Chinese Journal of Information Fusion
http://dx.doi.org/10.62762/CJIF.2025.868838

RESEARCH ARTICLE

AMRC-NET: A New Method for Recognition of Russian Handwritten Text Integrating Multipath Mechanism and Linguistic Features

Zifeng Dang[1], Jiahan Liu[1,*], Shuang Xiong[1], Jun Ma[1], Xunhuan Ren[1] and Viktar Yurevich Tsviatkou[1]

[1] Department of Infocommunication Technologies, Belarusian State University of Informatics and Radioelectronics, Minsk 220013, Belarus

Abstract

Russian handwritten text recognition presents significant challenges due to the complex morphology of the Cyrillic alphabet, prevalent cursive writing, and substantial writer variability. To address the limitations of existing methods in dynamic contextual modeling and language-specific feature adaptation, this paper proposes an end-to-end framework named AMRC-NET. This framework integrates a multi-path architecture with linguistic feature awareness through three core modules: a Context Enhancement Module for long-range dependency modeling, a Russian Alphabet Morphology Optimization Module for script-specific pattern capture, and a Multi-Path Adaptive Fusion Mechanism for dynamic output integration. Extensive experiments on the Cyrillic Handwriting Dataset demonstrate that AMRC-NET achieves 48.15% in recognition accuracy, significantly outperforming existing models such as STAR-Net and GRCNN. Ablation studies further reveal that the three modules contribute accuracy improvements of 4.82%, 2.03%, and 1.58%, respectively, confirming their individual effectiveness and synergistic integration. The model also exhibits enhanced generalization capability, as reflected by lower validation loss and a reduced overfitting gap, while visual analysis confirms its robustness in challenging cursive writing scenarios.

Keywords: Russian handwritten text recognition, AMRC-NET, attention mechanism, multi-scale feature fusion, adaptive model integration, optical character recognition.

1 Introduction

Handwritten Text Recognition (HTR) constitutes a core task in document image analysis and understanding, aiming to automatically transcribe text from handwritten images into editable textual formats [1]. This technology has substantial

Academic Editor:
Guoyan Wang

Submitted: 29 December 2025
Accepted: 11 February 2026
Published: 29 March 2026

Vol. 3, **No.** 1, 2026.
10.62762/CJIF.2025.868838

***Corresponding author:**
✉ Jiahan Liu
2362267196@qq.com

Citation
Dang, Z., Liu, J., Xiong, S., Ma, J., Ren, X., & Tsviatkou, V. Y. (2026). AMRC-NET: A New Method for Recognition of Russian Handwritten Text Integrating Multipath Mechanism and Linguistic Features. *Chinese Journal of Information Fusion*, 3(1), 62–73.

© 2026 by the Authors. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

application potential in diverse domains, including historical document digitization, postal address identification, automated form processing, and educational assistance systems. For Russian handwritten text recognition, two principal methodological categories have been established: traditional pipeline-based approaches and deep learning-driven end-to-end methods. This paper first delineates the technical characteristics of these two paradigms, conducts a systematic analysis of their respective merits and limitations, and subsequently proposes an optimized recognition framework to address the extant challenges in this domain.

Prior to the widespread proliferation of deep learning, traditional handwriting recognition methods predominantly adhered to a four-stage pipeline: ”preprocessing – character segmentation – manual feature extraction – statistical classification” [4]. Specifically, these approaches first manually extract discriminative features (e.g., directional gradients [5], projection profiles, or contour descriptors) from preprocessed images, followed by individual character classification using classical statistical classifiers such as Support Vector Machines (SVMs) [6], K-Nearest Neighbors (KNNs), or Hidden Markov Models (HMMs) [7, 8]. A notable advantage of this paradigm lies in its efficacy in small-scale, well-constrained scenarios–characterized by standardized writing styles and limited vocabularies [9, 10]–where manually designed features can effectively capture the intrinsic patterns of simple characters. Nevertheless, this approach is inherently constrained. First, its recognition performance is highly dependent on the rationality and effectiveness of artificially designed features. Upon exposure to variations in writing styles or degraded image quality (e.g., noise, blurring, or uneven illumination), model performance deteriorates drastically [11, 12], rendering it inadequate for complex real-world tasks. Second, the character segmentation step is inherently error-prone, and such errors propagate to subsequent stages, thereby significantly undermining the final recognition accuracy.

The advent and rapid advancement of deep learning technology [2] have led to the rise of end-to-end recognition methods based on neural networks, which have gradually become the mainstream paradigm and achieved remarkable progress in multi-language handwriting recognition [3]. Representative methodologies include STAR-Net [13] and GRCNN [14]. The core advantage of these approaches resides in their ability to obviate the need for manual feature engineering and independent character segmentation, enabling end-to-end training from input images to output text sequences. This not only substantially enhances the model’s adaptability to diverse writing styles but also reduces reliance on domain-specific expert knowledge. Despite these merits, existing deep learning-based end-to-end methods still exhibit prominent limitations, particularly for Russian handwritten text. First, regarding decoding mechanisms and feature representation: STAR-Net employs CTC-based static alignment decoding [15], which lacks the capacity to dynamically focus on task-relevant image regions; while GRCNN incorporates a gating mechanism to strengthen local feature modeling, it fails to effectively capture long-range contextual dependencies. Second, in terms of language adaptability: both methods adopt a language-agnostic general architecture during feature extraction, failing to implement targeted optimization for the unique morphological characteristics of Russian handwriting [16]. These limitations result in suboptimal recognition performance for Russian text, a field that remains relatively underdeveloped compared with the recognition of Latin script languages such as English and French.

To address these challenges, this paper proposes a novel Russian-specific handwritten text recognition model named AMRC-NET. The proposed framework integrates a context enhancement module, a Russian Alphabet Morphology Optimization Module, and a multi-path adaptive fusion mechanism, thereby enhancing the model’s capacity for recognizing Russian handwriting under complex conditions.

In particular, in view of the shortcomings of the ”monotonic alignment hypothesis” in CTC decoding that cannot dynamically focus on the relevant region when processing hyphens, the Context Enhancement Module (CEM) proposed achieves non-monotonic and adaptive feature alignment through the multi-head attention mechanism. This mechanism allows the model to dynamically allocate attention weights at different time steps, thereby capturing the long-range dependencies between characters and effectively alleviating the problem of blurring character boundaries caused by continuous writing.

