Self-supervised Contrastive Learning for Irrigation Detection in Satellite Imagery

Chitra S. Agastya [* 1 2] **Sirak Ghebremusse** [* 1] **Ian Anderson** [* 1] **Colorado Reed** [1] **Hossein Vahabi** [1]
Alberto Todeschini [1]

Abstract

Climate change has caused reductions in river runoffs and aquifer recharge resulting in an increasingly unsustainable crop water demand from reduced freshwater availability. Achieving food security while deploying water in a sustainable manner will continue to be a major challenge necessitating careful monitoring and tracking of agricultural water usage. Historically, monitoring water usage has been a slow and expensive manual process with many imperfections and abuses. Machine learning and remote sensing developments have increased the ability to automatically monitor irrigation patterns, but existing techniques often require curated and labelled irrigation data, which are expensive and time consuming to obtain and may not exist for impactful areas such as developing countries. In this paper, we explore an end-to-end real world application of irrigation detection with uncurated and unlabeled satellite imagery. We apply state-of-the-art self-supervised deep learning techniques to optical remote sensing data, and find that we are able to detect irrigation with up to nine times better precision, 90% better recall and 40% more generalization ability than the traditional supervised learning methods.

1. Introduction

Water is essential for global food production and security, agriculture being the largest consumer of freshwater globally (WWDR, 2014). While we need water to grow crops to address the growing demand for food, excessive use of water has negatively impacted ecosystem and adversely affected crop production and livelihood. Ground truth data on agricultural water demand, traditionally gathered from local surveys, are often inadequate and do not capture the temporal signature and dynamics of irrigation patterns (Graf, 2020). Irrigated land is difficult to track in underdeveloped countries due to lack of infrastructure and government policies around water usage.

Detecting irrigation is critical to understand water usage and promote better water management. Such data will potentially enable the study of climate change impact on agricultural water sources, monitor water usage, help detect water theft and illegal agriculture and inform policy decisions and regulations related to water compliance and management. This is a particularly hard problem to solve due to the lack of curated and labelled data available that are centered around irrigation systems. BigEarthNet-S2, for example, is a small fraction of the 26 TB of data produced daily by the European Space Agency, and required several million dollar grants to curate and label [3].

The small sample of labelled datasets currently available are from developed countries and supervised learning from these may not generalize to developing and under developed countries. In this report we investigate the problem of irrigation detection using self-supervised learning on high resolution multi-spectral satellite images that were collected for monitoring of land, vegetation, soil and water cover. We research whether pre-training with contrastive self-supervised learning from uncurated satellite images of one geography, followed by fine-tuning with a small fraction of publicly available labelled data from a different geography, has the generalization ability to both detect irrigated land from the first geography and also from entirely different geographical regions.

2. Methodology

We use the following machine learning pipeline to build an irrigation detection system: (1) We pre-train a deep neural network using a large amount of unlabeled data from a geographic area of interest, i.e. where we would like to do

---
[*]Equal contribution [1]School of Information, University of California, Berkeley, USA [2]IBM Chief Analytics Office, Armonk, New York, USA. Correspondence to: Chitra S. Agastya <chitra.agastya@berkeley.edu, Chitra.S.Agastya@ibm.com>, Sirak Ghebremusse <sirakzg@berkeley.edu>, Ian Anderson <imander@berkeley.edu>.

irrigation detection. (2) We fine-tune the network using publicly available labeled satellite image datasets that contain irrigation labels. (3) We apply the irrigation detection system to the geographic area of interest used for pre-training in (1.). For (1.) we use the SimCLR framework(Chen et al., 2020a;b) because it consistently outperformed previous methods for self-supervised and semi-supervised learning on ImageNet, without requiring specialized architectures.

3. Data

The data for our research are Sentinel2 surface reflectance images collected from three sources. We use Sentinel2 archive of level2A images of the Central Valley region of California as our unlabelled source for self-supervised pre-training. Central Valley is one of the most productive agricultural regions in the world [1]. The Central Valley is home to more than 7 million acres of permanently irrigated land and houses many major cities such as Sacramento and Fresno, providing a diverse mix of irrigated and non-irrigated land for our study [2].

BigEarthNet-S2 (Sumbul et al., 2019), a large scale benchmark archive for remote sensing image understanding from ten European countries, is used in our experiments for supervised training and evaluation. For evaluating model generalization, we use Sentinel2 images of ground truth coordinates spanning six different countries from croplands.org(Zhong et al., 2017; Oliphant et al., 2017; Gumma et al., 2017; Xiong et al., 2017).

4. Experiments

We established our baseline using traditional supervised learning approaches on the BigEarthNet-S2 data. Our target label is a binary variable indicating whether the land in the image is permanently irrigated or not. We conducted several experiments by varying different parameters: (i) size of the CNN architecture, (ii) training from scratch versus using pretrained ImageNet weights (iii) size of the training data. We use varying training data sizes of 190, 570, 1902, 4756, 9515 and 19024 records balanced between the two target classes. We also refer to these as data split percentages of 1, 3, 10, 25, 50 and 100 respectively.

4.1. SimCLR-S2 Model

We call our adaptation of the SimCLR framework to Sentinel2 images, *SimCLR-S2*. While the original SimCLR implementation uses only the 3 visible RGB channels, we have access to 10 channels (bands) in our data. Some of these, like infrared bands, provide information about the presence of moisture due to vegetation. We therefore consider using all 10 bands throughout training and inference.
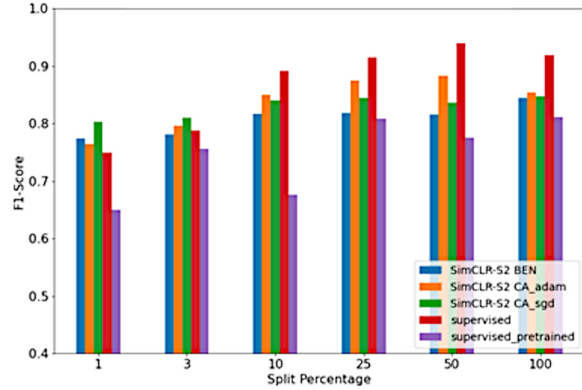

Figure 1. A comparison of stage-2 fine-tune results of SimCLR-S2 with supervised baseline for the different data sizes (given by split percentage with 1% having 190 training records). SimCLR-S2 models (i) consistently outperform supervised baseline scores pretrained with ImageNet weights and (ii) outperform baselines scores when trained from scratch on smaller sizes of annotated data (1% and 3% splits)

The SimCLR-S2 framework has three stages. The first stage takes an unlabeled dataset and creates a mini batch of images. For each image in the mini batch, we perform two image augmentations, picked at random from a list of transformation pairs, and compute the contrastive loss between any two images in the mini batch with an objective to minimize the contrastive loss across the entire dataset. In the second stage, the pretrained self-supervised model is fine tuned with a small labelled dataset for the task of identifying irrigated land versus non-irrigated land. In the third stage, the fine tuned model is then used as a teacher in a knowledge distillation process to teach labels for an unlabeled dataset on a student network of the same size or smaller. The student network bootstraps from the teacher network, learning nuances from the teacher, resulting in performance equal to or better than the teacher model, even on smaller networks.

