Differential Cross-Attention based Network for Robust Modulation Classification in OTFS Systems

Tonmoy Rajkhowa[1], Sanjeev Sharma[1], Kuntal Deka[2], and Tharmalingam Ratnarajah[3]
[1]Dept. of Electronics Engg., IIT (BHU) Varanasi, U.P., India, {tonmoyrajkhowa.rs.ece24, sanjeev.ece}@iitbhu.ac.in
[2]Dept. of Electrical and Electronics Engg., IIT Guwahati, Assam, India, kuntaldeka@iitg.ac.in
[3]Dept. of Electrical and Computer Engg., San Diego State University, San Diego, CA 92182-1309, USA t.ratnarajah@ieee.org

***Abstract*—Automatic modulation classification (AMC) is essential for intelligent next-generation wireless systems but degrades significantly in high-mobility doubly dispersive channels. Orthogonal time frequency space (OTFS) modulation, operating in delay-Doppler (DD) domain, improves the robustness by mitigating channel variations. However, existing deep learning-based AMC methods for 2D OTFS signals are largely limited to 1D representations, leading to loss of spatial information. To address this, a differential attention-based dual-stream network, termed DADS-Net, is proposed for OTFS-based AMC systems. DADS-Net processes the 2D in-phase/quadrature (I/Q) and amplitude/phase (A/P) representations of the DD domain in parallel streams. Each stream is enhanced using a channel-spatial attention (CSA) module, and a differential cross-attention mechanism (DCAM) is employed to fuse features by suppressing common noise and emphasizing discriminative patterns. Simulation results over the extended vehicular A (EVA) channel show that DADS-Net consistently improves classification performance over existing OTFS-AMC methods, particularly under high Doppler and low-to-moderate signal-to-noise ratio (SNR) conditions. While absolute accuracy remains limited due to the inherent complexity of OTFS-based AMC, DADS-Net provides stable and reliable gains, highlighting its robustness and practical applicability.**

***Index Terms*—Automatic modulation classification, differential cross-attention, orthogonal time-frequency space.**

I. Introduction

Automatic modulation classification (AMC) plays a key role in enabling adaptive transmission, spectrum awareness, and intelligent resource management in next-generation wireless communication systems [1]. In high-mobility environments, where the effects of delay and Doppler are prominent, the widely adopted 5G waveform, such as orthogonal frequency-division multiplexing (OFDM), fails severely due to inter-carrier interference [2]. In contrast, orthogonal time-frequency space (OTFS) modulation demonstrates resilience to delay and Doppler spreads [3], [4]. OTFS operates in the delay–Doppler (DD) domain and effectively mitigates channel variations.

Traditionally, AMC has relied on computationally intensive likelihood-based (LB) methods requiring prior channel-state information [1], [5]–[7], while feature-based (FB) approaches offer improved efficiency but degrade under low signal-to-noise ratio (SNR) or deep-fading conditions [8]–[11]. More recently, deep learning (DL)-based methods [12]–[14] have shown superior performance by learning directly from raw signals. Although AMC is well-studied for traditional waveforms, such as OFDM [1], its application to OTFS, a key 6G waveform, remains in its early stages. Existing OTFS-AMC approaches, including residual-stacked convolutional long short-term memory (Res-CLDNN) [10], random forest [15], and convolutional neural networks with squeeze-and-excitation (CNN-SE) [16], primarily rely on standard in-phase/quadrature (I/Q) and amplitude/phase (A/P) representations but often convert the inherently 2D delay–Doppler (DD) signals into 1D forms, resulting in loss of structural information.

Multi-modal AMC architectures leverage both in-phase/quadrature (I/Q) and amplitude/phase (A/P) representations to obtain richer and more discriminative features compared to uni-modal approaches [17]. However, existing methods rely on simplistic fusion strategies, such as concatenation, addition, or multiplication, which fail to effectively capture complementary and contrastive information, and are unable to suppress noise or enhance discriminative patterns. Furthermore, conventional softmax-based cross-attention mechanisms are computationally expensive and assign non-negligible weights to irrelevant regions that introduces attention noise, leading to confusion among similar modulation classes under low signal-to-noise ratio (SNR) or high Doppler conditions [18].

