Self-Supervised Learning for SAR Target Recognition with Multi-Task Pretext Training

Paper ID #348 (1571116074)

Md Al Siam, Dewan Fahim Noor

Department of Electrical & Computer Engineering
Tuskegee University

IEEE Southeast Conference 2025
March 27-30, 2025

TUSKEGEE UNIVERSITY

Outline

- Introduction
- Objectives and Key Contributions
- Related Works
- Proposed Framework
- Dataset Description
- Data Processing & Pretext Tasks
- CNN Architecture
- Downstream Classification
- Results
- Comparison with Literature
- Conclusion

Introduction

- **Synthetic Aperture Radar (SAR) Images**
  - **Operates in all weather conditions** – Penetrates clouds, rain, and fog for uninterrupted imaging
  - **Day and night functionality** – Use of microwave signals enables imaging in complete darkness
  - **Consistent high resolution** – Able to maintains clarity regardless of weather, lighting, or atmospheric conditions.

Figure 1: McDonnell Douglas MD-80 aircraft and Airbus A300-600R jetliners at the Roswell Air Center in New Mexico (Photo: breakingdefense.com)

Introduction

- **SAR Imaging Importance**
  - **Dual-Use Technology** – Essential for both military (e.g., reconnaissance) and civilian (e.g. disaster response) applications.

- **Key Challenges**
  - **Limited labeled data** - Scarcity of annotated SAR datasets
  - **Synthetic vs. Real Data Discrepancy** – Significant domain gap between synthetic and real-world SAR imager
  - **Target variability** – Object configurations image SAR image appearance
  - **Sensor variability** – sensor settings affects outputs.
  - **Environmental factors** - Dependence on Weather/meteorological conditions influence SAR image interpretation

Introduction

- **Synthetic Data Limitations**
  - Synthetic data cannot fully replicate real-world SAR phenomenology
  - Synthetic SAR data, even when carefully generated, fails to fully bridge the gap to real-world sensor data, leading to suboptimal performance

"Persistent domain gaps in even meticulously truthed synthetic data cannot provide satisfactory ATR performance"

Figure 2: Performance degradation when synthetic data exceeds 60% of training data.

Objectives and Key Contributions

- **Objectives**
  - Eliminate dependency on synthetic data
  - Improve performance with limited labeled data
  - Develop a self-supervised learning (SSL) framework for SAR target recognition

- **Key Contributions**
  - Developed a novel self-supervised learning framework
  - Attained competitive performance on the SAMPLE dataset with significant reduction of data requirements (68.4% of available measured data)
  - Elimination of dependency on synthetic data while achieving state-of-the-art performance

Related Works

- **Traditional SAR ATR Methods:**
  - Template matching and pattern recognition
  - Limited adaptability in uncontrolled settings *(Fu et al. 2018)*
- **Machine Learning Approaches:**
  - Feature extraction (geometric, electromagnetic, transform features)
  - Classifiers like SVM, KNN, and neural networks *(Li et al. 2023)*
- **Deep Learning in SAR ATR:**
  - Early CNN architectures achieved > 92% accuracy on MSTAR dataset
  - Hybrid methods combining CNNs with physics-based features (Morgan et al. 2015, Chen et al. 2016)
- **Data Augmentation and Transfer Learning:**
  - Techniques to address limited labeled data (Ding et al. 2017, Furukawa et al. 2017)
  - Synthetic data generation and domain adaptation (Marmanis et al. 2017, Huang et al. 2019)

Proposed Framework

- **Self-Supervised Learning Framework:**
  - Multi-task pretext training to learn robust features
  - Eliminates need for synthetic data
- **Key Components:**
  - **Input Processing:** Normalized grayscale SAR images
  - **Pretext Tasks:** Nine transformation tasks (e.g., rotation, flipping, denoising)
  - **CNN Architecture:** Four convolutional blocks, dense layer for feature embedding
  - **Downstream Classification:** Multiple classifiers (SVM, XGBoost, Random Forest)
- **Why Self-Supervised Learning?**
  - SSL learns robust features from unlabeled data
  - Reduces reliance on costly labeled data

