



ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И МАШИННОЕ ОБУЧЕНИЕ/ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

DOI: <https://doi.org/10.60797/IRJ.2026.172.87> EDN: HDITKY**SIGNATURE FORGERY DETECTION USING DEEP LEARNING**

Research article

Paramonova A.D.^{1,*}¹ORCID : 0009-0008-9568-4527;¹National Research University Higher School of Economics, Moscow, Russian Federation

* Corresponding author (paramonovaanna2005[at]gmail.com)

Suggested: 08.07.2026; Accepted: 29.09.2026; Published: 16.10.2026

Abstract

Handwritten signature is one of the most frequently used methods to identify a person. The field of offline signature verification (OSV) has been extensively studied in the past, with researchers mainly employing CNN- or Transformer-based novel architectures. There are only a few cases of using transfer learning in OSV and most of them are based on models from 2019 and earlier. Therefore, we propose to use pretrained hybrid models ConvNeXt (2022) and CoAtNet (2021), which combine the strengths of self-attention and convolution for transfer learning. Their architectural details and results of testing against models in image classification present great potential in the offline signature verification field. We employ several training schemes and verification methods to fully assess the capabilities of the proposed models in this forensic field. The experimental results have shown that, although the models do not produce sufficient results on Western language datasets, they compete favorably on UTSig, a Persian dataset, achieving 82.3% accuracy.

Keywords: offline signature verification, writer-independent verification, deep learning, transfer learning, ConvNeXt, CoAtNet.

ОБНАРУЖЕНИЕ ФАЛЬСИФИЦИРОВАННЫХ ПОДПИСЕЙ С ПРИМЕНЕНИЕМ ГЛУБИННОГО ОБУЧЕНИЯ

Научная статья

Парамонова А.Д.^{1,*}¹ORCID : 0009-0008-9568-4527;¹Высшая школа экономики, Москва, Российская Федерация

* Корреспондирующий автор (paramonovaanna2005[at]gmail.com)

Предложена: 08.07.2026; Принята: 29.09.2026; Опубликовано: 16.10.2026

Аннотация

Рукописная подпись — один из наиболее часто используемых методов для идентификации человека. Область верификации подписей хорошо изучена, однако в основном исследователи используют придуманные с нуля архитектуры, основанные на свёрточных нейронных сетях или трансформерах. Существует лишь несколько примеров использования трансферного обучения для этой задачи, и большинство из них основаны на моделях, датированных 2019 годом и ранее. В связи с этим мы предлагаем использовать для трансферного обучения предобученные модели ConvNeXt (2022) и CoAtNet (2021), которые сочетают в себе преимущества механизма самовнимания и свертки. Их архитектурные детали и результаты в задаче классификации изображений дают основания ожидать конкурентные результаты в рассматриваемой области. Мы используем несколько подходов к тренировке моделей и принятию решений на основании извлеченных признаков. В ходе экспериментов было выяснено, что, несмотря на отсутствие удовлетворительных результатов на базах данных западных языков, предлагаемые модели показывают конкурентоспособное качество на базе данных UTSig персидского языка с точностью 82.3%.

Ключевые слова: офлайн-верификация подписей, независимая от автора верификация, глубинное обучение, трансферное обучение, ConvNeXt, CoAtNet.

Introduction

Biometrics use physiological or behavioral attributes to verify an individual's identity. Unlike passwords or keys, biometric traits are unique and cannot be lost, making them a reliable means of authentication. Among behavioral attributes, handwritten signatures remain one of the most widely used means of authentication in financial institutions, insurance, monetary transactions, etc. Therefore, signature verification, i.e., determining whether a signature is genuine, remains a critical task. However, manual verification is tedious and time-consuming, which has led to the development of automated signature verification systems based on machine learning [1].

Signature verification systems are generally divided into online and offline approaches. Online systems capture dynamic features such as pen pressure, which makes forgery detection easier but requires special input devices and is therefore less common in practice. This study focuses on offline signature verification (OSV), where signatures are represented as images obtained from scanners, making it more difficult to detect a fake. Among the three main types of offline forgeries — random, simple, and skilled — this research focuses on skilled forgeries as the most challenging ones [2]. A Writer Independent (WI)

approach is adopted: the model captures global characteristics of genuine and forged signatures, avoiding retraining when new users are enrolled.