To address these limitations, this work proposes a differential cross-attention mechanism (DCAM)-based dual-stream framework, termed DADS-Net, for robust feature fusion in OTFS-AMC under high mobility. DADS-Net operates directly on the 2D OTFS frames in DD domain. It extracts spatio-temporal features from I/Q and A/P representations using a channel-spatial attention (CSA) module, and fuses them via DCAM. Inspired by differential amplifiers [18], DCAM suppresses common noise while emphasizing discriminative patterns, thereby improving generalization under diverse channel conditions. Given the inherent difficulty of OTFS-AMC in high-mobility scenarios, the proposed approach focuses on enhancing robustness and consistency rather than maximizing absolute classification accuracy. In summary, the primary contributions of this work are as follows:

- This work proposes DADS-Net that processes the I/Q and A/P components of 2D OTFS frames in dual-streams and effectively exploits their complementary information for enhanced feature extraction in AMC.
- The proposed DADS-Net employs a CSA module to refine features by emphasizing informative components and critical regions while preserving spatial locality. Unlike prior OTFS-AMC methods that convert 2D frames into 1D sequences, CSA operates directly on the 2D

OTFS structure, retaining essential contextual information. Additionally, the DCAM enhances feature fusion by subtracting two softmax attention maps, suppressing common noise and highlighting contrastive patterns.

- Extensive experiments over the realistic extended vehicular A (EVA) channel [19] show that the proposed DADS-Net consistently improves classification performance over existing OTFS-based AMC methods across mobility levels of 100 km/h and 500 km/h, as well as different delay–Doppler (DD) frame sizes.

The remainder of this paper is structured as follows: Section II presents the OTFS system model. Section III outlines the proposed DADS-Net architecture. Section IV highlights the experimental settings and the simulation results are presented in Section V. Finally, Section VI concludes the paper.

II. Preliminaries

A. OTFS System Model

OTFS is a 2D modulation scheme that maps information symbols onto the DD domain to combat the doubly-spread wireless channels (DDWCs) [3]. The DD grid of size $M \times N$ is indexed by delay and Doppler bins, with resolutions determined by system parameters, where larger M favors delay resolution and larger N supports Doppler resolution [4]. Each OTFS frame spans a duration $T_f = NT$ and bandwidth $M\Delta f$, with modulation and demodulation enabling conversion between DD and time domains. The information symbols $\mathbf{X}_{\mathrm{DD}}$ in the DD domain are transformed into the time–frequency (TF) domain representation $\mathbf{X}_{\mathrm{TF}}$ using the inverse symplectic finite Fourier transform (ISFFT) [4] as:

$$\mathbf{X}_{\mathrm{TF}}[n,m] = \frac{1}{MN}\sum_{k=0}^{N-1}\sum_{l=0}^{M-1}\mathbf{X}_{\mathrm{DD}}[k,l]\cdot e^{j2\pi\left(\frac{nk}{N}-\frac{ml}{M}\right)}. \quad (1)$$

Here, $n = 0,\ldots,N-1$ and $m = 0,\ldots,M-1$ denote the time and frequency indices. The TF-domain signal $\mathbf{X}_{\mathrm{TF}}$ is then transformed into the time-domain signal $s(t)$ via the Heisenberg transform, incorporating the transmit pulse shaping function $g(t)$ as:

$$s(t) = \sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\mathbf{X}_{\mathrm{TF}}[n,m]\cdot e^{j2\pi m\Delta f(t-nT)}g(t-nT). \quad (2)$$

The transmitted signal $s(t)$ passes through the DSWC with impulse response $h(\tau,\nu)$ and is received as $r(t)$

$$r(t) = \int_{\tau}\int_{\nu} h(\tau,\nu)e^{j2\pi\nu(t-\tau)}s(t-\tau)d\tau d\nu + z(t). \quad (3)$$

Here, $z(t)$ is the additive white Gaussian noise (AWGN).

At the receiver, the signal $r(t)$ is sampled with period $T_s = 1/B$ to obtain the discrete sequence $r[\widehat{n}T_s]$, where $\widehat{n} = 0,\ldots,NM-1$. This sequence is transformed into the TF-domain $\mathbf{Y}_{\mathrm{TF}}$ using the Wigner transform, and then mapped to the DD-domain $\mathbf{Y}_{\mathrm{DD}}$ via the symplectic finite Fourier transform (SFFT) as:

$$\mathbf{Y}_{\mathrm{DD}}[k,l] = \sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\mathbf{Y}_{\mathrm{TF}}[n,m]\cdot e^{-j2\pi\left(\frac{nk}{N}-\frac{ml}{M}\right)} + \mathbf{Z}[k,l], \quad (4)$$

