SAND2024-05336C

Sandia National Laboratories

Exceptional service in the national interest

Benchmarking Transferability Metrics for SAR ATR

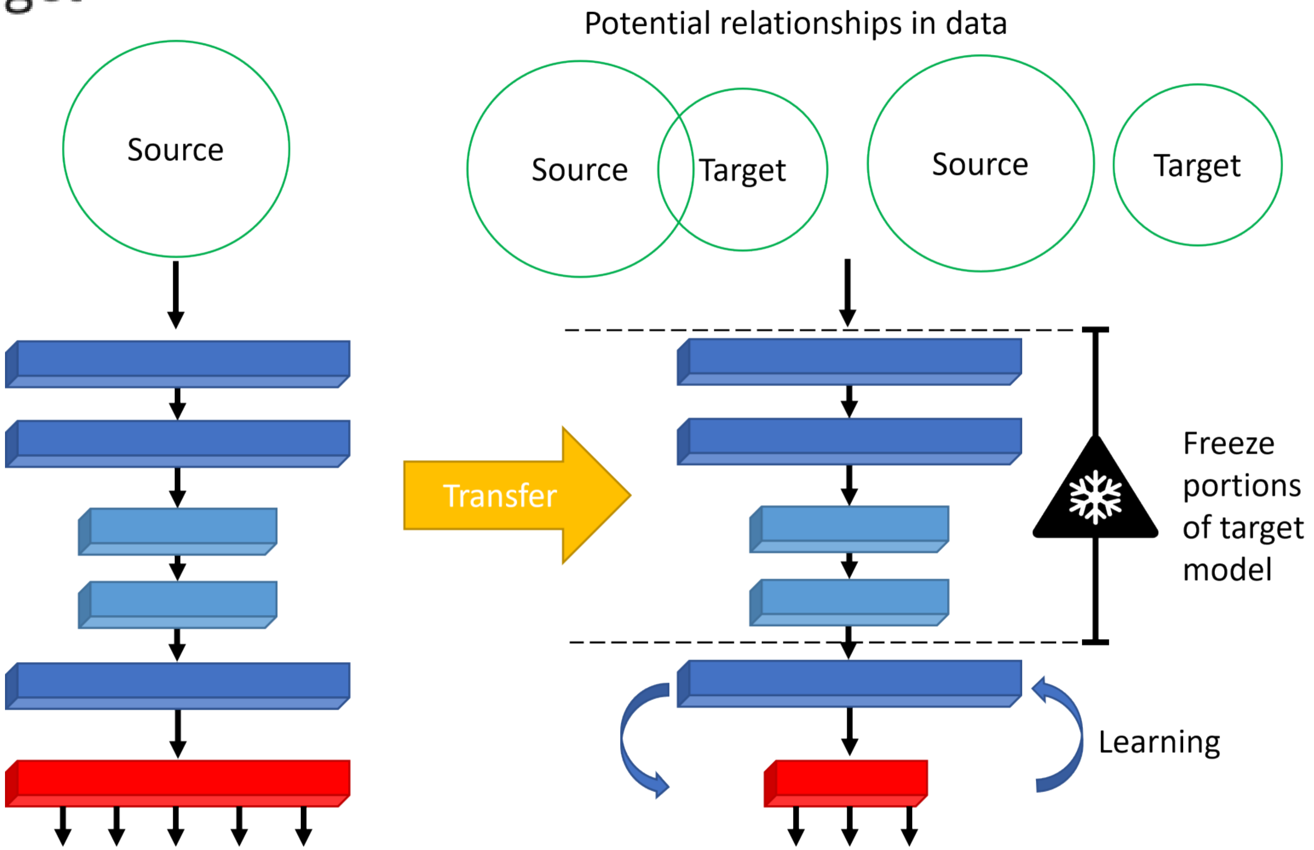
By: Johannes Bauer, Efrain H. Gonzalez, William M. Severa, and Craig M. Vineyard

U.S. DEPARTMENT OF ENERGY NNSA National Nuclear Security Administration

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Transfer Learning

- Source dataset & Target dataset (Ex: MNIST→SAMPLE)
- Pre-train on Source → Retrain on Target
- Advantages:
  - High performance on target
  - Less target data needed
  - Reduced training time
- Choices:
  - Type of model
  - Source dataset
  - Type of retraining



(Zhuang et al. , 2020), (Weiss et al., 2016), (Pan et al., 2009)

Some Transfer Learning Choices

- Dataset:
  - Homogenous transfer (Ex: MSTAR→SAMPLE)
  - Heterogenous transfer (Ex: CIFAR10→SAMPLE)
- Retraining:
  - Retrain Head
  - Fine-tune
- What makes a good source dataset?
- How do you choose a model?

(Zhuang et al. , 2020), (Nguyen et al., 2020)

Transfer Learning & SAR ATR

- Domain Limitations:
  - Dataset size! (Ex: CIFAR10 has 50,000 vs SAMPLE has 1,366)
  - Too many sensing interaction effects (azimuth angle, background, etc.)
- Non-SAR to SAR:
  - ImageNet to MSTAR
  - CIFAR10 to TerraSAR
- Synthetic to Measured
- Are metrics helpful for transfer learning in the SAR ATR domain?

