META-LEARNING AND SELF-SUPERVISED PRETRAINING FOR STORM EVENT IMAGERY TRANSLATION

Ileana Rugina*
MIT EECS
irugina@mit.edu

Rumen Dangovski*
MIT EECS
rumenrd@mit.edu

Mark Veillette
MIT Lincoln Lab
mark.veillette@ll.mit.edu

Pooya Khorrami
MIT Lincoln Lab
pooya.khorrami@ll.mit.edu

Brian Cheung
MIT CSAIL & BCS
cheungb@mit.edu

Olga Simek
MIT Lincoln Lab
osimek@ll.mit.edu

Marin Soljačić
MIT Physics
soljacic@mit.edu

ABSTRACT

Recent advances in deep learning have provided impressive results across a wide range of computational problems such as computer vision, natural language, or reinforcement learning. Many of these improvements are however constrained to problems with large-scale curated data-sets which require a lot of human labor to gather. Additionally, these models tend to generalize poorly under both slight distributional shifts and low-data regimes. In recent years, emerging fields such as meta-learning or self-supervised learning have been closing the gap between proof-of-concept results and real-life applications of machine learning by extending deep-learning to the semi-supervised and few-shot domains. We follow this line of work and explore spatio-temporal structure in a recently introduced image-to-image translation problem for storm event imagery in order to: *i)* formulate a novel multi-task few-shot image generation benchmark in the field of AI for Earth and Space Science and *ii)* explore data augmentations in contrastive pretraining for image translation downstream tasks. We present several baselines for the few-shot problem and discuss trade-offs between different approaches. Our code: https://github.com/irugina/meta-image-translation.

1 INTRODUCTION

Benchmarks such as ImageNet (Deng et al., 2009) in computer vision or SQuAD (Rajpurkar et al., 2016) in natural language processing have been pivotal in popularizing deep-learning techniques and demonstrating their power. More recently, works such as ObjectNet (Barbu et al., 2019) in vision have shown impressive results on these established benchmarks do not translate to good performance in real-world situations, where the datasets might be less structured or more diverse. There is a lot of interest in devising more challenging datasets, both of general interest as well as domain-specific applications, that more closely resemble real-world situations practitioners might encounter when trying to deploy machine learning models. Growing fields such as self-supervised (Le-Khac et al., 2020) or multi-task learning (Hospedales et al., 2020) reflect these interests and provide promising solutions to the aforementioned issues. However, the problem of model evaluation remains: for example, in few-shot learning model evaluation is currently largely constrained to Omniglot (Lake et al., 2019; 2015) (which has essentially been saturated), Miniimagenet (Vinyals et al., 2017) and Metadataset (Triantafillou et al., 2019). Similarly, contrastive pretraining techniques are generally evaluated on ImageNet (He et al., 2020; Chen et al., 2020).

*Equal contribution.

We address these known limitations in the broader field of AI by contributing a new computer vision multi-task problem and move away from classification problems towards the field of image-generation by leveraging a weather dataset (Veillette et al., 2020) to formulate a novel few-shot image-to-image translation problem. In doing so we use the spatio-temporal structure of this dataset in construction of the few-shot tasks. We further leverage this structure through contrastive pretraining by exploring novel data augmentations and show consistent improvements in sample quality. Our work has three main contributions: *i)* we introduce a novel few-shot image translation benchmark and provide several baselines for this problem; *ii)* we train generative adversarial networks using model-agnostic meta-learning (MAML) (Finn et al., 2017) and discuss the advantages and drawbacks of this approach; *iii)* we pretrain part of the generator parameters using contrastive learning and show consistent improvements in downstream image-generation performance.

2 Related Work

Storm Event Imagery The Storm Event Imagery (SEVIR) (Veillette et al., 2020) is a radar and satellite meteorology dataset. It is a collection of over 10,000 weather events, each of which tracks 5 sensor modalities within 384 km $\times$ 384 km patches for 4 hours. Appendix A further details these modalities, shows example samples, and defines domain-specific evaluation metrics. Veillette et al. (2020) suggested several machine learning problems that can be studied on SEVIR and provided baselines for two of these: nowcasting and synthetic weather radar generation. In both cases they train U-Net models to predict VIL information.

