Equivariant Representation Learning for Augmentation-based Self-Supervised Learning via Image Reconstruction

Qin Wang Kai Krajsek Hanno Scharr

IAS-8: Data Analytics and Machine Learning, Forschungszentrum Jülich, Germany
Jülich Supercomputing Centre (JSC), Forschungszentrum Jülich, Germany
{qi.wang, k.krajsek, h.scharr}@fz-juelich.de

Motivation

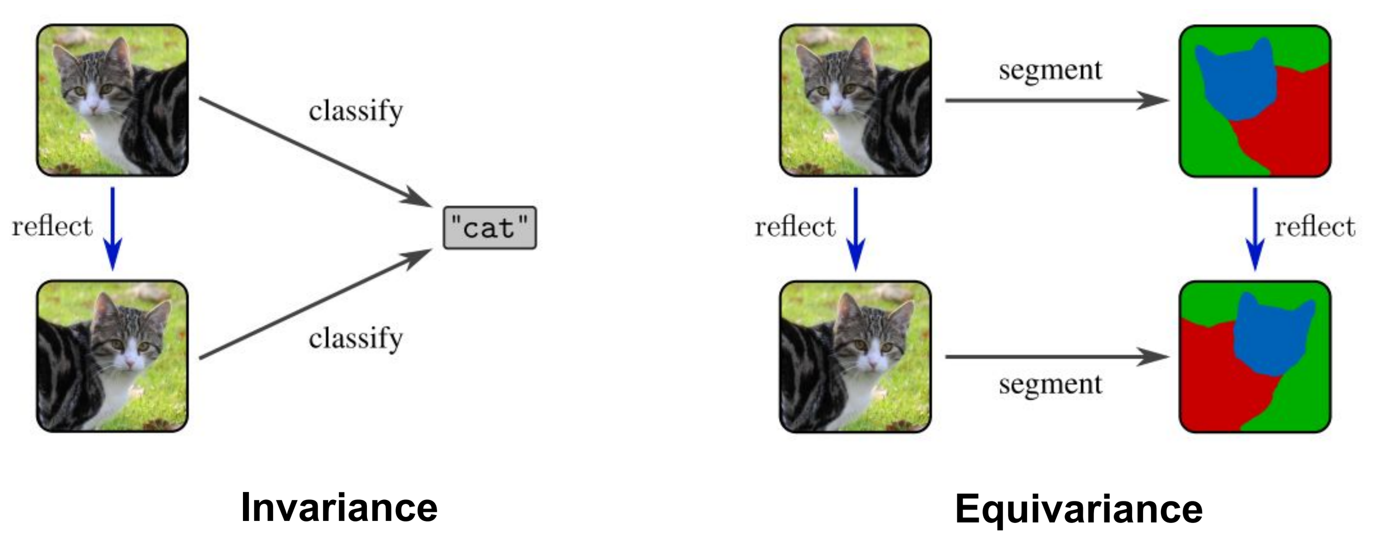


Figure 1. Commutative diagram for invariance and equivariance. Retrieved from *Equivariant and Coordinate Independent Convolutional Networks*[1], (p. v), by Maurice Weiler, 2023.

Equivariance in feature learning ensures that a model's learned representations remain consistent under various transformations, including 2D or 3D translations, rotations, scaling, and changes in color or illumination