The main contributions of this paper are summarized

as follows:

- We propose an end-to-end framework named AMRC-NET, which integrates a multi-path architecture with linguistic feature awareness, effectively addressing the challenges of cursive writing and morphological complexity in Russian handwritten text recognition.
- We introduce three core modules: a Context Enhancement Module for long-range dependency modeling, a Russian Alphabet Morphology Optimization Module for script-specific pattern capture, and a Multi-Path Adaptive Fusion Mechanism for dynamic output integration.
- We validate the proposed method through extensive experiments on the Cyrillic Handwriting Dataset. The results demonstrate that our model achieves a recognition accuracy of 48.15%, significantly outperforming existing models such as STAR-Net and GRCNN, and confirm the effectiveness of each module through ablation studies.

Beyond its immediate application, this work also offers broader potential, although the Russian alphabet morphology optimization module proposed in this paper is trained and verified on Russian data, its design concept is based on the universal morphological characteristics of the Cyrillic script, so it has the theoretical potential to migrate to other languages that use the Cyrillic alphabet (such as Belarusian, Ukrainian, etc.). This further highlights the reference value of this work in the broader field of morphological enrichment text recognition.

2 Related Work

This section reviews two representative end-to-end frameworks for handwritten text recognition: STAR-Net and GRCNN. It systematically delineates their design principles, analyzes their technical contributions, and critically examines their inherent limitations.

STAR-Net adopts an integrated "geometric correction–feature extraction–sequence decoding" architecture. It employs learnable Thin-Plate Spline (TPS) transformation to rectify text lines, thereby mitigating issues related to handwriting deformation and tilt [13]. Robust visual features are extracted using a Deep Residual Network [17], while sequential dependencies are modeled via a Bidirectional Long Short-Term Memory (BiLSTM) network [18]. Finally, end-to-end transcription is achieved through a Connectionist Temporal Classification (CTC) mechanism [15]. This framework pioneered the joint optimization of geometric correction and text recognition, significantly boosting the recognition performance for deformed text on standard benchmarks and establishing itself as a widely adopted baseline model in the field. However, it suffers from two notable limitations. First, the CTC decoding mechanism relies on a monotonic alignment assumption [15], which lacks dynamic attention capability over the input sequence. This makes it less effective in handling challenging cases such as connected characters, local blur, or scribbled writing. Second, the TPS correction module is highly sensitive to the accuracy of text-line boundary and control point estimation [13], making it prone to correction errors in low-quality images or complex backgrounds. Such errors introduce noise into subsequent stages, compromising recognition robustness.

In contrast to the explicit geometric correction in STAR-Net, GRCNN focuses on enhancing feature representation within the convolutional stage itself. It incorporates Gated Recurrent Convolutional Neural Networks to embed a recurrent mechanism into convolutional operations, thereby strengthening the model's perception of local structures like continuous strokes and stroke orders [14]. This effectively introduces local contextual information at the feature extraction stage [19, 22]. The extracted features are then processed by a BiLSTM for global dependency modeling, with CTC used for sequence decoding [15]. While GRCNN excels at local feature modeling, its recurrent connections are still confined to local receptive fields [14], leaving the modeling of long-range dependencies primarily to the backend BiLSTM. Furthermore, similar to STAR-Net, its dependence on CTC decoding limits performance on samples that violate the monotonic alignment assumption. More critically, both frameworks employ a general, language-agnostic architecture [16, 20], lacking targeted optimization for the specific morphological characteristics of languages like Russian, which restricts their adaptability in cross-lingual and diverse handwriting style scenarios.

In summary, while STAR-Net and GRCNN have advanced the field of handwritten text recognition, they share common limitations: reliance on a static or monotonic decoding mechanism (CTC);

