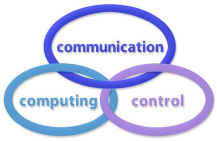
INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 21, Issue: 5, Month: October, Year: 2026
Article Number: 7060, https://doi.org/10.15837/ijccc.2026.5.7060



CCC Publications

Research on an OCR-Based Identification Model for Low-Quality Cross-Border Logistics Invoices

Junfeng Zhang, Wei Shen, Bohao Zheng

Junfeng Zhang
School of Computer Science and Technology
Zhejiang Sci-Tech University,
Hangzhou 310018, China
202230603137@mails.zstu.edu.cn

Wei Shen
School of Computer Science and Technology
Zhejiang Sci-Tech University,
Hangzhou 310018, China
shenwei@mail.zstu.edu.cn

Bohao Zheng
School of Computer Science and Technology
Zhejiang Sci-Tech University,
Hangzhou 310018, China
2579681128@qq.com

Abstract

Due to complex transportation conditions, cross-border logistics invoices often suffer from image blur and irregular layouts, which makes information extraction challenging. To address this issue, this paper proposes an OCR (Optical Character Recognition)-based recognition approach for cross-border logistics invoices. The proposed framework consists of two main stages: text detection and text recognition.

To improve precision and adaptability in both stages, targeted enhancements are introduced. In the text detection stage, the ACMix module is integrated into the backbone of DBNet to strengthen the detection of small and irregular text regions. By combining the local feature modeling capability of convolution with the global context modeling ability of attention, the detection performance on long text and complex backgrounds is effectively improved. In the text recognition stage, InceptionNeXt is adopted to replace ResNet45 in the visual backbone of ABINet, enhancing multi-scale feature extraction. By integrating Inception's multi-scale design with ConvNeXt's large-kernel optimization strategy, the proposed model achieves more robust feature representation for blurred and multi-scale text in cross-border documents.

A dataset consisting of 700 real-world cross-border logistics document images provided by the Zhejiang Postal Exchange Bureau is constructed, and extensive experiments are conducted on this dataset. Experimental results demonstrate that the proposed method outperforms existing approaches in OCR tasks for cross-border logistics documents, achieving an end-to-end (E2E) precision of 84.72%, a recall of 74.95%, and an Hmean of 79.54%.

Keywords: OCR, DBNet, ACMix, ABINet, InceptionNeXt.

1 Introduction

In recent years, with the continuous deepening of economic globalization, cross-border logistics has rapidly become a critical pillar of global economic activities. This field involves multiple countries and regions and is characterized by long transportation cycles, diverse transportation modes, and multiple operational stages, including customs clearance, international transportation, and destination customs processing.

In this context, the ability to efficiently and accurately extract information from large volumes of cross-border logistics documents is essential for ensuring smooth logistics operations and reliable automated processing. Traditionally, such information has been processed primarily through manual means. However, with the rapid advancement of computer science and artificial intelligence technologies, automated information extraction has become increasingly feasible, offering the potential to significantly reduce manual workload, improve processing efficiency, and better meet the growing demands of cross-border logistics.

Currently, information extraction from cross-border logistics documents faces several significant challenges. First, document formats vary widely across different countries and logistics providers, often involving complex layouts. Second, long-distance transportation increases the risk of document damage and distortion. Third, privacy protection regulations limit the availability of real-world document image datasets. In addition, factors such as shooting angles, lighting conditions, and waterproof coatings frequently result in degraded image quality (as shown in Fig. 1).

Despite continuous advances in computer technology, these challenges still cause cross-border logistics document processing to rely heavily on manual labor, thereby restricting the level of automation.

Figure 1: The document formats are diverse and the image quality is poor

These challenges have limited the amount of research on information extraction from cross-border logistics documents. Among existing studies, SwFB [1] achieves end-to-end information extraction from document images to structured data, effectively alleviating the error accumulation problem of stepwise processing pipelines. However, this method relies on predefined information types and can only extract specific fields. Extending it to additional information types requires modifications to both the data annotations and the model architecture.

In contrast, optical character recognition (OCR) technology has been extensively utilized for recognizing domestic logistics documents, invoices, tickets, and other printed materials. In this paper, the task of cross-border logistics document information extraction is decomposed into two primary components: text detection and text recognition. By enhancing models in these two domains, we aim to achieve automated extraction of relevant information.