OSV remains challenging due to high intra-personal and the resulting low inter-personal signature variations. Existing OSV methods commonly rely on convolutional neural networks (CNNs), which effectively extract multi-level local features through convolutional and pooling layers [2], [3], [4]. With the emergence of Transformers, recent studies have incorporated self-attention mechanisms to capture global image context [5]. Hybrid CNN-Transformer models combine the strengths of both approaches by extracting local signature details while preserving broader spatial dependencies. Several researchers have used such architectures to achieve competitive results compared to SigNet [3], TransOSV [5], ViT [6], and other advanced systems [7]. Simultaneously, transfer learning remains relatively underexplored in OSV, despite its ability to address the problem of limited signature datasets, improve generalization, and accelerate training. Although existing transfer learning approaches have shown promising results [8], [9], many of the models used in this setting are comparatively outdated and do not reflect recent progress in hybrid computer vision architectures.

Thus, the objective of this study is to develop and evaluate a transfer learning-based approach for the WI OSV task using modern hybrid computer vision architectures, aimed at improving the detection of skilled forgeries under limited-data conditions.

The novel contributions of this work are as follows:

1. Hybrid ConvNeXt [10] and CoAtNet [11] models are employed for OSV using a transfer learning approach. To the best of our knowledge, these models have not been previously used in the field.
2. The proposed architectures are evaluated under two training schemes: standalone classification and Siamese networks to assess their applicability to OSV from both direct classification and similarity-learning perspectives. Several verification strategies are compared, including Euclidean distance-based verification and cosine similarity-based verification.

Research methods and principles

Generally, OSV methods consist of three stages: data pre-processing, feature extraction, and signature verification.

2.1. Data preprocessing

Building OSV systems has proven to be a challenge since data acquisition is done under different conditions. To address variability in background artifacts and remove irrelevant patterns, several preprocessing steps are performed. In the initial phase, the Otsu threshold algorithm [12] is used for background elimination. Background pixels are converted to white, while grayscale values are retained for the foreground. The image is centered on an 840×1360 canvas by calculating the center of mass after binarization. Signatures are resized to 256×256 , then randomly cropped to 224×224 for training and center-cropped for evaluation. Original and preprocessed images are shown in Figure 1.



Figure 1 - Sample signature:
a) before preprocessing; b) after preprocessing
DOI: <https://doi.org/10.60797/IRJ.2026.172.87.1>

2.2. Feature extraction

For feature extraction, ConvNeXt and CoAtNet are pretrained on ImageNet-21K [13] – a large public dataset containing 14 million images. The first layers of the models are adapted to grayscale input by averaging pretrained weights across the RGB channels, eliminating color information that is irrelevant to signature verification.

2.2.1. ConvNeXt

ConvNeXt is a model made up of CNN modules that incorporates various Transformer-inspired techniques [10]. Its publicly available configurations are summarized in Table 1:

Table 1 - ConvNeXt configurations

DOI: <https://doi.org/10.60797/IRJ.2026.172.87.2>

Size	# Params, millions	# Flops, billions
Tiny	29	4.5
Small	50	8.7

Size	# Params, millions	# Flops, billions
Base	89	15.4
Large	198	34.4

The model outperforms several state-of-the-art vision architectures, including Swin Transformer, DeiT, and ViT, while remaining competitive with EfficientNet and EfficientNetV2 and preserving the simplicity of CNNs. Since ViT has already shown promising results in OSV, ConvNeXt may also be effective for this task. Its successful adaptation to related visual recognition problems, such as emotion recognition [14], further supports its investigation in offline signature verification.

2.2.2. CoAtNet

CoAtNet is based on integrating convolution and self-attention within computational blocks and stacking them into a unified network [11]. The available model configurations are summarized in Table 2:

Table 2 - CoAtNet configurations

DOI: <https://doi.org/10.60797/IRJ.2026.172.87.3>

Size	# Params, millions	# Flops, billions
Base	75	15.7
Large	168	34.7

CoAtNet outperforms ViT, Swin, and EfficientNet on ImageNet-1K, ImageNet-21K, and JFT-300M, demonstrating better scaling efficiency than pure CNNs and Transformers. These results motivate its application to OSV.

