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

Johan Vik Mathisen

May 16, 2024

Our work, as well as out models can be divided into two stages. Firstly we investigate the effect of non-contrastive SSL on a proven tokenization model, VQVAE [**VQVAE**], with the goal of learning more expressive representations. The expressiveness is measured in terms of the models ability to reconstruct unseen data, as well at the performance of learned latent representations on a downstream classification task. The SSL models we consider are, as introduced in section 3 , BarlowTwins and VIbCReg.

make hyperlink

Secondly we investigate the effects of NC-VQ-VAE on prior learning by training a MaskGIT model on top of the tokenization models.

Additionally we provide ablations investigating robustness to augmentations and the effect of augmentations reconstruction weight.

0.1 Stage 1

As mentioned in the section on representation learning, one needs to determine a set of tasks one wish to evaluate on, in order to say anything about the quality of the representations. We evaluate the representations based on two tasks

0.1.1 Reconstruction

0.1.2 Classification

0.1.3 Codebook investigations

In the two tokenization models, how does the codebooks differ? Look at codebook utlization. Histograms across dimensions?

0.1.4 Visual inspection

0.2 Stage 2

0.2.1 Evaluation metrics

- **IS**:
- **FID**:
- **Visual inspection**:
- **Token usage**:
- **Generating distribution**:

0.3 Ablation studies

0.3.1 Augmentation Reconstruction Weight

Here are the results of the ablation on the effect of "Augmentation Reconstruction Weight" on Stage 1. "Augmentation Reconstruction Weight" is the weight given to the reconstruction loss on the augmented branch. Tested weights 0.05, 0.1, 0.15 and 0.2. Augmentations [Window Warp, Amplitude Resize] and [Slice and Shuffle]. The weight has little effect on linear probe accuracy across the four datasets tested, and the two sets of augmentations. The effect on Validation reconstruction loss is small for all except FordA. It seems, not very surprisingly, that the choice of augmentations are of (much) greater importance.

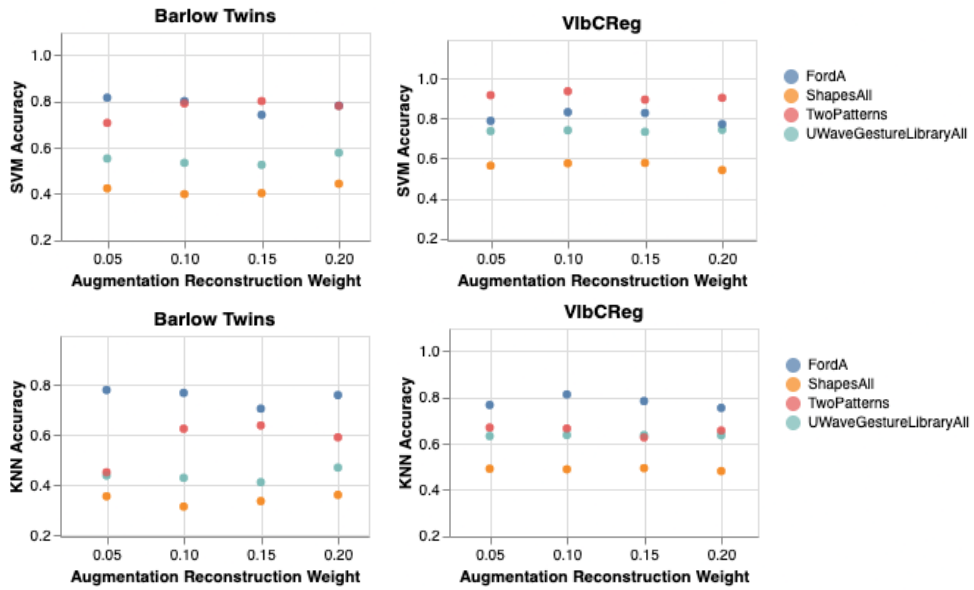

Figure 1: Augmentations: Window Warp and amplitude resize. Averaged across 2 runs. Trained for 250 epochs

0.3.2 Augmentation robustness

S1 - S2 - Augs: Choose datasets such that half were thought to be well fitting for slice and shuffle and half for amplitude resize + window Warp. This was after seeing FordA/B performing well with SS, and looking at the augmented views.
FordA/B, Electric devices, ShapesALL for SS
TwoP, UWave, symbols, Mallar for ampRes + winwarp

Visual inspection: Plot training samples + spectrogram, compare to augmented view. Compare these against others from other classes.
Does the resulting improvement in stage 1 transfer to stage 2?

