ITC 3/55
Information Technology and Control
Vol. 55 / No. 3/ 2026
pp. 1357-1363
DOI 10.5755/j01.itc.55.3.44966

Dual-Angle Spatiotemporal Network (DAST-Net): A Hybrid LSTM-Transformer Architecture for Online Signature Verification in Human-Computer Interface Systems

Received 2026/05/07 | Accepted after revision 2026/08/05

HOW TO CITE: Hao, Y., Xu, D., Zheng, Z., Liu, Y. (2026) Dual-Angle Spatiotemporal Network (DAST-Net): A Hybrid LSTM-Transformer Architecture for Online Signature Verification in Human-Computer Interface Systems. *Information Technology and Control*, 55(3), 1357-1363. https://doi.org/10.5755/j01.itc.55.3.44966

Dual-Angle Spatiotemporal Network (DAST-Net): A Hybrid LSTM-Transformer Architecture for Online Signature Verification in Human-Computer Interface Systems

Yanli Hao*, Dahao Xu, Zuyang Zheng
Department of Forensic Science, Fujian Police College, Fuzhou 350007, China; e-mails: haoyl-0427@163.com (Y. Hao); 2492224916@qq.com (D. Xu); 285982951@qq.com (Z. Zheng)

Yingxia Liu
Department of Forensic Science, Chongqing Police College, Chongqing 401331, China; e-mail: liuyingxia27@163.com (Y. Liu)

Corresponding author: haoyl-0427@163.com

Secure identity verification in human-computer interface (HCI) systems is critical for protecting information system access. Existing methods often fail to integrate fine-grained temporal dynamics with long-range structural dependencies and largely neglect sensor-derived pen kinematic data. This paper presents the Dual-Angle Spatiotemporal Network (DAST-Net), a hybrid architecture for signature verification. The framework incorporates azimuth and altitude angles, captured by standard HCI sensors,

as novel dynamic cues. A stacked LSTM models short-term temporal patterns, while a Transformer encoder captures cross-stroke correlations. Evaluated on three datasets (self-constructed, SVC2004-Task2, MYCT-100), DAST-Net achieves 99.43% accuracy and a 0.89% equal error rate (EER), representing a relative reduction of 36–63% over state-of-the-art baselines. These results demonstrate that fusing dual-angle features within a hybrid LSTM-Transformer framework provides a robust and implementable solution for enhancing information system security in HCI applications.

KEYWORDS: Online Signature Verification, Spatiotemporal Network, Dual-Angle Features, LSTM-Transformer, Human-Computer Interaction

Introduction

Identity authentication faces increasing forgery threats in the current digital era. Traditional methods such as passwords and PINs have proven vulnerable to various attacks. Online handwritten signatures offer inherent biometric uniqueness and legal validity. They have emerged as a pivotal technology in digital forensics, financial transactions, and electronic contracting [8]. Unlike knowledge-based methods, handwritten signatures capture distinctive neuromuscular coordination patterns. This makes them inherently difficult to replicate accurately.

Deep learning has significantly influenced the evolution of online signature verification. Early research concentrated on static feature extraction, analyzing spatial attributes such as contour geometry and curvature [2, 9]. While computationally efficient, these approaches overlook the temporal dimension of handwriting. Subsequent studies explored dynamic features including velocity profiles and pressure variations [7-8]. These capture time-varying aspects but often fail to encode holistic structural relationships. More recently, multimodal fusion strategies have been pursued to enhance verification accuracy [4, 11, 14-15].

Despite these advances, two fundamental challenges persist. First, modality silos limit performance: unimodal systems relying exclusively on either spatial or temporal features fail to capture their complementary nature. Second, multimodal frameworks struggle to balance short-term dynamics with long-range structure. Local dynamics-such as abrupt velocity changes-and global properties-such as cross-stroke correlations-are frequently modeled in isolation. This reduces robustness against skilled forgeries, where impostors may replicate either visual appearance or temporal patterns but rarely both.

Conventional feature extraction methods have notable limitations. Static approaches lack temporal information, making them vulnerable to visual forgery [1, 9]. Dynamic techniques such as DTW and HMM capture temporal variations but overlook subtle biomechanical cues [6, 10]. Pen inclination angles-azimuth and altitude-offer a valuable alternative. These angles encode fine-grained motor habits that are deeply ingrained during natural writing. They reflect individualized grip postures that remain stable across genuine signatures yet are exceptionally difficult to forge [1, 3]. Deep learning has advanced the field through automatic feature learning. CNNs extract hierarchical spatial representations. RNNs and LSTMs model sequential dependencies [5, 8, 13], while Transformers capture long-range contextual relationships [4, 10]. Nevertheless, each architecture addresses only part of the verification challenge.