where $\mathbf{Z}[k,l]$ is the DD-domain AWGN signal. In general, the relationship between transmitted information symbols $\mathbf{x}_{\mathrm{DD}} = \mathrm{vec}(\mathbf{X}_{\mathrm{DD}}) \in \mathcal{C}^{\mathrm{MN}\times 1}$ and received information symbols $\mathbf{y}_{\mathrm{DD}} = \mathrm{vec}(\mathbf{Y}_{\mathrm{DD}}) \in \mathcal{C}^{\mathrm{MN}\times 1}$ can be expressed as:

$$\mathbf{y}_{\mathrm{DD}} = \mathbf{H}_{\mathrm{DD}}\mathbf{x}_{\mathrm{DD}} + \mathbf{z}_{\mathrm{DD}} \in \mathcal{C}^{MN\times 1}, \quad (5)$$

where $\mathbf{H}_{\mathrm{DD}} \in \mathcal{C}^{MN\times MN}$ is the doubly-spread DD-domain channel and $\mathbf{z}_{\mathrm{DD}} = \mathrm{vec}(\mathbf{Z})$ is the AWGN vector. Here, $\mathrm{vec}(\cdot)$ denotes the vectorization operation.

B. Signal Pre-processing

The received signal $\mathbf{Y}_{\mathrm{DD}}$ is used as input to DADS-Net and represented over S OTFS frames as $\mathbf{Y}_{\mathrm{DD}} = [\mathbf{Y}_i]_{i=1}^{S}$, where each frame $\mathbf{Y}_i \in \mathbb{R}^{M\times N}$. Each frame is decomposed into I/Q components as $\mathbf{Y}_{i(\mathrm{I/Q})} = \{\Re(\mathbf{Y}_i), \Im(\mathbf{Y}_i)\}$, with $\Re(\cdot)$ and $\Im(\cdot)$ denoting real and imaginary parts. Similar to RGB channels, the I/Q representation captures the complex baseband structure for learning spatial features. Additionally, amplitude (A) and phase (P) are obtained as $A = \sqrt{\Re(\mathbf{Y}_i)^2 + \Im(\mathbf{Y}_i)^2}$ and $\Phi = \tan^{-1}\left(\frac{\Im(\mathbf{Y}_i)}{\Re(\mathbf{Y}_i)}\right)$, providing complementary magnitude and phase information to enhance classification robustness.

III. Proposed Methodology

DADS-Net, illustrated in Fig. 1, processes the I/Q and A/P components in parallel streams while preserving the 2D OTFS structure. The CSA module extracts spatio-temporal features from each stream, which are then fused by the DCAM and passed through a dense layer for modulation classification.

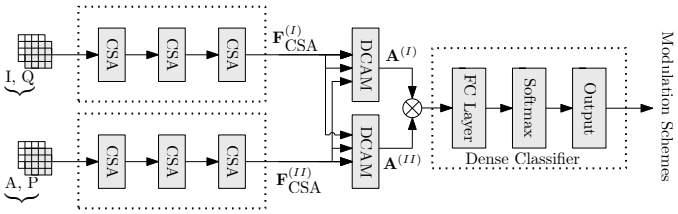

Fig. 1: Block diagram of DADS-Net illustrating the CSA and DCAM modules.

A. Channel-Spatial Attention (CSA)

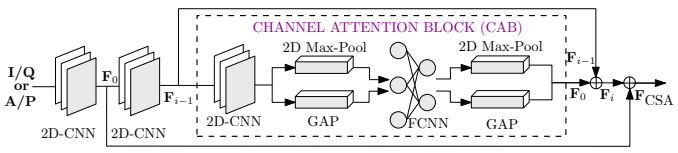

Fig. 2: Architecture of a CSA module block, indicating the channel attention and spatial attention block.

The CSA module processes OTFS frames in their native 2D form preserving spatial structure. The I/Q or A/P inputs are first passed through 2D CNNs to extract features $\mathbf{F}_0$

and $\mathbf{F}_{i-1}$, which are refined using a channel-attention block (CAB) with max-pooling and global average pooling (GAP) to model inter-channel dependencies. A fully connected network generates channel-attention maps to enhance relevant features. A spatial-attention mechanism further emphasizes important regions using similar pooling operations. The combined attention outputs are fused and added to $\mathbf{F}_i$ to produce $\mathbf{F}_{\text{CSA}}$, capturing both spatial and channel-wise information. With residual connections and 2D processing, CSA provides an efficient alternative to transformer-based attention [20].