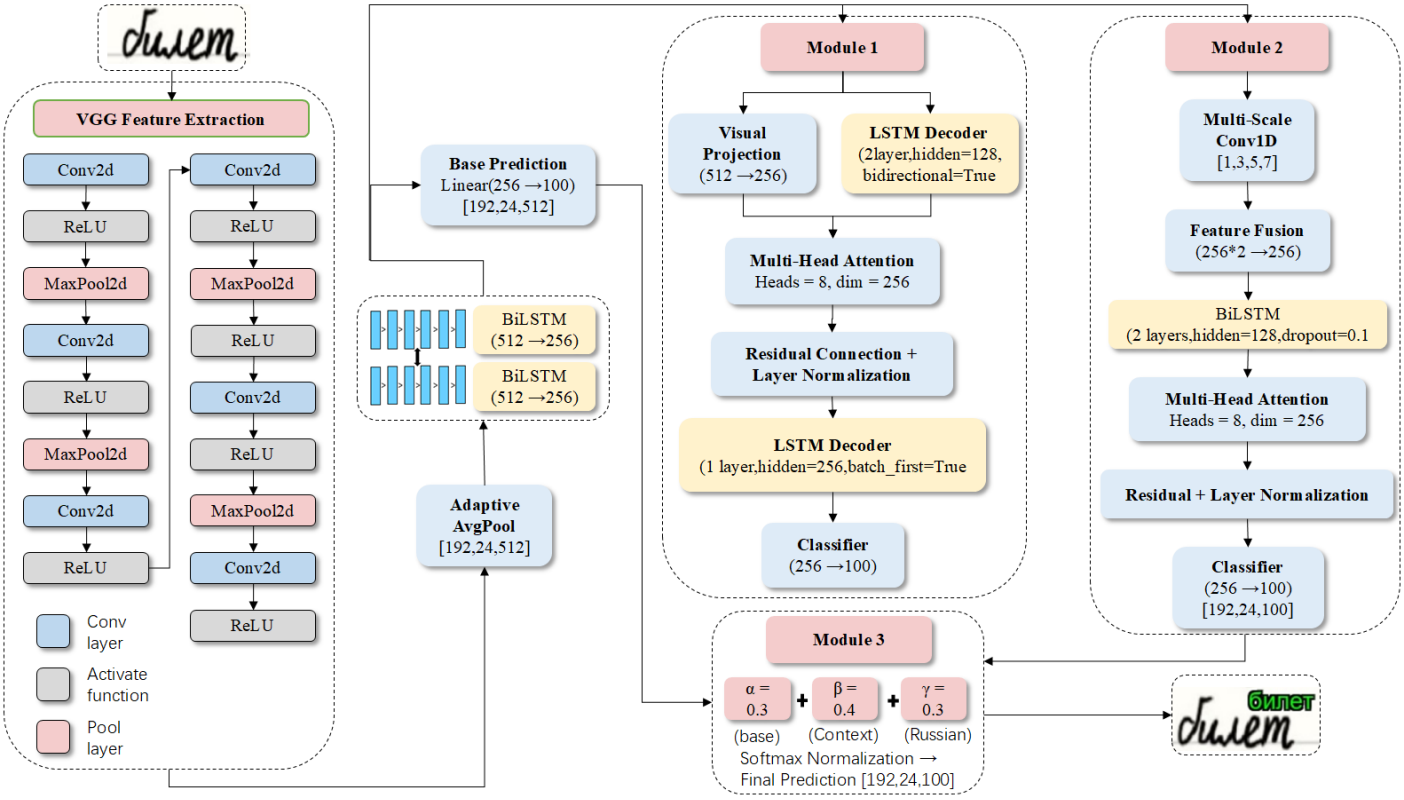


Figure 1. Overall framework diagram of the model.

use of generalized feature extractors not tailored to specific linguistic properties; and a single-path decision process that may lack robustness. To address these gaps, this paper proposes a novel recognition framework for Russian handwritten text. Our approach centers on three key innovations: dynamic context modeling, language-specific feature adaptation (with a focus on Russian), and multi-path decision fusion. Correspondingly, we introduce a context-enhanced decoder, a module optimized for Russian letter morphology, and a multi-path adaptive fusion mechanism. This integrated framework is designed to systematically improve recognition performance and generalization capability in complex writing scenarios.

3 Methodology

To simultaneously enhance sequence context modeling and optimize the morphological features of Russian handwriting, this paper proposes an end-to-end recognition model, AMRC-NET, whose overall architecture is illustrated in Figure 1. The model consists of three core components: a Context Enhancement Module, an Alphabet Morphology Optimization Module, and an Adaptive Fusion Decision-Making Module.

The backbone feature extraction network based on VGG, mainly based on its applicability consideration in handwritten text recognition tasks. Compared with deeper architectures such as ResNet, VGG can effectively extract local details and contour features of characters with low computational cost by stacking a simple 3×3 convolutional layer [21], avoiding the introduction of redundant parameters due to excessive network depth. In addition, VGG has a regular structure and is easy to integrate with other modules (such as LSTM and attention mechanism), which is suitable as the basic feature extraction stage of the multi-path fusion framework, ensuring the stability of feature expression while maintaining the lightweight of the model.The network extracts the basic feature sequence of the shape [batch_size, 24, 512] from the input image to capture its spatial structure and preliminary timing information. This feature sequence is transmitted as a shared input in parallel to two dedicated branches: the Context Enhancement Module (for modeling long-distance dependencies) and the Russian Letter Morphology Module (for extracting language-related features). Subsequently, the output of these two modules is input to the adaptive fusion decision module together with the original backbone features for feature fusion and the final recognition results are generated.

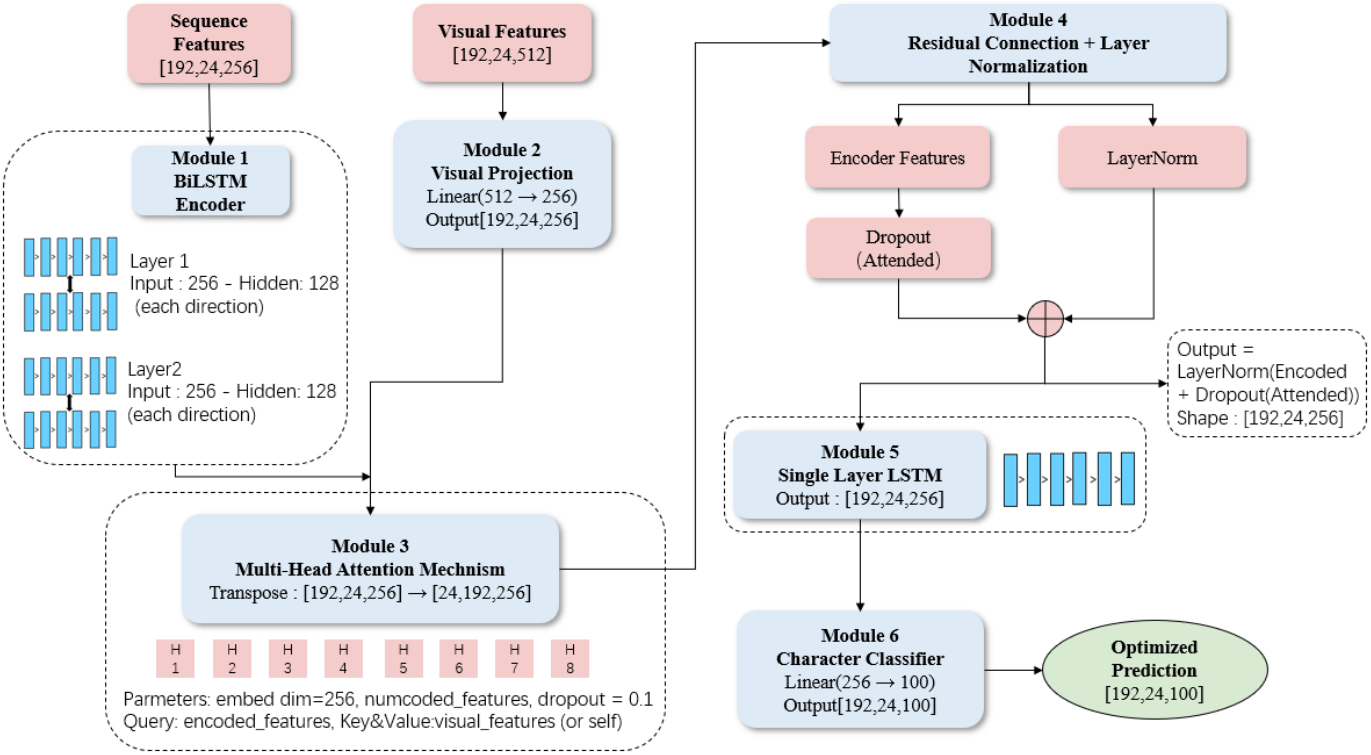


Figure 2. Context Enhancement Module.