(Krizhevsky et al. , 2009), (Lewis et al. , 2019), (Al Mufti et al. , 2018), (Kang et al., 2016)

Methods: Model & Training

- Model: ResNet18
- Bayesian search for hyperparameters
- Fine-tune: full model retraining
- 30 pre-trained models per source dataset
- 30 fine-tuned models per combination

Parameter	Value
Input Size	64×64
Image resize	Center Crop
Batch Size	192
Pre-training Learning Rate	0.001
Fine-tuning Learning Rate	0.00001-0.01
Random Equalization Probability	0.95
Epochs	10-100
Optimizer	Adam
Weight Initialization	Kaiming Normal

Methods: Datasets

SAR Datasets

- SAMPLE
- MSTAR
- MSTAR (1st half)
- MSTAR (2nd half)
- UNICORN (SAR)

Non-SAR Datasets

- MNIST
- CIFAR10
- UNICORN (EO)
- Overhead MNIST

(Krizhevsky et al. , 2009), (Lewis et al. , 2019), (Deng, 2012), (Leong et al., 2019), (Noever et al., 2021)

Methods: SAR Dataset Discussion

Synthetic To Measured

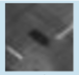
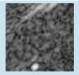
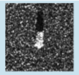
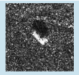
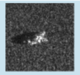
- SAMPLE: Training set vs Test set
  - SAMPLE→MSTAR1
  - SAMPLE→MSTAR2

Fine-tune Train to Test Datasets

- MSTAR (1st half)
  - 15th – 17th degree
- MSTAR (2nd half)
  - 15th – 17th degree
- SAMPLE
  - Synthetic to Measured

(Lewis et al. , 2019)

Methods: Dataset Specifics

Dataset	CIFAR10	MNIST	OverheadMNIST	UNICORN		MSTAR		SAMPLE	
				EO	SAR	MSTAR (1st Half)	MSTAR (2nd Half)	Synthetic	Measured
Class	Airplane	0	Storage tanks	Sedan	Sedan	BTR60		M1	M1
	Automobile	1	Parking lots	SUV	SUV	BRDM2		M2	M2
	Bird	2	Ships	Pickup truck	Pickup truck	D7		M35	M35
	Cat	3	Helicopter	Van	Van	T62		M548	M548
	Deer	4	Car	Box truck	Box truck	ZIL131		M60	M60
	Dog	5	Stadium	Motorcycle	Motorcycle		BMP2	BMP2	BMP2
	Frog	6	Oil gas field	Flatbed truck	Flatbed truck		BTR70	BTR70	BTR70
	Horse	7	Runway mark	Bus	Bus		T72	T72	T72
	Ship	8	Plane	Pickup truck w/ trailer	Pickup truck w/ trailer		2S1	2S1	2S1
	Truck	9	Harbour	Flatbed truck w/ trailer	Flatbed truck w/ trailer		ZSU234	ZSU234	ZSU234
Image size	32x32	28x28	28x28	31x31	55x55	128x128	128x128	128x128	128x128
Amount of training data	50000	60000	8519	459262	459262	1448	2220	1366	NA
Amount of testing data	10000	10000	1065	2000	2000	1290	1913	NA	1366
Example Image	 Airplane	 9	 Car	 Pickup truck	 Pickup truck	 BTR60	 2S1	 T72	 T72

(Krizhevsky et al. , 2009), (Lewis et al. , 2019), (Deng, 2012), (Leong et al., 2019), (Noever et al., 2021)

Methods: Dataset Combinations

Non-SAR to SAR	
Source	**Target**
CIFAR10	MSTAR1
CIFAR10	MSTAR2
CIFAR10	SAMPLE
MNIST	MSTAR1
MNIST	MSTAR2
MNIST	SAMPLE
OMNIST	MSTAR1
OMNIST	MSTAR2
OMNIST	SAMPLE
UNICORN (EO)	MSTAR1
UNICORN	MSTAR2

SAR to SAR	
Source	**Target**
MSTAR	SAMPLE
MSTAR1	MSTAR2
MSTAR1	SAMPLE
SAMPLE	MSTAR1
SAMPLE	MSTAR2
UNICORN (SAR)	MSTAR1
UNICORN (SAR)	MSTAR2
UNICORN (SAR)	SAMPLE

Transferability Metrics

- **H-score:** information-theoretic metric that is used to calculate transferability from a source task S to a target task T for some feature function f_S related to S. Higher values are better.

$$\mathcal{H}(f_S) = tr\left(cov(f_S(X))^{-1} cov\left(\mathbb{E}\left[f_S(X)|Y\right]\right)\right)$$

- **Log Maximum Evidence (LogMe):** a measure of the relationship between the source model features and the labels of the target dataset. A normalization of the logarithm of the marginalized likelihood. Higher values are better.

- **Gaussian Bhattacharyya Coefficient (GBC):** measures the ease with which classes in the target dataset can be separated when using the features from the source model. Higher values are better.

$$\mathrm{GBC} = -\sum_{i,j:\ i \neq j} \int \sqrt{p_{ci}(x) p_{cj}(x)} dx$$

(Alvarez-Melis et al. , 2020), (Bao et al. , 2019), (You et al., 2021)

Transferability Metrics

- **Optimal Transport Dataset Distance (OTDD):** Euclidean distance between features & OT for distance between labels, then OT for distance between datasets

$$\mathrm{OTDD}(D_A, D_B) = \min_{\pi \in \Pi(\alpha,\beta)} \int_{\mathcal{Z}\times\mathcal{Z}} d_{\mathcal{Z}}(z, z')^p \, \mathrm{d}\pi(z, z')$$