Text detection methods can generally be categorized into regression-based and segmentation-based approaches.

Regression-based methods detect text by predicting the geometric shape of text regions. EAST [2] is an FCN-based [3] text detection method that predicts text regions at the pixel level, enabling efficient detection. However, when applied to cross-border logistics documents containing long text, EAST tends to split text lines into multiple fragments, which compromises the integrity of the extracted information.

SAST [4] extends EAST by improving text box shape modeling and achieves robust performance in detecting inclined and multi-oriented text. Nevertheless, it struggles with dense text scenarios commonly found in cross-border logistics documents, often resulting in missed detections.

TextBoxes++ [5] improves text detection by modifying the anchor box ratios of SSD [6] and introducing rotated anchors to better handle multi-oriented text. While this design is effective for multi-scale text detection, its reliance on predefined anchor boxes limits its adaptability to the diverse layouts of cross-border logistics documents, leading to reduced detection accuracy.

Segmentation-based methods extract text regions via semantic segmentation and are generally more robust to low-quality images compared to regression-based methods.

CRAFT [7] employs character-level region detection with character-aware supervision to localize text regions and link adjacent characters into text instances. This approach demonstrates strong robustness when handling low-quality images in cross-border logistics documents. However, its reliance on character-level predictions introduces substantial post-processing overhead, resulting in relatively slow inference speed.

PSENet [8] adopts a progressive scale expansion strategy to gradually merge text regions through multi-scale segmentation. This design enables effective detection of long and curved text, making PSENet well-suited for dense text scenarios. Nevertheless, the region merging process required during post-processing leads to limited inference efficiency.

PAN [9] introduces text centerline detection and employs pixel aggregation to reduce post-processing complexity. As a result, PAN achieves fast inference speed and performs well in real-time applications, particularly for small text detection. However, its performance degrades when applied to long text instances, such as address information in cross-border logistics documents.

Mask TextSpotter [10] integrates instance segmentation based on Mask R-CNN with text recognition to achieve end-to-end text spotting. This method performs well in detecting curved and densely distributed text, making it suitable for end-to-end OCR tasks. However, its high computational complexity leads to relatively slow inference.

FCENet [11] represents text contours using Fourier transform-based embeddings, enabling effective detection of irregularly shaped text. Nevertheless, detecting low-contrast, damaged, and densely distributed text in cross-border logistics documents remains challenging.

DBNet [12] introduces a differentiable binarization mechanism that replaces traditional hard thresholding with a sigmoid function, enabling end-to-end optimization and automatic threshold learning. Combined with a feature pyramid network for multi-scale feature fusion, DBNet achieves high detection accuracy and efficiency for text regions of varying sizes. However, small text instances may be weakly represented on feature maps, and multi-scale fusion can dilute fine-grained features. Moreover, the limited receptive field of convolutional kernels restricts the model's ability to capture global contextual information, particularly for long or irregularly shaped text. These limitations motivate further improvements for text detection in cross-border logistics documents.

In the field of text recognition, CRNN [13] is a widely adopted framework that extracts visual features using a convolutional neural network (CNN) [14], models sequential dependencies with a recurrent neural network (RNN) [15], and applies Connectionist Temporal Classification (CTC) [16] loss for alignment-free sequence prediction. The CTC algorithm employs a forward–backward dynamic programming strategy to compute the probabilities of all possible alignment paths and decodes the most probable sequence. Owing to its independence from character-level annotations and relatively fast inference speed, CRNN has been widely used in early text recognition tasks.

However, CRNN mainly relies on visual features and implicitly assumes a left-to-right text layout, which limits its robustness to distorted, degraded, or irregular text commonly found in cross-border

logistics documents. Moreover, the CTC framework does not explicitly model contextual semantic information, leading to reduced recognition accuracy when handling low-quality text.

To address the limitations of CTC in recognizing irregular text, RARE [17] introduced the Spatial Transformer Network (STN) [18] into text recognition, enabling geometric correction to alleviate text distortions. Building upon RARE, ASTER [19] incorporated both STN and an attention mechanism [20], further improving recognition performance for curved text.