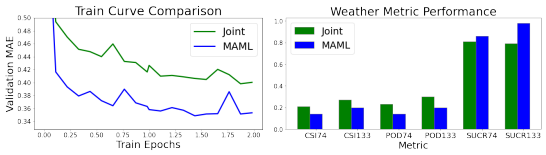
Generative Adversarial Networks in Low Data Regimes There has been a lot of interest in training GANs in low-data settings, where the discriminator can easily overfit and training does not progress. (Zhao et al., 2020a). The only prior work we are aware of on the topic of few-shot multi-task image generation with second-order gradient updates is that by (Clouâtre & Demers, 2019). They optimize using Reptile (Nichol et al., 2018), a first-order approximation to MAML, and evaluate on the MNIST and Omniglot datasets. They also introduce a dataset which presents a very clear delimitation between different tasks and more generally does not exhibit the challenges of modeling real-world phenomena because the examples are icons rather than realistic images. Zhao et al. (2020a) apply augmentations to both real and generated samples and require the transformations to be differentiable in order to backpropagate to the generator with good results using as little as 10% of the available samples. Consistency Regularization (CR) is a semi-supervised training technique introduced to GANs by (Zhang et al., 2019) and later extended by (Zhao et al., 2020b). These techniques regularize the discriminator and generator network using data augmentations on the generated samples or latent variables.

Machine Learning for Forecasting We focus on leveraging and building upon the work in (Veillette et al., 2020) because of the associated public large dataset. However, there has been wider interest in applying deep learning to improve the efficiency and performance of weather nowcasting systems. Ravuri et al. (2021) introduced deep generative models that nowcast radar data for 90 minutes and are trained to minize three loss functions: one temporal and one spatial discriminator terms to ensure spatiotemporal consistency, as well as a grid cell regularization term that improves performance by penalizing errors at the grid cell resolution level. In contrast, we explore different avenues towards similar results, and encourage spatiotemporal consistency through either explicit task construction or contrastive pretraining.

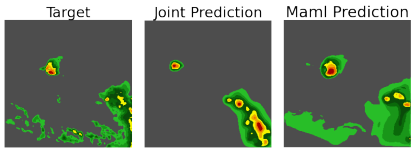
3 Few-Shot Benchmark and Our Baselines

Benchmark Construction We leverage the SEVIR (Veillette et al., 2020) dataset to construct a few-shot multi-task image-to-image translation problem where each task corresponds to one event. From the 49 available frames we keep the first N_{support} frames to form the task's support set and the next N_{query} to be the query. Throughout the following experiments we set $N_{\text{support}} = N_{\text{query}} = 10$.

Methods We solve the aforementioned task using either first-order or second-order gradient descent methods on U-Nets (Ronneberger et al., 2015) trained using either reconstruction or adversarial objectives. Note that in the case when we train GANs using MAML we are searching for a good



(a) **Validation mean absolute error throughout training and test-set evaluation of weather-specific metrics.** MAML optimization leads to better train objective and SUCR but lower CSI and POD.



(b) **Target VIL test-frame and generated samples.** Task-adaptation helps recognize sparse VIL regions.

Figure 1: **Reconstruction Loss Results:** Objective evolution through training; final evaluation on domain-specific metrics; examples of generated samples.

initialization for multiple related saddle-point problems. Despite this challenging task, we still obtain good performance. We write out the algorithm for adversarial meta-learning in Appendix and note that the other three training regimes are simplified version of our novel contribution above. Note the introduction of two hyperparameters that influence training: λ specifies the importance of reconstruction loss relative to adversarial term while η controls how much we adapt to each task in MAML's inner optimization.

Experimental Details We run experiments using a single 32GB Nvidia Volta V100 GPU. For MAML optimization (Arnold et al., 2020) we use meta-batch sizes of 2, 3 or 4 events. For the corresponding joint training baselines we used $N_{\text{support}} + N_{\text{query}}$ frames from each event and comparable number of events to keep comparisons fair.

Results We test our multi-task few-shot formulation and demonstrate MAML provides empirical gains by comparing the performance of models trained using either meta-learning algorithms or joint training for both reconstruction and adversarial loss objectives. We find throughout all our experiments that meta-learning minimizes the reconstruction error compared to joint training. On the other hand, achieving better performance on the training objective does not always translate to higher weather evaluation metrics.