3.1 Context Enhancement Module

As shown in Figure 2, the Context Enhancement Module adopts the structure of dual parallel processing of sequence features and visual features. The module uses sequence features with dimensions of [192, 24, 256] and visual features with dimensions of [192, 24, 512] as inputs, and realizes feature enhancement and classification prediction through six core processing stages.

In the module, the sequence features are contextually encoded by a dual-layer bidirectional LSTM encoder, and the visual features are reduced to 256 dimensions by linear projection to align with the sequence features. Subsequently, the two were fused through an 8-head attention mechanism, with an attention dimension of 256 and a Dropout rate of 0.1. The attention output is further optimized by a single-layer LSTM decoder after residual connection and layer normalization. Finally, the linear classifier maps the features to the 100-dimensional character category space and outputs the optimized prediction results with dimensions [192, 24, 100], realizing the effective fusion and context enhancement of visual and sequence information.

3.2 Alphabet Morphology Optimization Module

As shown in Figure 3, the Russian Alphabet Morphology Optimization module adopts a hierarchical feature processing architecture to model and optimize the morphological features of Russian handwritten text. The module uses sequence features with dimensions of [192, 24, 256] and visual features with dimensions of [192,24,512] as inputs, and realizes morphological feature enhancement and recognition optimization through seven stages of processing.

The module projects visual features to 256 dimensions and fuses them with sequence features to form enhanced visual representations. Morphological features are then extracted using a multi-scale 1D convolutional module with kernel sizes of 1, 3, 5, and 7, which are selected to capture typical stroke-length variations in Russian handwriting: size-1 kernels model local pixel-level details, size-3 capture short strokes (e.g., dots and flicks), size-5 cover medium-length strokes (e.g., horizontals and verticals), and size-7 handle longer continuous strokes or inter-character ligatures. This hierarchical multi-scale design enables robust feature extraction across varying writing styles.

After extraction, the multi-scale outputs are concatenated, transposed, and further fused with the

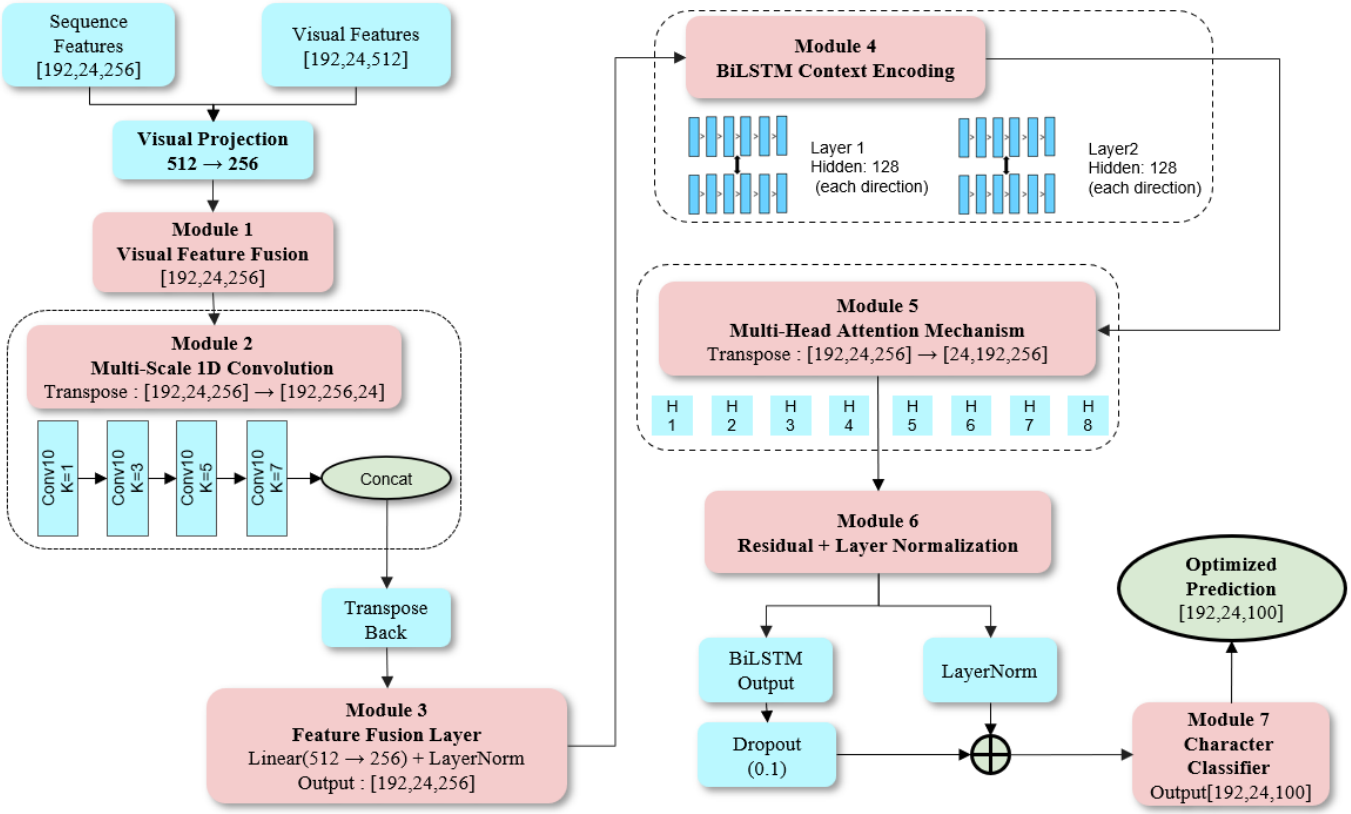


Figure 3. Adaptive fusion decision-making Module.

original sequence features via linear projection and layer normalization, resulting in a 256-dimensional feature vector. This vector is encoded by a two-layer bidirectional LSTM (128 hidden units per layer) and refined through multi-head attention with residual connections and layer normalization. Finally, a linear classifier maps the enhanced features to a 100-dimensional character space, producing Russian-optimized predictions of shape [192, 24, 100] for targeted morphological modeling and recognition.