In 2019, SAR [21] was proposed, employing a two-dimensional attention mechanism to directly attend to spatial feature maps and using LSTM for sequence modeling, thereby enhancing its capability to recognize irregular text. In the same year, NRTR (Non-Recurrent Transformer for Text Recognition) [22] replaced traditional RNN-based sequence modeling with a Transformer architecture. NRTR adopts an encoder–decoder structure with self-attention to model global dependencies and utilizes positional encoding to preserve sequence order. Benefiting from its high degree of parallelism, NRTR is particularly effective for long-text recognition. Nevertheless, these approaches still encounter difficulties when dealing with low-quality, noisy, or blurred text in cross-border logistics documents.

In 2020, SEED [23] was introduced, integrating visual and language models to enhance text recognition by exploiting semantic information. Subsequently, ABINet (Autonomous Bidirectional Iterative Network) [24] further decoupled text recognition into a visual model for feature extraction and a language model for semantic correction. Through multiple rounds of iterative interaction between visual and linguistic representations, ABINet achieves improved recognition accuracy. However, since the language model relies on the outputs of the visual model, the overall recognition performance is highly dependent on the quality of visual features. For low-quality cross-border logistics documents, ensuring robust visual feature extraction is therefore crucial for reliable text recognition.

In summary, the proposed approach enhances DBNet and ABINet to tackle the challenges associated with information extraction from cross-border logistics documents. The improved models are utilized to effectively extract information from low-quality cross-border logistics documents. The specific contributions of this work are summarized as follows:

- Introduced the ACMix [25] module into DBNet to enhance its capability in detecting long text and irregular text.
- Enhanced the visual model of ABINet using InceptionNeXt[26], improving the quality of features input to the language model.
- Constructed a dataset using 700 real document images legally obtained from the Zhejiang Postal Exchange Bureau, augmented to 2,000 images through Gaussian noise and Gaussian blur. The dataset was split into training and testing sets at a 4:1 ratio.
- Conducted comparative experiments on this dataset, evaluating the proposed improved model against other models. The results demonstrate that the enhanced model achieves notable improvements in cross-border logistics document recognition.

2 Related work

2.1 ACMix

ACMix (as shown in Fig. 2) is a neural network module that combines convolution and attention mechanisms. It can simultaneously use the local feature extraction ability of convolution and the global information modeling ability of the attention mechanism, so as to improve the model's small target and complex background text detection ability. In ACMix, the input image is first projected through three independent 1×1 convolution kernels, generating an intermediate feature set. This set serves as the input features for both the convolutional path and the self-attention path. In the attention path, the calculation is performed as shown in Equation (1).