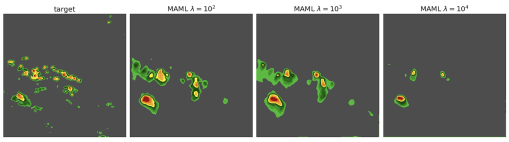
Reconstruction Loss We compare joint training with MAML that uses a single adaptation step for each event, evaluate model performance using weather metrics with two different thresholds (74 and 133). We summarize our evaluation results in Figure 1a. We find that even though U-Nets trained with MAML achieve better performance on the optimization objective, these improvements do not consistently translate to gains on weather-specific evaluation. In particular, we see that finetuning to specific tasks leads to better precision but worse recall and IOU. Limitations given by training with reconstruction loss, such as blurry outputs, remain.

Figure 1b exhibits one synthetic VIL imagery generated through either method, as well as the ground-truth data. The task adaptation mechanism helps in this case to recognize that there are some storm events in the lower-left corner, although it is not very effective at predicting the correct shape of these low-intensity precipitations on a fine-grained scale.

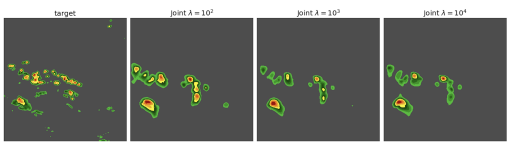
Adversarial Loss We train generative adversarial networks using the second-order MAML procedure and the joint training baseline. We compare the evolution of the reconstruction error throughout training in Figure 6 (in the Appendix) and notice MAML significantly helps in minimizing the training objective. We used $\lambda = 10^2$ and $\eta = 10^{-4}$ for this MAML curve.

Next, we evaluate on meteorological metrics for all values of λ and η, and summarize our results in Table 1 and 2 (in the Appendix) for joint and MAML training, respectively. For joint adversarial training, especially when evaluating with lower thresholds, we see the critical success index is fairly constant as we vary λ while increasing λ leads to lower recall and higher precision. This suggests that placing more weight on the reconstruction loss leads to predicting fewer high-valued VIL pixels.

For MAML adversarial training we do not identify any clear trend between hyperparameters λ and η and the values of meteorological metrics on the test-split. We believe this shows training is more unstable in this regime: instabilities are further exacerbated when optimizing a bilevel Nash equilib-



(a) **Adversarial MAML - generated samples.** Fine-tuning helps identify low-intensity VIL regions.



(b) **Adversarial Joint - generated samples.** Reconstruction loss biases the model towards sparser predictions.

Figure 2: Generated samples from training with the adversarial loss.

rium problem with gradient descent, as we did above. A comparison between Tables 1 and 2 shows that, similarly to the case of reconstruction loss, MAML optimization leads to higher precision and lower recall. After visually inspecting the generated samples we find that some of the models seem to exhibit mode collapse where the generated samples are not even realistic, while some of them do resemble the ground-truth. We present examples of samples successfully generated by models trained with MAML on adversarial loss below and note there is a large variance in the fraction of realistic samples across different models. This is not reflected in any of the evaluation metrics: we believe this further underscores that in image generation the correlation between good evaluation performance and high sample quality is rather weak.

Figures 2b and 2a compare samples generated by models trained on adversarial loss through either joint or MAML-based procedures for different values of λ. The MAML models all used an inner SGD learning rate of 10^{-5}. We see that in this case the intuitions from the reconstruction loss setting are still valid and the task-adaptation inherent to MAML enables it to correctly generate low-intensity VIL data that joint-setting misses out on. We also confirm the aforementioned trend of higher λ values leading to lower VIL values.

4 Self-supervised Pretraining

Method We follow recent work in self-supervised pretraining which applies contrastive learning to convolutional networks before finetuning on classification tasks and improves downstream performance and data efficiency. We ask if these improvements extrapolate to our image-to-image setup. The main distinction between our scenario and those in previous work is that we can initialize only a fraction of our parameters through contrastive pretraining. We restrict our attention to the U-Net encoder parameters during the pretraining stage and follow the same network architecture as in Section 3. Our experiments are inspired by the large-scale study on unsupervised spatiotemporal representation learning, conducted by (Feichtenhofer et al., 2021). In particular, we focus on MoCoV3 (Chen et al., 2021), which is a state-of-the-art contrastive learning method, because (Feichtenhofer et al., 2021) identify the momentum contrast (MoCo) contrastive learning method as the most useful for spatiotemporal data.

