

You're your own best teacher: A Self-Supervised Learning Approach For Expressive Representations

Johan Vik Mathisen

May 24, 2024

0.1 Stage 1

0.1.1 Reconstruction

TODO: Table for val recons

0.1.2 Classification

Mean linear probe accuracy

Dataset	Baseline		SSL Method											
	Regular		Barlow Twins								VibCReg			
	None		Warp		Slice		Gauss		Warp		Slice		Gauss	
	KNN	SVM	KNN	SVM	KNN	SVM	KNN	SVM	KNN	SVM	KNN	SVM	KNN	SVM
FordA	0.70	0.74	0.83	0.84	0.91	0.89	0.80	0.83	0.80	0.74	0.87	0.86	0.76	0.78
ElectricDevices	0.35	0.41	0.35	0.44	0.38	0.41	0.40	0.42	0.33	0.38	0.36	0.39	0.39	0.43
StarLightCurves	0.87	0.89	0.93	0.93	0.94	0.94	0.88	0.88	0.92	0.94	0.91	0.93	0.89	0.89
Wafer	0.93	0.89	0.96	0.94	0.96	0.94	0.96	0.93	0.97	0.94	0.96	0.92	0.97	0.92
ECG5000	0.80	0.83	0.85	0.81	0.88	0.84	0.86	0.84	0.86	0.82	0.88	0.84	0.84	0.82
TwoPatterns	0.34	0.53	0.69	0.91	0.66	0.82	0.47	0.71	0.64	0.90	0.68	0.80	0.55	0.72
UWaveGestureLibraryAll	0.31	0.40	0.62	0.70	0.56	0.63	0.40	0.54	0.62	0.73	0.55	0.66	0.44	0.55
FordB	0.58	0.60	0.64	0.67	0.74	0.76	0.64	0.68	0.63	0.64	0.70	0.70	0.61	0.64
ShapesAll	0.29	0.30	0.49	0.55	0.53	0.60	0.40	0.48	0.48	0.56	0.54	0.60	0.40	0.46
SonyAIBORobotSurface1	0.56	0.68	0.54	0.70	0.61	0.74	0.53	0.70	0.48	0.74	0.58	0.71	0.54	0.69
SonyAIBORobotSurface2	0.81	0.86	0.77	0.79	0.80	0.80	0.80	0.81	0.77	0.85	0.80	0.85	0.80	0.85
Symbols	0.50	0.60	0.59	0.60	0.50	0.66	0.59	0.66	0.45	0.61	0.42	0.62	0.43	0.63
Mallat	0.63	0.77	0.72	0.81	0.76	0.83	0.68	0.78	0.79	0.87	0.77	0.85	0.69	0.86

Table 1: Summary of mean linear probe accuracy by SSL Method and Augmentation. Average across 4 seeds. Best result for KNN and SVM are highlighted in bold.

Max linear probe accuracy

Dataset	Baseline		SSL Method											
	Regular		Barlow Twins								VibCReg			
	None		Warp		Slice		Gauss		Warp		Slice		Gauss	
	KNN	SVM	KNN	SVM	KNN	SVM	KNN	SVM	KNN	SVM	KNN	SVM	KNN	SVM
FordA	0.75	0.78	0.84	0.88	0.93	0.92	0.85	0.87	0.81	0.77	0.88	0.90	0.86	0.85
ElectricDevices	0.35	0.43	0.36	0.45	0.39	0.43	0.45	0.46	0.34	0.42	0.39	0.42	0.42	0.45
StarLightCurves	0.89	0.91	0.94	0.95	0.96	0.96	0.90	0.91	0.95	0.95	0.93	0.95	0.90	0.90
Wafer	0.94	0.89	0.97	0.95	0.97	0.95	0.97	0.93	0.97	0.95	0.97	0.95	0.97	0.94
ECG5000	0.83	0.84	0.88	0.86	0.90	0.88	0.90	0.88	0.88	0.85	0.89	0.86	0.86	0.85
TwoPatterns	0.37	0.62	0.75	0.96	0.68	0.85	0.55	0.75	0.70	0.92	0.71	0.81	0.63	0.76
UWaveGestureLibraryAll	0.34	0.43	0.67	0.74	0.60	0.67	0.43	0.54	0.67	0.76	0.58	0.67	0.48	0.58
FordB	0.60	0.63	0.67	0.71	0.76	0.80	0.69	0.74	0.67	0.65	0.74	0.77	0.63	0.68
ShapesAll	0.33	0.34	0.53	0.59	0.59	0.65	0.44	0.50	0.50	0.56	0.57	0.63	0.44	0.48
SonyAIBORobotSurface1	0.67	0.80	0.61	0.77	0.76	0.80	0.60	0.74	0.51	0.79	0.63	0.75	0.63	0.75
SonyAIBORobotSurface2	0.84	0.89	0.80	0.86	0.82	0.84	0.83	0.82	0.81	0.88	0.81	0.88	0.83	0.87
Symbols	0.56	0.66	0.65	0.69	0.55	0.73	0.64	0.71	0.51	0.65	0.45	0.67	0.46	0.69
Mallat	0.54	0.88	0.57	0.87	0.74	0.89	0.66	0.80	0.74	0.92	0.72	0.88	0.62	0.90

Table 2: Summary of max linear probe accuracy by SSL Method and Augmentation. Maximum value across 4 seeds. Best result for KNN and SVM are highlighted in bold.

TODO: Plots that illustrate.

Visual inspection

TODO: PCA-TSNE-UMAP plots

TODO: Latent space plots

0.2 Stage 2

0.2.1 Generative quality

TODO: FID-IS table - mean and max

TODO: Plots that illustrate.

Class conditional sampling

TODO: CAS table

TODO: Plots that illustrate.

Visual inspection

Generated vs real.

0.2.2 Thoughts

Better inception score and CAS of our models indicate that the class separability learned in latent space makes the conditional distributions more distinct easier to classify. The FID is variable, but in many cases better, which indicated that the generative distributions are closer to the ground truth.

Gaussian noise aug seems to result in a lot easier the BT/VibCReg loss to minimize.

Slice and shuffle is harder to minimize, but could seem to push representations for different classes further apart resulting in better linear probes.

Talk about the difficulty/ease in minimizing the SSL loss for the different augmentations. Does this affect linear probes / reconstruction / FID / IS / Prior loss

Mention that during experiments with our stage 2 modification, embed / fine-tune, we observed that the val prior loss with our modification was higher, but with similar shape as without. If we had time and computational resources to re-run the experiments, then we would omit the stage 2 modification. The FID/IS in our main experiments are in many cases better than baseline VQVAE, despite higher val prior loss.

For datasets of smaller size with classes of different characteristics (a clear distributional difference in visual inspection [Sony2 and Symbols]) NC-VQVAE seems to perform better both in terms of FID and IS.

The biases introduced by augmentations in stage 1 seems to be included in the generated samples to some degree. In particular datasets with high frequency components, when applying Gaussian noise (easier to spot), has substantially better FID score.

Is there correlation between CAS and linear probe accuracy??

0.3 Discussion

The added flexibility of NC-VQVAE, with possibility of choosing dataset specific augmentations, can in some applications be beneficial.

0.4 Further work

[1] suggest that focus on augmentations is of great importance. The hunt for good augmentations in the time series domain is ongoing and should probably get more attention.

HF-LF split - augmentations tailored for HF and LF, as they often have quite different characteristics.

Wavelet transform to improve HF-LF split.

Further optimize the relationship between aug recon loss and choice of augmentations.