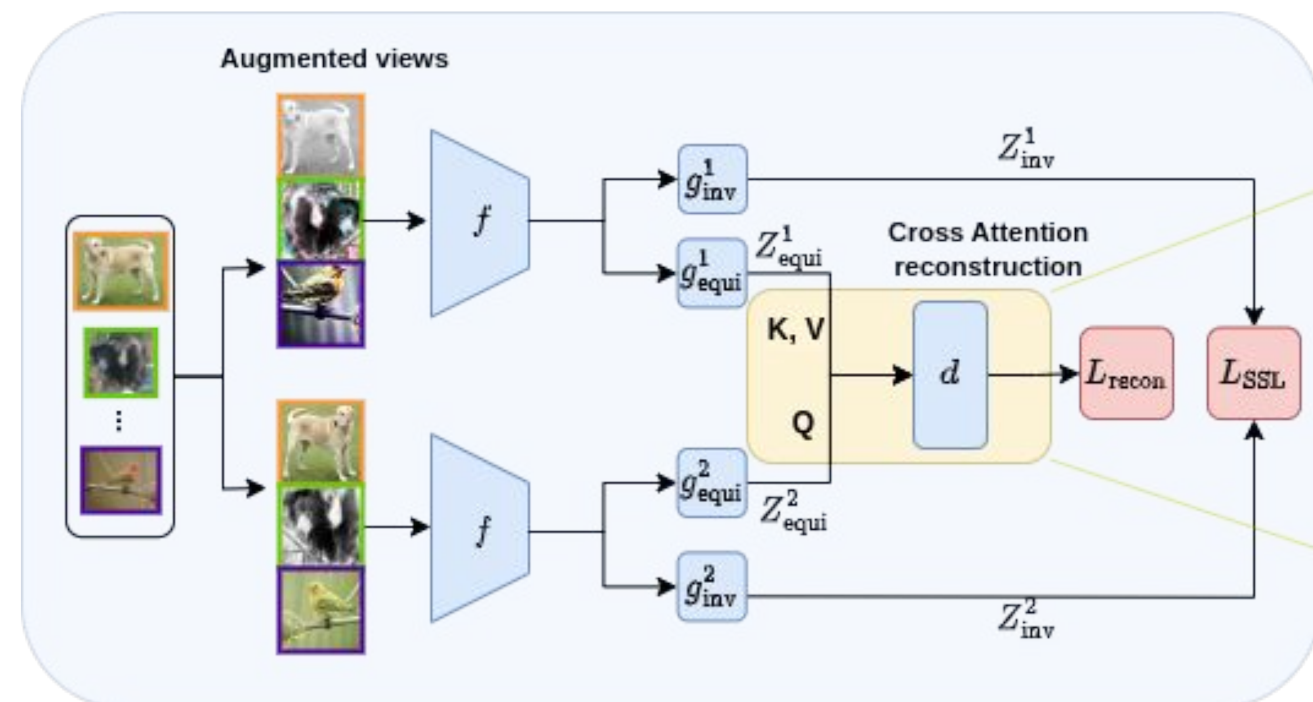
Current state-of-the-art method that introduce equivariance to SSL, SIE

- has been tested only on small, artificial datasets (3DIEBench), limiting its proven applicability to real-world scenarios.
- It requires prior knowledge of transformations to learn equivariant features, which may not always be available or easily determinable.
- It struggles when dealing with images that have undergone unknown transformations.

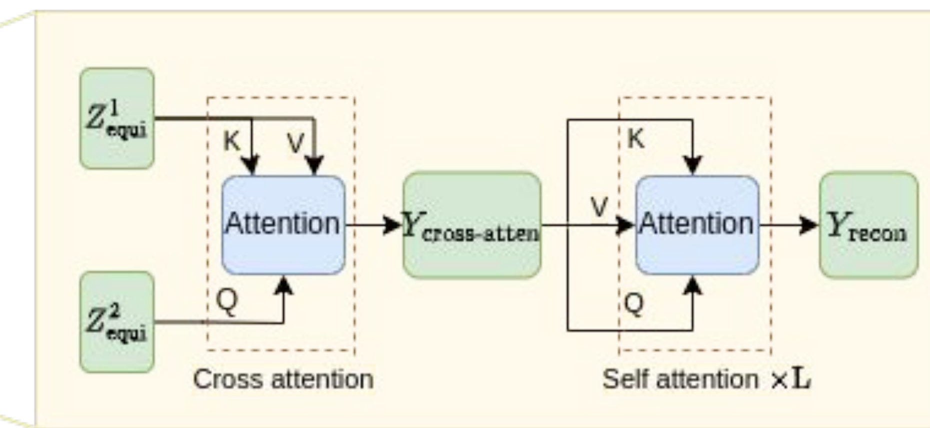
Contribution

- We introduce reconstruction as an auxilary task to learn equivariance, addressing the limitations of augmentation-based self-supervised learning.
- We demonstrate the effectiveness of our method on both artificial (3DIEBench) and natural (ImageNet) datasets, showing comparable (3DIEBench) and improved performance (ImageNet) compared to existing baselines.
- We provide extensive evaluations on various image transformations, including rotation, color jittering, translation, and scaling, demonstrating the robustness of our learned representations.

Method



Split Invariant and Equivariant Representations



$$Y_{\text{cross-atten}} = \mathbf{softmax}((W_Q Z^2_{\text{equi}})(W_K Z^1_{\text{equi}})^{\mathrm{T}})(W_V Z^1_{\text{equi}})$$

Cross-Attention Reconstruction

- The framework divides the representations extracted from the encoder into two parts: one invariant and the other equivariant. The invariant part uses augmentation-based SSL loss to encourage the network to learn invariant features.
- To facilitate the learning of equivariant features from the images, we introduce an auxiliary reconstruction task. The reconstruction is performed using a decoder, **d**, which consists of a cross-attention layer followed by **L** self-attention layers.

Experimental Results

Evaluation on 3DIEBench dataset.