Pretraining objective. For a given representation q of a query frame from the dataset, a positive key representation k^+ and a negative key representation k^-, the loss function increases the similarity between q and k^+, and decreases the similarity between q and k^-. All representations are normalized on the unit sphere. In particular, the loss is the InfoNCE loss (Oord et al., 2018) is $\hat{\mathcal{L}}_q = -\log \frac{e^{p(q)\cdot \mathrm{sg}(k^+)/\tau}}{e^{p(q)\cdot \mathrm{sg}(k^+)/\tau} + \sum_{k^-} e^{p(q)\cdot \mathrm{sg}(k^-)/\tau}}$ for a temperature parameter τ and a predictor MLP p, which is a two layer MLP, with input dimension 128, hidden dimension 2048, output dimension 128, BatchNorm and ReLU in the hidden layer activation, and where "sg" is the stopgradient operation. Following (Chen et al., 2021), the gradients are not backpropagated through $k^{\{+,-\}}$ and the encoder representations both for keys and queries are obtained after a composition of the backbone and the projector (which is a two layer MLP, with dimensions [256, 2048, 128] with BatchNorm and ReLUs in between the hidden layers, and ending with a BatchNorm with no trainable affine parameters). Additionally, the branch for key representations follows the momentum update policy $\theta_\mathrm{k} \leftarrow m\theta_\mathrm{k} + (1-m)\theta_\mathrm{q}$ from (He et al., 2020) with momentum parameter $m = 0.999$, where θ_k are the weights in the key branch and θ_q are the weights in the query branch (see Section B). We further consider using the temporal structure of SEVIR for augmentations, as follows. Each event consists

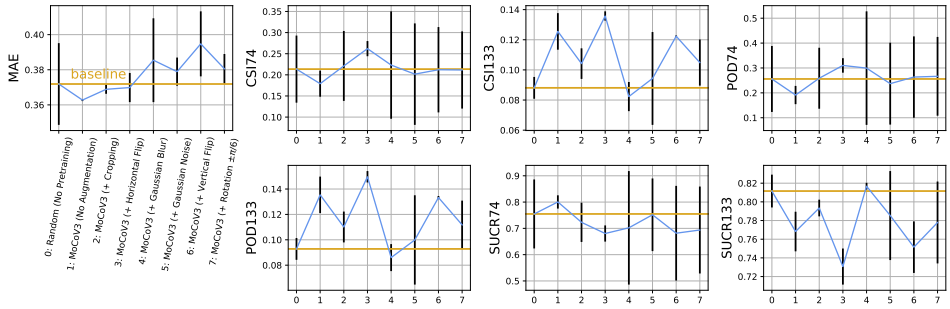
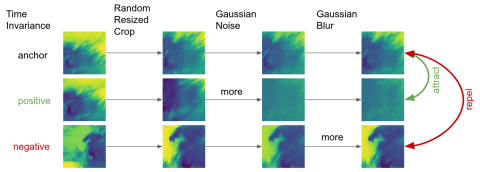
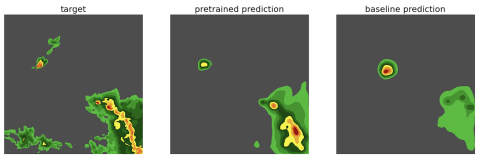


Figure 3: **Contrastive Learning for SEVIR.** For mean absolute error lower is better. For every other evaluation measure, higher is better.



(a) **Augmentations for the contrastive learning experiment** By indicating "more" we show examples of a larger magnitude of the augmentation being applied.



(b) **Pretrained encoder - generated samples** Pretrained models better identify the sparse high VIL values.

Figure 4: Contrastive pretraining for SEVIR.

of 49 frames, so we anchor every even frame as query frame and use every odd frame as key frame. For each query frame, to obtain q and k^+ we apply the following stochastic transformations to the frame twice: random resized crops using scale (0.8, 1.0); random horizontal flips with probability 0.5, pixel-wise gaussian noise sampled from the normal distribution $\mathcal{N}(0, 0.1)$ with probability 0.5, gaussian blur with kernel size 19, random vertical flips with probability 0.2, random rotation by angle unformly chosen in $(-\pi/6, \pi/6)$. The rest of the augmentation arguments follow the default in the Torchvision library (`https://pytorch.org/vision/stable/transforms.html`). In Figure 4a we present a conceptual visualization of the transforms. To obtain k^- we apply the above stochastic transformations to the corresponding key frame once (see Section D). lists design challenges when using data augmentation.