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \tag{1}$$

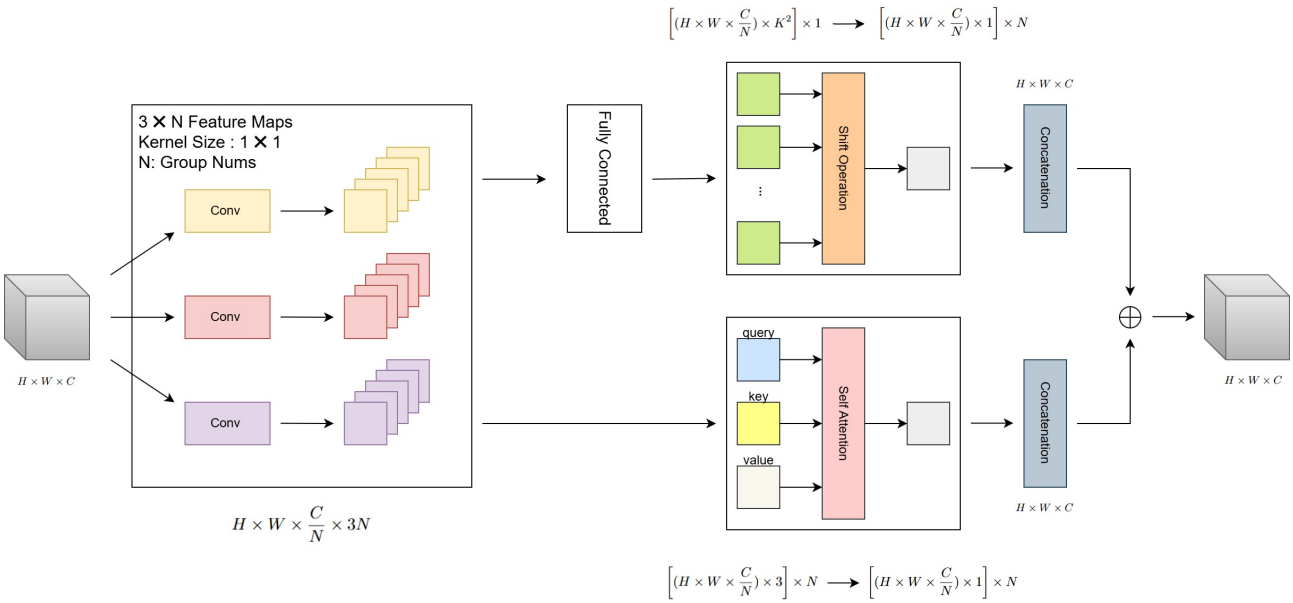


Figure 2: The overall structure of the ACMix module

where: the three projected features serve as Q (Query), K (Key), and V (Value), respectively. The relevance between different positions is measured by computing the dot product of Q and K ($\mathbf{QK}^\top$). The scaling factor $\sqrt{d_k}$ is used to prevent excessively large dot product values from causing gradient vanishing, where d_k is the dimension of K. The attention weights are then normalized using the Softmax function, yielding the attention distribution. Finally, a weighted sum of V is computed to obtain the global attention features.

In the convolutional path, the calculation is performed as shown in Equation (2).

$$\mathbf{X}_{\text{Conv}} = \mathbf{W}_{\text{Conv}} * \mathbf{X} \tag{2}$$

The standard $k\times$k convolution operation is used to extract local contextual features from the input while preserving the spatial resolution of the input features. Finally, the outputs from the self-attention path and the convolutional path are fused using Equation (3).

$$\mathbf{Y} = \alpha \cdot \mathbf{X}_{\text{Conv}} + (1-\alpha) \cdot \text{Attention}(Q, K, V) \tag{3}$$

where: α is a learnable parameter, $\alpha \in [0, 1]$, used to adjust the relative weights between convolution and attention. During training, α gradually learns the optimal fusion ratio.

2.2 InceptionNeXt

InceptionNeXt (as depicted in Fig. 3) is an efficient CNN model designed for vision tasks. It integrates the strengths of Inception and ConvNeXt. Specifically, in multi-scale deep convolution, it employs a 3×3 convolution kernel alongside asymmetric large kernels of size 11×1 and 1×11. These kernels process features at different scales simultaneously via parallel branches, thereby enhancing the model's receptive field and information extraction capability. The formulation is presented in Equation (4).

$$Y = \text{Concat}(Y_1, Y_2, Y_3, Y_4) \tag{4}$$

where:

$$\begin{aligned}
Y_1 &= \text{Conv}_{3\times3}(X) \quad \text{(Local Detail Information Extraction)};\\
Y_2 &= \text{Conv}_{11\times1}(X) \quad \text{(Emphasizing Horizontal Feature Extraction)};\\
Y_3 &= \text{Conv}_{1\times11}(X) \quad \text{(Emphasizing Vertical Feature Extraction)};\\
Y_4 &= \text{Conv}_{1\times1}(X) \quad \text{(Preserving Original Features)};
\end{aligned}$$

Depthwise convolution, where convolution is performed independently on each channel, the calculation is performed as shown in Equation (5).

$$Y_i = Y_{\mathrm{DW}}(i,j,c) = \sum_{m=0}^{K-1}\sum_{n=0}^{K-1} X(i+m, j+n, c) \cdot W_{\mathrm{DW}}(m,n,c) + X(i,j,c) \tag{5}$$

where: $X(i+m, j+n, c)$:The pixel value at position c in the input channel. $W_{\mathrm{DW}}(m,n,c)$:Depthwise convolution kernel(3×3 or 5×5); $X(i,j,c)$is Residual connection.

Finally, the convolution results from all paths are weighted and fused, as formulated in Equation (6).

$$Y = \sum_i \alpha_i Y_i, \quad \text{where} \quad \sum_i \alpha_i = 1 \tag{6}$$

where: Y_i: Feature maps extracted from different convolution paths;α_i: Learnable weight parameters corresponding to different paths, ensuring that the total contribution weights sum to 1.