Proposed Framework

1. Input Processing:
Converting raw SAR images to normalized grayscale format

2. Pretext Tasks:
Use nine distinct transformation tasks for subsequent facilitate self-supervised learning

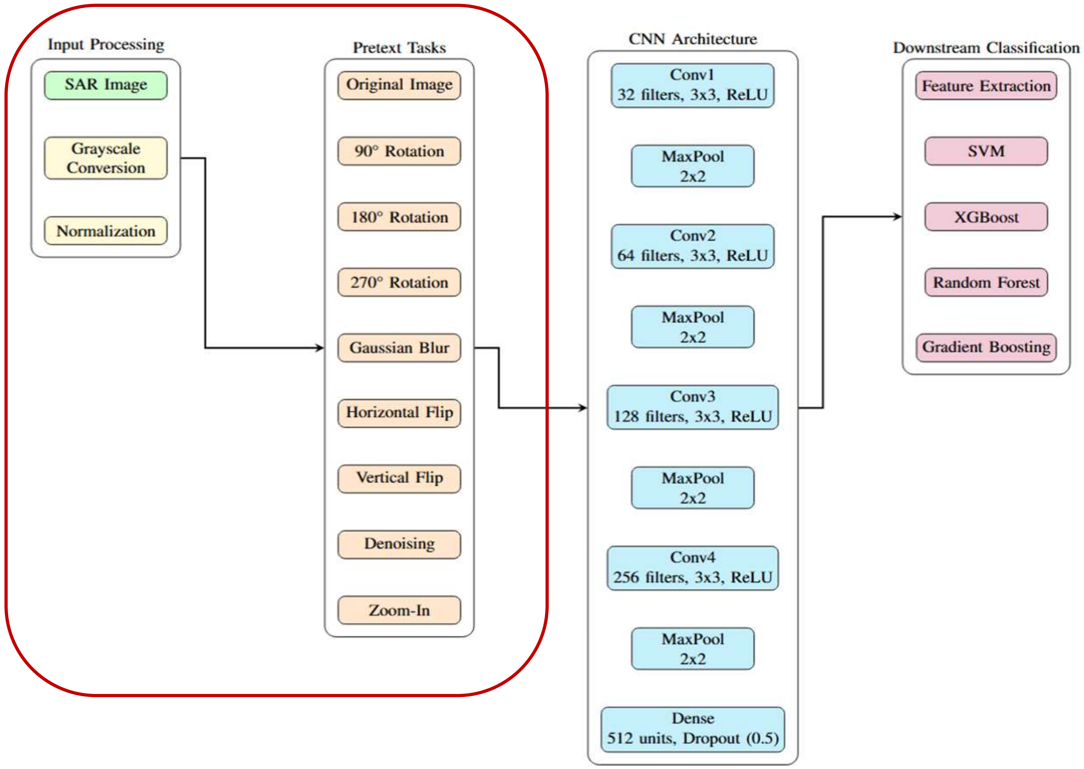

Figure 3: Proposed SSL Framework

Proposed Framework

3. CNN Architecture:
Specialized network with four convolutional blocks and dense layer for feature learning

4. Downstream Classification:
Multiple classifier evaluation using extracted features

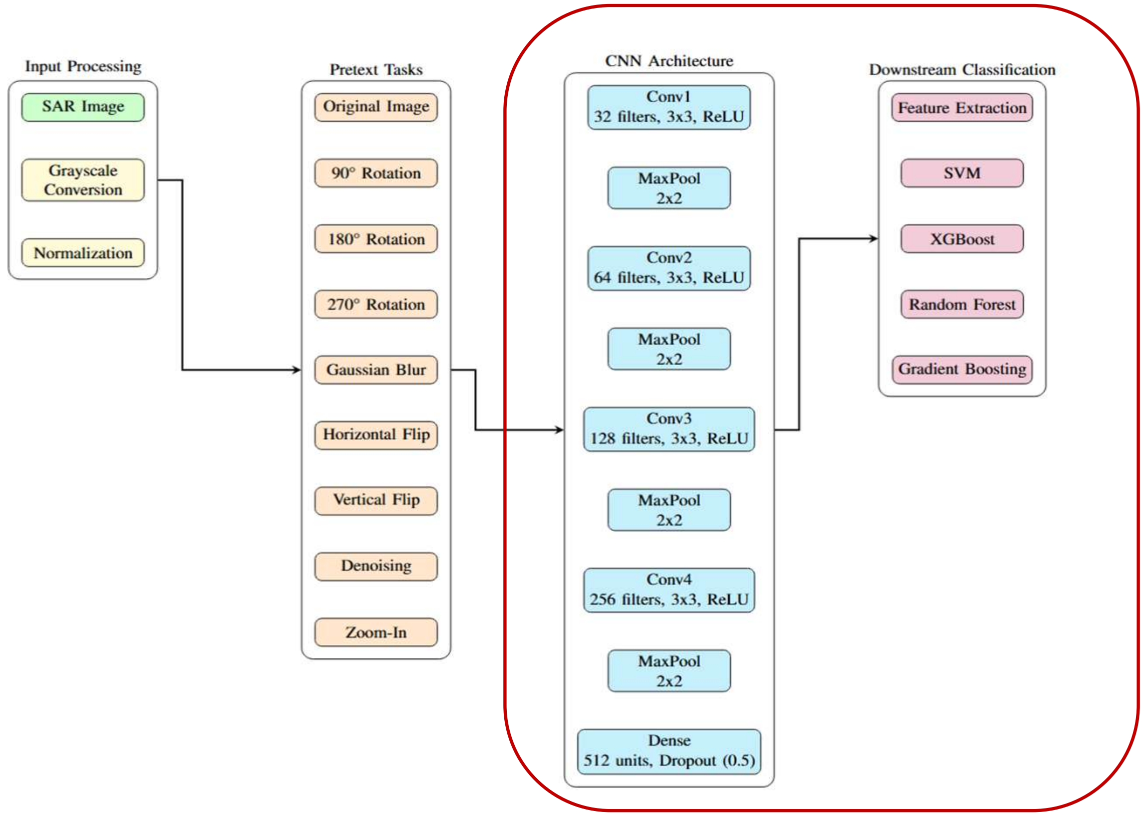

Figure 3: Proposed SSL Framework

Dataset Description

- **SAMPLE Dataset:**
  - Synthetic and Measured Paired and Labeled Experiment Dataset
  - Synthetic and measured SAR images of 10 military vehicles by Lewis et al.