Results In Figure 3 we report our results. Firstly, for mean absolute error we find marginal yet somewhat consistent gains up to level 3 augmentation. Secondly, we also evaluate on meteorological metrics. We find that even though pretraining has a marginal effect on the reconstruction loss train objective, it often provides important gains on domain-specific evaluation criteria. We highlight the large improvement in CSI133 and POD133. We observe that up to level 4 MoCoV3 augmentation we obtain improvements throughout all measures with the contrastive pretraining. Finally, Figure 4b demonstrates that pretraining the U-Net encoder leads to better performance in high-VIL regions.

5 Conclusion

We formulated a novel few-shot multi-task image-to-image translation problem leveraging spatio-temporal structure in a large-scale storm event dataset. We provided several benchmarks for this problem and considered two optimization procedures (joint training and gradient-based meta-learning) and two loss functions (reconstruction and adversarial). We trained U-Nets in all these regimes and presented each model's performance, as well as evaluated on various domain-specific metrics. We discussed the advantages and disadvantages of each of these. In this process we also explored a training method unexplored until now to the best of our knowledge: meta-learning adversarial GANs with second-order gradient updates. Additionally, we explored pretraining U-Net encoder parameters using various augmentations in both the spatial and temporal domains.

REFERENCES

Sébastien M R Arnold, Praateek Mahajan, Debajyoti Datta, Ian Bunner, and Konstantinos Saitas Zarkias. learn2learn: A library for Meta-Learning research. August 2020. URL http://arxiv.org/abs/2008.12284.

Andrei Barbu, David Mayo, Julian Alverio, William Luo, Christopher Wang, Dan Gutfreund, Josh Tenenbaum, and Boris Katz. Objectnet: A large-scale bias-controlled dataset for pushing the limits of object recognition models. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/file/97af07a14cacba681feacf3012730892-Paper.pdf.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. *arXiv preprint arXiv:2002.05709*, 2020.

Xinlei Chen, Saining Xie, and Kaiming He. An empirical study of training self-supervised visual transformers. *arXiv e-prints*, pp. arXiv–2104, 2021.

Louis Clouâtre and Marc Demers. FIGR: few-shot image generation with reptile. *CoRR*, abs/1901.02199, 2019. URL http://arxiv.org/abs/1901.02199.

J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.

Christoph Feichtenhofer, Haoqi Fan, Bo Xiong, Ross Girshick, and Kaiming He. A large-scale study on unsupervised spatiotemporal representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3299–3309, 2021.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In Doina Precup and Yee Whye Teh (eds.), *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1126–1135, International Convention Centre, Sydney, Australia, 06–11 Aug 2017. PMLR. URL http://proceedings.mlr.press/v70/finn17a.html.

Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9729–9738, 2020.

Timothy Hospedales, Antreas Antoniou, Paul Micaelli, and Amos Storkey. Meta-learning in neural networks: A survey, 2020.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

B. Lake, R. Salakhutdinov, and J. Tenenbaum. Human-level concept learning through probabilistic program induction. *Science*, 350:1332 – 1338, 2015.

Brenden M. Lake, Ruslan Salakhutdinov, and Joshua B. Tenenbaum. The omniglot challenge: a 3-year progress report, 2019.

Phuc H. Le-Khac, Graham Healy, and Alan F. Smeaton. Contrastive representation learning: A framework and review. *IEEE Access*, 8:193907–193934, 2020. ISSN 2169-3536. doi: 10.1109/access.2020.3031549. URL http://dx.doi.org/10.1109/ACCESS.2020.3031549.

Alex Nichol, Joshua Achiam, and John Schulman. On first-order meta-learning algorithms, 2018.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. Squad: 100,000+ questions for machine comprehension of text, 2016.

