Can Self-Supervised Representation Learning Methods Withstand Distribution Shifts and Corruptions?

LULEÅ UNIVERSITY OF TECHNOLOGY

Fraunhofer HHI

Prakash Chandra Chhipa[1], Johan Rodahl Holmgren[1], Kanjar De[1,2], Rajkumar Saini[1], and Marcus Liwicki[1]

[1]Machine Learning Group, Luleå University of Technology, Sweden, prakash.chandra.chhipa@ltu.se

[2]Video Coding Systems, Fraunhofer Heinrich-Hertz-Institut, Berlin, Germany

ICCV23 PARIS

OOD-CV Workshop

Motivation

- Research is needed to learn invariant SSL representations capable of handling distribution shifts and corruptions; this study provides a ground in this direction by sharing insights into the robustness performance of a large-scale dataset.
- We considered the most popular SSL paradigms, namely contrastive learning, knowledge distillation, mutual information maximization, and clustering. We exhaustively evaluated the corruptions, and their severity levels present in ImageNet-C dataset to understand the resilience of each method
- We compare the robustness performance across multiple metrics, including qualitative analysis.

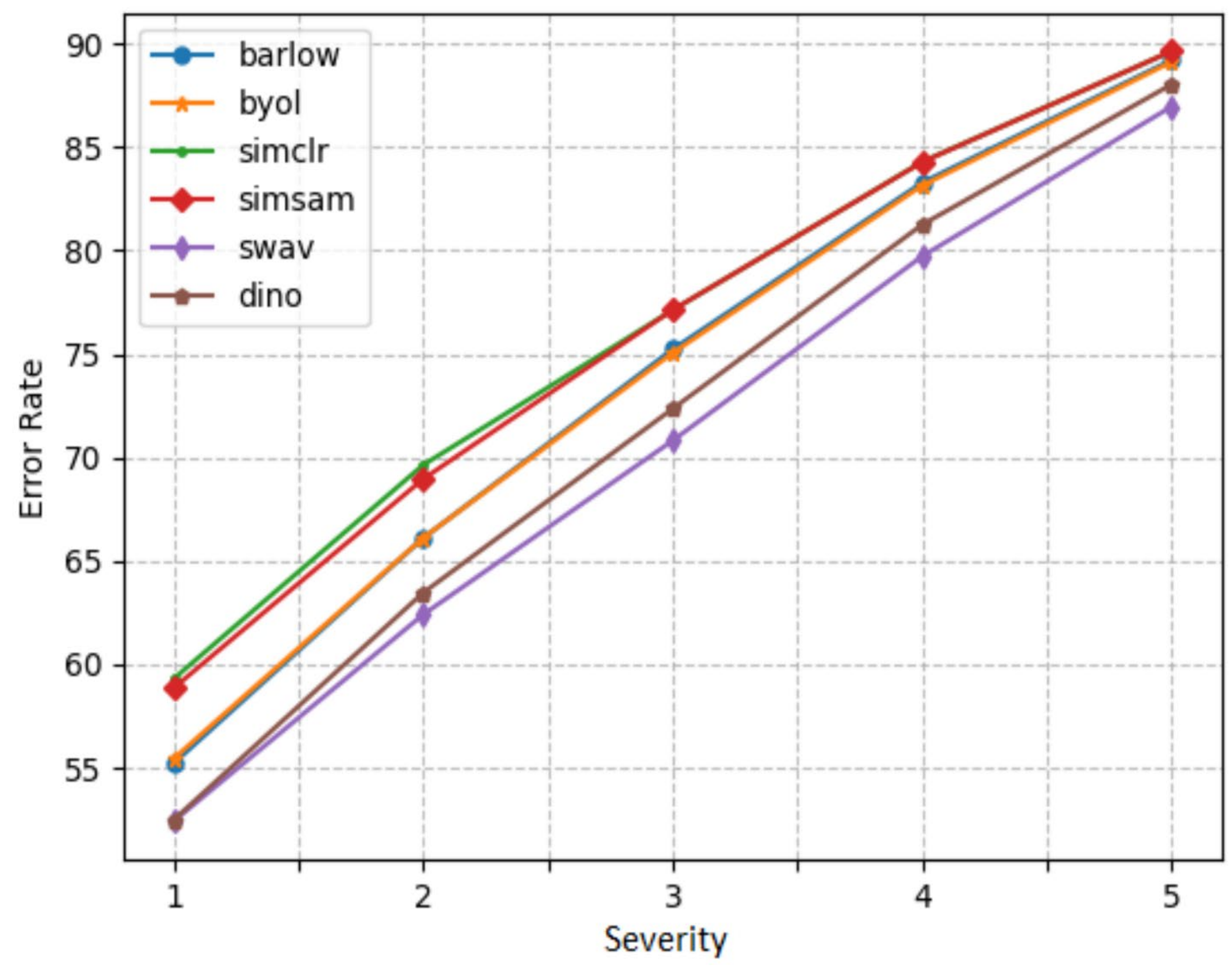

Fig1: Error rates vs. severity levels across ImageNet-C

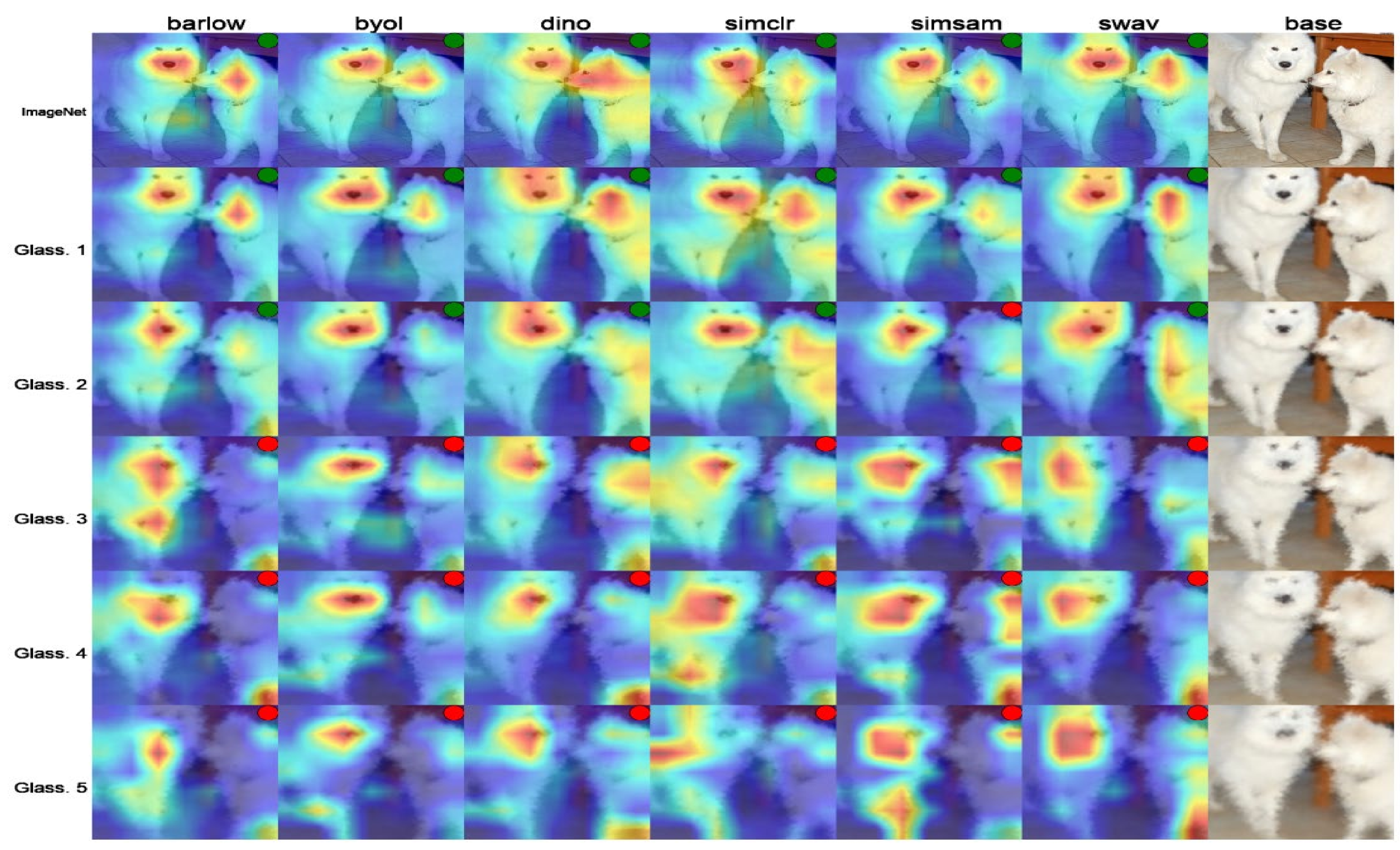

Fig2: Glass blur on dogs; markers - correct (green) and incorrect (red) classifications.