The module contains three parallel prediction paths. The basic prediction path directly inputs the sequence features into the linear classifier to generate a baseline prediction P_1. The general context optimization path first encodes the sequence features through double-layer bidirectional LSTM, and then enhances them with multi-head attention mechanism and LSTM decoder, and finally outputs the general optimization prediction P_2. The Russian specialized optimization path first integrates visual and sequence features, and then outputs the Russian optimization prediction P_3 through multi-scale convolution and bidirectional LSTM coding, combined with the attention mechanism. The system uses the learnable weight parameters ω_α, ω_β, and ω_γ to Softmax normalization of the prediction results of the three paths to ensure that the sum of weights is 1. Finally, the final prediction is generated by the weighted fusion formula:

$$P_{\text{final}} = \omega_\alpha \cdot P_1 + \omega_\beta \cdot P_2 + \omega_\gamma \cdot P_3 \tag{1}$$

The mechanism can dynamically adjust the contribution weight of each path according to the complexity of the input sample and language characteristics, realizing the effective balance between basic recognition stability, general context understanding ability and language specialized modeling, and significantly improving the overall recognition performance of the system in complex Russian handwriting scenarios.

4 Experiments and Discussion

4.1 Experiment Setup and dataset description

The experiments were conducted on a computational server equipped with an NVIDIA A100 GPU and CUDA 12.1 to ensure reproducibility and reliability. All models were implemented using Python 3.9 and the PyTorch 2.5.1 framework.

We utilized the Cyrillic Handwriting Dataset for this study. The dataset was split into a training set

Table 1. Comparison of ablation experiments.

Methods	Accuracy	Validation Loss	Training Loss	Overfitting Gap
VGG-BILSTM-CTC	39.72%	1.62	2.81×10^{-3}	1.61
VGG-BILSTM-CTC + CEM	44.54%	1.29	2.10×10^{-3}	1.29
VGG-BILSTM-CTC + RAMOM	46.57%	1.26	2.27×10^{-3}	1.26
VGG-BILSTM-CTC + CEM + RAMOM + MPAFM	**48.15%**	**1.23**	1.93×10^{-3}	**1.22**

*(*Proposed AMRC-NET model*)*

containing 57,827 samples and a hold-out test set containing 16,002 samples, encompassing a diverse range of Russian handwriting styles. Additionally, twenty percent of the training data was randomly allocated as a validation set for hyperparameter tuning and early stopping. All input images were normalized and resized to a fixed dimension of 32 by 100 pixels.

During training, Key hyperparameters were set as follows: a batch size of 192 and total iterations of 5,000. The model was designed to recognize 187-character classes. Performance was evaluated on the validation set every 50 iterations to monitor the training progress.

4.2 Evaluation Metrics

We adopt Accuracy as the primary evaluation metric to quantify the overall correctness of the model. Formally, it is calculated as the proportion of correct predictions over all predictions, expressed by the following formula:

$$\text{Accuracy} = \frac{\sum_{i=1}^{N} I(p_i = g_t)}{N} \times 100\% \tag{2}$$

where N is the total sample size; p_i denotes the predicted string for the i-th sample; g_t is the ground-truth label string for the t-th sample and $I(\cdot)$ is Indicator function, which returns 1 if the condition inside is true and 0 otherwise.

In addition to accuracy, the model's performance was assessed using two standard metrics for text recognition: the Character Error Rate (CER) and the Word Error Rate (WER). The CER is derived from the Levenshtein distance. This distance measures the minimum number of editing operations, specifically substitutions (S), insertions (I), or deletions (D), required to transform the predicted text into the ground truth. The final CER value is obtained by normalizing this edit distance by the total number of characters N in the reference text, as defined by the following formula:

$$\text{CER} = \frac{S + I + D}{N_c} \tag{3}$$

Similarly, the Word Error Rate WER is calculated using the Levenshtein distance at the word level. It is defined as the minimum number of word edits, including substitutions (S_w), insertions (I_w), and deletions (D_w), needed to match the predicted sequence to the reference, divided by the total number of words in the reference text (N_w). The formula is as follows:

$$\text{WER} = \frac{S_w + I_w + D_w}{N_w} \tag{4}$$

Additionally, a character-level error rate metric proposed in [10] is adopted for evaluation. This method processes all recognition results to count both the total occurrence of each character (f_c) and the correct recognition count for each character (P_c). The error rate per character is then calculated with the following formula:

$$\text{CER}_c = \left(1 - \frac{P_c}{f_c}\right) \times 100\% \tag{5}$$

where c denotes a character, CER_c represents the error rate for that character, and f_c indicates its frequency in the dataset.

4.3 Ablation experiment

To evaluate the contribution of each component, we conduct an ablation study. The performance of the VGG-BILSTM-CTC baseline model, along with its incremental variants incorporating the CEM-Context Enhancement Module (CEM), the Russian Alphabet Morphology Optimization Module (RAMOM) and Multi-path Adaptive Fusion Mechanism (MPAFM) is comprehensively compared in Table 1.