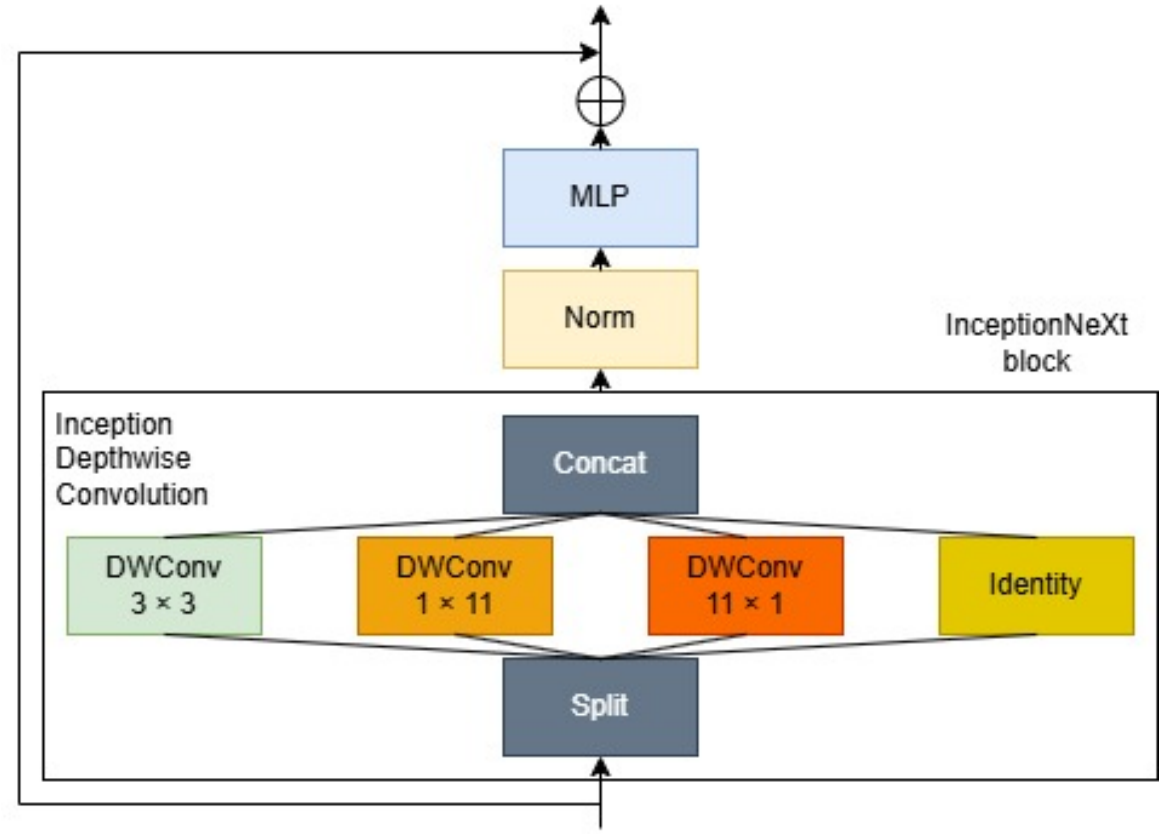


Figure 3: The overall structure of the InceptionNeXt module

3 Research method

This paper presents an end-to-end recognition system in which cross-border logistics document images are preprocessed and sequentially input into the text detection model and the text recognition model, ultimately producing the final recognition results.

3.1 Improved DBNet

In the text detection stage, an enhanced DBNet model is employed by integrating the ACMix module into the original architecture, as illustrated in Fig. 4. The introduction of ACMix enables joint modeling of local and global features, which is particularly beneficial for detecting long text

lines, irregular layouts, and small text instances commonly observed in cross-border logistics documents. Specifically, ACMix is embedded into the MobileNet backbone of DBNet to strengthen feature representation while preserving the lightweight design of the network. To achieve a balanced trade-off between detection performance and inference efficiency, the ACMix module is inserted only at the C3 stage of the backbone, which is empirically validated in the experimental analysis.