Since BigEarthNet is our only source of annotated data, we use 3% of its labeled data for evaluation. We performed several experiments with variations in: data used for unsupervised pretraining; data used in supervised fine-tuning; and CNN architectures used in self-training distillation stages. We then compare the results for these experiments with the supervised baseline results.

4.1.1. HYPER-PARAMETER TUNING

We evaluated training SimCLR-S2 with the SGD and Adam optimizers. With SGD optimization with cosine decay, the contrastive loss did not decrease throughout training,

while with the Adam optimizer, we observed decreasing contrastive loss curves. As a result, we selected the Adam optimizer for our pipeline. A learning rate of 0.0005 had the best performance with the Adam optimizer.

5. Evaluation Methods and Results

We conducted two types of evaluations to compare the performance of SimCLR-S2 model to that of the supervised baseline model. We evaluated the F-1 scores of predictions from both models on held-out data from BigEarthNet-S2 dataset. This test data was not a part of the training process. However, because both the test and training data are from the same region i.e. Europe, we evaluate the generalization ability of the self-supervised vs supervised models by comparing: (i) the precision scores on an unlabelled set of Sentinel2 images (ii) the recall scores on Sentinel2 images for known irrigated land co-ordinates from different countries.

5.1. Results on BigEarthNet-S2 hold out data

The results from our fine-tuning experiments showed that the SimCLR-S2 model outperformed the supervised baseline on smaller data sizes. This is evidenced in figure 1, where for 1% of data splits, the F-1 scores were better or on par with that of the supervised learning. Our SimCLR-S2 model consistently outperformed the supervised model that was pretrained with ImageNet weights.
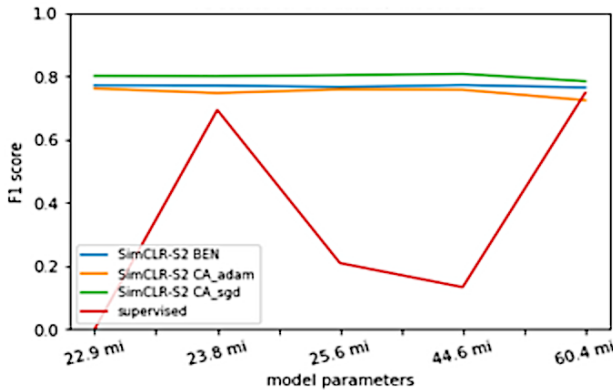
Xception, with 22.9 million parameters, saw an improved F-1 score for data splits from 3% and above, and with larger models accuracy over the fine-tuned teacher model increased across the board. We suspect that the freezing of the convolutional layers during the fine-tuning process resulted in their poorer performance compared to the student models, which had all layers enabled for distill learning. Distillation scores were better than those of supervised baselines for smaller data sizes on many architectures. As evidenced by Figure 2, for the 1% data split SimCLR-S2 outperformed the supervised baseline across the board.

5.2. Results on generalizability tests

We ran two case studies to check for generalizability of our SimCLR-S2 model. In one, we tested for model precision on an unlabelled test dataset and in the second one, we tested for model recall on a labelled test dataset. To demonstrate that the model generalizes well, we used data from diverse geographies different from the geography that the model was trained on.

5.2.1. PRECISION

For precision, we evaluated Sentinel2 images of unlabelled data from California, USA on our SimCLR-S2 fine-tuned model. We took the top 100 images that our model predicted as irrigated with at least 99% confidence. We then visually inspected these images to score the precision. We crowdsourced the visual inspection using Amazon Mechanical Turk and did an additional visual verification of the
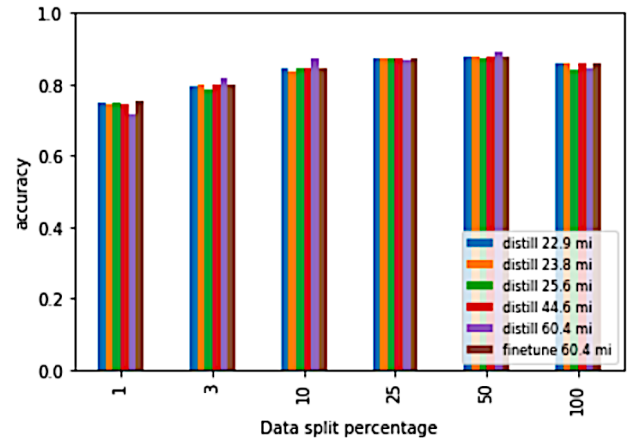

Figure 2. A comparison of stage-3 distillation results of SimCLR-S2 with supervised baseline for different architectures (in number of million parameters) for training data size of 190 records. SimCLR-S2 models outperform supervised baseline scores across architectures.

We took each ResNet152 model fine-tuned across the various data splits and performed distillation learning on five model architectures. We found that even the smallest model

Figure 3. A comparison of accuracy with and without distillation across different data sizes for different architectures (in number of million parameters).Most distilled models perform on par or better that a simple finetuned model (i.e. without distillation) across data sizes even on smaller architectures.

Table 1. A comparison of precision scores from SimCLR-S2 and supervised baseline on unseen data from different geography.

Training data size (num records)	Precision (SimCLR-S2)	Precision (supervised)
190	**0.99**	0.11
570	**1.00**	0.2
1902	**0.99**	0.36
4756	**0.98**	0.95
9515	**1.00**	0.78
19024	**1.00**	0.47

scores to ensure that the survey responses aligned with our expectations.

To compare these scores with our supervised baseline, we performed similar evaluation and visual inspection with our supervised baseline model. An interesting observation we made was that prediction confidence of the supervised baseline was far lower than that of the SimCLR-S2 model. While we could determine the top 100 predictions with a minimum confidence threshold of 99% with SimCLR-S2 models, with the supervised models we had to drop the threshold to 50% to obtain the top 100 predictions. *Table 1* shows the median precision scores. SimCLR-S2 model consistently outperformed supervised baseline in all the data splits indicating that our generalized well on unseen data (see Figure 5 in the supplementary material).

5.2.2. RECALL

For recall, we chose to evaluate our models' prediction against global crowdsourced ground truth data from croplands.org. We sampled 100 irrigated cropland coordinates spanning 6 different geographies: Brazil, India, Indonesia, Myanmar, Tunisia and Vietnam. Since the ground truth data is not verified at source, before processing, we visually inspected the raw images to ensure that they looked like irrigated cropland. We then compared our model predictions to

Table 2. A comparison of recall scores on SimCLR-S2 and supervised baseline for irrigated cropland from diverse geographies

Country	Training data (num records)	Recall (SimCLR-S2)	Recall (supervised)
Brazil	190	**0.75**	0.5
India	190	**0.9**	0.67
Indonesia	570	**0.76**	0.07
Tunisia	570	0.78	**0.91**
Vietnam, Myanmar	190	**0.9**	0.00

the original target label to compute recall.

Our results demonstrated that SimCLR-S2 model generalized well across geographies. A model trained with just a fraction[1] of the overall annotated data for the task at hand, was able to outperform the supervised baseline model, in most cases, as evidenced by the recall scores in *table 2*.

