A fine opportunity to leverage pre-trained models adapted to Solar PV classification

Anonymous Author(s)
Affiliation
Address
email

Abstract

Renewable energy such as solar is key to ensuring access to affordable and sustainable energy generation. Surveying its adoption patterns globally is pivotal to measuring and evaluating renewable energy access and creating a more efficient and equitable grid. Leveraging high-resolution imagery to detect solar PVs has proven to be a more exhaustive way of covering all PVs, including residential PVs that can be very challenging to track via conventional surveying methods. While the literature has developed models to classify and segment PV installations, residential PV is still challenging to identify using medium resolution ($\approx$ 30 cm/pixel or above) remote-sensing products. This work explores different fine-tuning (FT) strategies of pre-trained *ViT* models for classification tasks in smaller dataset settings. While FT offers an opportunity for fast and computationally efficient model deployment, practitioners have to be cautious about the effects of fine-tuning on OOD classification and how advances in text attention mechanisms do not necessarily map to image architectures. Moreover, the LoRA technique (Low-Rank Adaptation) is identified as an efficient method for fine-tuning, enhancing the model's adaptability to specific tasks while preserving its generalizability. Despite these advancements, achieving robust OOD classification in a foundational model context remains a challenging task.

1 Introduction

PV adoption deployment and access are crucial for achieving affordable and clean energy access and meeting decarbonization goals across the globe. Understanding the distribution of solar PV allows governments, businesses, and individuals to identify regions with the greatest potential for solar energy generation and efficiently integrate solar PV into the electrical grid. This is particularly relevant as annual growth in the solar industry will average 15% [NREL, 2023]. Hence, there is a need to explore more efficient ways of tracking and monitoring it .

Furthermore, PV adoption is not uniform often showcasing underlying race and income inequalities [Sunter et al., 2019, Lukanov and Krieger, 2019, Kwan, 2012]. Measuring adoption helps to alleviate the deepening of "energy privileges" [Stokes et al., 2023] and the misallocation of tax benefits that lead to inequitable adoption of renewable energy. While field surveying is available[1], its time and spatial coverage is often inadequate to capture longitudinal changes in solar adoption, especially in fast-adopting markets like the United States.

Existing projects in the literature have built longitudinal PV adoption data with significant spatial and time coverage (i.e. *DeepSolar* [Wang et al., 2022, Yu et al., 2018]). Nonetheless, some of

[1]The US Department of Energy's *OpenPV* is the largest database of PV installation in the US using crowd-sourced data, but it was discontinued in 2019.

Submitted to Computational Sustainability Workshop at NeurIPS 2023. Do not distribute.

these projects have some limitations such as high-resolution ($\leq$ 20 cm spatial resolution) data requirements that are not publicly available. Second, given the sparsity of PV adoption, there are no country-wide datasets, leading to urban biases [Wang et al., 2022]. Lastly, deploying these models can be cumbersome and prohibiting for researchers and policy-makers in the developing world, where computing resources and labels are limited. Although *DeepSolar*[Yu et al., 2018] and *DeepSolar++*[Wang et al., 2022] models have achieved high performances in tasks such detecting solar PVs (precision (recall) of 93.1% (87.5%) in residential areas) and predicting installation years (rate of 93.9 ± 1.0 over a random sample of 23 counties), their coverage is still limited to the US and dependency on high-resolution satellite/aerial imagery.

Fine-tuning pre-trained models for downstream tasks has demonstrated superior performance when compared to training from scratch in the context of language models. This approach has emerged as the prevailing and widely accepted strategy for addressing downstream classification and generation tasks within the domain of language modeling, as supported by relevant scholarly works [Wei et al., 2021, Zhang and Bowman, 2018]. Recent advances in fine-tuning, such as *LoRa* [Hu et al., 2021], have also streamlined the fine-tuning process by reducing the parameter space allowing faster inferences and task adaptation. This, combined with different optimization techniques such as contrastive learning [Chen et al., 2020] and self-supervised learning [Chen et al., 2021], have boosted accuracy on different vision classification benchmarks: a ResNet-50 pre-trained with ImageNet, improves CIFAR-10 classification from 95% to 98% [Chen et al., 2020]. Fine-tuning also alleviates some of the financial and environmental limitations related to training from scratch. Patterson et al. [2021], Schwartz et al. [2019]. These are particularly relevant in the developing world where access to GPU computing is limited and cloud-based options can be burdensome. Fine-tuning can achieve competing performances with fewer training labels and shorter computation times.

