Dual-stream contrastive predictive network with joint handcrafted feature view for SAR ship classification

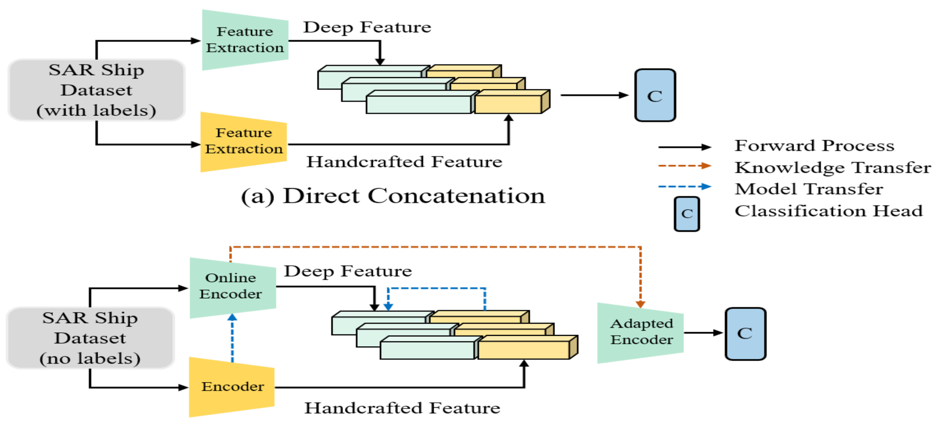
Xianting Feng*, Hao Zheng*, Zhigang Hu§, Liu Yang, Meiguang Zheng

Abstract

Most existing synthetic aperture radar (SAR) ship classification technologies heavily rely on correctly labeled data, **ignoring the discriminate features of unlabeled SAR ship images.** Even though researchers try to enrich CNN-based features by introducing traditional handcrafted features, **existing methods easily cause information redundancy and fail to capture the interaction between them.** To address these issues, we propose **a novel dual-stream contrastive predictive network (DCPNet)**, which consists of two asymmetric tasks and a false negative sample elimination module.

Introduction

Synthetic Aperture Radar (SAR) is an active microwave remote sensing imaging system, and **SAR ship classification** has always been a hot research topic.



(a) Direct Concatenation

(b) Knowledge Transfer

Problem:

1) Difficult to collect a large-scale labeled data of remote sensing images due to the acquisition and annotation costs.

2) Direct concatenation of features may create redundancy and can't capture the interaction between features.

Solution:

1) Handcrafted feature complement the deep feature due to its characteristics about being able to focus on the specific physical information.

2) CL methods can learn discriminative features between multiple representations.

Contributions:

1) We utilize CL for the first time to learn complementary information between handcrafted and deep features.

2) DCPNet focuses on the connections between samples, which enables the full utilization of unlabeled data, and realizes the reuse of handcrafted knowledge.

Conclusion

Through two-stage comparative experiments, it is concluded that the performance of the pre-training framework DCPNet in SAR ship classification tasks is superior than the existing CL methods, and the accuracy of supervised models are also effectively improved. Otherwise, DCPNet only achieves knowledge transfer for single handcrafted feature, so further research is needed on how to aggregate information between multiple handcrafted features.