6. Conclusion

SimCLR-S2 is successful in detecting irrigation in multi-spectral images, as evidenced by our results. Through our experiments, we showed that SimCLR-S2 can be used on satellite imagery consisting of several more channels than in typical computer vision applications; and performs better than supervised models in certain scenarios. Our results demonstrated that the SimCLR paradigm consistently outperformed supervised learning using significantly smaller sizes of labelled data, and that these improvements can be distilled into smaller models with fewer parameters. We also show that the SimCLR-S2 model generalizes well across diverse geographies. Where manual annotation is expensive and time consuming, we showed that real-world, un-curated classification tasks of satellite images benefit from contrastive self-supervised learning to perform image classification using a significantly smaller fraction of labelled images, while still achieving better results than supervised learning methods.

We identify a few areas for future work. While we performed all training with a fixed number of training epochs (50), previous works done by Chen et al(2020b) indicated they trained on as many as 800 epochs for unsupervised pre-training. We also acknowledge that during our investigation into the most effective augmentations to pair for training, a larger subset of the training data, and more epochs, may have been beneficial. The SimCLR-S2 model, with these improvements, could pave the path for a new state of the art in detecting and tracking global irrigation data using high resolution satellite imagery.

References

California Central Valley, American Museum of Natural History. URL https://www.amnh.org/learn-teach/curriculum-collections/grace/grace-tracking-water-from-space/california-central-valley.

Central Valley, U.S. Geological Survey. Geographic

---

[1]BigEarthNet-S2 has over 575,000 images, 13,589 of which are irrigated cropland. With SimCLR-S2, the smallest training data only uses 0.033% of overall images and 0.69% of irrigated images for supervised fine-tuning.

Names Post Phase I Map Revisions. Various editions. 01-Jan-2000., 1990. URL https://geonames.usgs.gov/apex/f?p=GNISPQ:3:::NO::P3_FID:252254.

Three ERC grants for the TU Berlin, 2018. URL https://www.pressestelle.tu-berlin.de/menue/tub_medien/publikationen/medieninformationen/2018/mai_2018/medieninformation_nr_732018/.

Bachman, P., Devon Hjelm, R., and Buchwalter, W. Learning Representations by Maximizing Mutual Information Across Views. *arXiv e-prints*, art. arXiv:1906.00910, June 2019.

Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. A Simple Framework for Contrastive Learning of Visual Representations. *arXiv e-prints*, art. arXiv:2002.05709, February 2020a.

Chen, T., Kornblith, S., Swersky, K., Norouzi, M., and Hinton, G. Big Self-Supervised Models are Strong Semi-Supervised Learners. *arXiv e-prints*, art. arXiv:2006.10029, June 2020b.

Goyal, P., Caron, M., Lefaudeux, B., Xu, M., Wang, P., Pai, V., Singh, M., Liptchinsky, V., Misra, I., Joulin, A., and Bojanowski, P. Self-supervised Pretraining of Visual Features in the Wild. *arXiv e-prints*, art. arXiv:2103.01988, March 2021.

Graf, L. Mapping and monitoring irrigated agriculture from space. Technical report, United Nations Office of Outer Space Affairs, 2020.

Gumma, M., Thenkabail, P., Teluguntla, P., Oliphant, A.J., X., J., Congalton, R. G., Yadav, K., Phalke, A., and Smith, C. Nasa making earth system data records for use in research environments (measures) global food security-support analysis data (gfsad) @ 30-m for south asia, afghanistan and iran: Cropland extent product (gfsad30saafgirce), 2017. URL https://doi.org/10.5067/MEaSUREs/GFSAD/GFSAD30SAAFGIRCE.001.

He, K., Fan, H., Wu, Y., Xie, S., and Girshick, R. Momentum Contrast for Unsupervised Visual Representation Learning. *arXiv e-prints*, art. arXiv:1911.05722, November 2019.

Hénaff, O. J., Srinivas, A., De Fauw, J., Razavi, A., Doersch, C., Eslami, S. M. A., and van den Oord, A. Data-Efficient Image Recognition with Contrastive Predictive Coding. *arXiv e-prints*, art. arXiv:1905.09272, May 2019.

Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., and Krishnan, D. Supervised Contrastive Learning. *arXiv e-prints*, art. arXiv:2004.11362, April 2020.

Oliphant, A., Thenkabail, P., Teluguntla, P., Xiong, J., Congalton, R., Yadav, K., Massey, R., Gumma, M., and Smith, C. Nasa making earth system data records for use in research environments (measures) global food security-support analysis data (gfsad) @ 30-m for southeast & northeast asia: Cropland extent product (gfsad30seace), 2017. URL https://doi.org/10.5067/MEaSUREs/GFSAD/GFSAD30SEACE.001.

Reed, C. J., Metzger, S., Srinivas, A., Darrell, T., and Keutzer, K. Selfaugment: Automatic augmentation policies for self-supervised learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2021a.

Reed, C. J., Yue, X., Nrusimha, A., Ebrahimi, S., Vijaykumar, V., Mao, R., Li, B., Zhang, S., Guillory, D., Metzger, S., et al. Self-supervised pretraining improves self-supervised pretraining. *arXiv preprint arXiv:2103.12718*, 2021b.

Sumbul, G., Charfuelan, M., Demir, B., and Markl, V. Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, Jul 2019. doi: 10.1109/igarss.2019.8900532. URL http://dx.doi.org/10.1109/IGARSS.2019.8900532.

van den Oord, A., Li, Y., and Vinyals, O. Representation Learning with Contrastive Predictive Coding. *arXiv e-prints*, art. arXiv:1807.03748, July 2018.

WWDR. The united nations world water development report 2014: water and energy; facts and figures, 2014. URL https://unesdoc.unesco.org/ark:/48223/pf0000226961.

Xiong, J., Thenkabail, P. S., Tilton, J., Gumma, M. K., Teluguntla, P., Congalton, R. G., Yadav, K., Dungan, J., Oliphant, A. J., Poehnelt, J., Smith, C., and Massey, R. Nasa making earth system data records for use in research environments (measures) global food security-support analysis data (gfsad) @ 30-m africa: Cropland extent product (gfsad30afce), 2017. URL https://doi.org/10.5067/MEaSUREs/GFSAD/GFSAD30AFCE.001.

Zhong, Y., Giri, C., Thenkabail, P., Teluguntla, P., Congalton, R. G., Yadav, K., Oliphant, A. J., Xiong, J., Poehnelt, J., , and Smith, C. Nasa making earth system data records for use in research environments (measures) global food security-support analysis data (gfsad)

@ 30-m for south america: Cropland extent product (gfsad30sace), 2017. URL https://doi.org/10.5067/MEaSUREs/GFSAD/GFSAD30SACE.001.