From Table 1, it can be observed that when compared to the baseline vgg-bilstm-ctc, our method improves

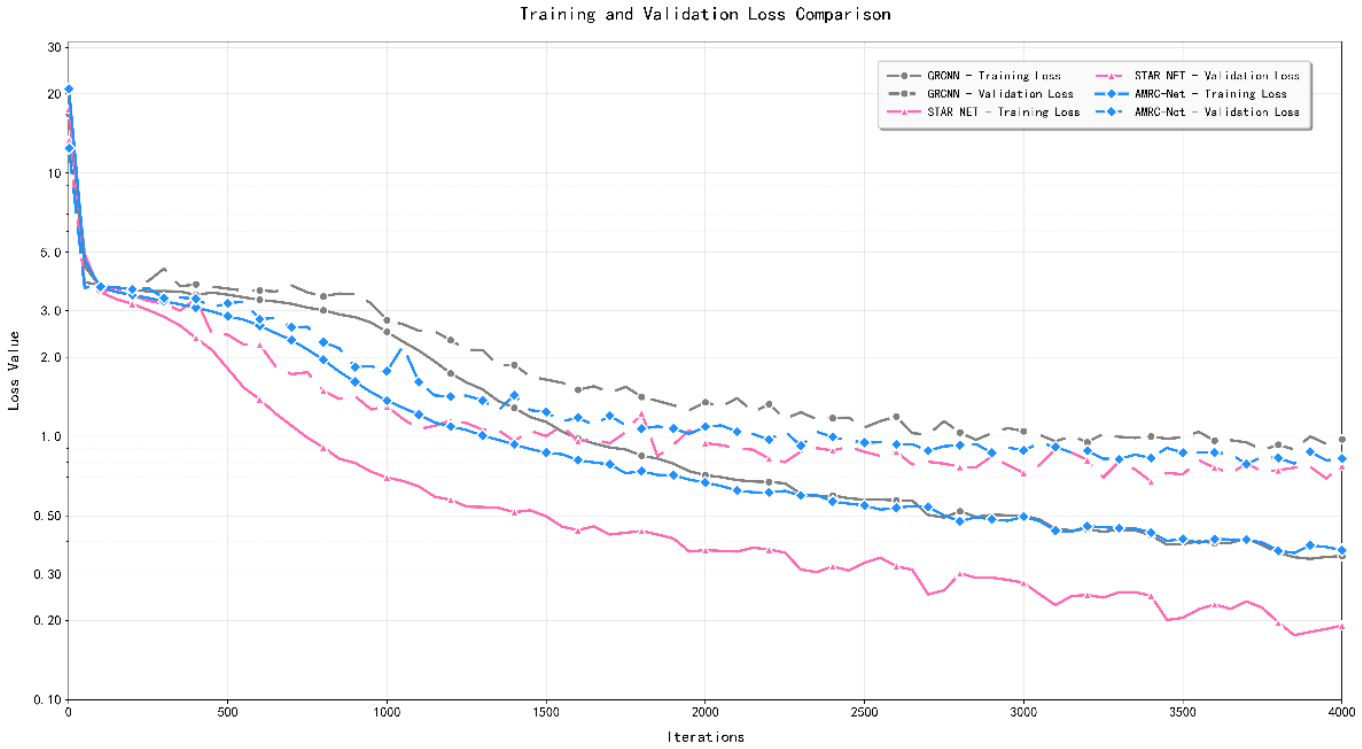


Figure 4. Comparison of the loss value of the three algorithms.

accuracy by 8.43 percentage points, which rising from 39.72% to 48.15%, reduces validation loss from 1.62 to 1.23, and narrows the overfitting gap from 1.61 to 1.22, demonstrating better generalization. The context enhancement module contributes 4.82% accuracy gain, the Russian alphabet module adds 2.03%, and the fusion mechanism further improves by 1.58%. Training loss also decreases from 2.81×10^{-3} to 1.93×10^{-3}, indicating stable convergence. The proposed multi-module fusion effectively enhances OCR accuracy and mitigates overfitting in complex Russian text recognition.

4.4 Loss Curve Analysis

The training and validation loss curves of various models over the first 4000 iterations are compared and depicted in Figure 4.

During the first 4000 iterations, AMRC-NET demonstrates competitive performance with a validation loss of 0.79, which is substantially lower than GRCNN's 0.98. Although STAR-NET achieves a slightly lower validation loss of 0.77, AMRC-NET exhibits the smallest gap between its training and validation losses, indicating superior generalization capability and reduced overfitting compared to the other models. AMRC-NET's training loss steadily decreases from 20.86 to 0.48 without severe overfitting, whereas GRCNN suffers from pronounced overfitting and STAR-NET shows a relatively larger performance gap. The smaller gap in AMRC-NET primarily stems from its Multi-Path Adaptive Fusion Mechanism (MPAFM). This mechanism dynamically integrates three complementary prediction paths—baseline prediction, context-optimized prediction, and Russian morphology-optimized prediction—through learnable weights. By balancing diverse feature representations and avoiding over-reliance on any single decision path, MPAFM provides implicit regularization, enhances robustness to handwriting variations, and effectively mitigates overfitting. In contrast, the more rigid single-path architectures of STAR-NET and GRCNN lack such adaptive regularization, making them more prone to overfitting on the training data. [24].

4.5 Comprehensive Experiment on Cyrillic Handwriting Dataset

A qualitative comparison of the trained models is presented in Table 2 using sample handwritten Russian text images. The table is structured as follows: the first two columns show the ground-truth texts and input images, followed by the recognition outputs of AMRC-Net, GRCNN, and STAR-NET. In the results, character errors are color-coded: red indicates a character recognition error, green indicates that its adjacent characters have been removed, and

Table 2. Visual Comparison of the prediction of different models.

GroundTruth	Input Image	GRCNN	STAR-NET	AMRC-NET
кофе		кope	кофе	кофе
друзья		Друзья	друзья	друзья
бʌлет		блет	облʌет	бʌлет
веспоӣ		Веспоӣ	веспоӣ	веспоӣ
фермер		фермер	фермер	фермер

blue indicates that extraneous characters have been inserted.

It can be seen from Table 2 that the proposed AMRC-NET model achieves perfect recognition across all five sample words. In contrast, the others exhibit characteristic error patterns. GRCNN demonstrates the weakest performance, correctly transcribing only the word фермер. It makes several errors on other samples: the word кофе is misread as кope, representing a character substitution error; the word бʌлет is recognized as блет, which is a character deletion error; and the word веспоӣ is grossly misrecognized as Becko, a case involving multiple substitution errors. STAR-NET shows improved but inconsistent accuracy, correctly identifying three words. However, it introduces an insertion error by recognizing the word бʌлет as облʌет. These visual observations indicate that our proposed architecture possesses greater robustness against the shape variations and spacing irregularities inherent in cursive handwriting, effectively mitigating the substitution, deletion, and insertion errors prevalent in the baseline models. In the results, character errors are color-coded: red indicates a character recognition error, green indicates that its adjacent characters have been removed, and blue indicates that extraneous characters have been inserted.