B. Differential Cross-Attention Mechanism (DCAM)

DADS-Net introduces a novel differential cross-attention mechanism (DCAM) that enhances conventional multi-head attention in transformers. A block diagram of the DCAM is shown in Fig. 3. It suppresses attention noise and amplifies relevant information. Inspired by differential amplifiers that cancel out noise, DCAM computes the difference between two softmax cross-attention maps to filter out common, irrelevant patterns. This results in sparser and more focused attention that improves key information retrieval, generalization, and robustness while requiring fewer parameters than standard transformers [18]. Unlike traditional cross-attention and static fusion methods such as addition, concatenation, or multiplication, DCAM dynamically emphasizes discriminative differences between complementary input modalities, such as I/Q and A/P, while suppressing redundant or correlated noise. This differential fusion strategy strengthens contrastive feature alignment and captures inter-modal dependencies more effectively, enhancing AMC performance.

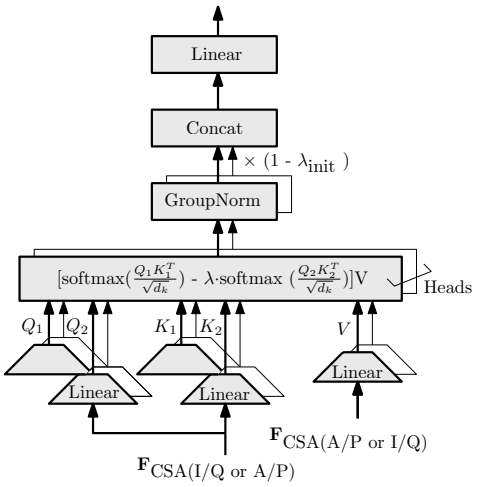

Fig. 3: Architecture of the DCAM, illustrating the difference of softmax attention maps used for noise suppression and contrast enhancement.

The feature maps $\mathbf{F}_{\text{CSA}} \in \{\mathbf{F}^{(I)}_{\text{CSA}}, \mathbf{F}^{(II)}_{\text{CSA}}\}$, where $\mathbf{F}^{(I)}_{\text{CSA}}$ and $\mathbf{F}^{(II)}_{\text{CSA}}$ denotes the feature maps extracted from I/Q and A/P components by CSA, are fused by DCAM, as illustrated in Fig. 3. These $\mathbf{F}^{(I)}_{\text{CSA}}$ and $\mathbf{F}^{(II)}_{\text{CSA}}$ are projected through linear layers to form separate queries $(\mathbf{Q}^{(I)}_1, \mathbf{Q}^{(I)}_2, \mathbf{Q}^{(II)}_1, \mathbf{Q}^{(II)}_2)$, keys $(\mathbf{K}^{(I)}_1, \mathbf{K}^{(I)}_2, \mathbf{K}^{(II)}_1, \mathbf{K}^{(II)}_2)$ and complementary values $(\mathbf{V}^{(I)}, \mathbf{V}^{(II)})$ matrices corresponding to each streams. For clarity, the projections for both streams are defined in a symmetric manner, where each stream generates its own query and key pairs while exchanging value representations. For I/Q stream, these operations can be expressed as:

$$\begin{aligned} \mathbf{Q}^{(I)}_1 &= \mathbf{W}_{\mathbf{Q}^{(I)}_1} \cdot \mathbf{F}^{(I)}_{\text{CSA}}, \quad \mathbf{Q}^{(I)}_2 = \mathbf{W}_{\mathbf{Q}^{(I)}_2} \cdot \mathbf{F}^{(I)}_{\text{CSA}} \\ \mathbf{K}^{(I)}_1 &= \mathbf{W}_{\mathbf{K}^{(I)}_1} \cdot \mathbf{F}^{(I)}_{\text{CSA}}, \quad \mathbf{K}^{(I)}_2 = \mathbf{W}_{\mathbf{K}^{(I)}_2} \cdot \mathbf{F}^{(I)}_{\text{CSA}} \\ \mathbf{V}^{(II)} &= \mathbf{W}_{\mathbf{V}^{(II)}} \cdot \mathbf{F}^{(II)}_{\text{CSA}}, \end{aligned} \tag{6}$$

where $\mathbf{W}_{\mathbf{Q}^{(I)}_1}, \mathbf{W}_{\mathbf{Q}^{(I)}_2}, \mathbf{W}_{\mathbf{K}^{(I)}_1}, \mathbf{W}_{\mathbf{K}^{(I)}_2}$, and $\mathbf{W}_{\mathbf{V}^{(II)}}$ are learnable weight matrices. Similarly, for the A/P stream, the queries, keys, and complementary value matrices are:

$$\begin{aligned} \mathbf{Q}^{(II)}_1 &= \mathbf{W}_{\mathbf{Q}^{(II)}_1} \cdot \mathbf{F}^{(II)}_{\text{CSA}}, \quad \mathbf{Q}^{(II)}_2 = \mathbf{W}_{\mathbf{Q}^{(II)}_2} \cdot \mathbf{F}^{(II)}_{\text{CSA}} \\ \mathbf{K}^{(II)}_1 &= \mathbf{W}_{\mathbf{K}^{(II)}_1} \cdot \mathbf{F}^{(II)}_{\text{CSA}}, \quad \mathbf{K}^{(II)}_2 = \mathbf{W}_{\mathbf{K}^{(II)}_2} \cdot \mathbf{F}^{(II)}_{\text{CSA}} \\ \mathbf{V}^{(I)} &= \mathbf{W}_{\mathbf{V}^{(I)}} \cdot \mathbf{F}^{(I)}_{\text{CSA}}, \end{aligned} \tag{7}$$

where $\mathbf{W}_{\mathbf{Q}^{(II)}_1}, \mathbf{W}_{\mathbf{Q}^{(II)}_2}, \mathbf{W}_{\mathbf{K}^{(II)}_1}, \mathbf{W}_{\mathbf{K}^{(II)}_2}$, and $\mathbf{W}_{\mathbf{V}^{(I)}}$ are learnable weight matrices. The attention maps $\mathbf{A}^{(I)}$ and $\mathbf{A}^{(II)}$ are then calculated via the difference of scaled-dot product as:

$$\begin{aligned} \mathbf{A}^{(I)} &= \left[\mathrm{f}\left(\frac{\mathbf{Q}^{(I)}_1 \cdot \mathbf{K}^{(I)T}_1}{\sqrt{d_{\mathbf{K}}}} \right) - \lambda \mathrm{f}\left(\frac{\mathbf{Q}^{(I)}_2 \cdot \mathbf{K}^{(I)T}_2}{\sqrt{d_{\mathbf{K}}}} \right) \right] \mathbf{V}^{(II)} \\ \mathbf{A}^{(II)} &= \left[\mathrm{f}\left(\frac{\mathbf{Q}^{(II)}_1 \cdot \mathbf{K}^{(II)T}_1}{\sqrt{d_{\mathbf{K}}}} \right) - \lambda \mathrm{f}\left(\frac{\mathbf{Q}^{(II)}_2 \cdot \mathbf{K}^{(II)T}_2}{\sqrt{d_{\mathbf{K}}}} \right) \right] \mathbf{V}^{(I)}, \end{aligned} \tag{8}$$

where f(·) is the softmax function, T denotes the matrix transpose, and $d_{\mathbf{K}}$ represents the dimension of the key vector and is used for normalization. $\lambda \in [0, 1]$ is the learnable parameter that controls the influence of either stream's attention. These attention maps are obtained for one attention head. For multiple differential-attention heads, the attention maps from each head are normalized (group-normalized) and scaled as $\mathbf{A}^{(I)'} - \mathbf{A}^{(I)} \cdot (1 - \lambda_{\text{init}})$ and $\mathbf{A}^{(II)'} - \mathbf{A}^{(II)} \cdot (1 - \lambda_{\text{init}})$, where $\lambda_{\text{init}} \in [0, 1]$. This formulation learns a contrastive attention map that suppresses common components between the two streams while preserving discriminative features, which is beneficial in OTFS systems where channel distortions affect both streams similarly. The group-normalized attention maps from multiple heads are concatenated and passed through a linear layer. The resulting outputs from both streams are then multiplied and fed into a fully connected layer for final classification.

IV. Experimental Settings

A. Dataset Generation

The EVA channel model [19] is used, and the dataset is synthetically generated in Python. Each sample is an $M \times N$ OTFS frame for a given modulation scheme, simulated over SNRs from -20 dB to 20 dB (with 2 dB steps) at mobilities of 100 km/h and 500 km/h. A total of 193,536 samples are generated and split into training, validation, and testing in a 7:1:2 ratio. Details are provided in TABLE I.

TABLE I: OTFS dataset specification.