Suman Ravuri, Karel Lenc, Matthew Willson, Dmitry Kangin, Remi Lam, Piotr Mirowski, Megan Fitzsimons, Maria Athanassiadou, Sheleem Kashem, Sam Madge, Rachel Prudden, Amol Mandhane, Aidan Clark, Andrew Brock, Karen Simonyan, Raia Hadsell, Niall Robinson, Ellen Clancy, Alberto Arribas, and Shakir Mohamed. Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597(7878):672–677, Sep 2021. ISSN 1476-4687. doi: 10.1038/s41586-021-03854-z. URL https://doi.org/10.1038/s41586-021-03854-z.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.

Timothy J. Schmit, Paul Griffith, Mathew M. Gunshor, Jaime M. Daniels, Steven J. Goodman, and William J. Lebair. A closer look at the abi on the goes-r series. *Bulletin of the American Meteorological Society*, 98(4):681 – 698, 2017. doi: 10.1175/BAMS-D-15-00230.1. URL https://journals.ametsoc.org/view/journals/bams/98/4/bams-d-15-00230.1.xml.

Eleni Triantafillou, Tyler Zhu, Vincent Dumoulin, Pascal Lamblin, Kelvin Xu, Ross Goroshin, Carles Gelada, Kevin Swersky, Pierre-Antoine Manzagol, and Hugo Larochelle. Meta-dataset: A dataset of datasets for learning to learn from few examples. *CoRR*, abs/1903.03096, 2019. URL http://arxiv.org/abs/1903.03096.

Mark Veillette, Siddharth Samsi, and Chris Mattioli. Sevir : A storm event imagery dataset for deep learning applications in radar and satellite meteorology. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 22009–22019. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/fa78a16157fed00d7a80515818432169-Paper.pdf.

Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Koray Kavukcuoglu, and Daan Wierstra. Matching networks for one shot learning, 2017.

Han Zhang, Zizhao Zhang, Augustus Odena, and Honglak Lee. Consistency regularization for generative adversarial networks. *CoRR*, abs/1910.12027, 2019. URL http://arxiv.org/abs/1910.12027.

Shengyu Zhao, Zhijian Liu, Ji Lin, Jun-Yan Zhu, and Song Han. Differentiable augmentation for data-efficient gan training, 2020a.

Zhengli Zhao, Sameer Singh, Honglak Lee, Zizhao Zhang, Augustus Odena, and Han Zhang. Improved consistency regularization for gans, 2020b.

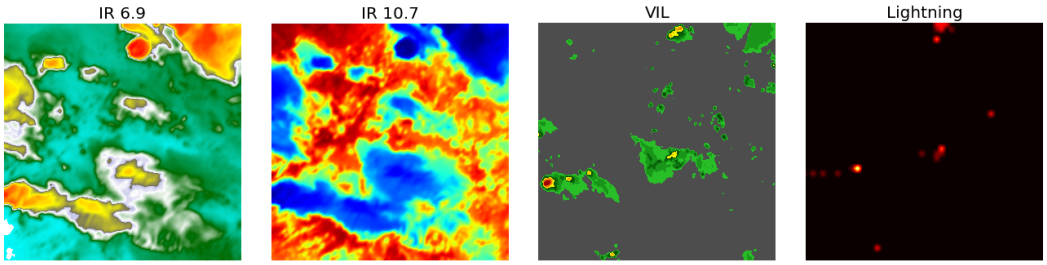


Figure 5: **Frame from The Storm Event Imagery (SEVIR) dataset. We use four of the five available modalities: 2 IR, VIL, and lightning information**

A Further SEVIR details

SEVIR events are uniformly sampled so that there are 49 frames for each 4 hour period, and the 5 channels consist of: *i)* 1 visible and 2 IR sensors from the GOES-16 advanced baseline (Schmit et al., 2017) *ii)* vertically integrated liquid (VIL) from NEXTRAD *iii)* lightning flashes from GOES-16 . Fig. 5 shows examples of the two IR, VIL, and lightning modalities. We disregard the visible channel.