3DIEBench	Classification	Rotation Prediction	Color Prediction
SIE(rot)	**0.820**	**0.724**	0.054
SIE(rot+color)	0.809	0.502	**0.980**
Ours	0.782	0.554	0.954

Our model strikes a balance, performing well across all tasks.

Evaluation on CIFAR10 dataset.

Cifar10	Rot Prediction	Color Prediction	Blur Radius	Trans Prediction
SIE(rot)	**0.989**	0.887	0.836	0.911
SIE(color)	0.813	**0.921**	0.825	0.822
SIE(blur)	0.814	0.833	**0.990**	0.807
SIE(trans)	0.876	0.812	0.810	**0.987**
SIE(all)	0.845	0.864	0.889	0.886
Ours	0.826	0.906	0.972	0.890

Evaluation on ImageNet.

ImageNet	Rotation	Color	Blur radius	Translation	Crop prediction	Flip
SIE(rot)	**0.990**	0.867	0.042	0.540	0.266	0.532
SIE(color)	0.078	**0.890**	0.097	0.355	0.178	0.333
SIE(blur)	0.153	0.883	**0.941**	0.189	0.412	0.415
SIE(trans)	0.213	0.885	0.023	**0.978**	0.368	0.511
SIE(crop)	0.273	0.819	0.018	0.450	**0.922**	0.485
SIE(flip)	0.155	0.798	0.056	0.312	0.266	**0.993**
VICReg[4]	0.318 ± 0.005	0.804 ± 0.016	0.101 ± 0.023	0.333 ± 0.008	0.423 ± 0.140	0.872 ± 0.070
SIE(all)	0.331 ± 0.007	0.899 ± 0.003	0.211 ± 0.005	0.925 ± 0.002	0.835 ± 0.008	0.945 ± 0.004
SIE(all, single each time)	0.435 ± 0.011	0.907 ± 0.009	0.377 ± 0.004	0.922 ± 0.010	0.829 ± 0.005	0.939 ± 0.007
Ours	0.862 ± 0.004	**0.921 ± 0.006**	0.823 ± 0.003	0.853 ± 0.005	0.912 ± 0.002	0.952 ± 0.008

- SIE models excel when pretrained with specific single transformations; their performance drops significantly for other transformations.
- Pretraining with randomly selected transformations (as in SIE (all, single each time)) improves results compared to SIE(all).
- Our models can handle unknown transformations better than SIE.

Transfer Learning Results

Transfer learning on classification and segmentation downstream tasks.

Methods	Cifar10 [13]	Cifar100 [13]	Food101 [14]	SUN397 [15]	DTD [16]	Pets [17]	Aircraft [18]
SIE(rot)	71.56	46.88	55.48	43.11	64.22	81.51	**50.21**
SIE(color)	67.99	48.78	57.19	42.32	60.87	80.27	41.15
SIE(crop)	80.84	49.35	59.24	52.38	61.82	84.63	47.35
Supervised	80.99	50.66	59.32	52.98	62.03	83.59	47.83
SIE(all)	79.91±0.18	53.12±0.05	58.42±0.20	56.11±0.08	63.56±0.11	**85.34±0.19**	46.88±0.23
Ours	**81.12±0.11**	**54.22±0.10**	**59.21±0.14**	**59.53±0.13**	**67.66±0.12**	84.32±0.09	49.75±0.22

ADE20K	mIOU	mAcc	aAcc
Supervised	0.268	0.328	0.751
SIE(all)	0.292	0.356	0.774
Ours	**0.312**	**0.379**	**0.802**

Utilisation of unknown transformations for learning equivariant representations

Cifar10	Rotation	Color	Blur Radius	Translation
Supervised	0.214	0.229	0.437	0.386
SIE(all)	0.402	0.395	0.511	0.479
Ours	**0.815**	**0.879**	**0.944**	**0.878**

- With the CIFAR10 dataset, we denote 80% of the training data as data subject to unknown transformations, and for 20% the transformations, including their parameters, are known.
- SIE as well as supervised are trained exclusively on the remaining 20% data with known transformations, whereas our method can leverage the entire dataset.

Acknowledgement

The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS[2] at Jülich Supercomputing Centre (JSC).

Reference List

[1] Maurice Weiler, Patrick Forr´e, Erik Verlinde, and Max Welling, Equivariant and Coordinate Inde-pendent Convolutional Networks, 2023.

[2] Jülich Supercomputing Centre. (2021). JUWELS Cluster and Booster: Exascale Pathfinder with Modular Supercomputing Architecture at Juelich Supercomputing Centre. *Journal of large-scale research facilities, 7,* A183. http://dx.doi.org/10.17815/jlsrf-7-183