Tongjing SUN, Haoran XU, Shishuo REN, Denghui ZHANG, 2026. An attention mechanism-based multi-domain feature fusion approach for active sonar target recognition. *ENGINEERING Information Technology & Electronic Engineering*. 27(2):250177. https://doi.org/10.1631/ENG.ITEE.2025.0177

An attention mechanism-based multi-domain feature fusion approach for active sonar target recognition

Key words: Acoustic target recognition; Neural network; Attention mechanism; Multi-domain feature fusion

Corresponding author: Tongjing SUN
E-mail: stj@hdu.edu.cn
ORCID: https://orcid.org/0000-0002-6647-5282

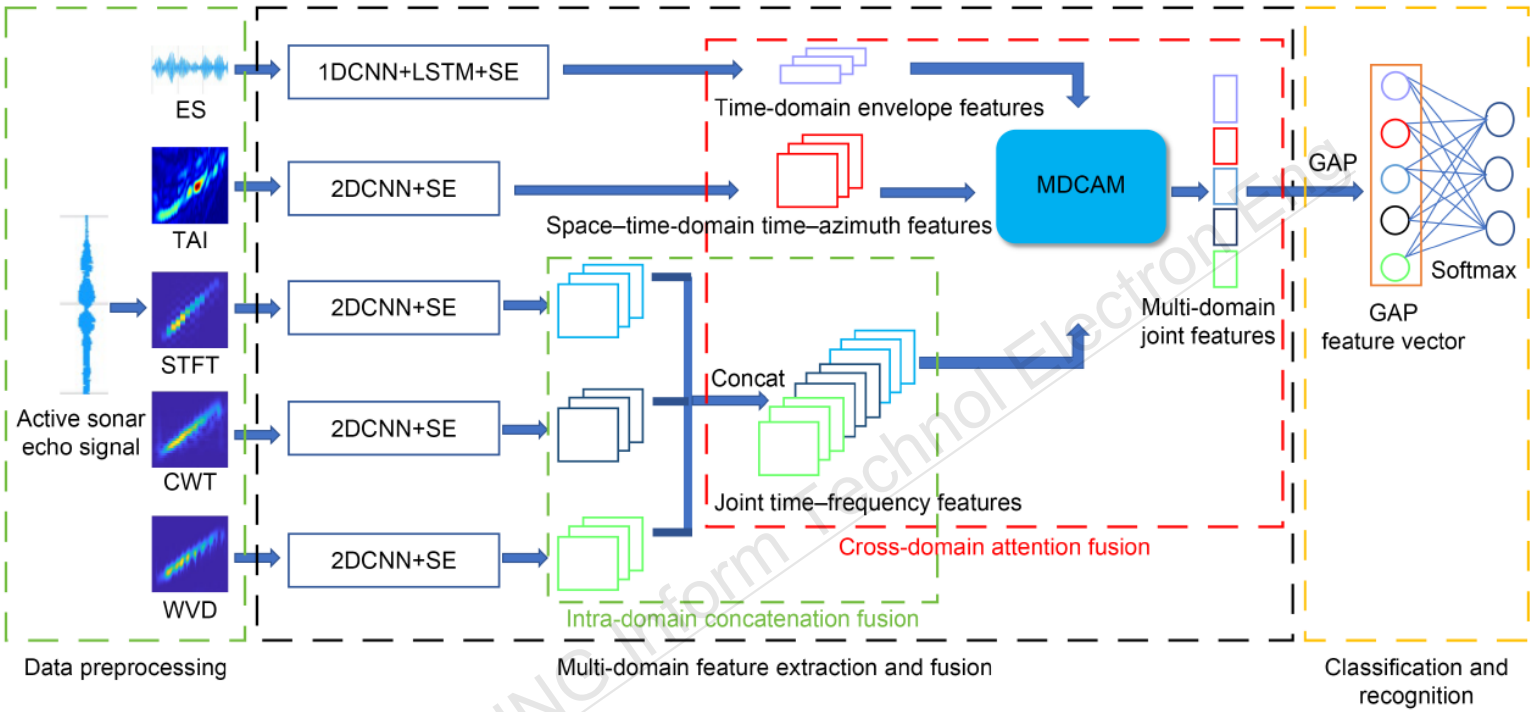
Motivation

1. The complex and changeable marine environment (e.g., multipath effect, Doppler distortion, and clutter interference) and scarcity of underwater acoustic samples pose significant challenges for active sonar target recognition, leading to distorted echo signals and insufficient feature extraction.
2. Existing deep learning-based fusion methods rely on simple concatenation strategies, which cause information redundancy and fail to effectively mine correlation information between multi-domain features, limiting feature representation capability.