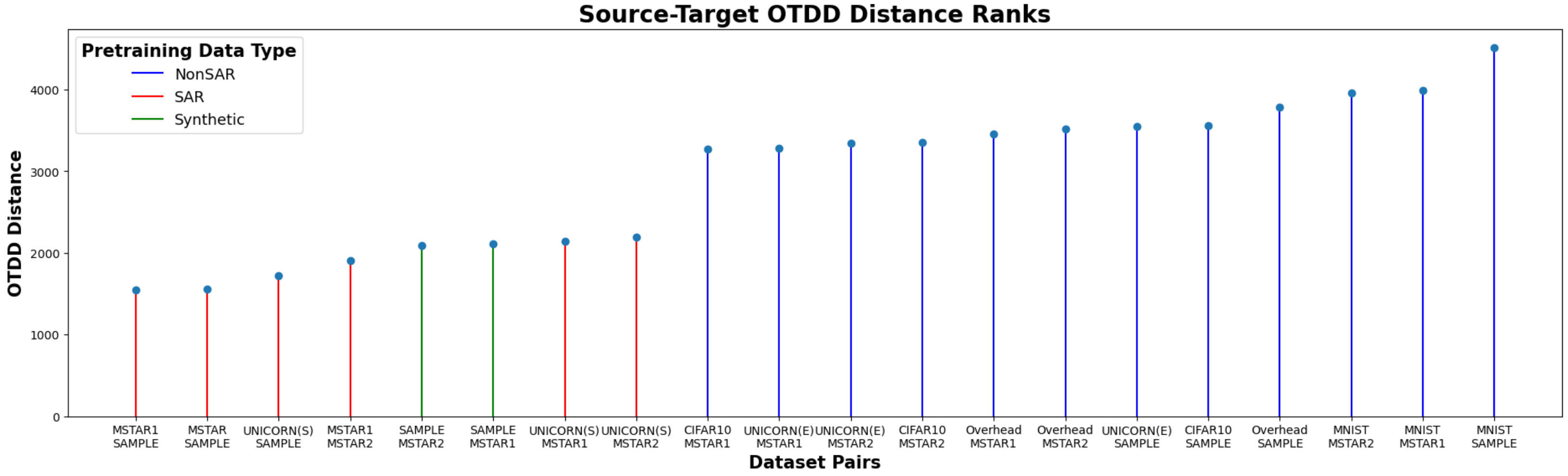
- **Log Expected Empirical Predication (LEEP):** measures a source model's ability to use the encoding of the source dataset to assign labels to a target dataset. Higher values are better.

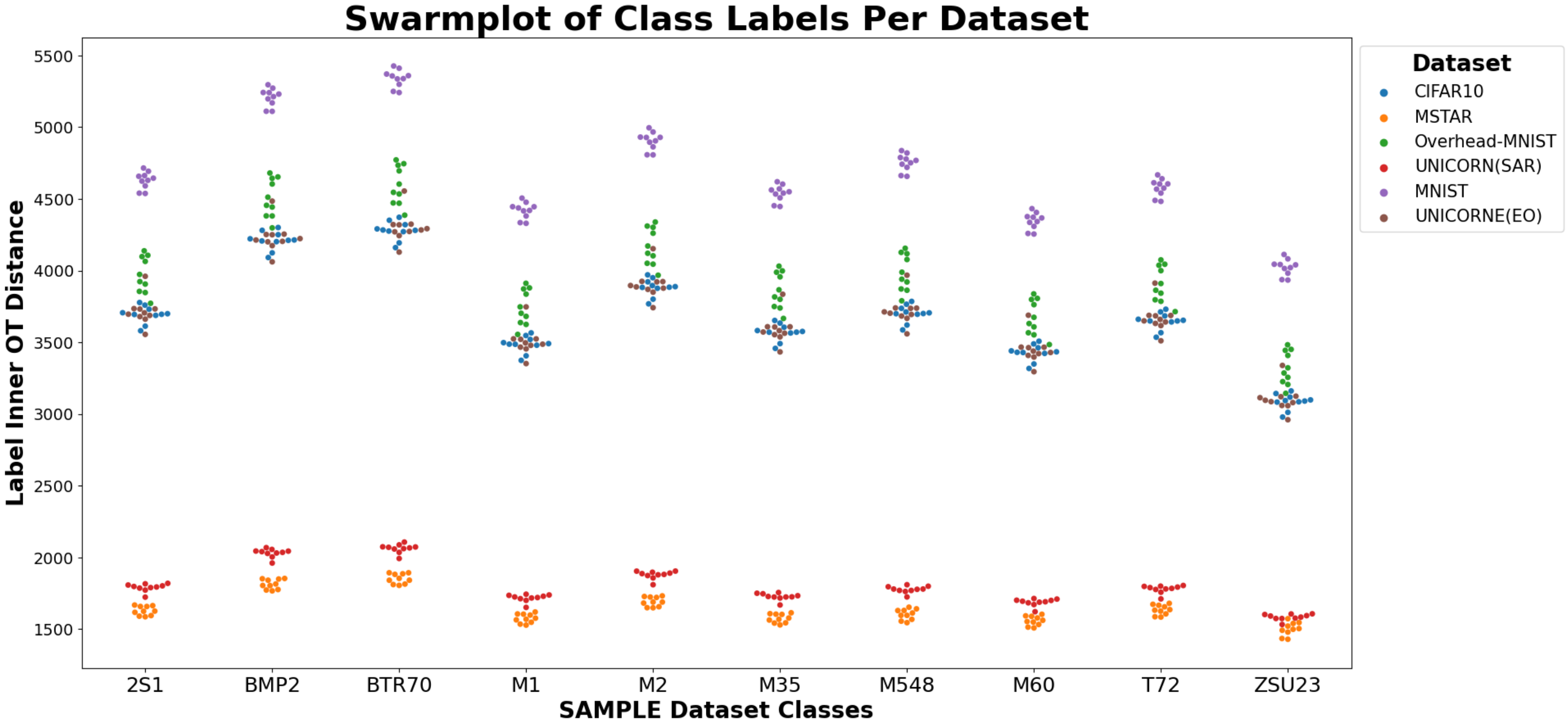
$$\mathrm{LEEP}(f, D) = \frac{1}{n} \sum_{i=1}^{n} \log \left(\sum_{z \in Z} \hat{P}\left(y_i | z\right) f\left(x_i\right)_z \right)$$