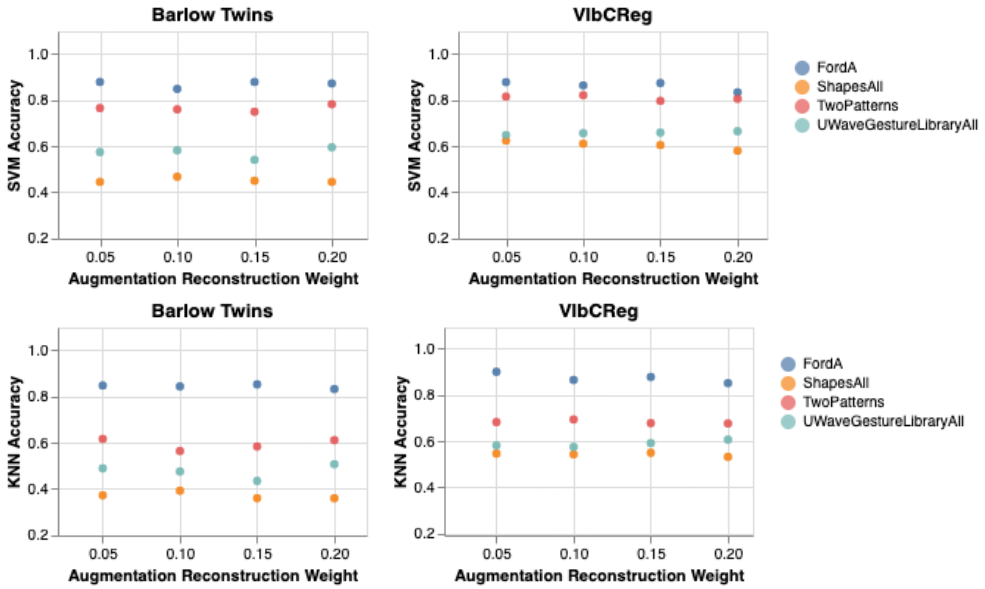

Figure 2: Augmentation: Slice and shuffle. Averaged across 2 runs. Trained for 250 epochs

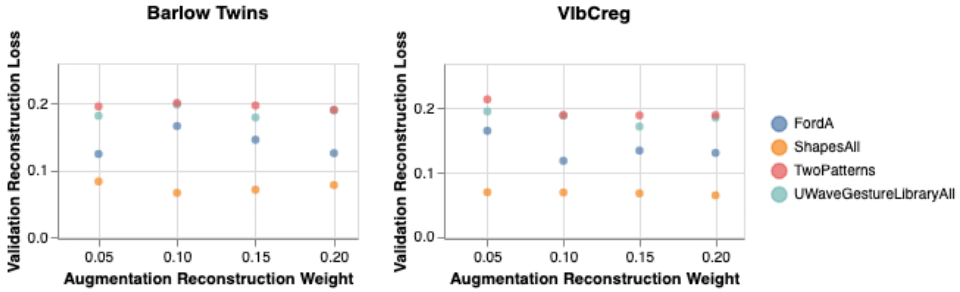

Figure 3: Augmentation: Window Warp and amplitude resize. Averaged across 2 runs. Trained for 250 epochs

TODO: Download the Wandb data.

Plot for each dataset and each augmentation: Color code according to SSL-model.

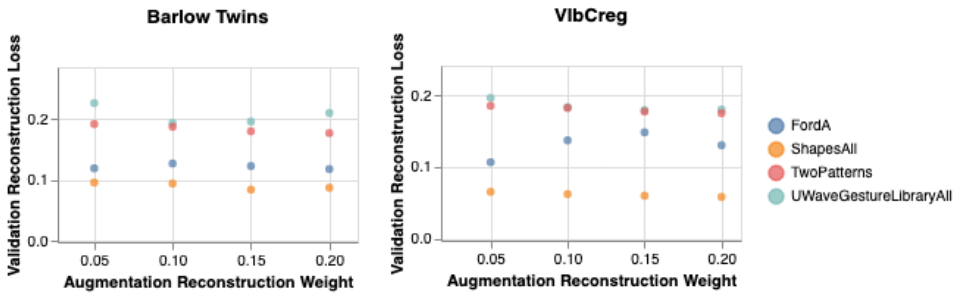

Figure 4: Augmentation: Slice and shuffle. Averaged across 2 runs. Trained for 250 epochs