Hybrid architectures have shown significant potential [3, 12, 14]. Combinations of CNNs with RNNs represent progress toward comprehensive spatiotemporal representation. However, existing approaches often rely on simple concatenation or late fusion, which fail to optimally balance contributions from different modalities. Moreover, despite the recognized value of pen inclination angles, these cues remain underutilized in contemporary deep learning systems [1, 3].

To address these challenges, this paper proposes the Dual-Angle Spatiotemporal Network (DAST-Net). This hybrid architecture integrates conventional features with pen inclination angles in a unified LSTM-Transformer framework. The framework comprises four integrated modules: a denoising autoencoder (DAE) for preprocessing, a stacked LSTM for local temporal dynamics, a Transformer encoder for long-range structural dependencies, and a hier-

archical weighted fusion strategy that emphasizes discriminative components-particularly azimuth and altitude variations [14]. The principal contributions are threefold: First, we introduce comprehensive integration of pen inclination angles as novel dynamic features that encode writer-specific patterns enhancing discriminative capacity against skilled forgeries. Second, we propose a hybrid LSTM-Transformer architecture that synergistically combines LSTM's fine-grained local dynamics with Transformer's global dependencies. Third, through extensive experiments on self-constructed data, SVC2004-Task2, and MYCT-100, we demonstrate that DAST-Net achieves 99.43% accuracy and 0.89% EER, reducing EER by 36-63% compared to existing methods. The remainder of this paper is organized as follows. Section II details materials and methods. Section III presents experimental results. Section IV provides discussion. Section V concludes the paper.

2. Materials And Methods

2.1. Data Acquisition and Preprocessing

Data Acquisition: Wacom Intuos Pro tablet (PTK470) with 100Hz sampling rate, 4096 pressure levels, and azimuth/altitude capture was used. Thirty participants (aged 20-45) provided 10 genuine signatures each and forged one other participant's signature, yielding 300 genuine and 300 forged samples.

Preprocessing: Min-max normalization applied to all features:

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}}. \quad (1)$$

Dynamic Time Warping (DTW) unifies signature sequence lengths:

$$DTW(S,T) = min\sqrt{\sum_{k=1}^{k}(s_{w_k} - t_{w_k})^2}. \quad (2)$$

Gaussian noise (μ=0, σ=0.01) added and denoising autoencoder (DAE) with 3 encoder layers (64→32→16) trained for noise robustness. To ensure realistic generalization evaluation, a user-independent partitioning strategy was adopted. The 30 users were divided into 24 users (80%) for training, 3 users (10%) for validation, and 3 users (10%) for testing. All signatures from a given user were confined to a single split, preventing data leakage.

2.2. Feature Extraction

DAST-Net extracts multimodal features. Static features include global geometric descriptors (height/width ratio, total stroke length, curvature) and local spatial features.

Dynamic features capture temporal dynamics and motor habits, including writing duration, segmental speed, pressure, and pen-inclination angles (azimuth and altitude). These angles create a unique "directional fingerprint" that is difficult to forge (see Table 1).

$$\theta_{\text{azimuth}} = \arctan\left(\frac{y_{\text{tilt}}}{x_{\text{tilt}}}\right) \times \frac{180^{\circ}}{\pi}. \quad (3)$$

$$\theta_{\text{altitude}} = \arcsin\left(\frac{z_{\text{tilt}}}{\sqrt{x_{\text{tilt}}^2 + y_{\text{tilt}}^2 + z_{\text{tilt}}^2}}\right) \times \frac{180^{\circ}}{\pi}. \quad (4)$$

Feature fusion via weighted summation:

$$F = \lambda F_{static} + (1 - \lambda) F_{dynamic}. \quad (5)$$

Optimal λ = 0.3 determined by 5-fold cross-validation.

Table 1
Static And Dynamic Features.

Feature Category	Description
Static Features	Height, width, height/width ratio, total stroke length, curvature, origin distances
Dynamic Features	Writing time, lifting time, down time, duration ratios, segmental speed, segment pressure
Pen Inclination	Azimuth angle, altitude angle

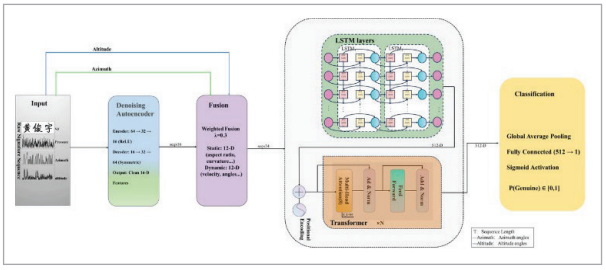
2.3 Network Architecture

DAST-Net comprises four modules (Figure 1)

Denoising Autoencoder (DAE): Pre-trained DAE processes raw input for noise-robust representations.