A comprehensive quantitative evaluation is then performed on the entire Cyrillic Handwriting Dataset, based on three evaluation metrics as described in Section 4.2: Accuracy, CER, and WER. The comparative results are presented in Table 3.

Table 3 indicates that the proposed AMRC-NET achieves the best overall performance across all three metrics. It attains the highest accuracy of 48.15%, which represents a 4.63 percentage point improvement over STAR-NET and a substantial 19.39 percentage points higher than GRCNN. This superiority is consistently reflected in the error rates: AMRC-NET yields the lowest Character Error Rate (CER) of 18.62% and the lowest Word Error Rate (WER) of 59.67%, outperforming STAR-NET (19.09% CER, 60.26% WER) and GRCNN (25.37% CER, 70.77% WER).

From the results shown in Table 3, AMRC-NET exhibits clear advantages in recognition capability and is the most effective model for this task.

Table 3. Comparison based on evaluation metric on entire dataset.

Methods	Accuracy	CER	WER
GRCNN	28.76%	25.37%	70.77%
STAR-NET	43.52%	19.09%	60.26%
AMRC-NET	**48.15%**	**18.62%**	**59.67%**

Additionally, we conducted a fine-grained analysis to examine the error distribution of each of the 33 Cyrillic letters across the three models. Table 4 summarizes the character-level accuracy statistics for all three models on the complete dataset using the CER_c metric. For ease of comparison, a smaller CER_c value indicates higher recognition accuracy for the corresponding character. Therefore, the minimum CER_c value for each character is highlighted in red in the table to visually indicate the best-performing model for that character.

From Table 4, it can be seen that the proposed AMRC-NET achieves the best performance on 24 out of the 33 letters, covering approximately 72.70% of the letters, particularly excelling on basic common letters such as ‘а’, ‘в’, ‘д’, and ‘е’, demonstrating strong

Table 4. CER_c comparison of Cyrillic characters across all three models.

Character	AMRC-NET	STAR-NET	GRCNN
а	12.71%	14.24%	21.99%
б	12.50%	11.29%	26.02%
в	6.83%	10.91%	9.69%
г	39.63%	29.41%	46.77%
д	19.22%	22.40%	31.62%
е	16.01%	19.20%	21.78%
ж	32.57%	44.44%	55.36%
з	13.57%	16.36%	21.15%
и	14.53%	16.81%	20.91%
й	9.16%	9.38%	9.86%
к	22.22%	29.73%	39.75%
л	17.79%	22.90%	27.64%
м	19.30%	23.58%	23.38%
н	16.63%	15.24%	21.66%
о	9.53%	11.42%	15.18%
п	17.95%	25.44%	21.03%
р	7.84%	11.73%	12.14%
с	15.64%	19.07%	23.83%
т	12.52%	11.35%	19.36%
у	6.78%	10.48%	12.00%
ф	20.54%	26.32%	37.04%
х	25.35%	34.78%	42.86%
ц	26.40%	23.81%	42.86%
ч	18.07%	20.51%	30.19%
ш	39.55%	20.00%	39.47%
щ	20.00%	0.00%	36.36%
ъ	57.69%	0.00%	20.00%
ы	19.35%	22.64%	28.18%
ь	14.40%	9.09%	31.90%
э	15.00%	28.57%	27.27%
ю	16.22%	20.00%	36.00%
я	19.89%	21.84%	24.16%
ё	73.33%	0.00%	40.08%

broad applicability. STAR-NET achieves optimal recognition on 9 letters due to its specific advantages, showing a 0% error rate on special characters such as 'щ', 'ъ', and 'ё', and outperforming the other models on letters such as 'г', 'Н', and 'т', reflecting its potential for recognizing niche characters. GRCNN does not achieve the best performance on any letter in this comparison, exhibits a high overall error rate, and underperforms the other two models on most letters, showing no clear competitive advantage.

However, the model ultimately achieved a recognition accuracy of 48.15%, reflecting the inherent complexity of Russian handwriting recognition tasks. The error mainly comes from three levels: at the visual level, the confusion of similar characters, the structural ambiguity caused by the continuous pen and deformation are the main obstacles; At the language level, the model lacks the ability to distinguish low-frequency special characters (such as "ъ" and "ё"), and lacks contextual understanding at the lexical and grammatical levels. At the data and preprocessing level, the diversity and balance of training samples are limited, coupled with the uneven image quality, which limits the generalization ability of the model. These analyses provide clear guidance for future research: it is necessary to integrate linguistic priors more deeply in feature representation, develop recognition modules that are more robust for deformation and continuous pen, and build more comprehensive training resources through data augmentation and synthesis techniques. The minimum CER_c value for each character is highlighted in red in the Table 4, providing a visual representation of the model with the best performance for that character.

5 Conclusion

This paper proposes a novel end-to-end model named AMRC-NET for Russian handwritten text recognition. The framework integrates three key components: a Context Enhancement Module for capturing long-range dependencies, a Russian Alphabet Morphology Optimization Module for script-specific feature extraction, and a Multi-Path Adaptive Fusion Mechanism for adaptive feature integration [23]. Together, these components establish a new paradigm for handling Russian script challenges in handwritten OCR. Extensive experiments on the Cyrillic Handwriting Dataset validate the effectiveness of AMRC-NET, which achieves a recognition accuracy of 48.15%, a significant improvement over strong baselines such as STAR-Net and GRCNN. Ablation studies further confirm the contribution of each module, with individual accuracy gains of 4.82%, 2.03%, and 1.58%, respectively, demonstrating their necessity and synergy.