3. Although attention mechanisms have shown potential in multi-domain fusion for computer vision and radar target recognition, most existing active sonar recognition methods adopt dual-domain interaction or single attention fusion, insufficiently exploring complex cross-domain dependencies and underutilizing feature complementarity.

Main idea

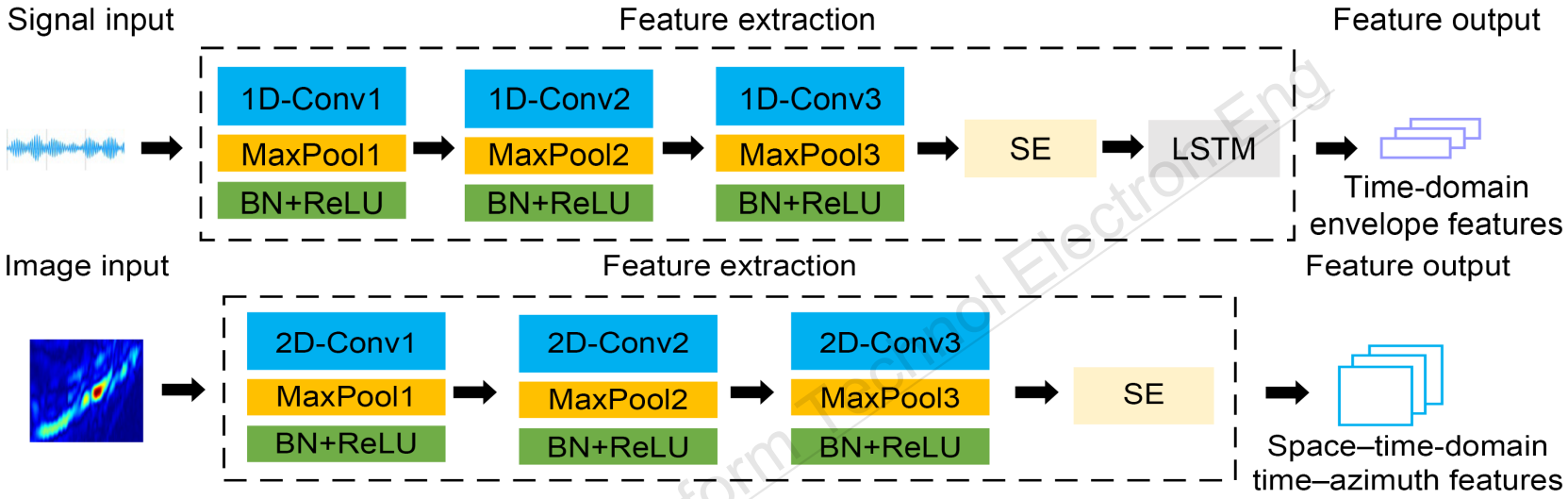
1. We preprocess active sonar echo signals into time-domain envelope sequences (ESs), space–time-domain time–azimuth images (TAIs), and time–frequency images (STFT/CWT/WVD), and construct differentiated networks (1DCNN-LSTM-SE for ES and 2DCNN-SE for TAI and time-frequency images) with squeeze and excitation (SE) channel attention to extract deep features from different domains.
2. Constructs a multi-domain cross-attention module (MDCAM), i.e., an extension of the Transformer's multi-head attention mechanism. It enables deep interaction and adaptive fusion of time-domain envelope, space–time-domain time–azimuth, and joint time–frequency features, effectively eliminating redundancy and promoting complementary information integration.