  - Azimuth angles: 10° to 80°, elevation angles: 14° to 17°
  - Publicly available on GitHub

- **Data Split:**
  - 920 measured images (68.4% of measured 1345 images)
  - 20% held out for testing (184 images)
  - 736 images for training (**54.72% of measured**)

Table 1: Distribution of SAMPLE Dataset

Class	Measured	Synthetic	Total
M60	176	176	352
2S1	174	174	348
BTR70	92	92	184
M548	128	128	256
M35	129	129	258
BMP2	107	107	214
M2	128	128	256
ZSU23	174	174	348
M1	129	129	258
T72	108	108	216
Total	1345	1345	2690

Dataset Description

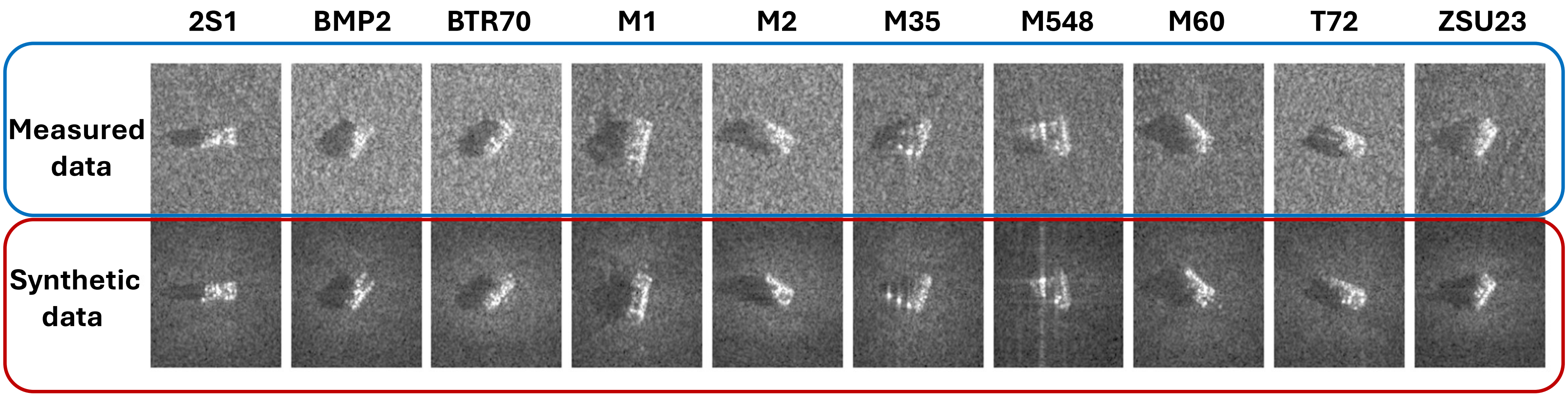

Figure 4: Measured Data (First Row) and Corresponding Synthetic Data (Second Row) in SAMPLE Dataset (Lewis et al., 2019)