(Pandy et al., 2022), (Nguyen et al., 2020)

Results: OTDD

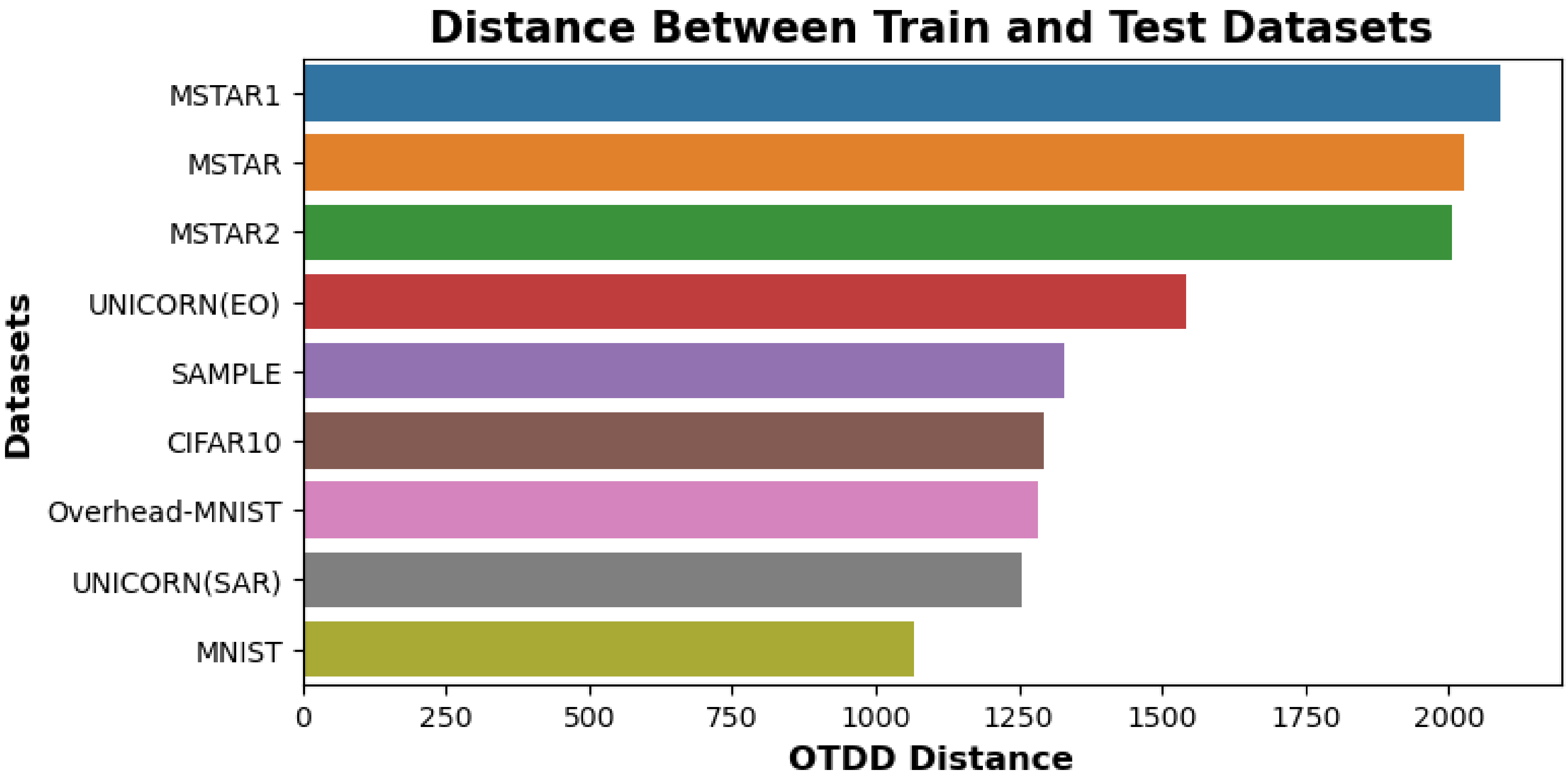
- SS closer than NS





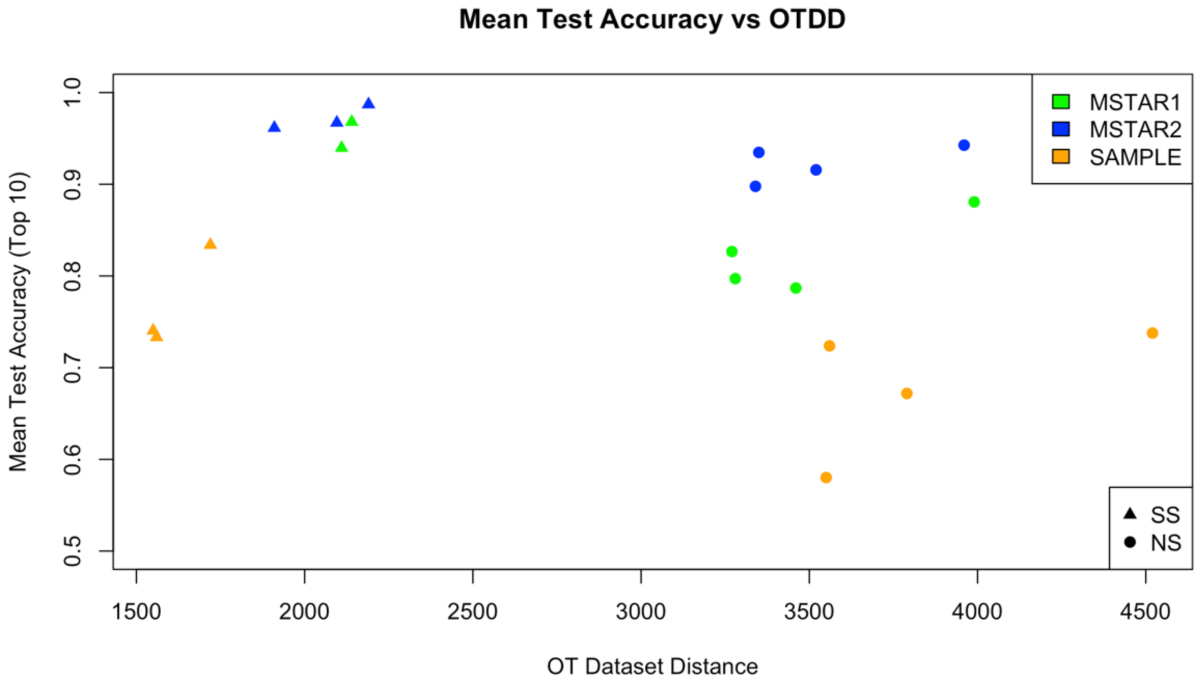
Results: OTDD

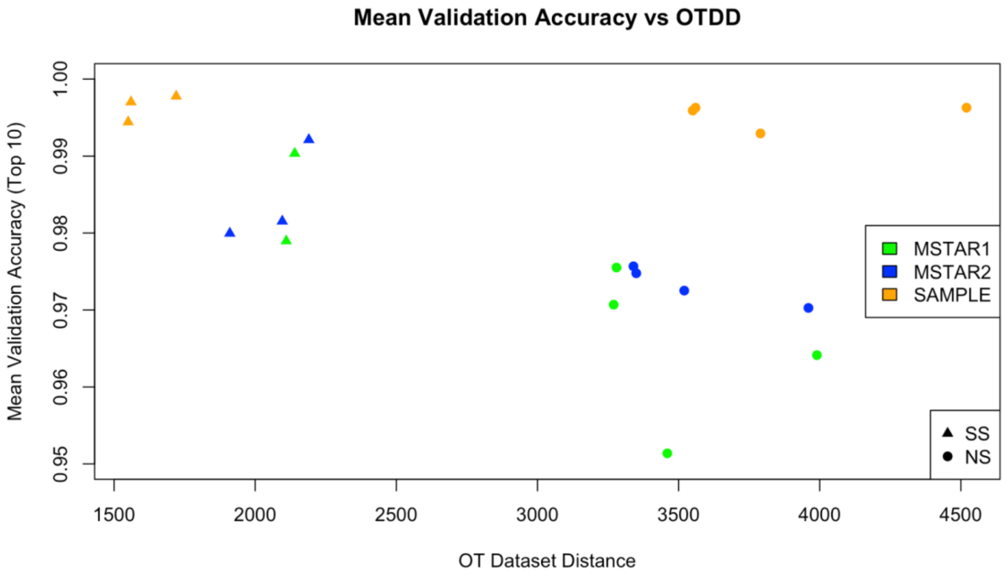