System model



1. Data preprocessing: converts raw echo signals into three domain representations——time-domain ESs, space–time-domain TAIs, and time–frequency images (STFT/CWT/WVD).
2. Multi-domain feature extraction and fusion: uses 1DCNN-LSTM-SE and 2DCNN-SE with channel attention to extract deep features, followed by intra-domain concatenation and cross-domain fusion via MDCAM.
3. Classification and recognition: employs global average pooling (GAP) and Softmax for efficient classification.

Multi-domain feature extraction



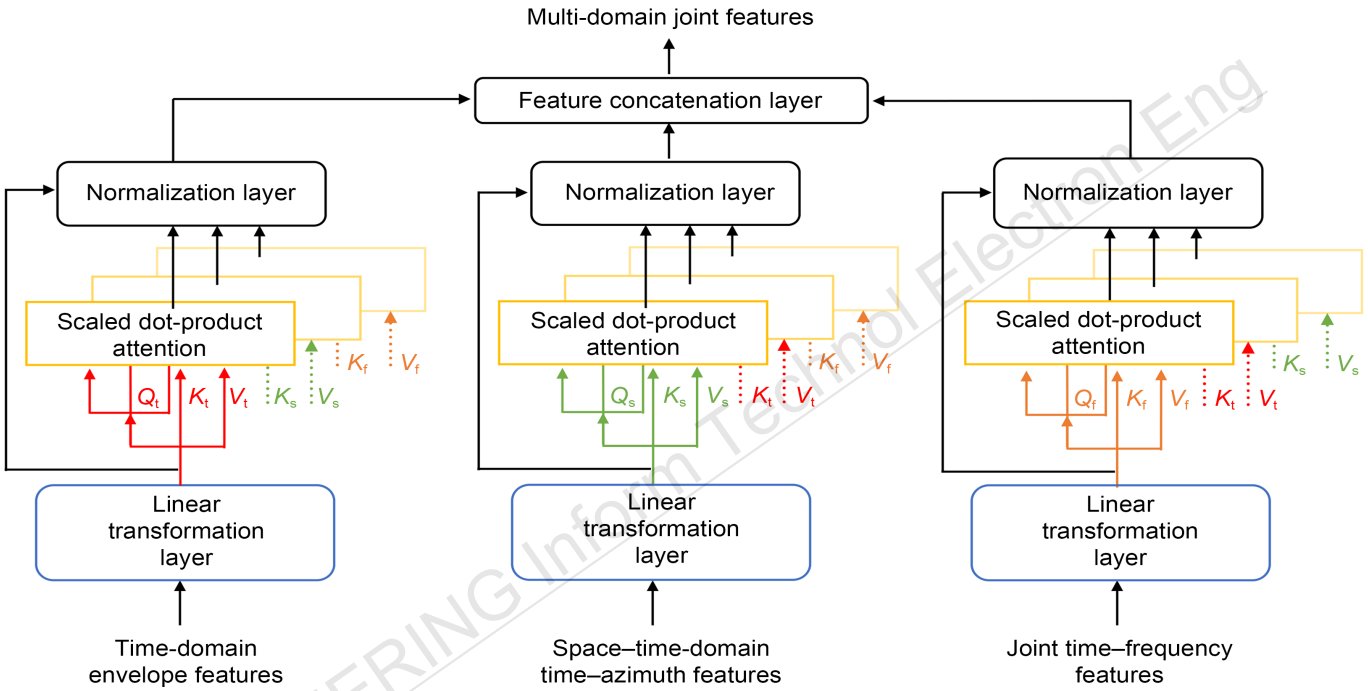
Differentiated network architectures

- Time-domain (ES): 1DCNN-LSTM-SE (captures local temporal features + long-term dependencies)
- Space–time/time–frequency domains (TAI/time–frequency images): 2DCNN-SE (extracts multi-scale image features)

SE channel attention mechanism

- Structure: Squeeze → Excitation → Reweighting
- Function: adaptively adjusts channel weights to strengthen target-relevant features and suppress redundancy