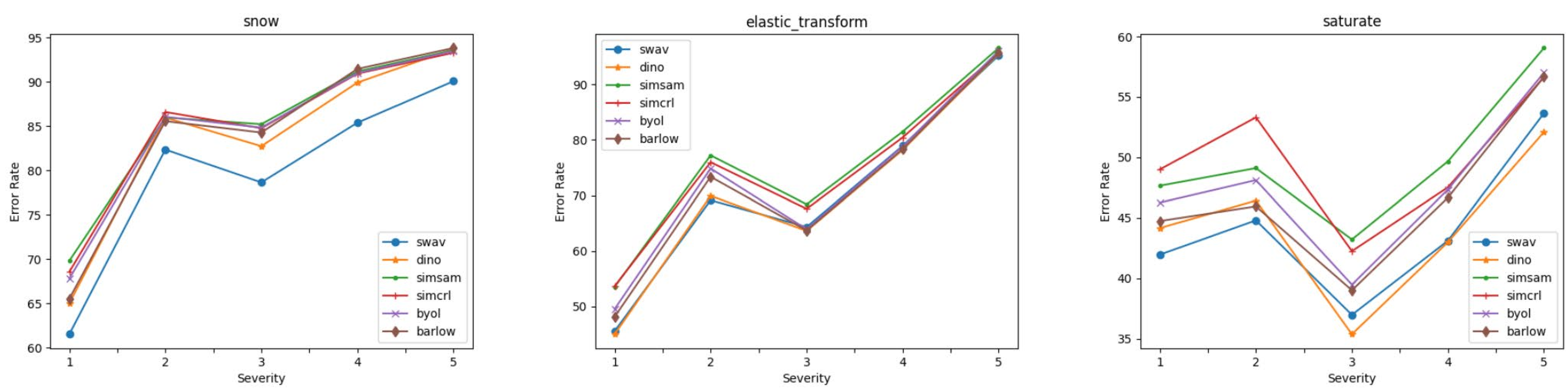

Fig2: Corruptions, namely, snow, saturate, and elastic (last row), SSL models perform poorly at severity level 2 than at severity level 3

Q&A

Q1: How do self-supervised representation learning (SSL) paradigms (contrastive learning, knowledge distillation, mutual information maximization, clustering) perform in terms of robustness when exposed to distribution shifts and image corruptions?

A1: Distribution shifts and image corruptions influence the robustness performance of the well-known SSL paradigms. The empirical analysis in this study shows that the error rates (averaged over all distribution shifts and image corruptions) increase with an increase in the severity levels of the distribution shifts and image corruptions.

Q2: To what extent can self-supervised representation learning methods maintain their robustness in the presence of distribution shifts, and what are the factors that limit their ability to do so?

A2: Extensive experiments reveal that SSL methods sustain robustness performance when subjected to lower levels of corruptions, and subsequently, the performance reduces when subjected to higher levels of corruptions. Higher corruptions may lead to massive distribution shifts, which may affect the robustness performance of learned representations.

Q3: What is the relationship between the robustness of different SSL paradigms and common categories of corruptions?

A3: Generally, robustness performance decreases for increased severity of corruptions; specifically, the weather group's robustness performance is poorer than that of other groups.

Q4: Do self-supervised representation learning methods deviate from the observed trend of error increase for certain corruptions, and what factors contribute to their robustness in the face of these corruptions?

A4: Yes; a few corruptions, namely, snow, elastic transform, and saturate, deviate from the observed trend supported by visual quality analysis.

Q5: To what extent does the presence of corruptions shift the focus of classifiers from overall representation to specific features?

A5: GradCam analysis reveals that there is a significant shift in the attention maps when the image is subjected to higher levels of corruption.

Q6: Do different backbones, such as Convolutional Neural Networks (CNNs) and Transformers, influence the behavior and robustness?

A6: Yes; the self-attention mechanism in transformer, in contrast to CNNs, does not embed any visual inductive bias of spatial locality.