- Difference in azimuth angle leads to a larger dataset distance

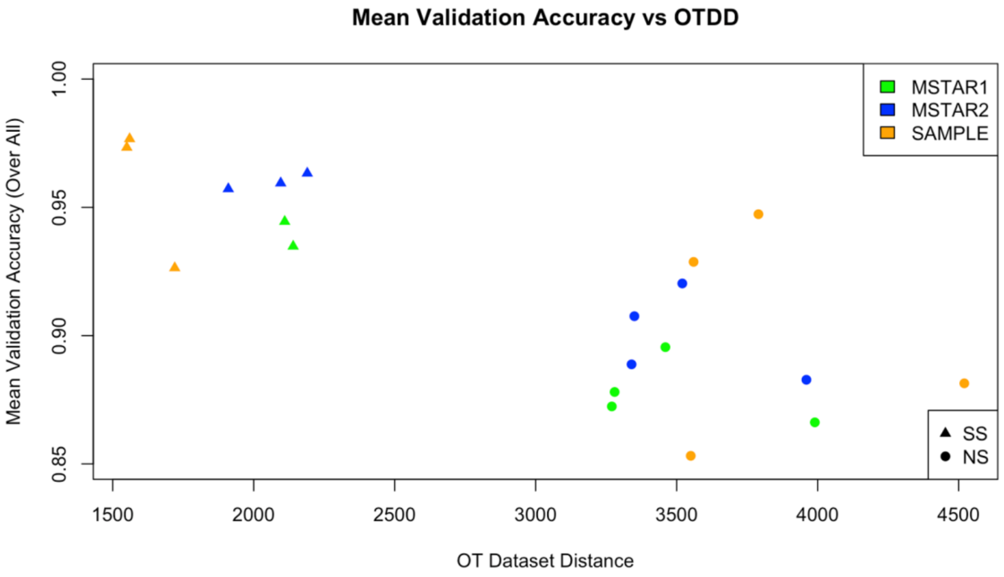


Results: Accuracy vs OTDD

- Lower OTDD = Higher Accuracy
- Top 10 fine-tuned models for best pre-trained models
- Overall: 30 fine-tuned models for best pre-trained models