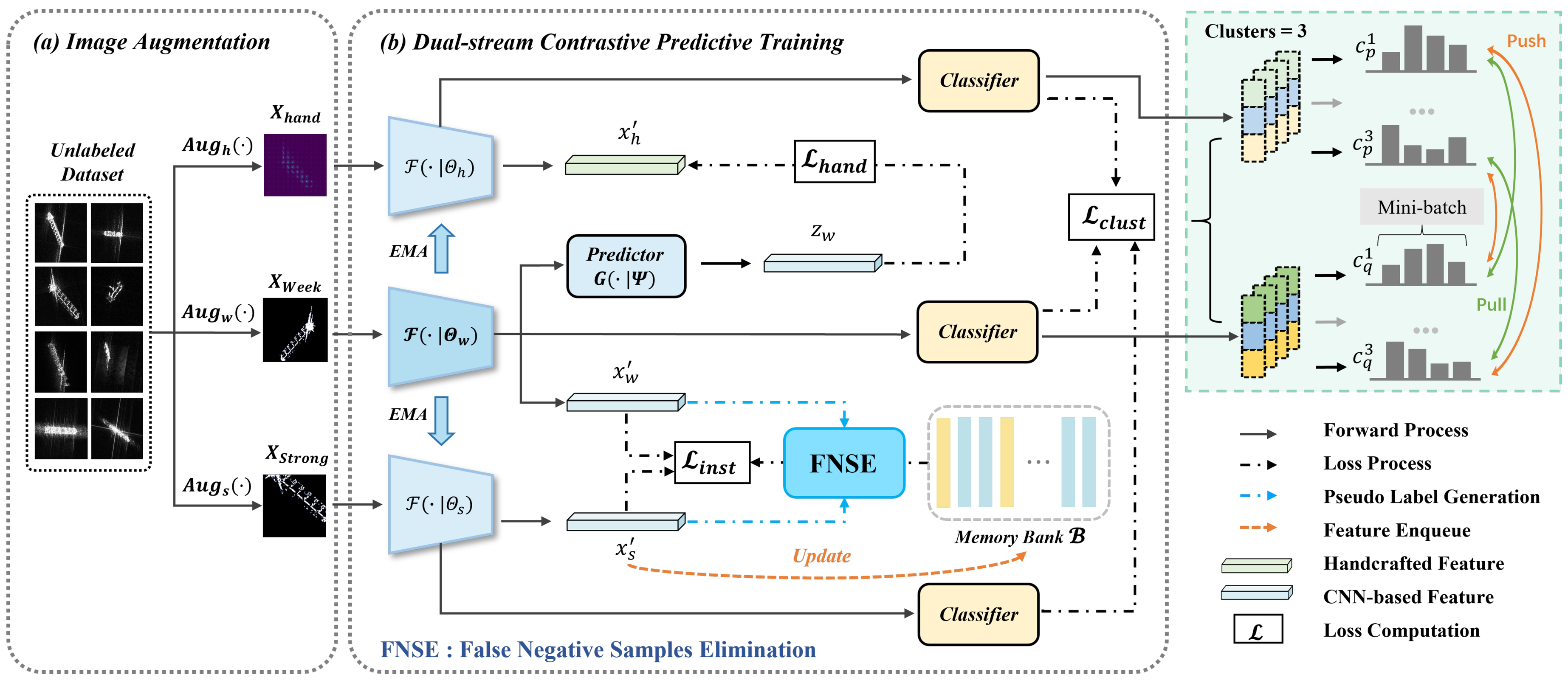
Acknowledgements

This work is supported by the Natural Science Foundation of China (Grant No: 62172442).

Contact Information

Email: xiantingfeng@csu.edu.cn
Tell: +86-15768217538

Methodology



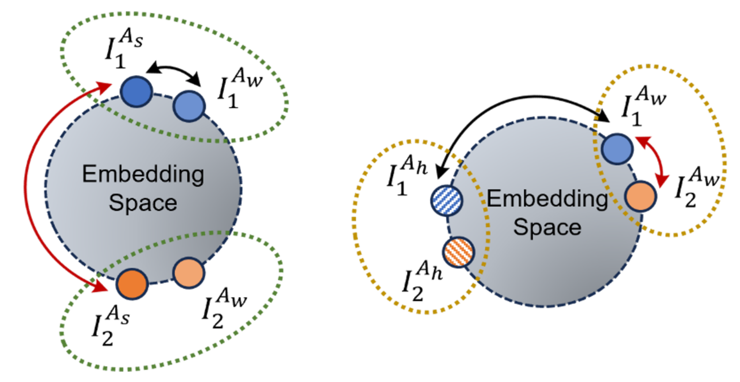
An overview of DCPNet:

* The early image augmentation stage
* The core dual-stream contrastive predictive training stage.

★ Handcrafted feature prediction tasks.

- Handcrafted features are far away from deep features in the embedding space.
- So forcing a comparison of two features at the instance level will affect the focus of feature extraction and make the model difficult to converge.
- The loss based on mean square error is defined as follows:

$$\mathcal{L}_{hand} = 2 - 2 \cdot \left(\frac{z_w}{\|z_w\|_2}\right)^T \left(\frac{x_h}{\|x_h\|_2}\right)$$



(a) Weak & Strong Augmentation

(b) Hand-crafted & Weak Augmentation

★ Instance-level image similarity comparison task.