Multi-domain cross-attention fusion



Two-stage fusion strategy

- Intra-domain fusion: Concatenate STFT/CWT/WVD features along channel dimension
- Cross-domain fusion: MDCAM (extends Transformer multi-head attention)

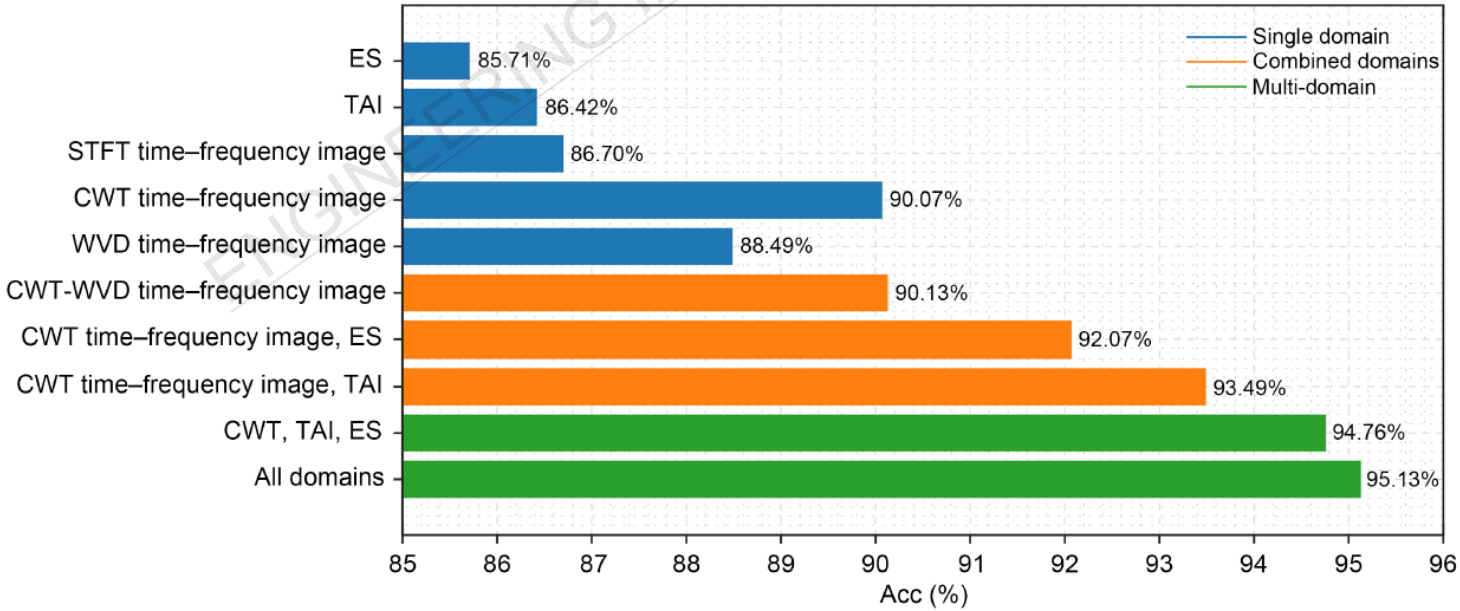
MDCAM core logic

- Generate ***Q***/***K***/***V*** matrices via linear transformation for each domain
- Compute self-attention (intra-domain dependencies) and cross-attention (inter-domain correlations)

Major results

Comparison with multi-domain fusion performance

Domain	Input data	Feature extraction	Feature fusion	Acc (%)
Single domain	CWT time–frequency image	2DCNN-SE		90.07
Combined domains	CWT time–frequency image and ES	1DCNN-LSTM-SE and 2DCNN-SE	Concat	90.60
	CWT time–frequency image and ES	1DCNN-LSTM-SE and 2DCNN-SE	MDCAM	92.07
	CWT time–frequency image and TAI	2DCNN-SE	Concat	91.65
	CWT time–frequency image and TAI	2DCNN-SE	MDCAM	93.49
Multi-domain	CWT image, TAI, and ES	1DCNN-LSTM-SE and 2DCNN-SE	Concat	92.52
	CWT image, TAI, and ES	1DCNN-LSTM-SE and 2DCNN-SE	MDCAM	94.76
	Time–frequency images*, TAI, and ES	1DCNN-LSTM-SE and 2DCNN-SE	Concat and MDCAM	95.13