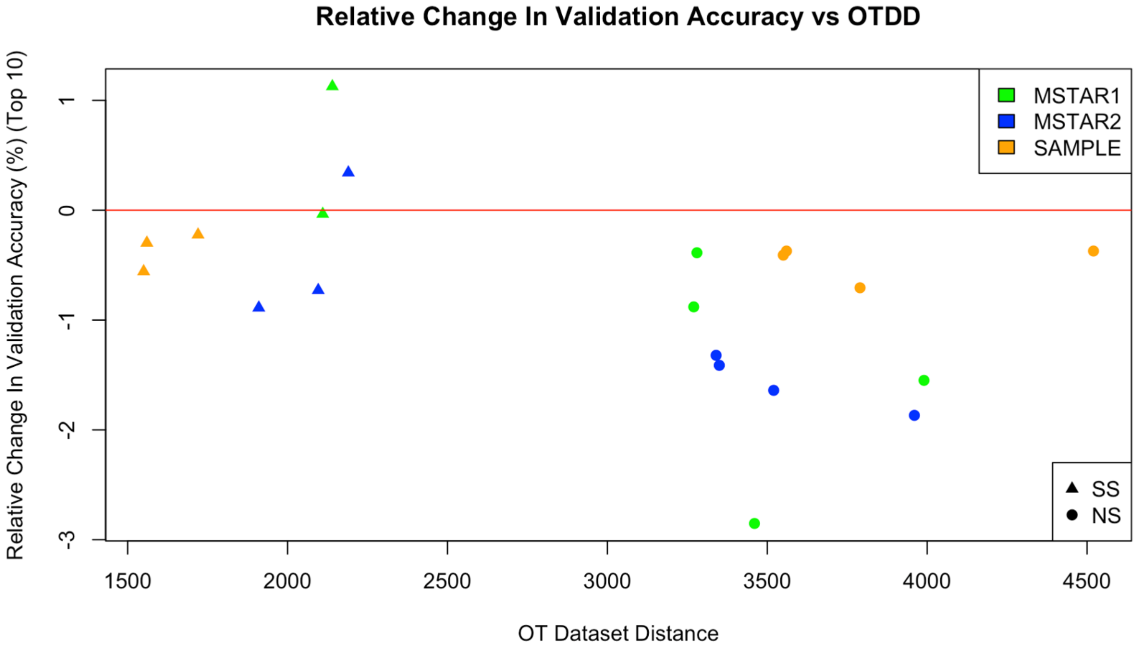


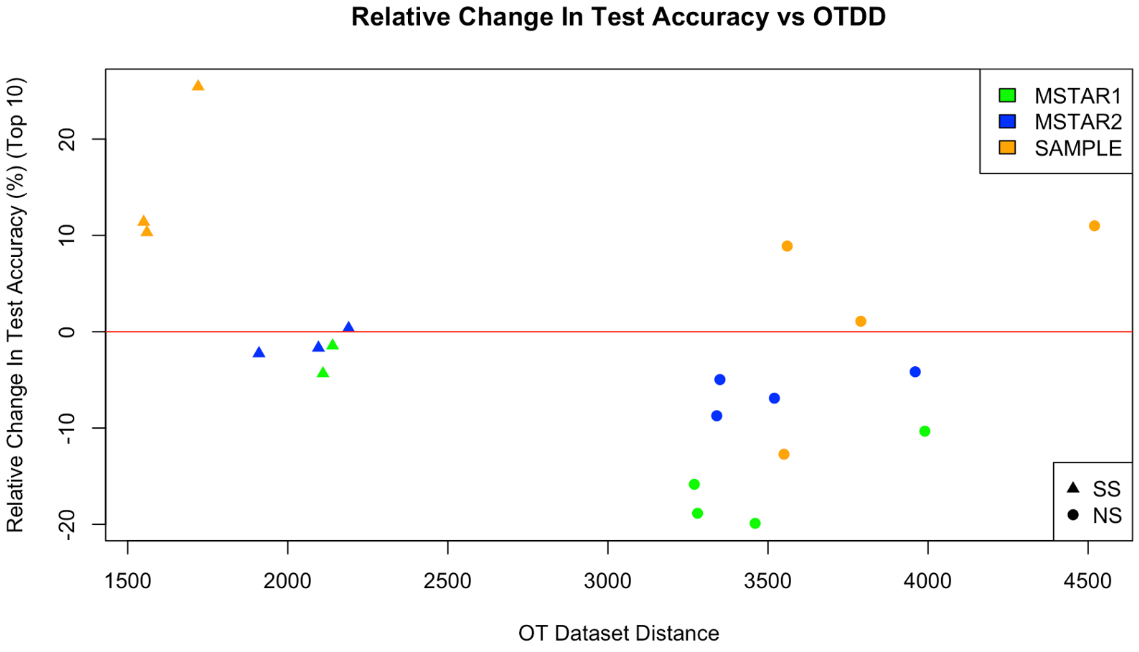




Results: Relative Change In Accuracy vs OTDD

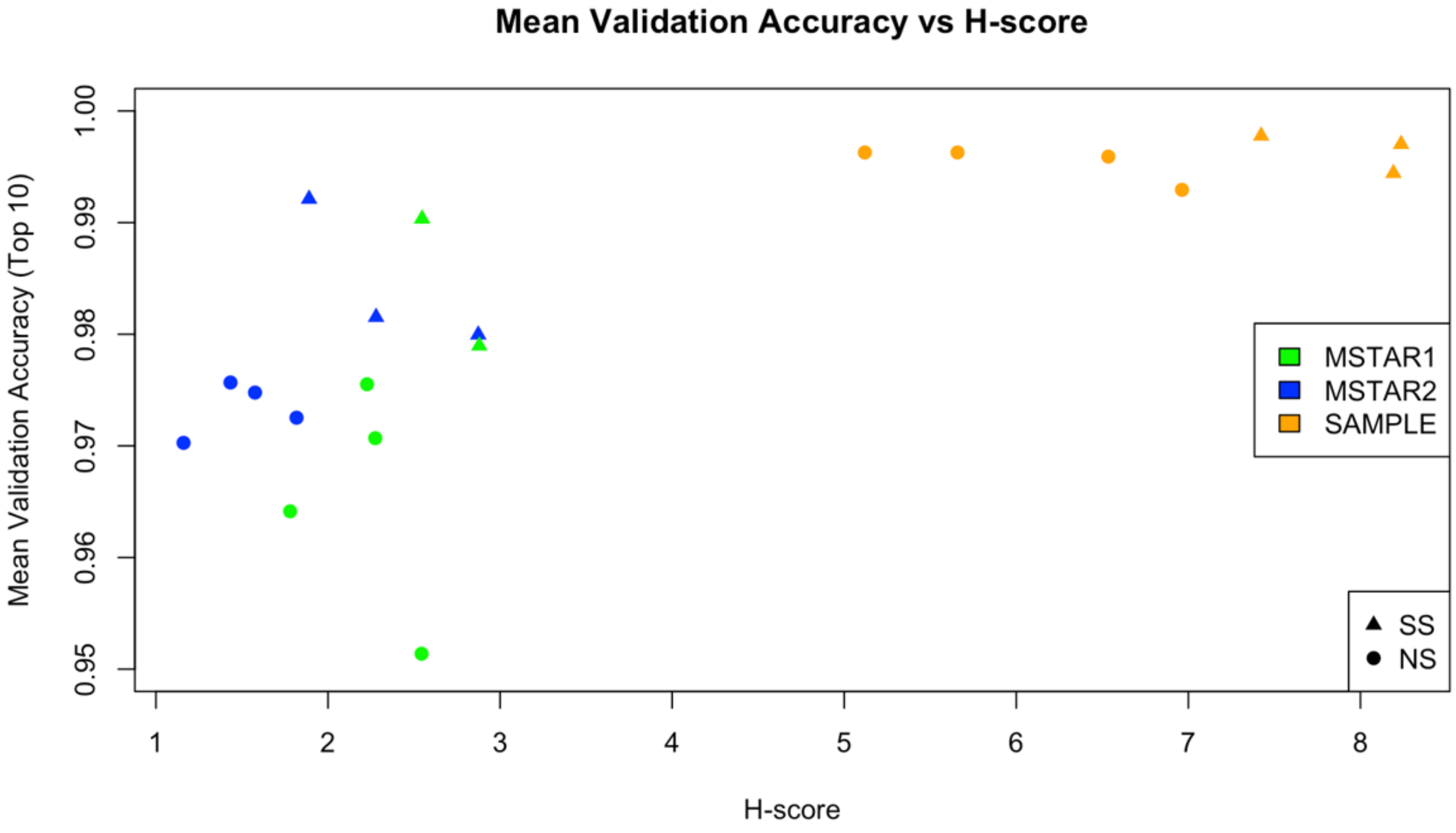
- Relative Change In Accuracy (%) = $100 \times \frac{Accuracy_{S \to T} - Accuracy_T}{Accuracy_T}$
- Positive = good, Negative = bad