Supplementary Material
Self-supervised Contrastive Learning for Irrigation Detection in Satellite Imagery

Chitra S. Agastya [* 1 2]   **Sirak Ghebremusse** [* 1]   **Ian Anderson** [* 1]   **Colorado Reed** [1]   **Hossein Vahabi** [1]
Alberto Todeschini [1]

1. Background

Vision models have traditionally been trained using labelled datasets, and the manual process of annotating data does not scale well. It is time consuming, expensive, often requiring expert knowledge and can be subjective to a user, making it a potential bottleneck to successfully training models. In this digital age, unlabelled data is ubiquitous, however unsupervised models do not perform as well as supervised models. Learning effective visual representations without human supervision has been a persistent problem. Discriminative approaches based on contrastive learning have recently shown great promise in achieving state-of-the art results (**?**), (**?**). These methods train networks to perform pretraining tasks where both inputs and labels are derived from unlabelled data.

2. Sentinel-2 Mission

Sentinel-2 (S2) is an Earth observation mission operated by the European Space Agency and acquires optical imagery at a high spatial resolution of 10 to 60 metres(m) over land and coastal waters. The S2 multispectral instrument includes 13 bands, which are captured at different spatial resolutions. S2 Super-resolution creates a 10 m resolution band for all the existing spectral bands (including those with 20 m and 60 m) using a trained convolutional neural network (CNN). This processing block's output is a multispectral 12 band, 10 m resolution GeoTIFF file. The first band i.e. B1 is discarded since it's only useful for atmospheric correction[1].

[*]Equal contribution   [1]School of Information, University of California, Berkeley, USA [2]IBM Chief Analytics Office, Armonk, New York, USA. Correspondence to: Chitra S. Agastya <chitra.agastya@berkeley.edu, Chitra.S.Agastya@ibm.com>, Ian Anderson <imander@berkeley.edu>, Sirak Ghebremusse <sirakzg@berkeley.edu>.

[1]https://up42.com/blog/tech/
sentinel-2-superresolution

3. Data Preprocessing

We download large swaths of satellite images measuring approximately 600 square kilometres in area. In all, we download images that cover roughly 50,000 square kilometres of area from Chicos to Bakersfield in Central Valley. We match the bands that we have from BigEarthNet, which excludes bands 1 and 10. Since each BigEarthNet image patch covers 1.44 square kilometres, we further split each of our downloaded images for all bands to 120 x 120 pixels.

We implemented our modeling pipeline using Tensorflow. Before feeding our data into our models, we processed the images to convert them into the TFrecord data format. We additionally calculated the mean and standard deviation for each band across the California dataset, and used the provided statistics for the BigEarthNet dataset. Applying this normalization step added the advantage of bringing each band's data to a similar range which allowed us to perform image augmentations across the whole image rather than having to apply unique or custom augmentations per channel. We performed further processing on the California dataset to avoid null values due to cloud cover and cloud shadows.
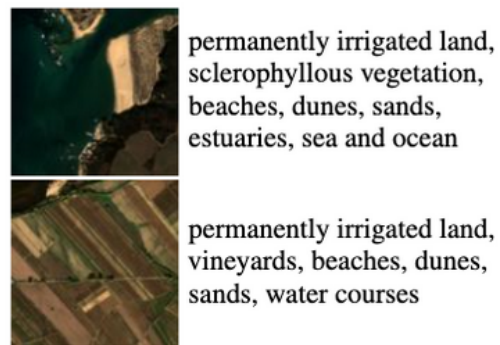

Figure 1. Sample of images and labels from BigEarthNet-S2 dataset

Table 1. Mapping of data split percentages to training records

SPLIT PERCENAGE	NUM. OF TRAINING RECORDS
1	190
3	570
10	1902
25	4756
50	9515
100	19024

4. Results

4.1. Additional Observations

A distribution of our performance scores across all of 336 working self-supervised models, showed that the SimCLR-S2 models brought the F-1, accuracy and AUC scores closer together across different data sizes and CNN architectures. The variability of these scores were much higher with supervised learning as evidenced in figure 2 in the appendix. All of our SimCLR-S2 models had an AUC score higher than 0.72, independent of training data or model sizes. The mean AUC score across all SimCLR-S2 models was around 0.83 indicating that SimCLR-S2 models were effective in detecting permanently irrigated land. Supervised learning on the other hand had models with AUC scores as low as 0.5. We also observed that the mean F-1 score between SimCLR-S2 and supervised models was the same.

A t-SNE heat map of our SimCLR-S2 pretraining model depicted in figure 14 showed clear indication of dense zones of clustering in both irrigated (image to the left) and non-irrigated classes (image to the right). We believe that training the models for longer than 50 epochs could potentially help with further separation of these clusters.
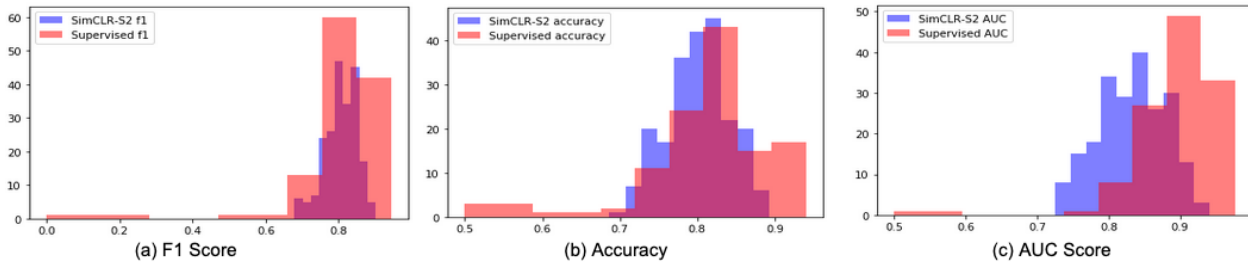
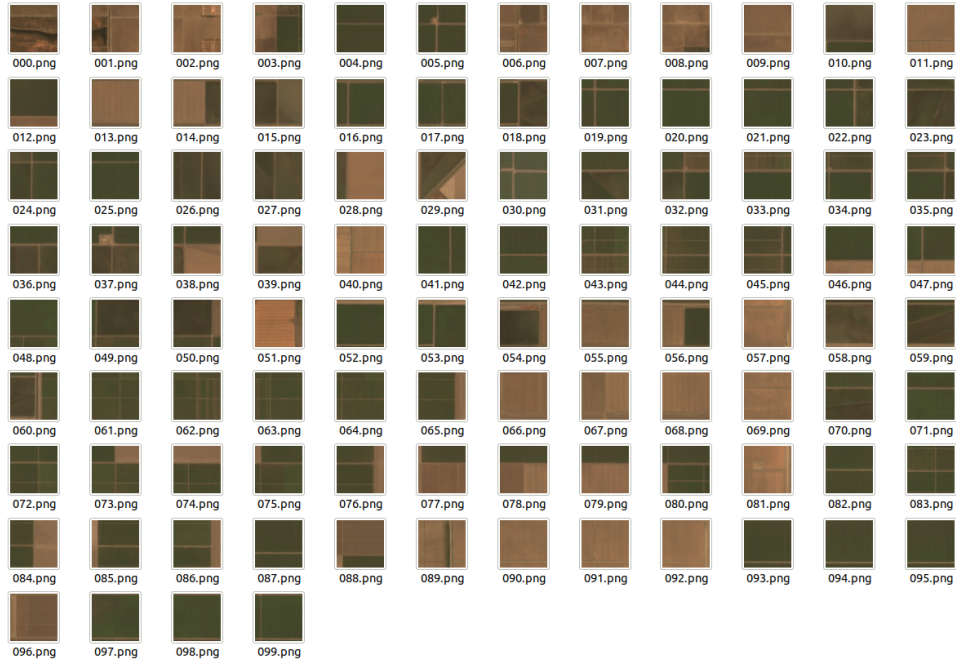*Figure 2.* Distribution of F-1, accuracy and AUC scores for the supervised and SimCLR-S2 experiments