We review common evaluation metrics used in the satellite and radar literature to analyse artificially-generated VIL imagery. They all compare the target and generated samples after binarizing them with an arbitrarily threshold in $[0, 255]$ and looking at counts in the associated confusion matrix. Let H denote the number of true positives, C denote the number of true negatives, M denote the number of false negatives and F the number of false positives. Veillette et al. (2020) define four evaluation metrics: Critical Succes Index **(CSI)** is equivalent to the intersection over union $\frac{H}{H+M+F}$; Probability of detection **(POD)** is equivalent to recall $\frac{H}{H+M}$; Succes Ratio **(SUCR)** is equivalent to precision $\frac{H}{H+F}$

B Further hyperparameters

Training We randomly split all SEVIR events into 9169 train, 1162 validation, and 1148 test tasks. Joint training baselines and MAML outer loop optimizations are both performed using the Adam optimizer (Kingma & Ba, 2017) with learning rate 0.0002 and momentum 0.5. We resize input modalities to all have 192×192 resolution and keep the target at 384×384. The generator's encoder has four convolutional blocks, and the decoder has five. All generator blocks except for the last decoder layer use ReLU activation functions. The very last layer uses linear activation functions to support z-score normalization for all four image modalities.

Contrastive Pretraining Our experiments use the following architectural choices: mini-batch, consisting of 3 events with 24 frames for queries and key 24 frames for keys each; 0.015 base learning rate for the baseline SimCLR and SimSiam and; 100 pretraining epochs; standard cosine decayed learning rate; 5 epochs for the linear warmup; 0.0005 weight decay value; SGD with momentum 0.9 optimizer. We report the joint training reconstruction loss experimental setup by finetuning the checkpoint obtained from pretraining. All parameters are in a Pytorch-like style.

C Methods

Below we present the meta-train loop for adversarial networks, which is a novel contribution of our work. For simplicity, we only present the variant with a single SGD inner-loop adaptation step. We train a U-Net generator G with model weights w_G jointly with an extranous patch discriminator D with model weights w_D using data $\mathcal{D} \in \mathbb{R}^{N_{\text{event}} \times N_{\text{frames}} \times C \times w \times h}$. We use batched alternating gradient descent as our optimization algorithm and consider batches $\mathcal{B} \in \mathbb{R}^{B \times N_{\text{frames}} \times C \times w \times h}$, where B is the meta-batch size. Each of these can be split along the second axis into the sup-

port and query sets, and along the third axis into the source (S) and target tensors (T) to create $S^{\text{support}} \in \mathbb{R}^{B\times N_{\text{support}}\times C_{\text{in}}\times w\times h}$, $S^{\text{query}} \in \mathbb{R}^{B\times N_{\text{query}}\times C_{\text{in}}\times w\times h}$, $T^{\text{support}} \in \mathbb{R}^{B\times N_{\text{support}}\times C_{\text{out}}\times w\times h}$, $T^{\text{query}} \in \mathbb{R}^{B\times N_{\text{query}}\times C_{\text{out}}\times w\times h}$. For any of these tensors $X \in \{S^{\text{support}},\ S^{\text{query}},\ T^{\text{support}},\ T^{\text{query}}\}$ we refer to the four-dimensional tensor given by the i^{th} task or event as X_i. We use such four-dimensional tensor quantities to evaluate the generator and discriminator loss functions:

$$\hat{\mathcal{L}}_G(t^{\text{generated}}, t, s; w_G, w_D) = -\log D(s, t^{\text{generated}}) + \lambda||t^{\text{generated}} - t||_1 \tag{1}$$

$$\hat{\mathcal{L}}_D(t^{\text{generated}}, t, s; w_G, w_D) = \frac{\log D(s, t^{\text{generated}}) - \log D(s, t)}{2}, \tag{2}$$

where $t^{\text{generated}} = G(s)$ is a generated target sample, t and s are corresponding input and output modalities, $||x||_1$ is the mean absolute error. Note the slight abuse of notation where by $\log D(x, y)$ with $x, y \in \mathbb{R}^{N\times C\times w\times h}$ we mean the average $\frac{1}{N}\sum_{i=1}^{N} \log D(x_i, y_i)$. This formulation also uses the trick of replacing $\max \log(1 - D(G(z)))$ with $\min \log D(G(z))$ to obtain a non-saturating generator objective. We wrote the loss functions above such that both players want to minimize their respective objectives.