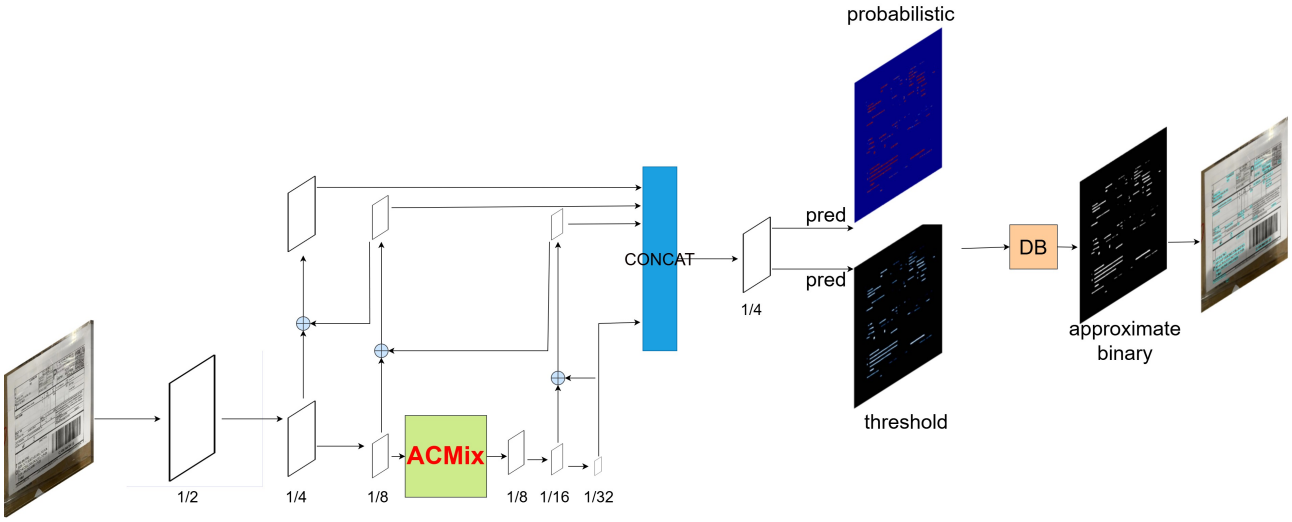


Figure 4: The overall architecture of DBNet with the ACMix module inserted at the C3 stage

3.2 Improved ABINet

In the text recognition stage, this paper proposes an enhanced ABINet model (as shown in Fig. 5). The primary improvement involves substituting the backbone network ResNet45 with InceptionNeXt to strengthen feature extraction capabilities. In ABINet, the language model utilizes the feature maps generated by the visual model for semantic training, thereby facilitating sequence inference and text recognition tasks.

For blurry images in cross-border logistics documents, the feature maps generated by the visual model often lack distinct semantic information. InceptionNeXt exhibits a superior multi-scale feature extraction capability, enabling it to effectively capture key features in complex scenarios, thereby enhancing the quality of input features for language model training.

4 Experiments and Results Analysis

4.1 Experimental Environment

The experimental environment used in this study is shown in Table 1.

The enhanced text detection model was pre-trained on the ICDAR2015 dataset and was subsequently fine-tuned for 800 epochs on the dataset used in this study, with an initial learning rate of 0.001. The improved text recognition model was pre-trained on the MJSynth and SynthText datasets and further fine-tuned for 10 epochs on the dataset specific to this study. The initial learning rate was set to 0.001 and later reduced to 0.0001.

This study legally acquired 700 real cross-border logistics document images from the Zhejiang Postal Exchange Bureau. The dataset was augmented to 2,000 images using Gaussian noise and Gaussian blur techniques and was split into training and testing sets at a ratio of 4:1.

4.2 Evaluation Metrics

To quantitatively evaluate the performance of different models in the OCR task for cross-border logistics documents, this study adopts standard evaluation metrics following the ICDAR2015 evaluation protocol.