LSTM for Local Modeling: Two-layer stacked LSTM with 128 hidden units per layer models short-term temporal dynamics (velocity and pressure fluctua-

Figure 1
DAST-Net architecture with dual-angle spatiotemporal fusion, consisting of four sequential modules: Denoising Autoencoder, Stacked LSTM, Transformer Encoder, and Hierarchical Fusion & Classification Layer.



tions). Dropout 0.2 applied. Output: 128-dimensional feature vectors.

Transformer for Long-Range Dependencies: 128-dimensional LSTM outputs projected to 512 dimensions. Six-layer Transformer encoder with 8 attention heads captures cross-stroke correlations and consistent pen inclination patterns. Positional encoding preserves temporal order.

Classification Layer: Final layer produces verification decision (genuine/forged).

2.4. Loss Function and Optimization

Hybrid loss combines binary cross-entropy (L_{CE}) and contrastive loss (L_{CON}):

$$L_{\mathrm{CE}}=-\sum_{i=1}^{N}\sum_{c=1}^{C} w_c \cdot y_{ic}\log(p_{ic}), \tag{6}$$

where $w_c(e.g.,w_{genuine}=0.5,w_{forged}f=0.5)$ mitigates class imbalance.

$$L_{\mathrm{CON}}=\frac{1}{2N}\sum_{i=1}^{N}[y_i d_i^2+(1-y_i)\max(0,m-d_i)^2]. \tag{7}$$

where d_i is the Euclidean distance between feature pairs, $y_i \in$ (1 for forged pairs), and m=1.0 is the margin hyperparameter.

$$L_{total}=\alpha L_{CE}+\beta L_{CON}. \tag{8}$$

Optimal weights: α=0.6, β=0.4. Adam optimizer with a learning rate of 0.001 and a weight decay of 0.01. L_2 regularization (1×10^{-5}) for 150 epochs, batch size 16. Data augmentation: time warping and magnitude scaling, see Table 2.

Table 2
Hyperparameters.

Parameter	Value
Optimizer	Adam (lr=0.001, weight decay=0.01)
Regularization	L_2 ($1e^{-5}$)
Loss Weights	α=0.6, β=0.4
Epochs/Batch	150/16
Augmentation	Time Warping + Magnitude Scaling

3. Results

3.1. Evaluation Metrics

Performance evaluated using Accuracy (ACC), False Acceptance Rate (FAR), False Rejection Rate (FRR), and Equal Error Rate (EER):

$$Acc=\frac{TP+TN}{TP+TN+FP+FN} \tag{9}$$

$$FAR=\frac{FP}{TN+FP} \tag{10}$$

$$FRR=\frac{FN}{TP+FN} \tag{11}$$

$$EER=FAR_T=FRR_T. \tag{12}$$

3.2. Model Performance

DAST-Net evaluated on self-constructed, SVC2004-Task2, and MYCT-100 datasets. Table 3 shows superior performance across all datasets, with 99.43% ACC and 0.89% EER on self-constructed data. Figure 2 confusion matrices confirm balanced low FAR and FRR.

Table 3
Performance Across Datasets.

Dataset	FAR (%)	FRR (%)	ACC (%)	EER (%)
Self-constructed	0.67	0.33	99.43	0.89
SVC2004-Task2	2.00	1.67	98.23	2.15
MYCT-100	1.00	0.67	98.76	1.02

3.3. Ablation Study

3.3.1. Feature Combination Impact

Three feature combinations evaluated (Table 4). Results (Table 5) show dual-angle features (Combination C) dramatically improve performance, reducing EER from 10.33% to 0.89% on self-built dataset. To evaluate the contribution of different feature categories, we construct three feature combinations for ablation analysis: Combination A uses static features only (12 dimensions); Combination B adds dynamic temporal features (velocity, pressure, and timing parameters) but excludes inclination angles (18 dimensions); and Combination C incorporates the full feature set including static features, dynamic temporal features, and pen inclination angles (azimuth and altitude), totaling 24 dimensions.

Figure 2

Confusion matrices for three datasets. (a) Self-constructed; (b) SVC2004-Task2; and (c) MYCT-100.

