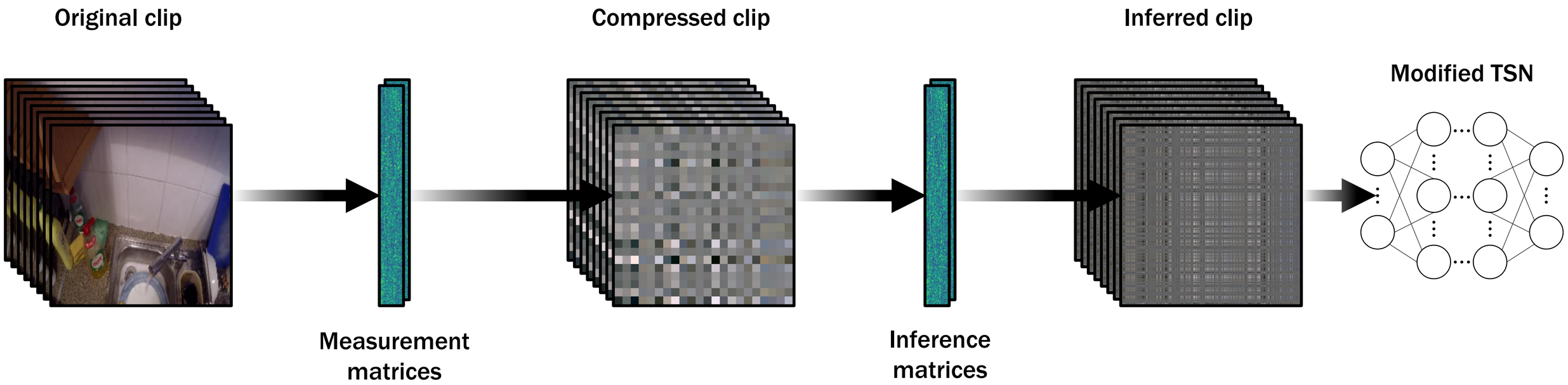
***Sam Pollard**, Supervisor: Michael Wray*

Motivation

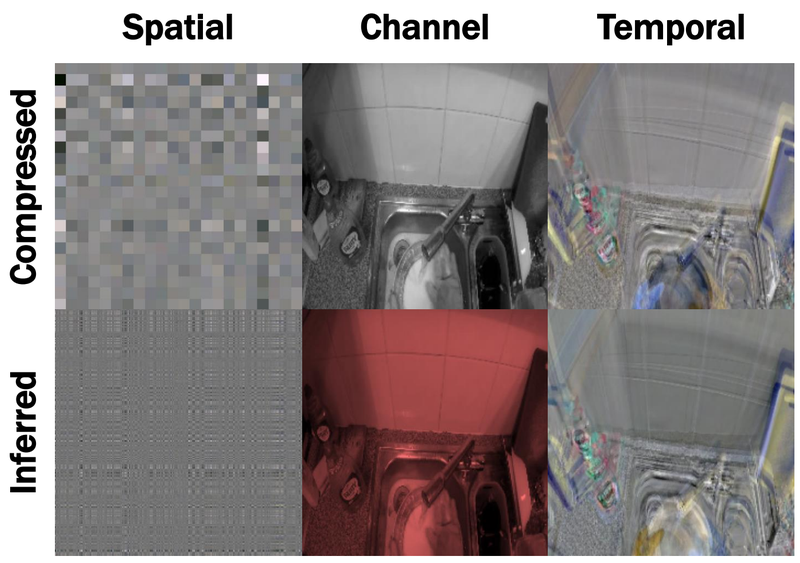
- We want to make egocentric action recognition more tractable for wearable devices.
- Compressed sensing takes M measurements of a signal of length N where ($M \ll N$) and aims to reconstruct fully.
- Instead we can train a neural network model with these measurement outputs.
- If we simulate this, how does it affect model performance?



