Distribution Estimation to Automate Transformation Policies for Self-Supervision

Seunghan Yang[1*], Debasmit Das[1*], Simyung Chang[1], Sungrack Yun[1], Fatih Porikli[1]
[1] Qualcomm AI Research
{seunghan, debadas, simychan, sungrack, fporikli}@qti.qualcomm.com

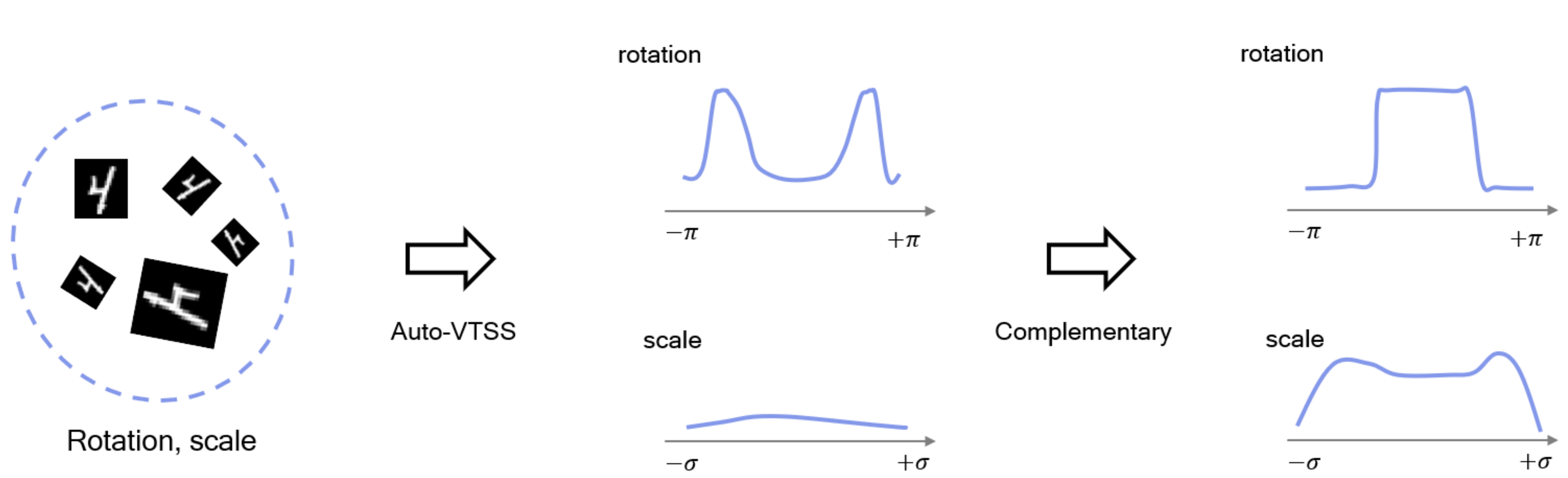
Which transformation would be helpful in self-supervised learning?

- In recent visual self-supervision works, an imitated classification objective, called pretext task, is established by assigning labels to transformed or augmented input images
- There are many papers with various pretext tasks involving transformations: rotation, shift, color, …
- However, we don't know which transformations are effective for a target task and dataset
- VTSS hypothesis [1]: **If the predicted instantiations of the transformations are already present in the dataset, the learned representation will be less useful**
- They consider only a manual way to find the useful transformation (difficult to do for every dataset and task)

Our contributions

- We introduce a learning framework based on a generative adversarial network to **automatically obtain the distribution of transformations present in a dataset**
- We construct the distribution complementary to the estimated distribution
- Finally, we generate **a useful pretext task for self-supervised learning with the transformation instances sampled from the complementary distribution**
- We show comprehensive study, analysis, and comparison of the representations learned from our transformation instances on multiple visual recognition benchmarks