Despite its promising results, this study has certain

limitations. First, the current model is primarily evaluated on constrained text lines; its performance on full-page documents with complex layouts remains to be explored. Second, the optimization for Cyrillic morphology, while effective, could be further deepened by incorporating more granular grammatical or syntactic cues. Future research will focus on: (1) extending the framework to handle unstructured document-level recognition, (2) exploring the integration of advanced language models for stronger semantic reasoning, and (3) adapting the core methodology to other cursive and morphologically rich scripts, such as Arabic or Devanagari.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported by the China Scholarship Council.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Kim, G., Govindaraju, V., & Srihari, S. N. (2000). An architecture for handwritten text recognition systems. *International Journal on Document Analysis and Recognition, 2*(1), 37-44. [CrossRef]

[2] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444. [CrossRef]

[3] Shi, B., Bai, X., & Yao, C. (2016). An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(11), 2298-2304. [CrossRef]

[4] Casey, R. G., & Lecolinet, E. (2002). A survey of methods and strategies in character segmentation. *IEEE transactions on pattern analysis and machine intelligence, 18*(7), 690-706. [CrossRef]

[5] Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (Vol. 1, pp. 886-893). IEEE. [CrossRef]

[6] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning, 20*(3), 273-297. [CrossRef]

[7] Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory, 13*(1), 21-27. [CrossRef]

[8] Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE, 77*(2), 257-286. [CrossRef]

[9] Tang, Y. Y., Cheriet, M., Liu, J., Said, J. N., & Suen, C. Y. (1999). Document analysis and recognition by computers. In *Handbook of Pattern Recognition and Computer Vision* (pp. 579-612). [CrossRef]

[10] Liu, C. L., Nakashima, K., Sako, H., & Fujisawa, H. (2004). Handwritten digit recognition: investigation of normalization and feature extraction techniques. *Pattern Recognition, 37*(2), 265-279. [CrossRef]

[11] Plamondon, R., & Srihari, S. N. (2000). Online and off-line handwriting recognition: a comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(1), 63-84. [CrossRef]

[12] Bunke, H. (2003, August). Recognition of cursive Roman handwriting: past, present and future. In *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.* (pp. 448-459). IEEE. [CrossRef]

[13] Liu, W., Chen, C., Wong, K. Y. K., Su, Z., & Han, J. (2016, September). Star-net: a spatial attention residue network for scene text recognition. In *BMVC* (Vol. 2, p. 7).

[14] Bluche, T., & Messina, R. (2017, November). Gated convolutional recurrent neural networks for multilingual handwriting recognition. In *2017 14th IAPR international conference on document analysis and recognition (ICDAR)* (Vol. 1, pp. 646-651). IEEE. [CrossRef]

[15] Graves, A., Fernández, S., Gomez, F., & Schmidhuber, J. (2006, June). Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the 23rd international conference on Machine learning* (pp. 369-376). [CrossRef]

[16] Espana-Boquera, S., Castro-Bleda, M. J., Gorbe-Moya, J., & Zamora-Martinez, F. (2010). Improving offline handwritten text recognition with hybrid HMM/ANN models. *IEEE transactions on pattern analysis and machine intelligence, 33*(4), 767-779. [CrossRef]

[17] He, K., Zhang, X., Ren, S., & Sun, J. (2016, June). Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770-778). IEEE. [CrossRef]

[18] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation, 9*(8), 1735-1780. [CrossRef]

[19] Chowdhury, A., & Vig, L. (2018). An efficient end-to-end neural model for handwritten text recognition. In *29th British Machine Vision Conference, BMVC 2018*.

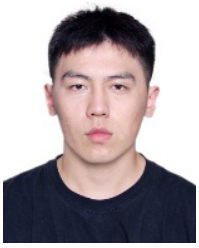
[20] Liu, C. L., & Suen, C. Y. (2009). A new benchmark on the recognition of handwritten Bangla and Farsi numeral characters. *Pattern Recognition, 42*(12), 3287-3295. [CrossRef]

[21] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).

[22] Kang, L., Riba, P., Rusiñol, M., Fornés, A., & Lladós, J. (2022). Pay attention to what you read: Non-recurrent handwritten text-line recognition. *Pattern Recognition, 129*, 108766. [CrossRef]

[23] Li, M., Lv, T., Chen, J., Cui, L., Lu, Y., Florencio, D., ... & Wei, F. (2023, June). Trocr: Transformer-based optical character recognition with pre-trained models. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 37, No. 11, pp. 13094-13102). [CrossRef]

[24] Wang, T., Xie, Z., Li, Z., Jin, L., & Chen, X. (2019). Radical aggregation network for few-shot offline handwritten Chinese character recognition. *Pattern Recognition Letters, 125*, 821-827. [CrossRef]



Zifeng Dang is a student at the Belarusian State University of Informatics and Radioelectronics and is currently studying for a bachelor's degree. His current research direction includes deep learning and handwriting recognition. This research was funded by the China Scholarship Council (CSC). (Email: 2093887901@qq.com)

Jiahan Liu is a student at the Belarusian State University of Informatics and Radioelectronics and is currently studying for an undergraduate degree at the university. Her current research direction includes deep learning and computer vision. This research was funded by the China Scholarship Council (CSC). (Email: 2362267196@qq.com)

Shuang Xiong is a student at the Belarusian State University of Informatics and Radioelectronics and is currently studying for an undergraduate degree at the university. Her current research direction includes deep learning and computer vision. This research was funded by the China Scholarship Council (CSC). (Email: xiongshuang806297@gmail.com)

Jun Ma received the B.S. degree from the Lanzhou University of Technology, in 2015, and the M.S. degree from the Belarusian State University of Informatics and Radioelectronics, in 2018, where he is currently pursuing the Ph.D. degree. His current research interests include image processing, machine learning, and computer vision. (Email: majun@bsuir.by)

Xunhuan Ren received the B.S. degree from the Lanzhou University of Technology, in 2015, M.S. degree and Ph.D. degree from the Belarusian State University of Informatics and Radioelectronics, in 2018 and 2024, respectively, where she is currently an associate professor of the department of infocommunication technologies. Her current research interests include image processing and information theory. (Email: renxunhuan@bsuir.by)

Viktar Yurevich Tsviatkou received the Ph.D. degree from the Belarusian State University of Informatics and Radioelectronics, in 1999. He works at the Belarusian State University of Informatics and Radioelectronics, where he is currently a Professor and the Dean of the department of infocommunication technologies, Faculty of Information Security. His research interests include digital image processing, pattern recognition, signal processing, and information theory.