for meta-train-batch $\mathcal{B} \in \mathbb{R}^{B\times N_{\text{frames}}\times C\times w\times h}$ **do**
  unpack $\mathcal{B} \in \mathbb{R}^{B\times N_{\text{frames}}\times C\times w\times h}$ along support/query, source/target into:
  $S^{\text{support}},\ S^{\text{query}},\ T^{\text{support}},\ T^{\text{query}}$
  init $l_G^{\text{batch}} = 0, l_D^{\text{batch}} = 0$
  for each event i out of B in meta-batch **do**
    forward pass $T_i^{\text{support; generated}} = G(S_i^{\text{support}})$
    $l_G^{\text{adapt}} = \mathcal{L}_G(T_i^{\text{support; generated}}, T_i^{\text{support}}, S_i^{\text{support}}; w_G, w_D)$ from Eq. 1
    $l_D^{\text{adapt}} = \mathcal{L}_D(T_i^{\text{support; generated}}, T_i^{\text{support}}, S_i^{\text{support}}; w_G, w_D)$ from Eq. 2
    task-specific parameters $\phi_G \leftarrow w_G - \eta\nabla_{w_G} l_G^{\text{adapt}}$
    task-specific parameters $\phi_D \leftarrow w_D - \eta\nabla_{w_D} l_D^{\text{adapt}}$
    forward pass $T_i^{\text{query; generated}} = G(S_i^{\text{query}})$
    $l_G = \mathcal{L}_G(T_i^{\text{query; generated}}, T_i^{\text{query}}, S_i^{\text{query}}; \phi_G, \phi_D)$ from Eq. 1
    $l_D = \mathcal{L}_D(T_i^{\text{query; generated}}, T_i^{\text{query}}, S_i^{\text{query}}; \phi_G, \phi_D)$ from Eq. 2
    update rolling sums $l_G^{\text{batch}} += l_G$ and $l_D^{\text{batch}} += l_D$
  end
  backpropagate 2nd order updates $\nabla_{w_G} l_G^{\text{batch}}$ and $\nabla_{w_D} l_D^{\text{batch}}$ to w_G and w_D
end
return good initializations w_G and w_D for both generator and discriminator.

Algorithm 1: One Epoch MAML-Train Loop for U-Net Generator with Adversarial Loss.

D Data augmentation for contrastive pretraining

Another difficulty particular to our setup is the problem of choosing data augmentations the input domain is invariant to because weather modalities have different invariances than natural images: for example, the popular color jitter transformation is not applicable here, because image-to-image translation is sensitive to color. From the standard augmentations, the ones we consider are: random resized crops, random horizontal flips, gaussian noise, gaussian blur, random vertical flips and random rotation.

E Additional experimental results

In Figure 6 and Tables 1 and 2 we present additional results that have been discussed in the main text.

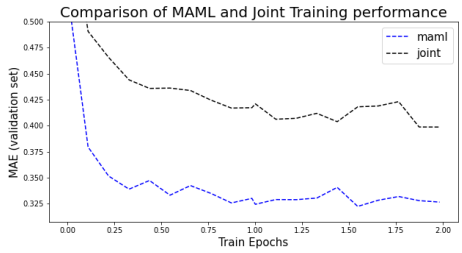


Figure 6: **Adversarial loss - train curve.** MAML outperforms Joint Training. Evaluation is done on validation set throughout training.

Table 1: **Adversarial Joint - evaluation.** Test-set evaluation on meteorological metrics.

thresh.	74			133		
metric	CSI	POD	SUCR	CSI	POD	SUCR
λ: 10^2	0.29	0.50	0.56	0.27	0.30	0.76
λ: 10^3	0.29	0.46	0.58	0.29	0.35	0.71
λ: 10^4	0.29	0.43	0.64	0.29	0.33	0.73

Table 2: **Adversarial MAML - evaluation.** Test-set evaluation on meteorological metrics. MAML models have higher precision and lower recall and IOU.

thresh.		74			133		
metric		CSI	POD	SUCR	CSI	POD	SUCR
η: 10^{-4}	λ: 10^2	0.14	0.16	0.93	0.24	0.26	0.90
	λ: 10^3	0.09	0.09	0.98	0.20	0.20	0.99
	λ: 10^4	0.13	0.21	0.91	0.21	0.32	0.87
η: 10^{-5}	λ: 10^2	0.19	0.23	0.87	0.23	0.27	0.84
	λ: 10^3	0.17	0.20	0.90	0.25	0.29	0.87
	λ: 10^4	0.12	0.15	0.93	0.22	0.26	0.91