Parameter	Value
Channel model	Extended Vehicular A (EVA)
Frequency spacing Δf	15 kHz
Fading model	Rayleigh
Vehicular speed v	100 *km/hr* & 500 *km/hr*
Total modulation schemes	BPSK, 8PSK, {4,8,16,32,64,128,256}-QAM
SNR range	-20 dB to 20 dB with 2 dB increments
Frames per SNR per scheme	1024
Number of delay bins M	16, 32
Number of Doppler bins N	16, 32
Delay values (in *nanoseconds*)	[0,30,150,310,370,710,1090,1730,2510]
Power delay profile (in dB)	[0,-1.5,-1.4,-3.6,-0.6,-9.1,-7.0,-12.0,-16.9]
Number of taps	9

B. Experimental Settings

DADS-Net is trained for 200 epochs with a batch size of 256 using the Adam optimizer and a fixed learning rate of 0.0001. Each stream has three CSA blocks with 64-filter convolutions (8×8, 4×4), while each CSA uses 32 filters and 4 attention heads. The DCAM uses 8 heads followed by a 32-neuron fully connected layer. A 10% dropout is applied, and the final FCNN outputs modulation class predictions.

V. Simulation Results and Discussions

TABLE II compares DADS-Net with LB-, FB-, and OTFS-based AMC methods across 100 km/h and 500 km/h for 16×16 and 32×32 frames. DADS-Net consistently achieves the best performance, with accuracies of 49.21%/48.97% (16×16) and 50.17%/49.63% (32×32) at 100/500 km/h, respectively. It is important to highlight that the absolute classification accuracy values are relatively low across all methods due to the challenging nature of OTFS-based AMC under high mobility and severe Doppler conditions. Although the numerical gains over the best baseline appear modest, the improvements are consistent across all evaluated conditions, which is critical in practical AMC systems where robustness and reliability are more important than isolated peak performance.

TABLE II: Proposed DADS-Net vs. prior OTFS-based AMC systems accuracy (%) comparison at 100 km/h and 500 km/h.

Frame-size	AMC System	100 km/h	500 km/h
16×16	ALRT [5]	14.2	13.1
	Random Forest [15]	41.23	40.17
	Decision Tree [15]	30.98	30.17
	Res-CLDNN [10]	46.21	45.39
	CNN-SE [16]	43.76	43.11
	2D-CNN [21]	49.01	48.52
	DADS-Net	**49.21**	**48.97**
32×32	ALRT [5]	15.92	13.2
	Random Forest [15]	41.93	41.71
	Decision Tree [15]	29.23	27.95
	Res-CLDNN [10]	41.96	41.01
	CNN-SE [16]	45.73	45.51
	2D-CNN [21]	49.44	48.42
	DADS-Net	**50.17**	**49.63**

Fig. 4 illustrates the classification accuracy vs. SNR for DADS-Net and prior OTFS-based AMC systems. Across both 16×16 and 32×32 frame-sizes, DADS-Net consistently achieves higher accuracy, particularly in low-to-moderate SNR regions, indicating its improved noise resilience and generalization capability. The performance gap widens with increasing SNR, underscoring DADS-Net's effective exploitation of spatial–channel correlations and its differential-attention mechanism for robust modulation recognition under dynamic channel conditions. Fig. 5 shows F1-score vs. SNR for each modulation. DADS-Net achieves near-perfect F1 for lower-order modulations (BPSK, QPSK, 8PSK) above 0 dB, while higher-order modulations (32QAM, 64QAM, 128QAM) improve steadily with SNR, demonstrating robustness in high-mobility, low-SNR conditions. Furthermore, Fig. 6 shows the t-SNE visualization of DADS-Net features for 32×32 frames at 500 km/h. Well-separated clusters for each modulation highlight its strong discriminative ability and robustness under high mobility. Finally, Fig. 7 shows DADS-Net confusion matrices for 32×32 frames at 500 km/h. The model achieves strong overall accuracy, with most predictions along the diagonal. Minor misclassifications occur for higher-order QAMs.

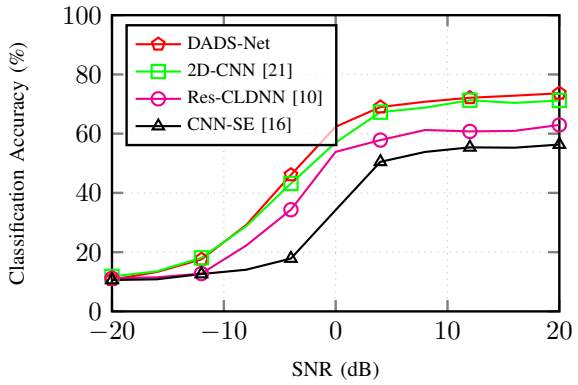

Fig. 4: Accuracy vs. SNR analysis between DADS-Net and prior OTFS-specific AMC systems, such as 2D-CNN [21], Res-CLDNN [10], and CNN-SE [16] across frame-size of 32×32, operating at 500 km/h.