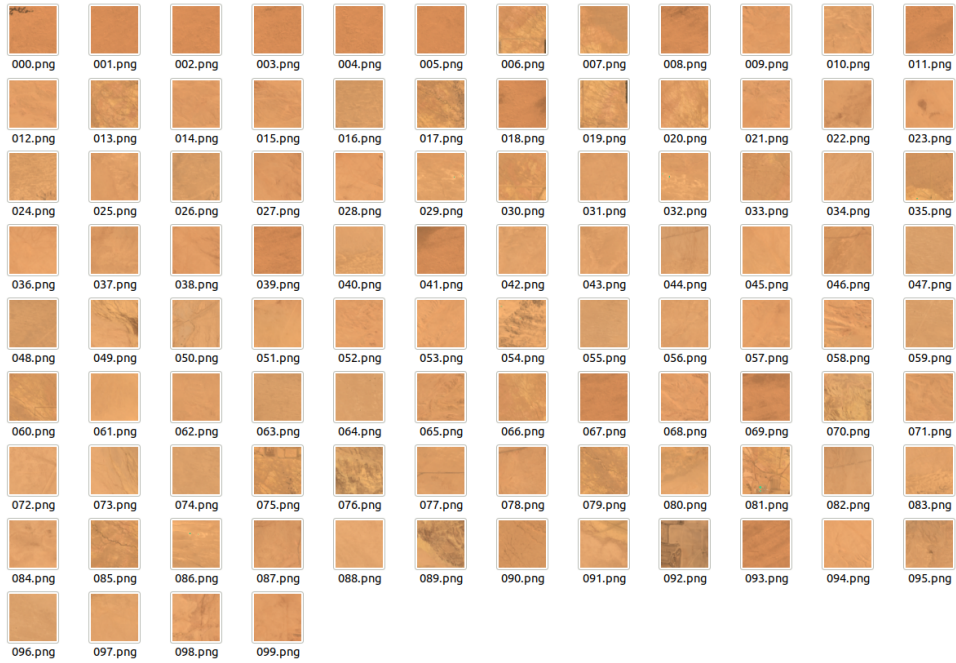*Figure 3.* Top 100 irrigated land predictions by SimCLR-S2 model on 1% data split

Figure 4. Top 100 irrigated land predictions by supervised baseline model on 1% data split

Figure 5. A comparison of top 100 irrigated land predictions by SimCLR-S2 vs supervised baseline on 1% training data size i.e. on 190 records. A visual inspection clearly shows that the SimCLR-S2 results outperform that of the supervised baseline
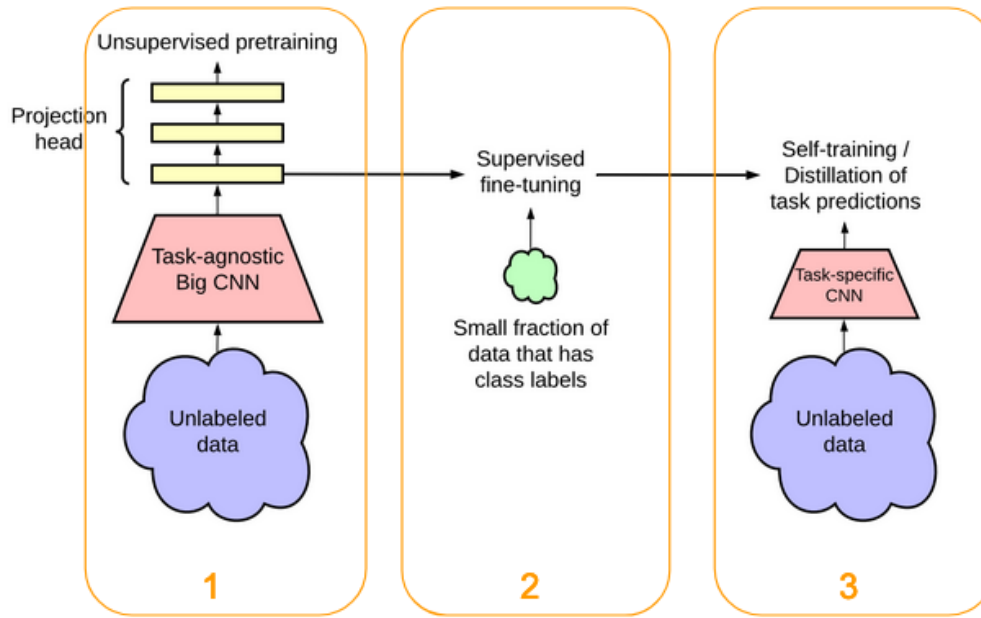
Figure 6. SimCLRv2 Framework showing the use of labelled and unlabelled data. SimCLR-S2 is implemented using this framework