Intuitively, fine-tuning all the layers of a neural network can adapt a pre-trained model faster to a new task and obtain better results than training only a few layers or reducing its training parameters' feature space. Nonetheless, previous work in language models (BERT [Devlin et al., 2019] and RoBERTa [Liu et al., 2019]) have shown that only a fourth of the last layers are required to keep similar levels of accuracy [Lee et al., 2019], other experiments have shown that linear-probing and freezing might be better alternatives to naïve transfer and Out-of Distribution (OOD) performance[Kumar et al., 2022] in contrastive model settings . Nonetheless, there is no evidence of recommendations on supervised fine-tuning examples and in OOD settings where data sources vary in terms of resolution and color space. Moreover, while *ViT* architectures have brought new gains, some of the previous rules for fine-tuning and transfer learning cannot be adopted from the CNN architectures [Chen et al., 2021].

In this paper, we test the ability of different fine-tuning strategies to adapt pre-trained vision transformer (*ViT*) models to a PV classification downstream task. Additionally, we also evaluate our best models for their ability to generalize in a OOD setting – geographical and spatial resolution domains. For this, we use two solar PV datasets in two areas of interest (AoI), China and the state of California (US), in different resolutions. While some recent literature [He et al., 2022, Goyal et al., 2022] have done similar experiments to the ones we present in this paper, this work inscribes into an open problem in Earth Observation (EO) and uses an applied example that differs in complexity from vision benchmarks in the literature. We show that (...)

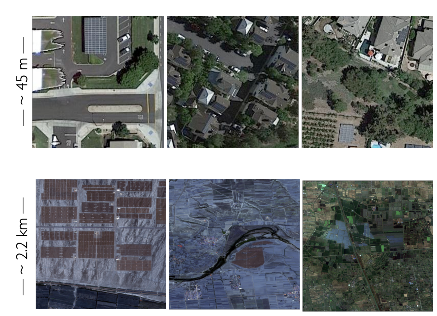

Figure 1: Labeled datasets for PV classification. Both image sets were collected around the same time: 2017 - 2019. The first high-resolution set corresponds to the Google Maps aerial imagery. The second set, Sentinel-2, has 10 times coarser resolution, but a higher revisit time, (better representation of the changes over time). The Sentinel-2 data is aggregated in time and cleaned to remove "bad" pixels and clouds. We have 4179 positive labels and 3966 negative labels across datasets.

2 Methodology

2.1 Data

Experiments will use two in-house remote-sensing datasets from different resolutions. On one hand, we collected GoogleMaps aerial data imagery for our two AoI, California (CA), and several Northern provinces of China (CN). These images have an approximate resolution of 0.2

m/pixel with three visible channels (RGB). On the other hand, we collected Sentinel-2 scenes with 10 m/pixel resolution (50 times coarser than the Google images) for the same locations in CN. While the Google imagery is static in time (latest), the Sentinel-2 data has an average revisit duration of 8 days for our AoI, thus we build a cloud-free median composite image by combining all the $2,560$ images collected from 2017 to 2019. We adopt '80-20-20' split policy for training, validation and testing sets.

The PVs labels for commercial and utility-scale installations come from two sources. For California, we used data from the *DeepSolar* database [Yu et al., 2018]. For China, we used data from Kruitwagen et al. [2021], a novel database with more than $60,000$ global PV installations. For both data sources, we extracted an image patch of 224×224 pixels around each of the labeled points and fed it to one of the most widely used *ViT* for image classification model pre-trained on `google/vit-base-patch16-224` that uses *ImageNet-21K*.