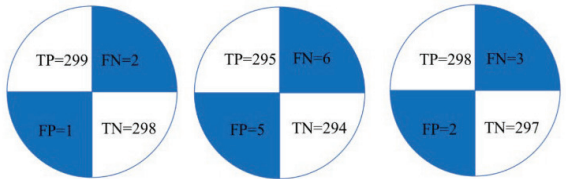


Table 4

Feature Combinations.

Comb.	Features
A	Static only (12-dim)
B	Static + dynamic (velocity, pressure) (18-dim)
C	Static + dynamic + azimuth & altitude (24-dim)

Table 5

Feature Combination Results.

Dataset	Combination	FAR (%)	FRR (%)	ACC (%)	EER (%)
Self-built	A	11.00	9.67	89.67	10.33
	B	6.33	5.67	94.00	6.00
	C	0.67	0.33	99.43	0.89
MYCT-100	A	12.33	10.67	88.33	11.50
	B	8.00	7.33	92.33	7.67
	C	1.00	0.67	98.76	1.02
SVC2004	A	15.00	13.67	86.00	14.33
	B	9.67	8.33	91.33	9.00
	C	2.00	1.67	98.23	2.15

3.3.2. Feature Space Discriminability Analysis

We quantitatively evaluated feature discriminability using intra-to-inter class distance ratios. Without inclination angles (Combination B), the intra-class

Table 6

Architecture Comparison.

Dataset	Model	Params ($\times 10^4$)	Time (ms)	ACC (%)
Self-built dataset	LSTM-Transformer	120.3	22.1	99.43
	fuseTSN	95.8	18.5	98.6
	CNN-LSTM	80.5	15.2	97.8
SVC2004-Task2	LSTM-Transformer	120.3	22.1	98.23
	fuseTSN	95.8	18.5	97.3
	CNN-LSTM	80.5	15.2	96.5
MYCT-100	LSTM-Transformer	120.3	22.1	98.76
	fuseTSN	95.8	18.5	98.0
	CNN-LSTM	80.5	15.2	97.2

distance was 0.42 and inter-class distance 0.54, yielding a ratio of 0.78—indicating substantial class overlap. After integrating azimuth and altitude angles (Combination C), intra-class distance reduced to 0.31 while inter-class distance increased to 0.97, achieving a ratio of 0.32—a 59% improvement. This enhanced separability is corroborated by ablation results (Table 5), where EER improved from 6.00% (Combination B) to 0.89% (Combination C)—an 85.2% relative reduction. These quantitative measures confirm that dual-angle features capture complementary discriminative information essential for robust signature verification.

3.3.3. Architecture Comparison

DAST-Net was compared with CNN-LSTM and fuseTSN. Table 6 shows DAST-Net consistently outperforms these baselines. On the MYCT-100 dataset, DAST-Net's EER is 63.6% lower than that of the CNN-LSTM model. Inference time (22.1ms) remains practical for real-time applications.

4. Discussion

Results robustly validate DAST-Net's effectiveness. Integrating dual-angle features with LSTM-Transformer addresses inadequate fusion of local/global information and neglect of fine-grained pen kinematics.

Ablation studies confirm azimuth and altitude angles capture unique "behavioral biometrics" highly discriminative against forgery. EER reduction from 10.33% to 0.89% demonstrates forgers struggle to replicate subtle motor habits related to pen orientation. This is further supported by the t-SNE visualization (Figure 4), which shows approximately 70–80% separation between genuine and forged samples after incorporating pen inclination angles.

LSTM-Transformer design outperforms CNN-LSTM and fuseTSN. LSTM effectively models local fluctuations while Transformer captures long-range cross-stroke correlations. This combined understanding is crucial for differentiating genuine signatures from skilled forgeries that replicate individual strokes but fail to maintain holistic structural integrity.

Limitations: The participant pool (30 individuals) is relatively small, which may limit generalizability. Future work will focus on larger-scale data collection and model compression techniques such as knowledge distillation to enable deployment on resource-constrained devices.

5. Conclusion

This paper proposed DAST-Net, a hybrid LSTM-Transformer architecture for online signature verification. The framework integrates azimuth and altitude pen inclination angles as key dynamic features alongside traditional descriptors. Evaluated on three benchmark datasets, the proposed model achieves **99.43% accuracy** and a **0.89% equal error rate (EER)**, significantly outperforming state-of-the-art baselines with a relative EER reduction of 36–63%. Ablation studies confirm that dual-angle features improve feature space separability, reducing EER from 10.33% to 0.89% on the self-constructed dataset. From an information systems perspective, DAST-Net provides a practical solution for enhancing authentication security within HCI environments, with compatibility for deployment in financial transaction systems, access control platforms, and electronic contracting services.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, author-ship, and/or publication of this article.