Table 2. Data strategy for SimCLR-S2 experiments

		Central Valley, California	BigEarthNet-S2
General Info	Year	2019-2020	2017-2018
	Image Size	120 x 120 x 10	120 x 120 x 10
	Source	sourced using Google Earth Engine API	http://bigearth.net/
Pre-Training	Data Size	50,194	575,579
	Labelled	No	No
Fine Tuning/ Supervised Learning	Data Size	n/a	Varying (230 to 23100)
	Labelled	n/a	Yes
Evaluation	Data Size	n/a	14,747
	Labelled	n/a	Yes
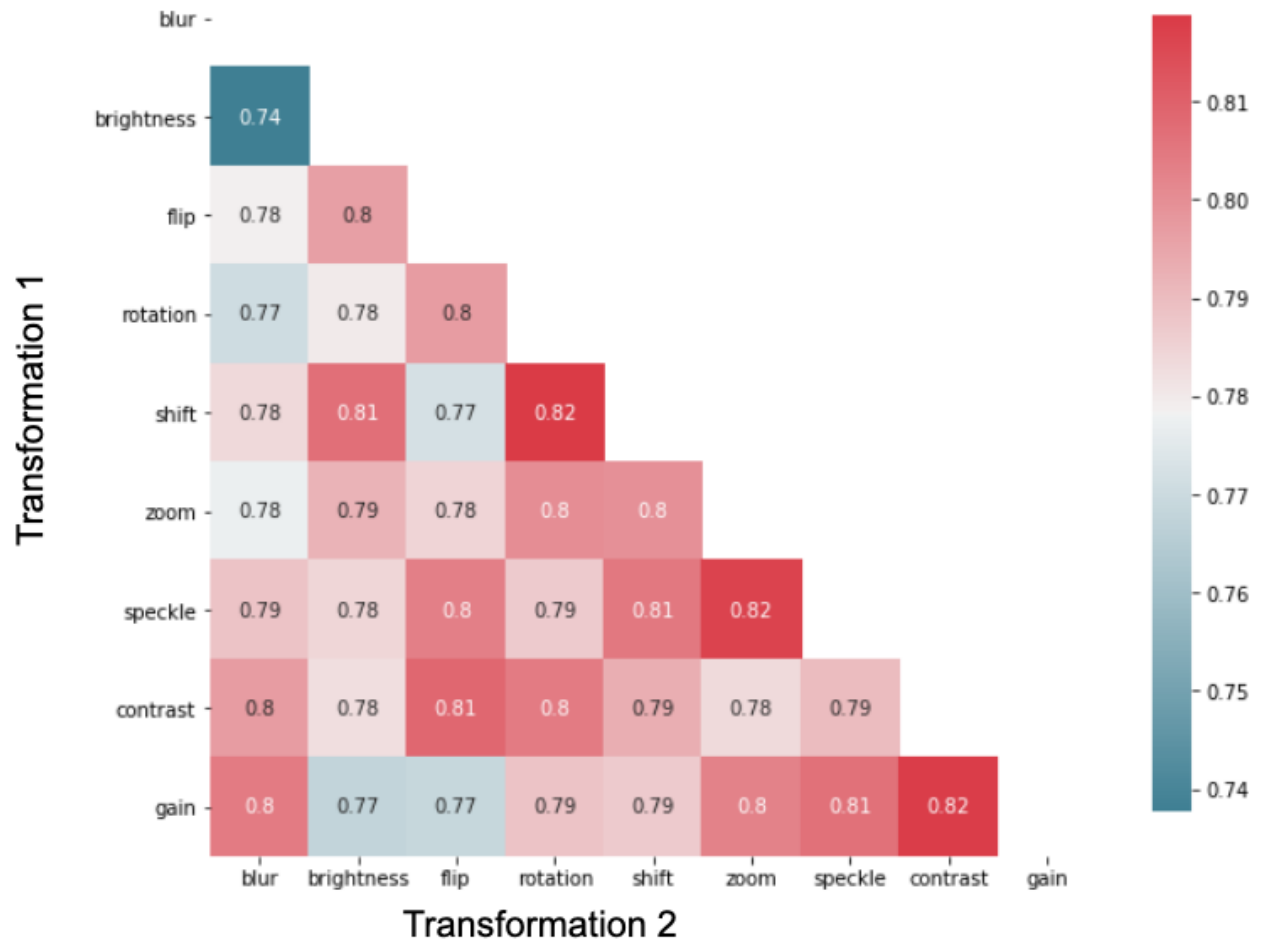
Figure 7. F-1 score for transformation pairs
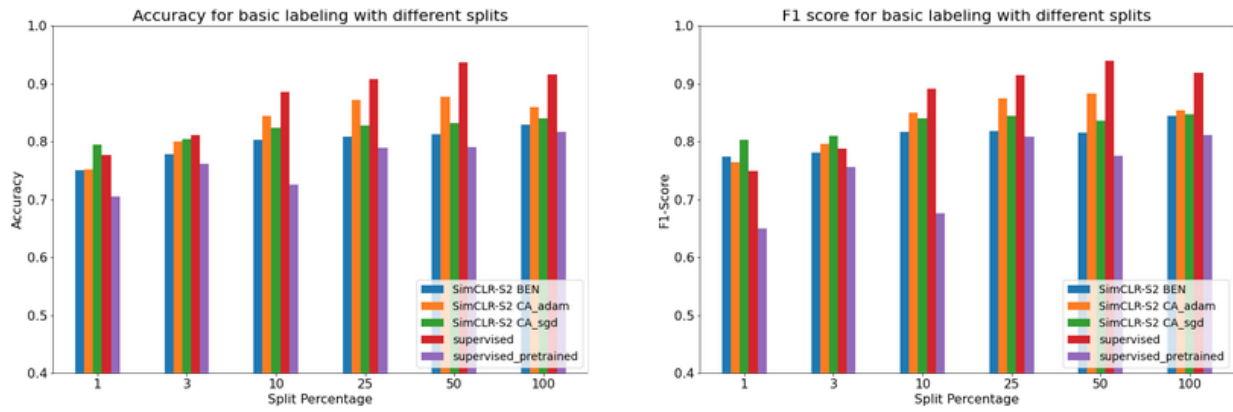
Figure 8. A comparison of stage 2 fine-tune results of SimCLR-S2 with supervised baseline
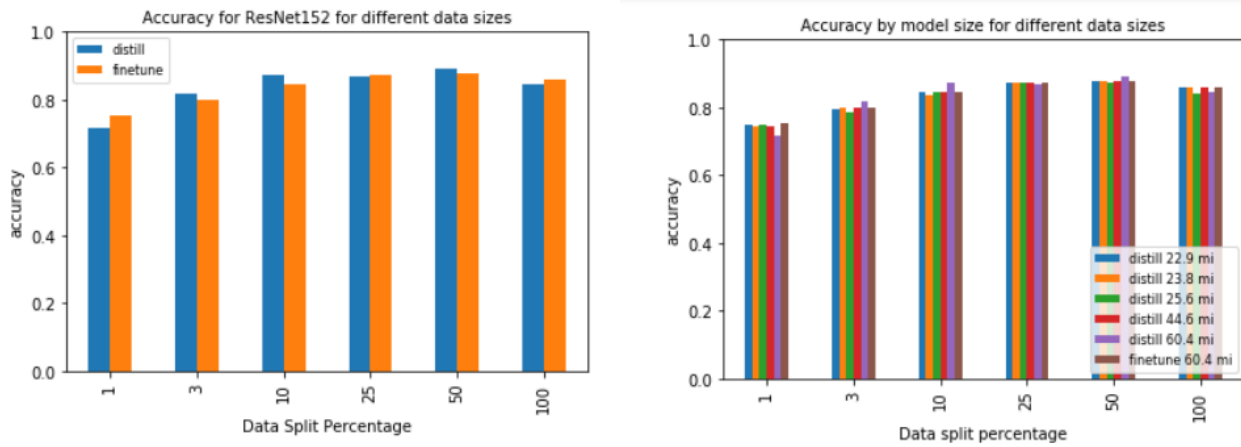
Figure 9. Distillation shows an improvement over fine-tune scores for many architectures and many data split percentages
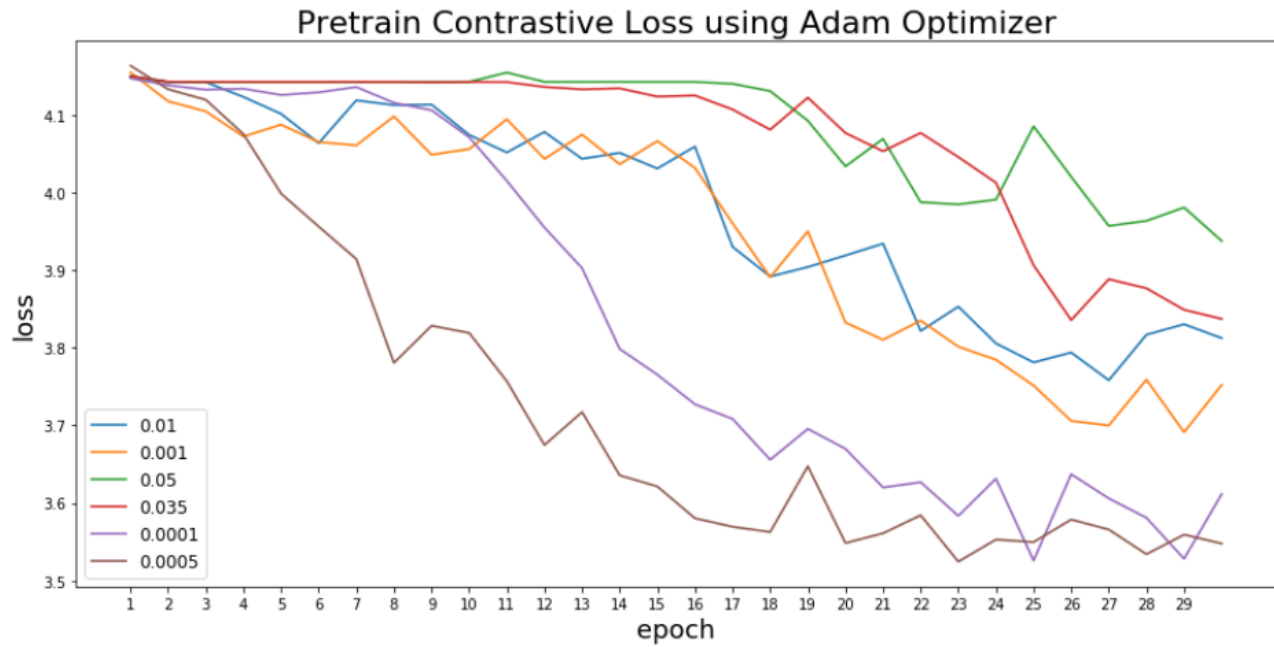
Pretrain Contrastive Loss using Adam Optimizer