Our concept diagram

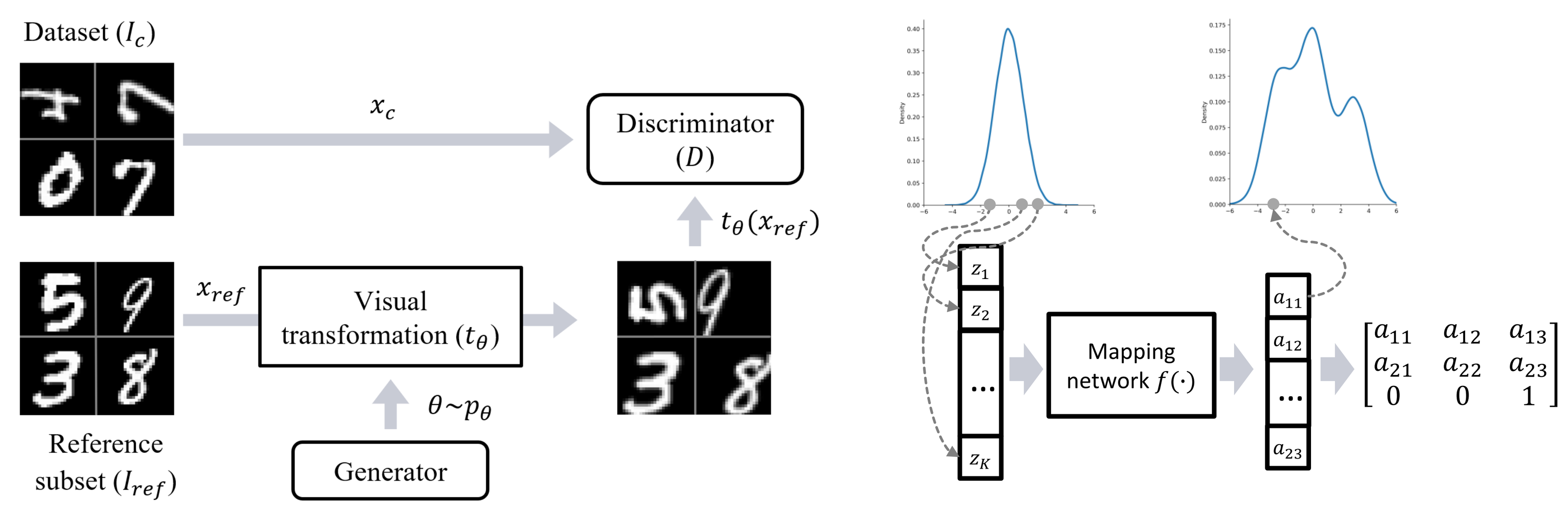


Automatically estimate the transformations present in a dataset, then use this information to obtain 'a complementary set'

Step I: Distribution estimation for visual transformations

Learning framework to estimate visual transformations in a dataset

1) We choose a reference subset that includes the most representative and frequent data (e.g., upright images)
2) Visual transformation parameters are sampled from generator, and a reference subset is transformed by the parameters
3) Samples from current dataset and transformed samples are inputs of discriminator, and through GAN-based loss functions two visual transformations become similar
4) Mapping network is trained to estimate the distribution of the current visual transformation



Our framework to estimate the visual transformation in a dataset / Generator producing visual transformation parameters through the mapping network that projects a known distribution to the desired distribution

Step II: Automating transformation policies

A complementary set of current visual transformations

1) To get the distribution, we use a sampling procedure, which we feed random vectors from the known distribution to the mapping network
2) All outputs are aggregated to from a histogram values, and we can obtain its complementary by subtracting from the peak value and normalizing it

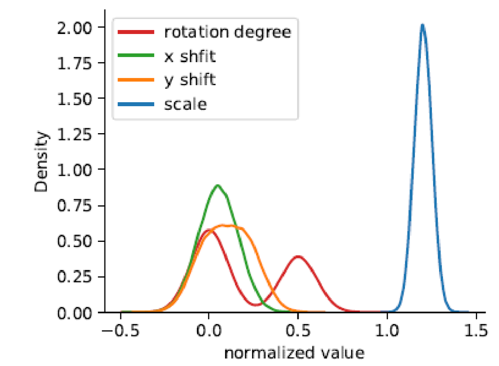
Transformation policies for self-supervision tasks

- In the manual policy, we sample the transformation instance from the parameters ranges where estimated distribution is 0 or low value
- In the automated policy, we exploit the inverse transform sampling method. We obtain the transformation instances by mapping samples from the uniform distribution to the inverse of CDF

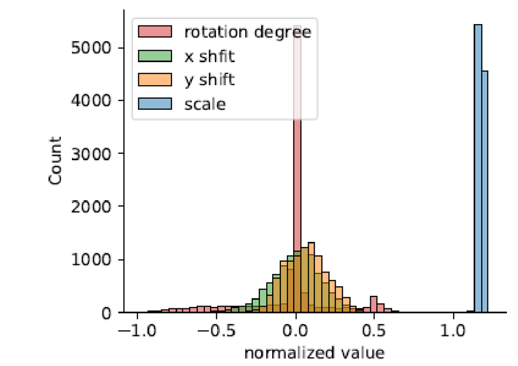
Experimental results

Distribution estimation and effect on downstream tasks

- Evaluation on a transformed MNIST dataset



Ground-truth distribution

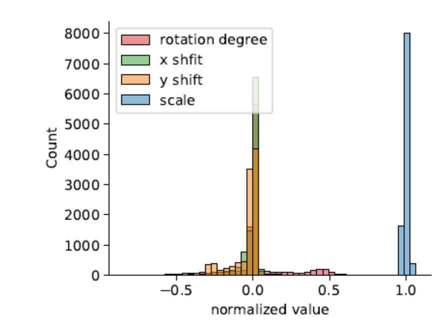


Estimated distribution

Task	Sup acc.
R: 0, 90	92.28
R: 0, 180	96.04
T: 0.1, 0.1	11.35
T: 0.5, 0.5	47.97
S: x1, x1.25, x1.5	33.57
S: x0.8, x1, x1.5	66.57
RST	97.37