DADS-Net requires 8.58 million FLOPs and 4.29 million MACs, with a total of 0.75 million trainable parameters. The model achieves an average latency of 1.05 ms per sample.

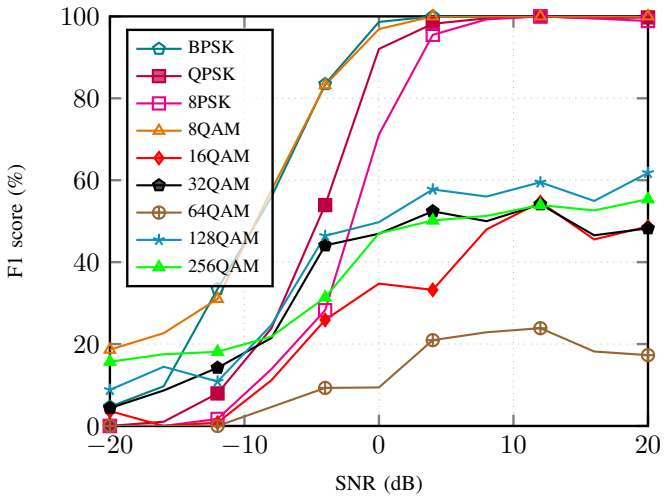

Fig. 5: F1-score vs. SNR for individual modulation schemes using the proposed DADS-Net with a 32×32 OTFS frame size operating at 500 km/h.

Fig. 6: t-SNE visualization of feature distributions for the proposed DADS-Net across 32×32 frame-size, at 500 km/h under the EVA channel.

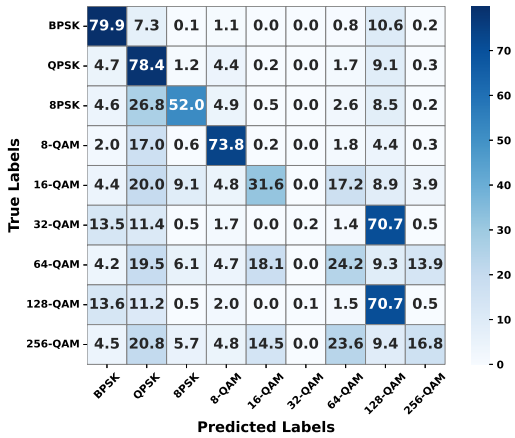

Fig. 7: Confusion matrix illustrating the overall performance of the proposed DADS-Net operating at 500 km/h across 32×32 frame-size.

Table III shows that single-stream DADS-Net (I/Q or A/P) achieves similar accuracy at 500 km/h, while the dual-stream model combining I/Q and A/P consistently outperforms them, reaching 49.63% for 32×32 frames, demonstrating the benefit of fusing complementary representations.

TABLE III: Performance analysis between single-stream DADS-Net without DCAM block employing I/Q or A/P representations and dual-stream DADS-Net at mobility of 500 km/h.

Architecture	**Features**	**Accuracy** (16×16)	**Accuracy** (32×32)
Single-stream	I/Q	48.01	48.18
Single-stream	A/P	48.26	48.39
Dual-stream	I/Q + A/P	48.97	49.63

VI. Conclusion

This paper presents DADS-Net, a dual-stream network with a DCAM for AMC in OTFS systems. By directly processing 2D OTFS frames, the model preserves structural information and effectively fuses in I/Q and A/P features. Results show that DADS-Net provides consistent performance improvements and enhanced robustness under high-mobility, high Doppler, and low SNR conditions. Although absolute accuracy remains limited due to the inherent complexity of OTFS-AMC, the proposed approach improves feature discrimination and noise resilience.

Acknowledgement

This research is supported by SPARC-UKIERI under Grant SPARCUKIERI/ 2024-2025/P3342.

References

[1] T. Rajkhowa, A. Singh, S. Sharma, K. Deka, and C. Adjih, "DSCAN-Net: A dual-stream network for classifying modulation schemes in OTFS systems," in *2025 Nat. Conf. on Commun. (NCC)*. IEEE, 2025.

[2] R. Hadani and A. Monk, "OTFS: A new generation of modulation addressing the challenges of 5G," *arXiv preprint arXiv:1802.02623*, 2018.

[3] R. Hadani, S. Rakib, M. Tsatsanis, A. Monk, A. J. Goldsmith, A. F. Molisch, and R. Calderbank, "Orthogonal time frequency space modulation," in *2017 IEEE Wireless Commun. and Netw. Conf. (WCNC)*. IEEE, 2017, pp. 1–6.