Figure 10. Adam optimizer with different learning rates

Pretrain Contrastive Loss (SGD vs Adam)
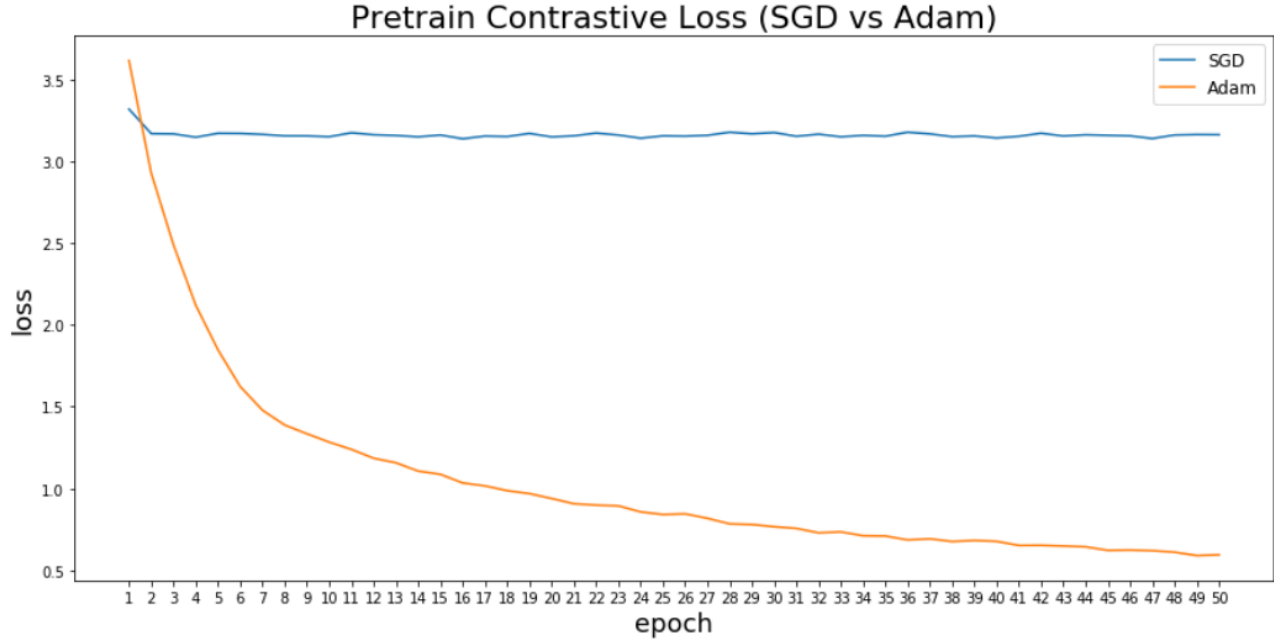
Figure 11. Pre-training loss with SGV vs Adam optimizer