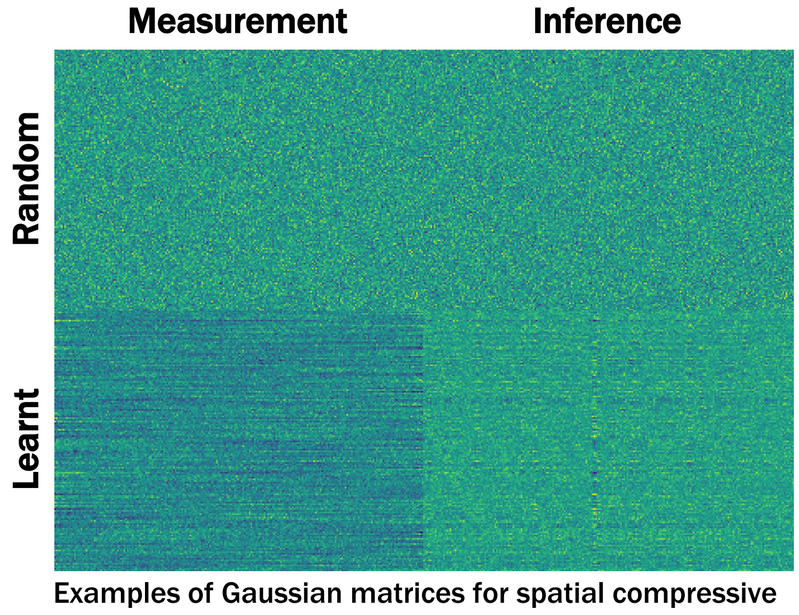
Implementation



Visualisations



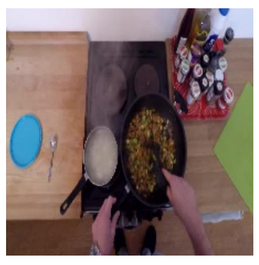
Types of compression with random Gaussian matrices and their resulting input to the TSN model. Respective measurement rates are 0.01, 0.33 and 0.125.



Examples of Gaussian matrices for spatial compressive learning at measurement rate of 0.5.

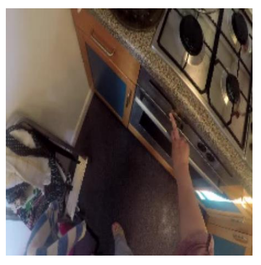
Results

Mode	Measurement Rate	Clip Dimensions	Matrix	Test Verb Accuracy	Test Noun Accuracy
Oracle	1	(8, 3, 224, 224)	None	46.03	56.43
Spatial	0.5	(8, 3, 158, 158)	Bernoulli	34.30	31.14
			Gaussian	32.69	31.19
			Bernoulli + Learnt	**36.82**	**32.80**
			Gaussian + Learnt	32.90	28.78
	0.25	(8, 3, 112, 112)	Bernoulli	32.74	20.74
			Gaussian	26.21	25.40
			Bernoulli + Learnt	**36.87**	**35.42**
			Gaussian + Learnt	30.60	27.01
	0.1	(8, 3, 71, 71)	Bernoulli	30.81	26.21
			Gaussian	**37.62**	**31.78**
			Bernoulli + Learnt	34.51	25.40
			Gaussian + Learnt	31.35	30.71
	0.01	(8, 3, 22, 22)	Bernoulli	26.74	24.17
			Gaussian	26.80	20.31
			Bernoulli + Learnt	31.46	26.21
			Gaussian + Learnt	**34.03**	**30.60**
Channel	0.33	(8, 1, 224, 224)	Bernoulli	45.12	48.29
			Gaussian	37.30	41.05
			Bernoulli + Learnt	**50.70**	50.43
			Gaussian + Learnt	48.18	**53.00**
Temporal	0.5	(4, 3, 224, 224)	Bernoulli	**45.71**	41.75
			Gaussian	43.89	**50.11**
			Bernoulli + Learnt	44.86	36.01
			Gaussian + Learnt	43.19	46.25
	0.25	(2, 3, 224, 224)	Bernoulli	**42.93**	44.16
			Gaussian	37.62	35.32
			Bernoulli + Learnt	38.64	31.62
			Gaussian + Learnt	42.07	**46.09**
	0.125	(1, 3, 224, 224)	Bernoulli	39.34	27.12
			Gaussian	41.80	39.12
			Bernoulli + Learnt	**46.09**	**39.28**
			Gaussian + Learnt	33.44	27.81



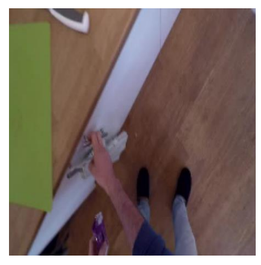
mix – 99.3%
put – 0.361%
take – 0.0620%

vegetable – 96.4%
rice – 1.62%
spatula – 1.37%



set – 27.2%
close – 16.8%
take – 12.9%

alarm – 26.0%
glove – 23.3%
drawer – 20.5%



put – 88.3%
open – 5.07%
wash – 1.30%

board – 89.2%
grater – 3.76%
leaf – 2.60%

University of BRISTOL

References

[1] Wang, Limin, et al. "Temporal segment networks: Towards good practices for deep action recognition." European conference on computer vision. Springer, Cham, 2016.
[2] Price, Will, and Dima Damen. "An evaluation of action recognition models on epic-kitchens." arXiv preprint arXiv:1908.00867 (2019).
[3] Damen, Dima, et al. "Scaling egocentric vision: The epic-kitchens dataset." Proceedings of the European Conference on Computer Vision (ECCV). 2018.
[4] Tran, Dat Thanh, et al. "Multilinear compressive learning." IEEE transactions on neural networks and learning systems 32.4 (2020): 1512-1524.