2.3. Signature verification

During this stage, the query signature is compared with the reference in the embedding space. In forensic tasks, it is reasonable to look at the spatial relationships between the query and the original embeddings when identifying the likeness of the former to the latter. Thus, after training, the classifier head is removed, and each model is used solely for embedding extraction. Two distance metrics are evaluated to measure signature similarity.

2.3.1. Euclidean distance

One of the most frequent techniques of verification [7] in OSV tasks is based on Euclidean distance.

Accepting a pair of embeddings (reference, query), Euclidean distance is calculated between two signatures, following the method described in [7]. If it is greater than a predefined threshold, the query is said to be forged. The threshold is calculated as follows:

At first, m random genuine signatures from each user are selected. The distances between them and the reference signatures are calculated. The results are aggregated into one array of numbers with maximum and minimum defined as:

$$d_{\max} = \max(d_1, d_2, \dots, d_{n-1}), \quad d_{\min} = \min(d_1, d_2, \dots, d_{n-1}),$$

where n is $m \times (\text{\# signatures used per person})$. A set of candidate thresholds is then generated with a fixed step:

$$D = (d_{\min}, d_{\min} + \text{step}, \dots, d_{\max})$$

The optimal threshold d from D is selected on the training set, with mm varied from 1 to 12 and the step size fixed at 0.01.

2.3.2. Cosine similarity

Cosine similarity measures the angle between the embeddings vector. Therefore, it is less sensitive to pen pressure, resolution, and brightness, focusing instead on shape-related differences between embeddings. During experiments, m genuine signature embeddings are averaged into one. This improves robustness by preserving the main biological pattern while averaging out random variations. A query is classified as genuine if its cosine similarity exceeds the threshold.

Experimental setup

Experiments are conducted over 3 OSV benchmark datasets: GPDS Synthetic [15], CEDAR [16], and UTSig [17]. Datasets cover multiple languages, their statistics and examples are shown in Table 3 and Figure 2:

Table 3 - Data statistics

DOI: <https://doi.org/10.60797/IRJ.2026.172.87.4>

Dataset	Language	No. of signers	Genuine per person	Forged per person
GPDS Synthetic	Western	4000 (first 1000 used)	24	30
CEDAR	Western	55	24	24
UTSig	Persian	115	27	6

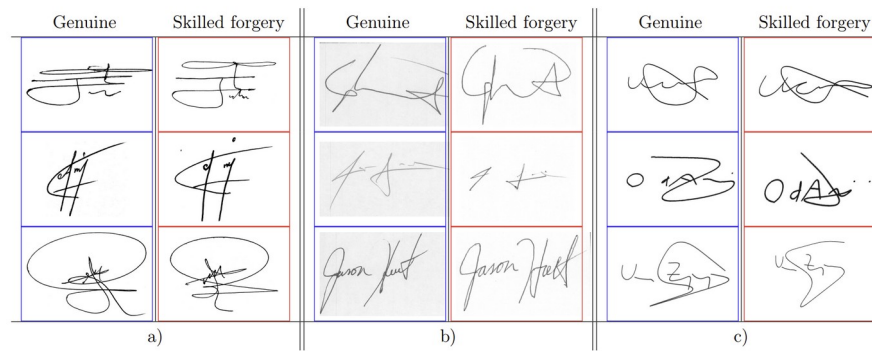


Figure 2 - Example of signature samples from different datasets:

a) UTSig; b) CEDAR; c) GPDS Synthetic

DOI: <https://doi.org/10.60797/IRJ.2026.172.87.5>

Models are trained only on GPDS Synthetic, while CEDAR and UTSig are used for external evaluation. Data is split strictly by users to prevent overfitting to user-specific features. GPDS users No.301–1000 are split into 90% training and 10% validation sets, while users No.1–300 form the test set. All CEDAR and UTSig users are used for evaluation.