2.2 Experiments

To test the best fine-tuning strategy for adaptation to the PV classification downstream task, we run the following experiments:

1. **Full network retraining** (*FT*): The most intuitive way of fine-tuning is retraining all the layers of the model with the downstream dataset. It is well known that this approach leads to better in-distribution accuracies. However, for OOD datasets when the shift is large, these may not perform very well. We use this as a benchmark to compare again other fine-tuning strategies. [$\sim$ 85M parameters]
2. **Layer freezing** (L_k): Freezing the first or last layers has become a common practice for fine-tuning in CNN vision models and other language tasks now. In our case, we freeze everything but the first two transformer blocks (L_2) for one set of experiments and similarly for the last two blocks (L_n). [$\sim$ 14M parameters]
3. **Linear probing** (*LP*): Following some of the literature Kumar et al. [2022], Chen et al. [2021] that posits larger gains from linear probing over fine-tuning in the presence of distribution shifts, we perform fine-tuning only the last MLP linear layer of each of the attention heads. [$\sim$ 28M parameters]
4. **Low-rank Matrix Factorization** (*LoRA*): The above fine-tuning methods still pose computational challenges. Approaches in NLP have shown that large-scale pre-trained models used for fine-tuning on different tasks rely on a small intrinsic dimension. In particular, we used *LoRA* Hu et al. [2021], which uses a low-rank decomposition where gradient updates are represented by $h = W_0x + \Delta W_x = W_0 + BAx$, where B and A are matrices in a reduced matrix space: $\mathbb{R}^{d\times r} \times \mathbb{R}^{r\times k}$, which can project the full original parameter space $\mathbb{R}^{d\times k}$. We use rank 4 for our experiments. We expect that we can achieve comparable performance by reducing the trainable parameters by more than 99%. [$\sim$ 300K parameters]

Each strategy will use a *ViT* (`google/vit-base-patch16-224`) with an Adam optimizer with the learning rate 1×10^{-5} with a linear-schedule and weight decay (L2 regularization) 1×10^{-3} to minimize over-fitting. All models are trained for 20 epochs.

3 Results

To compare the experiment results, we use a fine-tuned ResNet50 (pre-trained on *ImageNet-21K*) with our dataset as a baseline. This CNN achieved a F1 score of 75%. However, the score improved by 20% with the *ViT* model. The initial set of experiments indicated some overfitting so we added regularization to minimize that.

We summarize the experiment results in Table 1. We observe that *LoRA* generally performs better with the best trade-off in terms of computational cost. The F1 scores for high-resolution datasets are higher than Sentinel as expected. The performance on the California HR dataset is lower than that of China's HR dataset even though the former has more samples. The reason is California's dataset includes commercial and utility-scale PVs that are harder to detect as compared to China's dataset which has only utility-scale (commercial-scale PVs are smaller in size than utility-scale PVs).

Another interesting observation is the comparison of L_1 and L_n for China HR and S2. L_1 for China S2 performs better as opposed to high resolution datasets. This is because the data distribution domain of Sentinel 2 is significantly different than the high-resolution aerial imagery.

	FT	L_2	L_n	*LoRA*	*LP*
CA [HR]	0.90	0.84	0.89	0.89	**0.92**
CN [HR]	0.96	0.95	0.96	**0.97**	0.95
CN [S2]	0.83	0.84	0.79	**0.85**	0.75

Table 1: F1 Scores of fine-tuning experiments. Each column corresponds to a different fine-tuning strategy: FT (full fine-tuning), L_2 (First two attention blocks), L_n (Last two attention blocks), *LoRA* (Low-rank adaptation), and *LP* Linear Probing.

3.1 OOD evaluation

We also evaluated our best models on OOD datasets. The results are shown in Table 1. The model trained on China Sentinel 2 dataset (the best being LoRA) is used to test the other two datasets (China HR and Cal HR) and so on. China S2 trained model performs better on China HR than California HR by $\sim 30\%$ F1 score. This is because the PVs are more similar in China HR and S2 than PVs in China and California.

Trained	**OOD**	**Acc**	**F1**
CN-S2 [LoRa]	CN HR	**0.64**	**0.75**
CN-S2 [LoRa]	CA HR	0.46	0.46
CN-HR [LoRa]	CN S2	0.48	0.29
CN-HR [LoRa]	CA HR	0.51	0.33
CA-HR [Linear]	CN S2	0.44	0.17
CA-HR [Linear]	CN HR	0.51	0.37

Table 2: OOD inference performance for different fine-tuned models. Each row corresponds to a different combination of fine-tuned model and OOD dataset.