[4] K. Deka, A. Thomas, and S. Sharma, "OTFS-SCMA: A code-domain NOMA approach for orthogonal time frequency space modulation," *IEEE Trans. on Commun.*, vol. 69, no. 8, pp. 5043–5058, 2021.

[5] A. Polydoros, "Likelihood ratio tests for modulation classification," in *MILCOM 2000 Proc. 21st Century Mil. Commun. Architectures and Technol. for Inf. Superiority*, vol. 2. IEEE, 2000, pp. 670–674.

[6] F. Hameed, O. A. Dobre, and D. C. Popescu, "On the likelihood-based approach to modulation classification," *IEEE Trans. on Wireless Commun.*, vol. 8, no. 12, pp. 5884–5892, 2009.

[7] S. Huang, Y. Chen, J. He, S. Chang, and Z. Feng, "Channel-robust automatic modulation classification using companding spectral quotient cumulants," *IEEE Trans. on Veh. Technol.*, vol. 73, no. 11, pp. 17 749–17 753, 2024.

[8] A. Swami and B. M. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Trans. on Commun.*, vol. 48, no. 3, pp. 416–429, 2000.

[9] J. Li, Q. Meng, G. Zhang, Y. Sun, L. Qiu, and W. Ma, "Automatic modulation classification using support vector machines and error correcting output codes," in *IEEE 2nd Inf. Technol., Netw., Electron. and Automat. Control Conf. (ITNEC)*, 2017, pp. 60–63.

[10] A. Kumar, U. Satija *et al.*, "Residual stack-aided hybrid CNN-LSTM based automatic modulation classification for orthogonal time-frequency space system," *IEEE Commun. Lett.*, vol. 27, no. 12, pp. 3255–3259, 2023.

[11] Z. Zhang, H. Luo, C. Wang, C. Gan, and Y. Xiang, "Automatic modulation classification using CNN-LSTM based dual-stream structure," *IEEE Trans. on Veh. Technol.*, vol. 69, no. 11, pp. 13 521–13 531, 2020.

[12] Z. Zhang, C. Wang, C. Gan, S. Sun, and M. Wang, "Automatic modulation classification using convolutional neural network with features fusion of SPWVD and BJD," *IEEE Trans. on Signal and Inf. Process. over Netw.*, vol. 5, no. 3, pp. 469–478, 2019.

[13] Y. Chen, W. Shao, J. Liu, L. Yu, and Z. Qian, "Automatic modulation classification scheme based on LSTM with random erasing and attention mechanism," *IEEE Access*, vol. 8, pp. 154 290–154 300, 2020.

[14] P. Ghasemzadeh, S. Banerjee, M. Hempel, and H. Sharif, "A new framework for automatic modulation classification using deep belief networks," in *IEEE Int. Conf. on Commun. Workshops (ICC Workshops)*. IEEE, 2020, pp. 1–6.

[15] P. Maurya, K. Captain, A. Parmar, and P. K. Shah, "Modulation classification for OTFS-NOMA in heterogeneous user mobility profile," in *IEEE Int. Conf. on Signal Process. and Commun. (SPCOM)*, 2024, pp. 1–5.

[16] J. Zhou, X. Liao, and Z. Gao, "Deep learning-based automatic modulation recognition in OTFS and OFDM systems," in *IEEE 97th Veh. Technol. Conf. (VTC2023-Spring)*, 2023, pp. 1–5.

[17] Z. Zhang, H. Luo, C. Wang, C. Gan, and Y. Xiang, "Automatic modulation classification using CNN-LSTM based dual-stream structure," *IEEE Trans. on Veh. Technol.*, vol. 69, no. 11, pp. 13 521–13 531, 2020.

[18] T. Ye, L. Dong, Y. Xia, Y. Sun, Y. Zhu, G. Huang, and F. Wei, "Differential transformer," *arXiv preprint arXiv:2410.05258*, 2024.

[19] A. Singh, S. Sharma, K. Deka, and V. Bhatia, "DL-based OTFS signal detection in presence of hardware impairments," *IEEE Wireless Commun. Lett.*, vol. 12, no. 9, pp. 1533–1537, 2023.

[20] A. Vaswani, "Attention is all you need," *Advances in Neural Inf. Process. Syst.*, vol. 30, 2017.

[21] T. Rajkhowa, S. Sharma, and K. Deka, "Automatic modulation classification for hardware-impaired OTFS transceiver systems," *IEEE Wireless Commun. Lett.*, vol. 15, pp. 615–619, 2026.