Data Processing & Pretext Tasks

- **Transformation Tasks:**
  - Original Image Preservation
  - 90°, 180°, 270° Rotation
  - Gaussian Blur
  - Horizontal and Vertical Flip
  - Denoising (BM3D algorithm)
  - Zoom-In Transformation

- **Purpose:**
  - Learn viewpoint-invariant and orientation-invariant features
  - Enhance robustness to noise and resolution changes

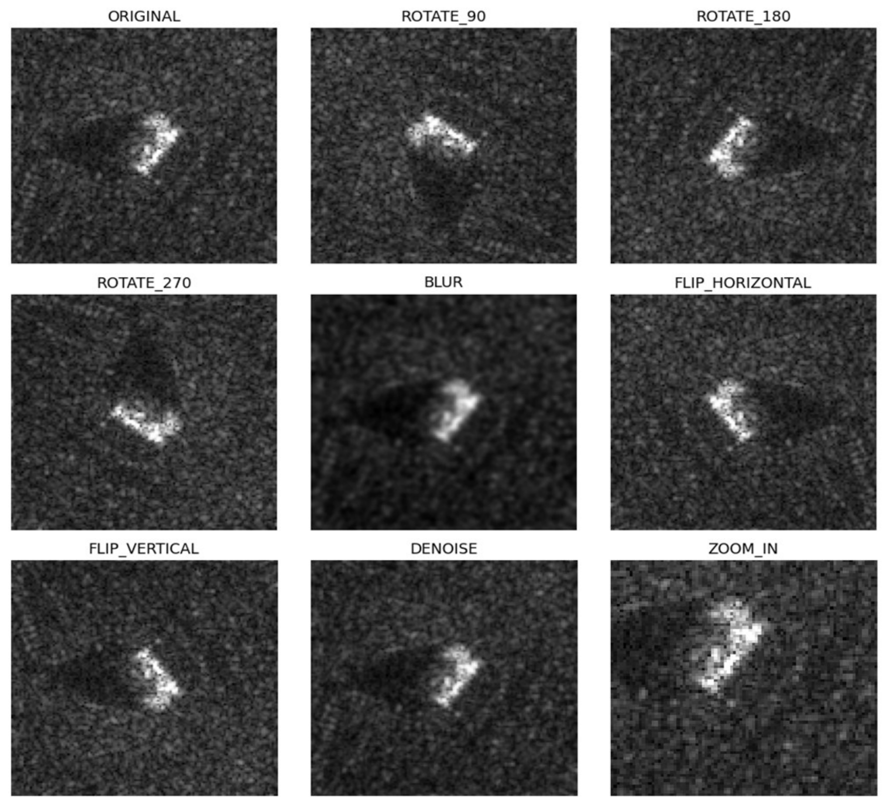

Figure 5: Example of Transformed Images

CNN Architecture

- **Network Design:**
  - Input: 128x128 grayscale images.
  - Four convolutional blocks (32, 64, 128, 256 filters)
  - MaxPooling (2x2) for spatial dimension reduction
  - Dense layer (512 units) for feature embedding
  - Dropout (0.50) to prevent overfitting
- **Training:**
  - Adam optimizer, learning rate 0.001
  - Sparse categorical cross-entropy loss
  - Early stopping with 3-epoch patience
  - 5-fold cross validation for testing

Table 2: CNN Architecture for Pretext Training

Layer	**Output Shape**	**Kernel**	**Activation**
Input	(128,128,1)	-	-
Conv2D	(128,128,32)	3×3	ReLU
BatchNorm	(128,128,32)	-	-
MaxPool2D	(64,64,32)	2×2	-
Conv2D	(64,64,64)	3×3	ReLU
BatchNorm	(64,64,64)	-	-
MaxPool2D	(32,32,64)	2×2	-
Conv2D	(32,32,128)	3×3	ReLU
BatchNorm	(32,32,128)	-	-
MaxPool2D	(16,16,128)	2×2	-
Conv2D	(16,16,256)	3×3	ReLU
BatchNorm	(16,16,256)	-	-
MaxPool2D	(8,8,256)	2×2	-
Dense	(8,8,512)	-	ReLU
Dropout(0.5)	(8,8,512)	-	-
GlobalAvgPool	(512)	-	-
Dense	(9)	-	Softmax