4 Discussion

We have analyzed various fine-tuning strategies for *ViT* architectures and found that full parameter retraining is often not required to achieve baseline performance. We also find that differing datasets and image characteristics (*i.e.* resolution, color space, etc.) call for different transfer learning methods, with *LoRA* being the best-performing strategy across datasets in our experiments. These results line up with some of the findings in the literature [Kumar et al., 2022, Chen et al., 2021]. As suggested by other papers [Goyal et al., 2022, Raghunathan et al., 2020], it is also relevant to think in terms of OOD examples and how different optimization processes between pre-training and fine-tuning can lead to different results (i.e. contrastive loss in pre-training, and cross-entropy during fine-tuning leads to sub-optimal results Goyal et al. [2022]). We would like to work towards improving the OOD evaluations as this can help us translate our PV classification parameters across different geographical regions, a especially crucial in the sustainability domain.

While other works have explored the use of few-shot meta-learners to achieve OOD performance in remote-sensing scenarios [Wang et al., 2020], this works differs from those approaches by caring solely about domain adaptation and wanting to explore a streamlined way of fine-tuning strategy for a common problem in PV detection. In this work, we have posed the opportunity of FT as a computationally simpler and less data-greedy alternative to training from scratch. We have shown that we can achieve comparable performance to our baselines, with a comparably smaller set of labels (6,200 labels). Despite our experiments combining different resolutions, we find that detecting small PV installations is still challenging, and fine-tuned pre-trained models are still not able to outperform models trained exclusively on high-resolution imagery.

References

T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. A Simple Framework for Contrastive Learning of Visual Representations. In *Proceedings of the 37th International Conference on Machine Learning*, pages 1597–1607. PMLR, Nov. 2020. URL `https://proceedings.mlr.press/v119/chen20j.html`. ISSN: 2640-3498.

X. Chen, S. Xie, and K. He. An Empirical Study of Training Self-Supervised Vision Transformers, Aug. 2021. URL `http://arxiv.org/abs/2104.02057`. arXiv:2104.02057 [cs].

J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, May 2019. URL `http://arxiv.org/abs/1810.04805`. arXiv:1810.04805 [cs].

S. Goyal, A. Kumar, S. Garg, Z. Kolter, and A. Raghunathan. Finetune like you pretrain: Improved finetuning of zero-shot vision models, Dec. 2022. URL `http://arxiv.org/abs/2212.00638`. arXiv:2212.00638 [cs].

X. He, C. Li, P. Zhang, J. Yang, and X. E. Wang. Parameter-efficient Model Adaptation for Vision Transformers, Dec. 2022. URL `http://arxiv.org/abs/2203.16329`. arXiv:2203.16329 [cs].

E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen. LoRA: Low-Rank Adaptation of Large Language Models, Oct. 2021. URL `http://arxiv.org/abs/2106.09685`. arXiv:2106.09685 [cs].

L. Kruitwagen, K. T. Story, J. Friedrich, L. Byers, S. Skillman, and C. Hepburn. A global inventory of photovoltaic solar energy generating units. *Nature*, 598(7882):604–610, Oct. 2021. ISSN 1476-4687. doi: 10.1038/s41586-021-03957-7. URL `https://www.nature.com/articles/s41586-021-03957-7`. Number: 7882 Publisher: Nature Publishing Group.

A. Kumar, A. Raghunathan, R. Jones, T. Ma, and P. Liang. Fine-Tuning can Distort Pretrained Features and Underperform Out-of-Distribution, Feb. 2022. URL `http://arxiv.org/abs/2202.10054`. arXiv:2202.10054 [cs].

C. L. Kwan. Influence of local environmental, social, economic and political variables on the spatial distribution of residential solar PV arrays across the United States. *Energy Policy*, 47:332–344, Aug. 2012. ISSN 0301-4215. doi: 10.1016/j.enpol.2012.04.074. URL `https://www.sciencedirect.com/science/article/pii/S0301421512003795`.

J. Lee, R. Tang, and J. Lin. What Would Elsa Do? Freezing Layers During Transformer Fine-Tuning, Nov. 2019. URL `http://arxiv.org/abs/1911.03090`. arXiv:1911.03090 [cs].

Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. RoBERTa: A Robustly Optimized BERT Pretraining Approach, July 2019. URL `http://arxiv.org/abs/1907.11692`. arXiv:1907.11692 [cs].

B. R. Lukanov and E. M. Krieger. Distributed solar and environmental justice: Exploring the demographic and socio-economic trends of residential PV adoption in California. *Energy Policy*, 134:110935, Nov. 2019. ISSN 0301-4215. doi: 10.1016/j.enpol.2019.110935. URL `https://www.sciencedirect.com/science/article/pii/S0301421519305221`.

NREL. Sumer 2023 Solar Industry Update. Technical report, National Reneweable Energy Laboratory, Aug. 2023. URL `https://www.nrel.gov/docs/fy23osti/87189.pdf`.

D. Patterson, J. Gonzalez, Q. Le, C. Liang, L.-M. Munguia, D. Rothchild, D. So, M. Texier, and J. Dean. Carbon Emissions and Large Neural Network Training, Apr. 2021. URL `http://arxiv.org/abs/2104.10350`. arXiv:2104.10350 [cs].

A. Raghunathan, S. M. Xie, F. Yang, J. Duchi, and P. Liang. Understanding and Mitigating the Tradeoff Between Robustness and Accuracy, July 2020. URL `http://arxiv.org/abs/2002.10716`. arXiv:2002.10716 [cs, stat].

R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni. Green AI, Aug. 2019. URL `http://arxiv.org/abs/1907.10597`. arXiv:1907.10597 [cs, stat].

L. C. Stokes, E. Franzblau, J. R. Lovering, and C. Miljanich. Prevalence and predictors of wind energy opposition in North America. *Proceedings of the National Academy of Sciences*, 120(40): e2302313120, Oct. 2023. doi: 10.1073/pnas.2302313120. URL https://www.pnas.org/doi/abs/10.1073/pnas.2302313120. Company: National Academy of Sciences Distributor: National Academy of Sciences Institution: National Academy of Sciences Label: National Academy of Sciences Publisher: Proceedings of the National Academy of Sciences.

D. A. Sunter, S. Castellanos, and D. M. Kammen. Disparities in rooftop photovoltaics deployment in the United States by race and ethnicity. *Nature Sustainability*, 2(1):71–76, Jan. 2019. ISSN 2398-9629. doi: 10.1038/s41893-018-0204-z. URL https://www.nature.com/articles/s41893-018-0204-z/. Number: 1 Publisher: Nature Publishing Group.

S. Wang, M. Rußwurm, M. Körner, and D. B. Lobell. Meta-Learning For Few-Shot Time Series Classification. In *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, pages 7041–7044, Sept. 2020. doi: 10.1109/IGARSS39084.2020.9441016. URL https://ieeexplore.ieee.org/document/9441016. ISSN: 2153-7003.

Z. Wang, M.-L. Arlt, C. Zanocco, A. Majumdar, and R. Rajagopal. DeepSolar++: Understanding residential solar adoption trajectories with computer vision and technology diffusion models. *Joule*, 6(11):2611–2625, Nov. 2022. ISSN 2542-4785, 2542-4351. doi: 10.1016/j.joule.2022.09.011. URL https://www.cell.com/joule/abstract/S2542-4351(22)00477-9. Publisher: Elsevier.

C. Wei, S. M. Xie, and T. Ma. Why Do Pretrained Language Models Help in Downstream Tasks? An Analysis of Head and Prompt Tuning. Nov. 2021. URL https://openreview.net/forum?id=MDMV2SxCboX.

J. Yu, Z. Wang, A. Majumdar, and R. Rajagopal. DeepSolar: A Machine Learning Framework to Efficiently Construct a Solar Deployment Database in the United States. *Joule*, 2(12):2605–2617, Dec. 2018. ISSN 2542-4351. doi: 10.1016/j.joule.2018.11.021. URL https://www.sciencedirect.com/science/article/pii/S2542435118305701.

K. Zhang and S. Bowman. Language Modeling Teaches You More than Translation Does: Lessons Learned Through Auxiliary Syntactic Task Analysis. In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 359–361, Brussels, Belgium, Nov. 2018. Association for Computational Linguistics. doi: 10.18653/v1/W18-5448. URL https://aclanthology.org/W18-5448.