

# Exploring Stylistic Invariance in Self-Supervised Contrastive Pretraining for Feature Extraction

Mark Alexander Vollmar

June 5, 2023

Self-supervised contrastive pretraining methods, such as SimCLR, MoCo, and others, have shown remarkable success in achieving invariance with respect to image augmentations. These methods generate multiple views from a source image, typically using ImageNet, to enhance feature extraction. In this thesis, we aim to investigate whether incorporating two views of the same object in different styles within a pipeline can yield a feature extraction invariant of stylistic representation. Specifically, we seek to train a model that generates similar embeddings for artistic and photographic representations of the same source image.

The thesis involves the following key steps:

- **Development of a Pretraining Pipeline:** We will create a pretraining pipeline for two contrastive methods: Barlow Twins [Zbo+21] and BYOL [Gri+20]. All pretraining will be done on ImageNet data.
- **Incorporation of Style Augmentations:** For each of the selected pretraining methods, we will integrate style augmentations into the training process. Specifically, we will explore offline augmentation using a stable diffusion style transfer method. This will enable the model to learn representations that are invariant to artistic styles.
- **Evaluation on Downstream Tasks:** To assess the effectiveness of the proposed approach, we will evaluate the trained models on multiple classification and detection downstream tasks in the artistic domain.

In addition to evaluating the proposed method, we will compare the performance against two other models:

- **Supervised Pretraining on ImageNet:** This will serve as a reference point to assess the impact of pretraining.
- **Pretraining with the unmodified methods of Barlow Twins and BYOL on ImageNet:** This will help gauge the benefits of incorporating style augmentations.

This thesis aims to contribute to the understanding of how stylistic invariance in feature extraction can be achieved by using self-supervised contrastive pretraining. The results obtained will provide insights into the effectiveness of incorporating stylistic representations in the pretraining pipeline.

## References

- [Gri+20] Jean-Bastien Grill et al. *Bootstrap your own latent: A new approach to self-supervised Learning*. 2020. arXiv: 2006.07733 [cs.LG].
- [Zbo+21] Jure Zbontar et al. *Barlow Twins: Self-Supervised Learning via Redundancy Reduction*. 2021. arXiv: 2103.03230 [cs.CV].