Performance on downstream task with various pretext tasks

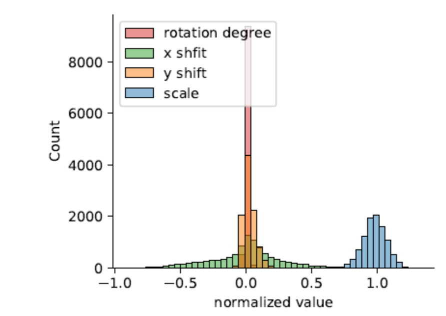
- Evaluation on the Fashion MNIST and SVHN dataset



Estimated distribution on Fashion MNIST

Task	Sup acc.
R: 0, 90, 180, 270	91.97
T: 0.2, 0.2	82.32
S: x1, x1.14, x1.33	79.03
RST	93.22

Performance on downstream task

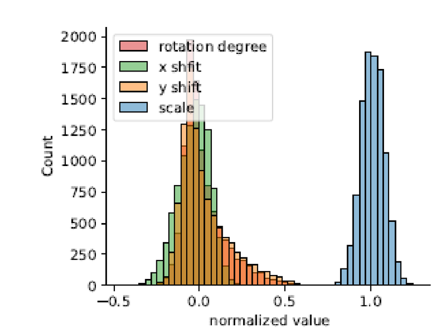


Estimated distribution on SVHN

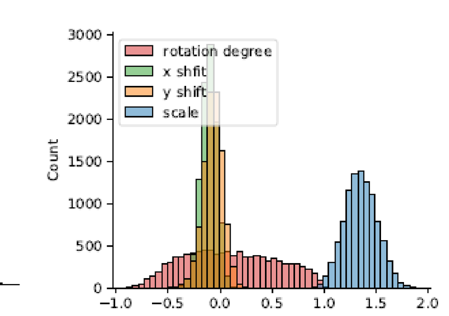
Task	Sup acc.
S: x1, x1.14, x1.33	54.30
S: x1, x1.23, x1.6	64.95
S: x1, x1.5, x2	67.03
T: 0.1, 0.1	92.34
T: 0.5, 0.5	93.63

Performance on downstream task

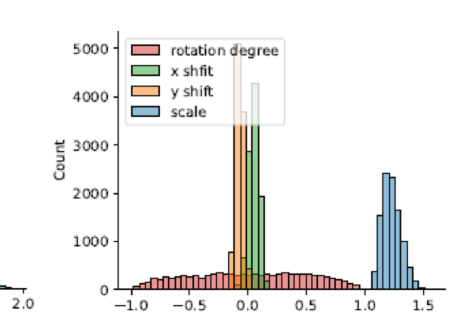
- More results of distribution estimation



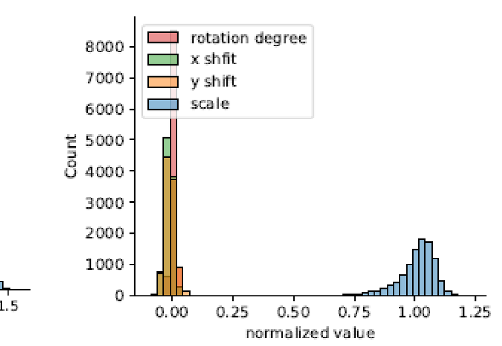
Estimated distribution on CropDisease



Estimated distribution on EuroSAT



Estimated distribution on ISIC2018



Estimated distribution on ChestX

- Extension to contrastive loss

Finetuning the model with both cross-entropy loss and supervised contrastive loss

We apply the proposed automated transformation policy to generating augmented versions used for contrastive loss

Method	ChestX	ISIC	EuroSAT	CropDiseases
Baseline	41.16	67.00	96.35	96.43
+ SimCLR with Rot.	38.24	68.43	97.57	97.48
+ SimCLR with Aff.	40.65	68.97	97.66	97.41
+ SimCLR with ATP	42.33	72.97	97.72	97.60

Effect of fine-tuning on cross-domain benchmark datasets. Rot., Aff., and ATP indicate random rotation, random affine transformation, and automated transformation policies, respectively

References

[1] Dipan K Pal, Sreena Nallamothu, and Marios Savvides (2020). "Towards a hypothesis on visual transformation based self-supervision". In: BMVC.