Downstream Classification

- **Classifiers:**
  - Support Vector Machine (linear kernel)
  - XGBoost (100 estimators)
  - Random Forest (100 estimators)
  - Gradient Boosting (100 estimators)

- **Feature Extraction:**
  - 512-dimensional feature vector from pretext model

Results

- **Performance Metrics:**

  - Accuracy, Precision, Recall, F1-score

  - Area Under ROC Curve (AUC)

  - True Positive Rates (TPR) at specific False Positive Rate (FPR) thresholds:
    - TPR at 1% FPR
    - TPR at 5% FPR
    - TPR at 10% FPR
    - TPR at 20% FPR

  - Emphasis on low FPR importance for SAR ATR application

Results

- Our SSL approach achieved comparable performance with **30% reduced data** and **no synthetic replacement**.
- **SVM**: Best average accuracy across folds (**89.78%**) and robust detection at low false positive rates.
- At 70% measured data, SSL achieved **96.64% TPR at 5% FPR**.
- Attained an average AUC of 0.9934
- Attained comparable or superior performance than **synthetic-data-dependent approaches**.

Table 3: Performance Metrics with SVM Downstream Classifier

Classifier	Split (n)	FPR 1%	FPR 5%	FPR 10%	FPR 20%	AUC	Accuracy	Precision	Recall	F1 Score
SVM	0	90.0	99.4	99.4	99.4	0.9973	91.11	92.24	91.11	91.22
	1	98.3	98.3	98.9	98.9	0.9928	96.11	96.28	96.11	96.09
	2	88.3	98.9	99.4	99.4	0.9967	90.56	90.90	90.56	90.33
	3	75.6	88.3	95.4	99.4	0.9840	81.11	81.36	81.11	80.27
	4	90.0	98.3	98.9	99.4	0.9962	90.00	90.69	90.00	89.71
	Average	**88.44**	**96.64**	**98.40**	**99.30**	**0.9934**	**89.78**	**90.29**	**89.78**	**89.52**

Comparison between Our Approach & Non-SSL Approach

- **Objectives:**
  - Validate SSL framework without synthetic data using the CNN architecture of Lewis et al. for pretext tasks
  - Use purely measured data; i.e., **no synthetic replacements**

Table 4: CNN Architecture of Lewis et al.

Layer	Output Shape	Kernel	Activation
Input	(64,64,1)	-	-
Conv2D	(64,64,16)	3×3	ReLU
MaxPool2D	(32,32,16)	2×2	-
Conv2D	(32,32,32)	3×3	ReLU
MaxPool2D	(16,16,32)	2×2	-
Conv2D	(16,16,64)	3×3	ReLU
MaxPool2D	(8,8,64)	2×2	-
Conv2D	(8,8,128)	3×3	ReLU
MaxPool2D	(4,4,128)	2×2	-
Flatten	(2048)	-	-
Dense	(1000)	-	ReLU
Dense	(500)	-	ReLU
Dense	(250)	-	ReLU
Dense	(9)	-	-

Comparison between Our Approach & Non-SSL Approach

- **Results:**
  - SSL achieved comparable performance with **reduced data** and **no synthetic replacement**
  - At 70% measured data, SSL achieved **93%** accuracy
  - Outperformed Lewis et al.'s approach that utilized the full available data
  - Achieved **99.09%** accuracy with **full measured data (SVM Downstream Classifier)**