Figure 12. Hyper parameter tuning for reducing contrastive loss in unsupervised pretraining step
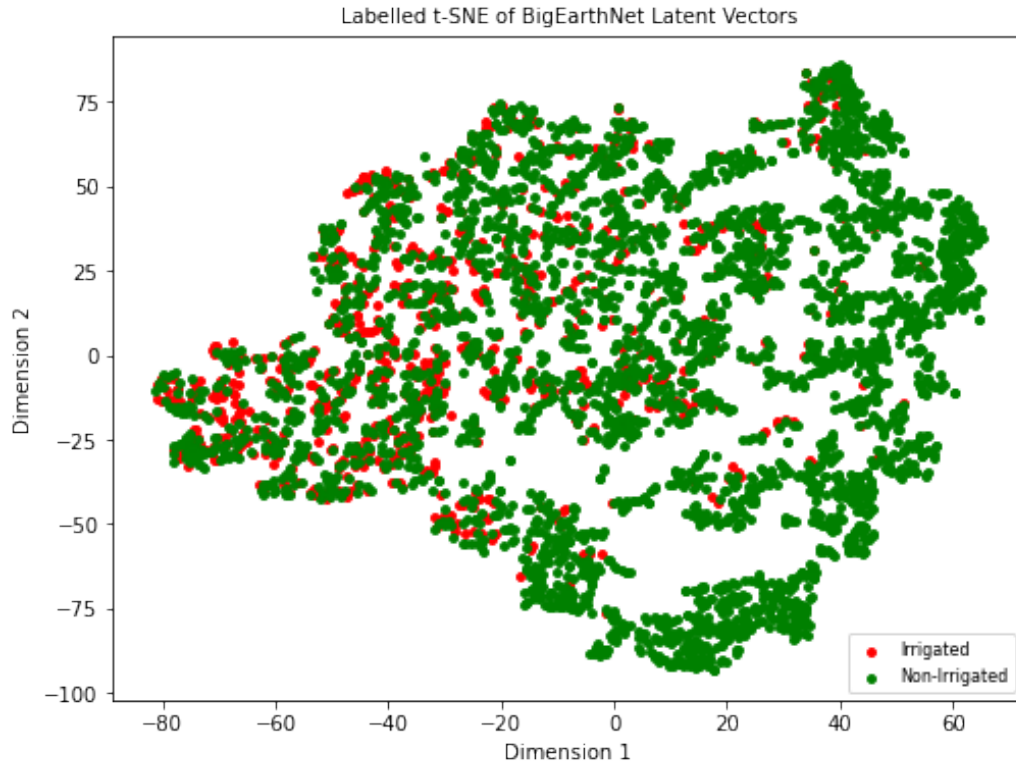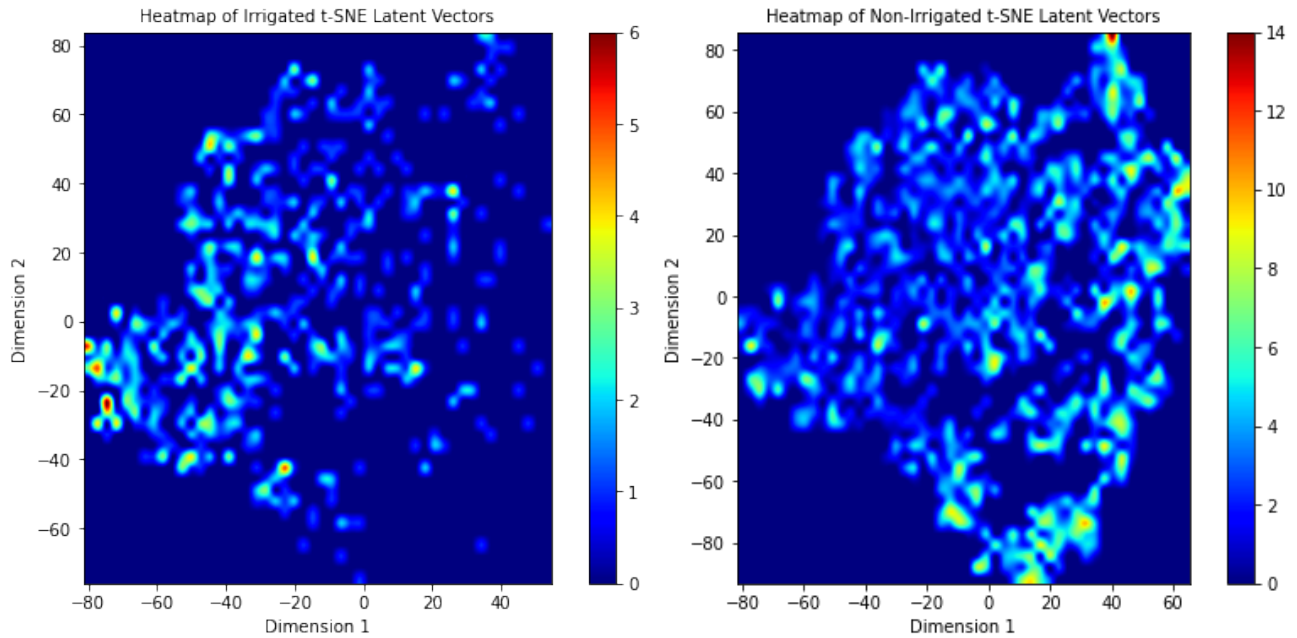
Figure 13. Labelled t-SNE of SimCLR-S2 latent vectors

Figure 14. Heatmap

Figure 15. Dimensionality reduction of the SimCLR-S2 latent vectors usng t-SNE, with labelled results and heatmaps indicating the results of contrastive loss on the latent space of our pretrained model.
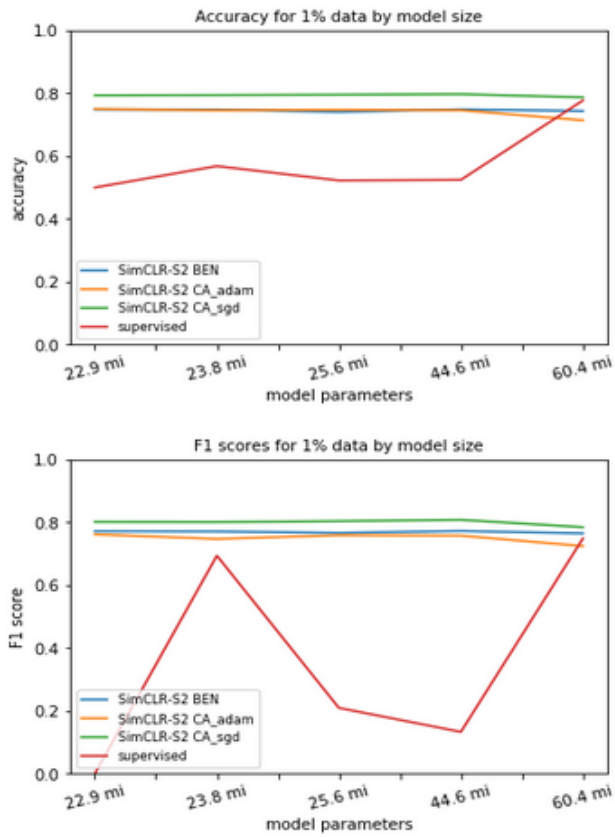
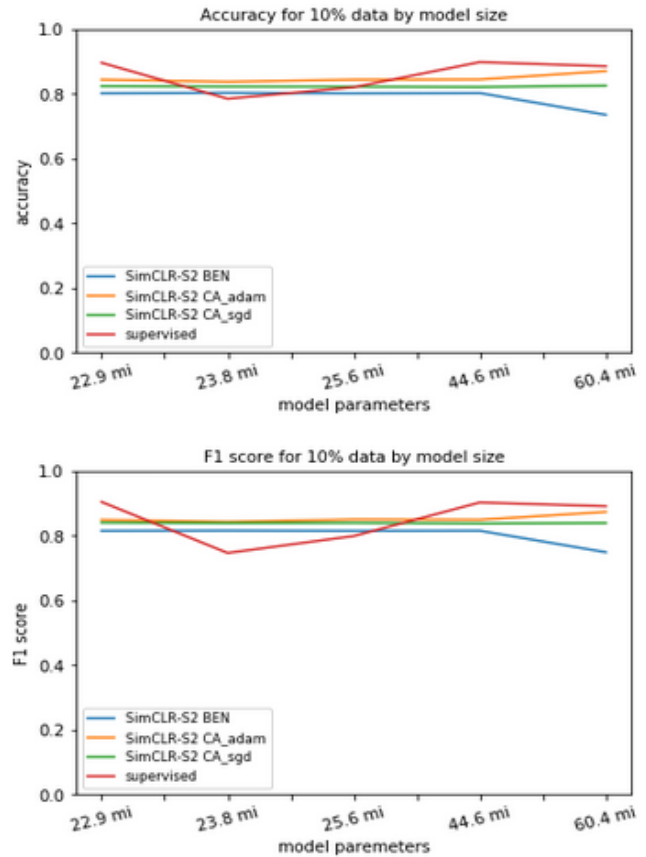*Figure 16.* Performance of distill vs supervised baseline scores for 1% data splits

Figure 17. Performance of distill vs supervised baseline scores for 10% data splits