Data Sharing Agreement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Funding

This work was supported by the 2025 Chongqing Language and Character Research Project under Grant No. yyk25112, and the 2025 College Student Innovation and Entrepreneurship Training Program Project entitled "Research on Online Handwritten Signature Verification Based on Multidimensional Features" under Grant No. S202511495033.

References

1. Diaz, M., Ferrer, M. A., Gil, J. M., Rodriguez, R., Zhang, P., Jin, L. Online Signature Verification Based on The Lagrange Formulation With 2D And 3D Robotic Models. Pattern Recognition, 2025, 164, Art. No. 111581. https://doi.org/10.1016/j.patcog.2025.111581
2. Hochreiter, S., Schmidhuber, J. Long Short-Term Memory. Neural Computation, 1997, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735
3. Hu, Y., Chen, L. Online Signature Verification Based on a Single Template Via Elastic Curve Matching. Sensors, 2019, 19(22), Art. No. 4858. https://doi.org/10.3390/s19224858
4. Jing-Yu, L., Yan-Ming, Z., Fei, Y., Cheng-Lin, L. Transformer-Based Stroke Relation Encoding for Online Handwriting and Sketches. Pattern Recognition, 2024, 148, Art. No. 110131. https://doi.org/10.1016/j.patcog.2023.110131
5. Lai, S., Jin, L., Yang, W. Online Signature Verification Using Recurrent Neural Network and Length-Normalized Path Signature. Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, Japan, Nov. 2017, Pp. 51-56. https://doi.org/10.1109/ICDAR.2017.73
6. Nakamura, S., Saeed, K., Szymkowski, P., Nishiuchi, N. Proposal of Online Handwritten Signature Authentication Using Minutiae Matching for Kanji Characters. Proceedings of the International Conference on Computer Information Systems and Industrial Management (CISIM), 2024, 17-29. https://doi.org/10.1007/978-3-031-71115-2_2
7. Pandey, A., Bhavsar, A., Nigam, A., Acharya, D., Singh, P., Bhardwaj, A., Diwakar, M. SIGN-GPT: Generating Writer-Specific Online Signatures for Digital Verification. Human-Centric Computing and Information Sciences, 2025, 15, Art. No. 50. Doi: 10.22967/HCIS.2025.15.050.
8. Pascual-Gaspar, J. M., Faundez-Zanuy, M., Vivaracho, C. Fast On-Line Signature Recognition Based on VQ With Time Modeling. Engineering Applications of Artificial Intelligence, 2011, 24(2), 368-377. https://doi.org/10.1016/j.engappai.2010.10.015
9. Plamondon, R., Gorette, L. Automatic Signature Verification and Writer Identification-The State of The Art. Pattern Recognition, 1989, 22(2), 107-131. https://doi.org/10.1016/0031-3203(89)90059-9
10. Roszczewska, K., Niewiadomska-Szynkiewicz, E. Online Signature Biometrics for Mobile Devices. Sensors, 2024, 24(11), Art. No. 3524. https://doi.org/10.3390/s24113524
11. Sekhar, V. C., Gautam, A., Viswanath, P., Sreeja, S. R., Sai, G. R. K. TSOSVNet: Teacher-Student Collaborative Knowledge Distillation for Online Signature Verification. Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Paris, France, Oct. 2023, 742-751. https://doi.org/10.1109/ICCVW60793.2023.00082
12. Singhal, M., Shinghal, K. An Efficient, Fast and Accurate Online Signature Verification Using Blended Feature Vector and Deep Learning. IETE Journal of Research, 2024, 70(9), 7354-7364. https://doi.org/10.1080/03772063.2024.2351567
13. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Ortega-Garcia, J. Exploring Recurrent Neural Networks for On-Line Handwritten Signature Biometrics. IEEE Access, 2018, 6, 5128-5138. https://doi.org/10.1109/ACCESS.2018.2793966
14. Vorugunti, C. S., Pulabaigari, V., Mukherjee, S. OSVFuseNet: Online Signature Verification by Feature Fusion and Depth-Wise Separable Convolution Based Deep Learning. Neurocomputing, 2020, 409, 157-172. https://doi.org/10.1016/j.neucom.2020.05.072
15. Xie, L. Y., Wu, Z. C., Zhang, X., Li, Y. Synchronous Spatio-Temporal Signature Verification Via Fusion Triplet Supervised Network. Engineering Applications of Artificial Intelligence, 2023, 123, Art. No. 106378. https://doi.org/10.1016/j.engappai.2023.106378

This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 (CC BY 4.0) License (http://creativecommons.org/licenses/by/4.0/).