Two training approaches are evaluated. First, standalone models are trained with Cross-Entropy Loss to classify individual signatures and assess their ability to extract discriminative and relevant features from individual images.

Since real-world verification usually involves a reference signature, in the second approach the discussed models are used as Siamese backbones. This improves generalization by focusing on relative feature relationships rather than class labels. Reference and query signatures are processed with shared weights and trained using Triplet Loss. For each user, triplets consist of an anchor, a randomly selected genuine positive, and a forged negative. Each genuine signature is used once as an anchor, producing 15,120 training triplets.

All models are trained on an NVIDIA Tesla V100 using Adam optimizer and a batch size of 32 over 50 epochs with a constant learning rate of $1e-5$.

The following metrics are considered in evaluating the performance of verification systems.

1. Accuracy denotes the ratio of correctly identified signatures to all the available test data.

2. EER displays the error rate at which FAR (False acceptance rate) equals FRR (False rejection rate).

Both of these metrics are said to be the most informative in the signature verification field [18].

Results and Discussion

Results are shown only for best-performing model sizes based on accuracy.

Compared with state-of-the-art models, CoAtNet and ConvNeXt generally achieve lower performance on CEDAR and GPDS Synthetic datasets (Table 4, Table 5 respectively).

Table 4 - Results on CEDAR

DOI: <https://doi.org/10.60797/IRJ.2026.172.87.6>

Method	Model	Accuracy, %	EER, %
Euclidean distance	SigNet [3]	100	0
	IDN [19]	–	3.62
	2C2S [20]	100	0
	SigCNN [24]	100	0
	HTCSigNet [7]	100	0
	ConvNeXt-L	82.9	16.7
	CoAtNet-L	82.8	16.8
	Siamese ConvNeXt-B	83.1	15.8
Cosine similarity	Siamese CoAtNet-B	83.4	15.3
	ConvNeXt-L	83.1	16.5
	CoAtNet-L	83.0	16.6
	Siamese ConvNeXt-B	86.1	13.7
	Siamese CoAtNet-B	85.0	14.7

Table 5 - Results on GPDS Synthetic
DOI: <https://doi.org/10.60797/IRJ.2026.172.87.7>

Method	Model	Accuracy, %	EER, %
Euclidean distance	SigNet [3]	77.76	–
	SigCNN [24]	–	7.34
	MSDN [21]	–	8.24
	TransOSV [5]	–	10.6
	HTCSigNet [7]	95.7	4.6
	ConvNeXt-B	69.0	30.9
	ConvNeXt-L	75.68	22.7
	CoAtNet-B	68.0	31.5
	CoAtNet-L	72.0	32.5
	Siamese ConvNeXt-B	79.2	20.6
	Siamese CoAtNet-B	79.5	20.2
Cosine similarity	ConvNeXt-B	70.7	29.0
	ConvNeXt-L	70.4	29.1
	CoAtNet-B	69.4	29.6
	CoAtNet-L	71.4	28.3
	Siamese ConvNeXt-B	80.0	19.9
	Siamese CoAtNet-B	78.9	20.9

However, the Siamese ConvNeXt-B and CoAtNet-B models reach accuracies of 80.0% and 78.9%, respectively, outperforming SigNet on GPDS Synthetic. Despite its rare use in OSV, cosine similarity yields better results for most models than the more frequently used Euclidean distance on GPDS Synthetic. The reason for this may lie in the specifics of the database. Contrary to the others used in this work, GPDS Synthetic dataset contains images with high variations in pen pressure as demonstrated in Figure 2. Therefore, as cosine similarity is less sensitive to such features, it produces greater performance for the discussed models.

However, when tested on UTSig dataset (Table 6), models outperform other architectures with Siamese ConvNeXt-B achieving the best accuracy of 82.3% and EER of 16.6.