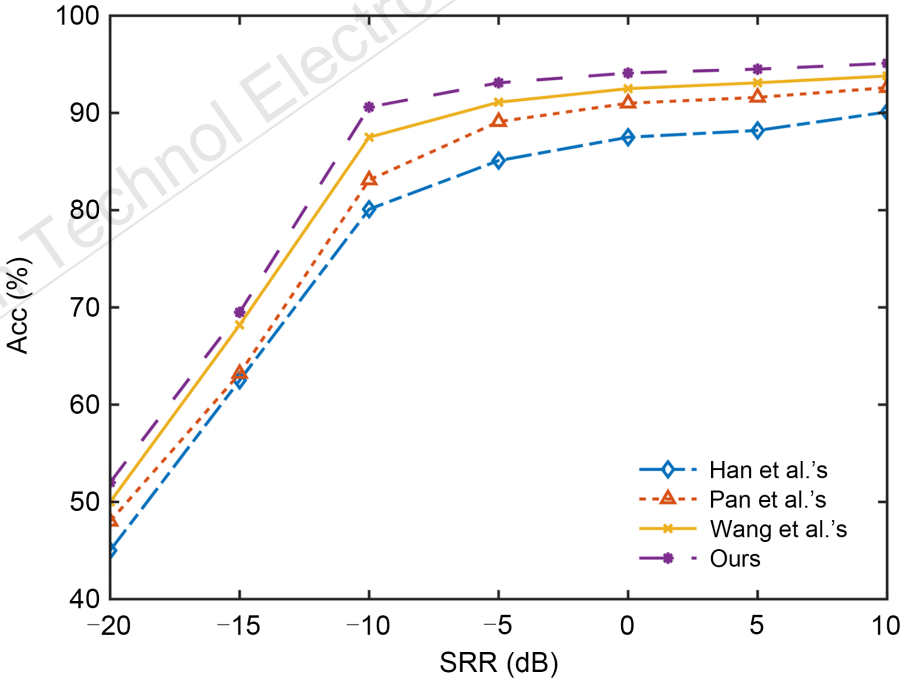


Major results

Comparison with state-of-the-art models

Model	Acc (%)	R (%)	F1 (%)	Parameter count ($\times 10^6$)	FLOP ($\times 10^9$)
VGG16	87.30	88.16	87.81	134.27	3.72
ViT	88.58	87.42	88.09	85.67	3.42
Han et al.'s	90.49	88.45	88.23	44.66	2.13
Pan et al.'s	93.07	90.87	90.45	129.02	3.65
Wang et al.'s	93.85	90.83	91.04	138.56	3.93
Ours	95.13	92.21	92.25	85.40	2.84

- Our method achieves the highest accuracy (95.13%) among all competitors.
- Significantly lower computational cost (FLOP) ensures high efficiency.



- Maintains >90% accuracy even at −10 dB (severe noise).
- Superior environmental adaptability compared to baseline models.

Conclusions

We have proposed a novel attention-based multi-domain feature fusion approach for active sonar target recognition, effectively addressing feature redundancy and enhancing inter-domain interaction:

- Designed MDCAM for adaptive cross-domain fusion, which significantly improved feature discriminability and classification accuracy.
- Achieved state-of-the-art performance (95.13% accuracy) with strong generalization ability in complex and low-SRR underwater environments.

Tongjing SUN is a professor and Master Supervisor of Hangzhou Dianzi University. She received a PhD degree from Harbin Engineering University in 2005. She is currently mainly engaged in the theoretical research on the echo characteristics of targets in complex marine environments, underwater acoustic signal processing and information fusion, target classification and recognition, etc.

Haoran XU received the B.S. degree in automation from Nanchang Hangkong University, Nanchang, China, in 2023. He is currently pursuing the M.S. degree in control science and engineering with the School of Automation, Hangzhou Dianzi University, Hangzhou, China. His current research interests include underwater acoustic signal processing and multimodal feature fusion.