- For each sample, different augmentations of the same image should be brought "nearby" in an embedding space since they likely contain similar semantic content or correlated features.
- The instance-level contrastive loss is defined as follows:

$$\mathcal{L}_{instance} = -\frac{1}{N}\sum_{i=1}^{N} log \frac{e^{s(x_w^i, x_s^i)/\tau}}{e^{s(x_w^i, x_s^i)/\tau} + \sum_{j=1}^{K'} \mathbb{1}_{[i \neq j]} e^{s(x_w^i, k_-^j)/\tau}}$$

★ Cluster-level image cluster consistency task.

- The same batch of images should have similar category distributions under different augmentations.
- We utilizes a consistency loss to promote the compactness within classes in feature space.
- Using $(c_p^i, c_q^i) \in \mathbb{R}^{N\times 1}$ to denote the distribution representation of cluster i under two augmentation strategies p and q, and $p, q \in \{A_{weak}, A_{strong}, A_{hand-crafted}\}$. The loss is defined as follows:

$$\mathcal{L}_{cluster} = -\frac{1}{M}\sum_{i=1}^{M} log \frac{e^{\frac{s(c_p^i, c_q^i)}{\tau}}}{e^{\frac{s(c_p^i, c_q^i)}{\tau}} + \sum_{j=1}^{M} \mathbb{1}_{[i \neq j]} e^{\frac{s(c_p^i, c_q^j)}{\tau}}}$$

and $\mathbb{1}_{[i \neq j]}$ represents distribution representation that does not belong to the same ship category.

Experiment

Table 1. Evaluation accuracy(%) of applying two evaluation method "Fine-tuning" and "KNN-way" for DCPNet and other advanced self-supervised methods.

Dataset	Method	Pre-training ep.	Fine-tuning			KNN-way	
			Ft1.	Ft2.	Ftall.	ResNet18	ResNet50
OpenSARShip	MoCo	200	66.05±1.74	<u>71.50±0.47</u>	70.33±1.21	67.67±0.95	67.85±0.35
	BYOL	200	<u>68.44±0.29</u>	68.84±1.80	**73.03±0.69**	-	-
	SimSiam	200	66.64±1.04	66.91±1.55	<u>72.79±0.90</u>	68.78±3.02	66.64±1.04
	DCPNet(Ours)	20	60.61±2.57	65.68±2.65	67.92±3.05	<u>69.29±0.29</u>	<u>69.92±0.95</u>
	DCPNet(Ours)	200	**70.37±1.01**	**73.66±1.01**	71.50±1.39	**70.66±0.42**	**69.94±0.60**
FUSARShip	MoCo	200	57.73±1.22	<u>74.94±0.60</u>	74.69±0.91	<u>70.13±0.58</u>	**70.29±1.01**
	BYOL	200	<u>60.21±0.74</u>	73.62±1.40	<u>85.25±0.39</u>	-	-
	SimSiam	200	56.29±1.37	61.62±0.89	80.71±0.31	65.92±0.85	56.67±0.35
	DCPNet(Ours)	20	56.59±0.83	72.66±0.48	72.49±0.88	61.08±0.99	55.97±0.66
	DCPNet(Ours)	200	**69.33±1.62**	**83.18±0.35**	**87.94±0.76**	**72.33±2.10**	<u>66.67±0.54</u>

◆ When the test accuracy of KNN-way is close to supervised models, extracted features are distinguishable enough to be classified through "ft1." or "ft2.".

Table 2. Fine-tuning accuracy(%) of DCPNet and state-of-the-art supervised learning baseline.

Method	Train ep.	OpenSARShip	FUSARShip
ResNet-18	100	71.70±0.94	79.71±0.73
ResNet-34	100	71.84±1.23	80.73±0.41
ResNet-50	100	72.15±1.20	80.96±0.47
DCPNet(Best)	100	**73.66±1.01**	**87.94±0.76**

◆ DCPNet can train an encoder that learns prior physical knowledge of handcrafted features without causing redundancy.

◆ DCPNet can obtain a feature set with better generalization and discrimination.