Table 6 - Results on UTSig
DOI: <https://doi.org/10.60797/IRJ.2026.172.87.8>

Method	Model	Accuracy, %	EER, %
Euclidean distance	SigNet [3]	67.0	32.3
	LwR [22]	72.7	29.7
	Top [22]	64.2	34.5
	HTCSigNet [7]	66.19	8.15
	Attention-based Siamese Networks Model [23]	81.9	–
	ConvNeXt-L	80.25	19.5
	CoAtNet-L	80.36	19.4
Cosine similarity	Siamese ConvNeXt-B	82.3	16.6
	Siamese CoAtNet-B	81.76	17.7

A possible explanation lies in the specifics of the datasets. As shown in Figure 2, in general UTSig dataset contains signatures with more complex elements than the smooth and sometimes fragmented curves of GPDS Synthetic due to the characteristics of the Persian script. Therefore, due to architectural details (self-attention in CoAtNet and hierarchical feature extraction in ConvNeXt) the proposed models adapt more easily to diverse Persian glyphs, whereas on simpler instances of Western languages they are prone to overfit to irrelevant details. The reason for underperformance on the CEDAR dataset might be the low resolution of images in this dataset, to which the models are not adapted yet, although further experiments are needed to confirm this.

It is noticeable that Siamese networks perform the best on all of the used datasets. Both ConvNeXt and CoAtNet have been designed for image classification, which does not require attention to such intricate details as signature verification does. Since standalone model training is based on models being capable of extracting only features relevant to the task, the chosen



models underperform, overfitting to superficial patterns. However, using Siamese networks and Triplet loss, one can refocus these models for a similarity learning task instead of classification

Conclusion

This work presents a transfer learning-based approach to Writer Independent offline signature verification using the modern ConvNeXt and CoAtNet architectures. Their performance is observed under standalone classification and Siamese training schemes. Euclidean distance and cosine similarity are used for embedding-based verification.

The main findings are as follows:

1. ConvNeXt and CoAtNet models achieve their best results on UTSig with 82.3% and 81.76% accuracy with Siamese training, but mainly underperform the other state-of-the-art models on the remaining datasets with the exception of SigNet.
2. The applicability of the proposed models depends on signature complexity and image quality: they perform better on structurally complex Persian signatures but may overfit simpler Western-style or low-resolution samples.
3. The choice of verification metric for such models should reflect the variability of the data. Cosine similarity achieves better results on GPDS Synthetic since it is generally more robust than Euclidean distance to variations in pressure, brightness, resolution, and scanning conditions.

The scientific significance of this work lies in defining the applicability and limitations of ConvNeXt and CoAtNet in OSV and examining how their performance depends on the training scheme, verification metric, and dataset properties.

Future research may focus on simplifying these architectures to avoid overfitting on signatures with lesser details.

Благодарности

Исследование выполнено с использованием суперкомпьютерного комплекса НИУ ВШЭ.

Конфликт интересов

Не указан.

Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

Acknowledgement

This research was supported in part through computational resources of HPC facilities at HSE University.

Conflict of Interest

None declared.

Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

Список литературы на английском языке / References in English

1. Bhavani S.D. A multi-dimensional review on handwritten signature verification: strengths and gaps / S.D. Bhavani, R.K. Bharathi // *Multimedia Tools and Applications*. — 2023. — 83. — P. 2853–2894.
2. Singla A. Exploring offline signature verification techniques: a survey based on methods and future directions / A. Singla, A. Mittal // *Multimedia Tools and Applications*. — 2024. — 84. — P. 2835–2875.
3. Dey S. SigNet: convolutional siamese network for writer independent offline signature verification / S. Dey, A. Dutta, JI Toledo // *arXiv preprint*. — 2017. — URL: <https://arxiv.org/abs/1707.02131> (accessed: 10.04.25)
4. Zheng Y. Learning the micro deformations by max-pooling for offline signature verification / Y. Zheng, B.K. Iwana, M.I. Malik et al. // *Pattern Recognition*. — 2021. — 118.
5. Li H. TransOSV: Offline Signature Verification with Transformers / H. Li, P. Wei, Z. Ma et al. // *Pattern Recognition*. — 2024. — 145.
6. Dosovitskiy A. An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale / L. Beyer, A. Kolesnikov, D. Weissenborn et al. // *International Conference on Learning Representations (ICLR)*. — 2021. — 10.
7. Zheng L. HTCSigNet: A Hybrid Transformer and Convolution Signature Network for offline signature verification / L. Zheng, Da Wu, Y. Zheng et al. // *Pattern Recognition*. — 2025. — 159.
8. Rath Upadhyay R. Handwritten signature verification system using hybrid transfer learning approach / R. Rath Upadhyay, K.K. Singh // *Evolving Systems*. — 2024. — 15. — P. 2313–2322.
9. Salama W.M. Signature Verification Based on Deep Learning / W.M. Salama // *Alexandria Journal of Science and Technology*. — 2023. — 1. — P. 55–63.
10. Liu Z. A ConvNet for the 2020s / Z. Liu, H. Mao, C.Y. Wu [et al.] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. — New Orleans, LA, USA : IEEE Computer Society, 2022. — P. 11966–11976. — DOI: 10.1109/CVPR52688.2022.01167.
11. Dai Z. CoAtNet: Marrying Convolution and Attention for All Data Sizes / Z. Dai, H. Liu, Q.V. Le et al. // *Advances in Neural Information Processing Systems*. — Red Hook : Curran Associates, Inc., 2021. — P. 3965–3977.
12. Otsu N. A Threshold Selection Method from Gray-Level Histograms / N. Otsu // *IEEE Transactions on Systems, Man, and Cybernetics*. — 1979. — 9. — P. 62–66.
13. Deng J. ImageNet: A largescale hierarchical image database / J. Deng, W. Dong, R. Socher et al. // *IEEE Conference on Computer Vision and Pattern Recognition*. — Los Alamitos : IEEE Computer Society, 2009.
14. El Boudouri Y. EmoNeXt: an Adapted ConvNeXt for Facial Emotion Recognition / Y. El Boudouri, A. Bohi // *IEEE 25th International Workshop on Multimedia Signal Processing (MMSP)*. — Piscataway : IEEE, 2023.
15. Ferrer M.A. Synthetic off-line signature image generation / M.A. Ferrer, M. Diaz-Cabrera, A. Morales // *International Conference on Biometrics (ICB)*. — Piscataway : IEEE, 2013. — P. 1–7.
16. Srinivasan H. Machine Learning for Signature Verification / H. Srinivasan, S.N. Srihari, M.J. Beal // *Computer Vision, Graphics and Image Processing*. — Berlin, Heidelberg : Springer, 2006.



17. Soleimani A. UTSig: A Persian Offline Signature Dataset / A. Soleimani, K. Fouladi, B.N. Araabi // IET Biometrics. — 2017. — 6. — P. 1–8.
18. Sharma N. A Comprehensive Study on Offline Signature Verification / N. Sharma, S. Gupta, P. Mehta // Journal of Physics: Conference Series. — 2021. — 1969.
19. Wei P. Inverse discriminative networks for handwritten signature verification / P. Wei, H. Li, P. Hu // 25th International Conference on Pattern Recognition. — Los Alamitos : IEEE Computer Society, 2021. — P. 3853–3859.
20. Ren J.-X. 2C2S: A two-channel and two-stream transformer based framework for offline signature verification / J.-X. Ren, Y.-J. Xiong, H. Zhan et al. // Engineering Applications of Artificial Intelligence. — 2023. — 118.
21. Liu L. Offline signature verification using a region based deep metric learning network / L. Liu, L. Huang, Y. Chen // Pattern Recognition. — 2021. — 118.
22. Ji X. Paired contrastive feature for highly reliable offline signature verification / X. Ji, D. Suehiro, S. Uchida // Pattern Recognition. — 2023. — 144.
23. Xiong Y.-J. Attention-based multiple siamese networks with primary representation guiding for offline signature verification / Y.-J. Xiong, S.-Y. Cheng, J.-X. Ren // International Journal on Document Analysis and Recognition (IJDAR). — 2024. — 27.
24. Wan Q. Learning Metric Features for Writer-Independent Signature Verification using Dual Triplet Loss / Q. Wan, Q. Zou // 25th International Conference on Pattern Recognition. — Los Alamitos : IEEE Computer Society, 2003.