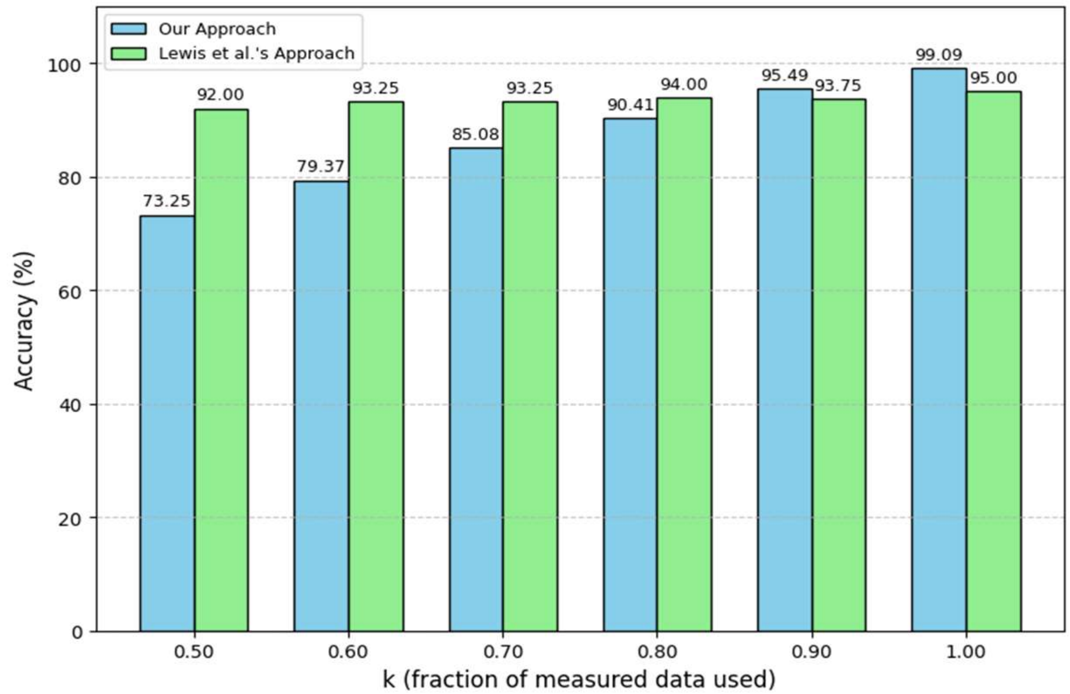

Figure 6: Accuracy Comparison between Our Approach (SSL-based) and Lewis et al.'s Approach (Non-SSL)

Conclusion

- **Key Contributions:**
  - Novel SSL framework for SAR target recognition
  - Eliminated dependency on synthetic data
  - Competitive performance with limited labeled data

- **Future Works:**
  - Incorporate domain-specific pretext tasks (e.g., radar frequencies)
  - Extend framework to other SAR applications with limited labeled data
  - Potential computational optimizations for large datasets

- **Impact:**
  - Advances SAR ATR and lays groundwork for domain-specific SSL research

Acknowledgement

This work is supported by the funds provided by the National Science Foundation and by DoD OUSD (R&E) under Cooperative Agreement PHY-2229929 (The NSF AI Institute for Artificial and Natural Intelligence)

References

[1] J. Li, Z. Yu, L. Yu, P. Cheng, J. Chen, and C. Chi, "A comprehensivesurvey on sar atr in deep-learning era," Remote Sensing, vol. 15, no. 5,p. 1454, 2023.

[2] Fu, K.; Dou, F.Z.; Li, H.C.; Diao, W.H.; Sun, X.; Xu, G.L. Aircraft recognition in SAR images based on scattering structure featureand template matching. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2018, 11, 4206–4217

[3] Morgan, David AE. "Deep convolutional neural networks for ATR from SAR imagery." In Algorithms for Synthetic Aperture Radar Imagery XXII, vol. 9475, pp. 116-128. SPIE, 2015.

[4] Chen, Sizhe, Haipeng Wang, Feng Xu, and Ya-Qiu Jin. "Target classification using the deep convolutional networks for SAR images." IEEE transactions on geoscience and remote sensing 54, no. 8 (2016): 4806-4817.

[5] Ding, Jun, Bo Chen, Hongwei Liu, and Mengyuan Huang. "Convolutional neural network with data augmentation for SAR target recognition." IEEE Geoscience and remote sensing letters 13, no. 3 (2016): 364-368.

[6] Furukawa, Hidetoshi. "Deep learning for target classification from SAR imagery: Data augmentation and translation invariance." arXiv preprint arXiv:1708.07920 (2017).

[7] Marmanis, Dimitrios, Wei Yao, Fathalrahman Adam, Mihai Datcu, Peter Reinartz, Konrad Schindler, Jan Dirk Wegner, and Uwe Stilla. "Artificial generation of big data for improving image classification: A generative adversarial network approach on SAR data." arXiv preprint arXiv:1711.02010 (2017).

[8] Lu, Changchong, and Weihai Li. "Ship classification in high-resolution SAR images via transfer learning with small training dataset." Sensors 19, no. 1 (2018): 63.

Thank You!

- Questions?