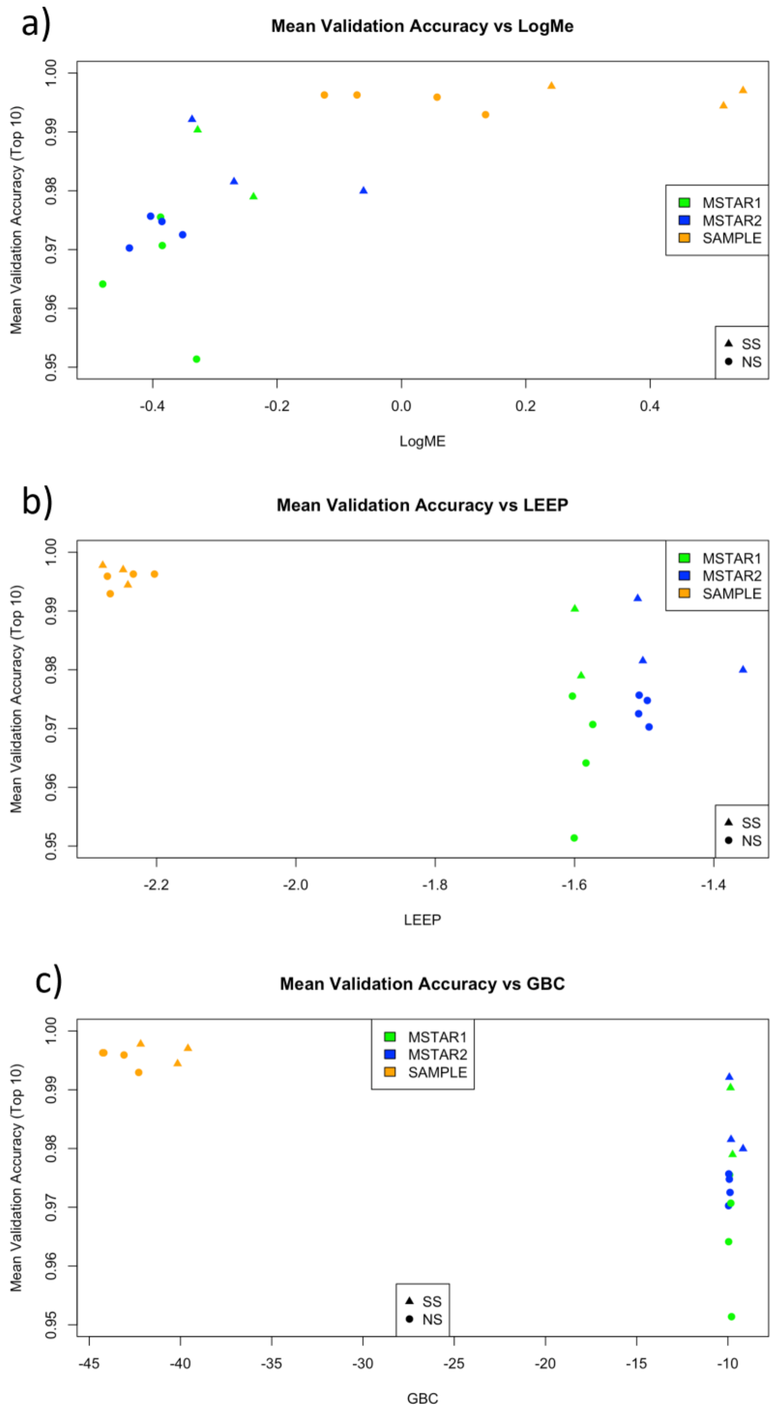
Results: H-score

- Higher H-score = Higher Accuracy



Results: LogMe & LEEP & GBC

- Higher LogMe = Higher Accuracy
- LEEP and GBC patterns unclear
- GBC assumes that per class distribution is Gaussian
- LEEP affected by number of classes



(Pandy et al., 2022), (Ibrahim et al., 2023)

Caveats & Conclusion

- Caveat 1: Averages over searched models
- Caveat 2: Only one neural network architecture
- Caveat 3: Transferability Metrics based on Training set to Training set transfer
- Caveat 4: OTDD values used were based on an upper bound
- Caveat 5: OTDD calculation require images of same size
- Caveat 6: Models that performed well on validation data skewed results
- Caveat 7: Models that performed poorly on test data skewed results

- Conclusion: Transfer learning is not straightforward

Thank You To:

- You all for listening
- The organizers of this event
- Collaborators: Johannes Bauer, William M. Severa, and Craig M. Vineyard
- Sandia National Laboratories

References:

[1] Maha Al Mufti, Esra Al Hadhrami, Bilal Taha, and Naoufel Werghi. Sar automatic target recognition using transfer learning approach. In *2018 International Conference on Intelligent Autonomous Systems (ICoIAS)*, pages 1–4, 2018.

[2] David Alvarez-Melis and Nicolò Fusi. Geometric dataset distances via optimal transport, 2020.

[3] Yajie Bao, Yang Li, Shao-Lun Huang, Lin Zhang, Lizhong Zheng, Amir Zamir, and Leonidas Guibas. An information-theoretic approach to transferability in task transfer learning. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 2309–2313, 2019.

[4] Li Deng. The mnist database of handwritten digit images for machine learning research [best of the web]. *IEEE Signal Processing Magazine*, 29(6):141–142, 2012.

[5] Shibal Ibrahim, Natalia Ponomareva, and Rahul Mazumder. Newer is not always better: Rethinking transferability metrics, their peculiarities, stability and performance. In *Machine Learning and Knowledge Discovery in Databases*, pages 693–709. Springer International Publishing, 2023.

[6] Chenyao Kang and Chu He. Sar image classification based on the multi-layer network and transfer learning of mid-level representations. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1146–1149, 2016.

[7] Alex Krizhevsky. Learning multiple layers of features from tiny images. 2009.

References:

[8] Colin Leong, Todd Rovito, Olga Mendoza-Schrock, Christopher Menart, Jason Bowser, Linda Moore, Steve Scarborough, Michael Minardi, and David Hascher. Unified coincident optical and radar for recognition (unicorn) 2008 dataset, 2019.

[9] Benjamin P. Lewis, Theresa Scarnati, Elizabeth Sudkamp, John W. Nehrbass, Stephen Rosencrantz, and Edmund G. Zelnio. A sar dataset for atr development: the synthetic and measured paired labeled experiment (sample). In *Defense + Commercial Sensing*, 2019.

[10] Cuong V. Nguyen, Tal Hassner, Matthias Seeger, and Cedric Archambeau. Leep: A new measure to evaluate transferability of learned representations, 2020.

[11] David Noever and Samantha E. Miller Noever. Overhead mnist: A benchmark satellite dataset, 2021.

[12] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

[13] Michal Pándy, Andrea Agostinelli, Jasper Uijlings, Vittorio Ferrari, and Thomas Mensink. Transferability estimation using bhattacharyya class separability. 2022.

[14] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big data*, 3:1–40, 2016.

[15] Kaichao You, Yong Liu, Jianmin Wang, and Mingsheng Long. Logme: Practical assessment of pre-trained models for transfer learning. 2021.

[16] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. 2020.