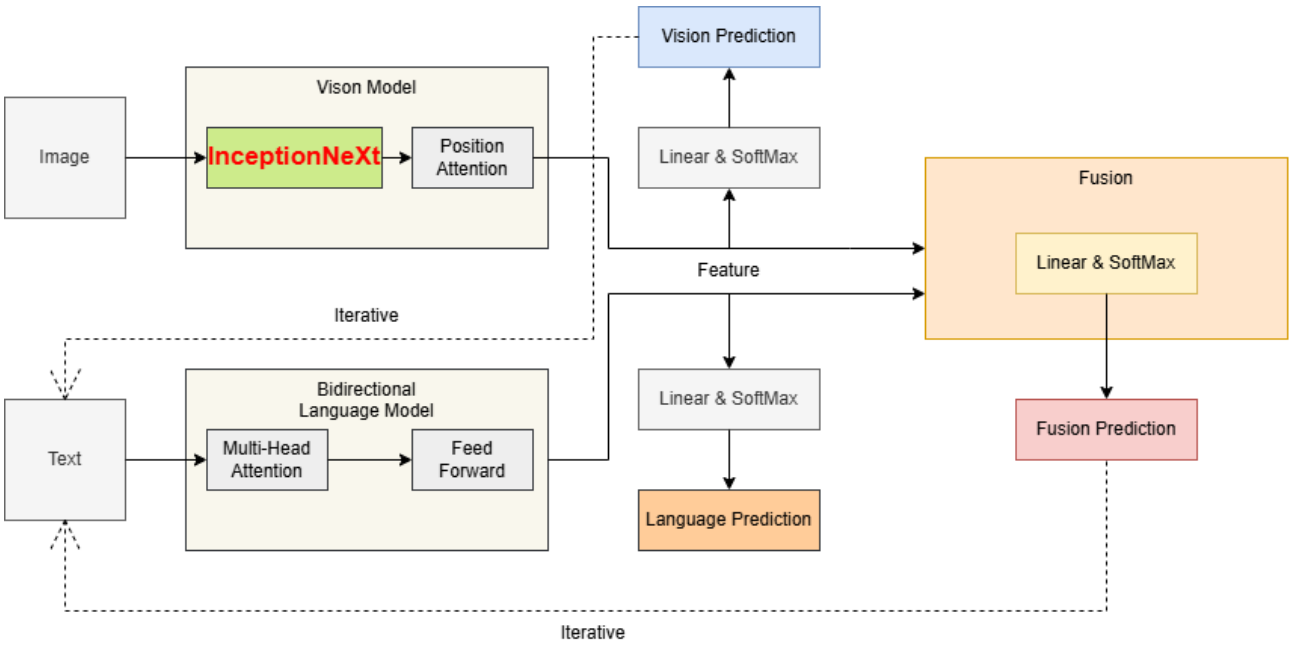


Figure 5: The overall architecture of ABINet with an InceptionNeXt-based visual backbone

Table 1: Experimental Environment Configuration

Configuration Items	Specifications
OS	Ubuntu 20.04
CPU	12 vCPU Intel Xeon Silver 4214R
Mem	32GB
GPU	RTX 3080Ti (12GB)
Pytorch	1.10.0
Python	3.8
CUDA	11.3
Numpy	1.21.4

For the text detection task, all experimental results are assessed using the official ICDAR2015 DetEval evaluation script, with default parameter settings. A predicted text region is considered a correct detection only when it is matched one-to-one with the corresponding ground-truth region and the Intersection over Union (IoU) is no less than 0.5. Based on this criterion, precision, recall, and Hmean are computed to measure detection performance.

For the text recognition task, recognition accuracy is adopted as the evaluation metric. A prediction is regarded as correct if the recognized text string exactly matches the annotated ground truth, ignoring case sensitivity and trimming leading and trailing spaces.

For end-to-end evaluation, a prediction is counted as an end-to-end hit only when both the text detection and text recognition results are correct according to the above criteria. Based on this evaluation protocol, end-to-end precision, recall, and Hmean are calculated to reflect the overall performance of the OCR system.

Based on the above evaluation criteria, the end-to-end precision (P), recall (R), and Hmean are computed as follows:

$$P = \frac{TP}{TP + FP} \tag{7}$$

$$R = \frac{TP}{TP + FN} \tag{8}$$

$$\text{Hmean} = \frac{2 \times P \times R}{P + R} \tag{9}$$

where TP, FP, and FN denote the numbers of true positives, false positives, and false negatives, respectively, under the end-to-end evaluation protocol.

4.3 Experimental Results Analysis

To analyze the impact of the ACMix module on text detection performance and inference efficiency at different stages of the DBNet backbone, ACMix is inserted into the C1–C4 stages of the network, respectively, and comparative experiments are conducted on the ICDAR2015 test set, as shown in Figure 6.

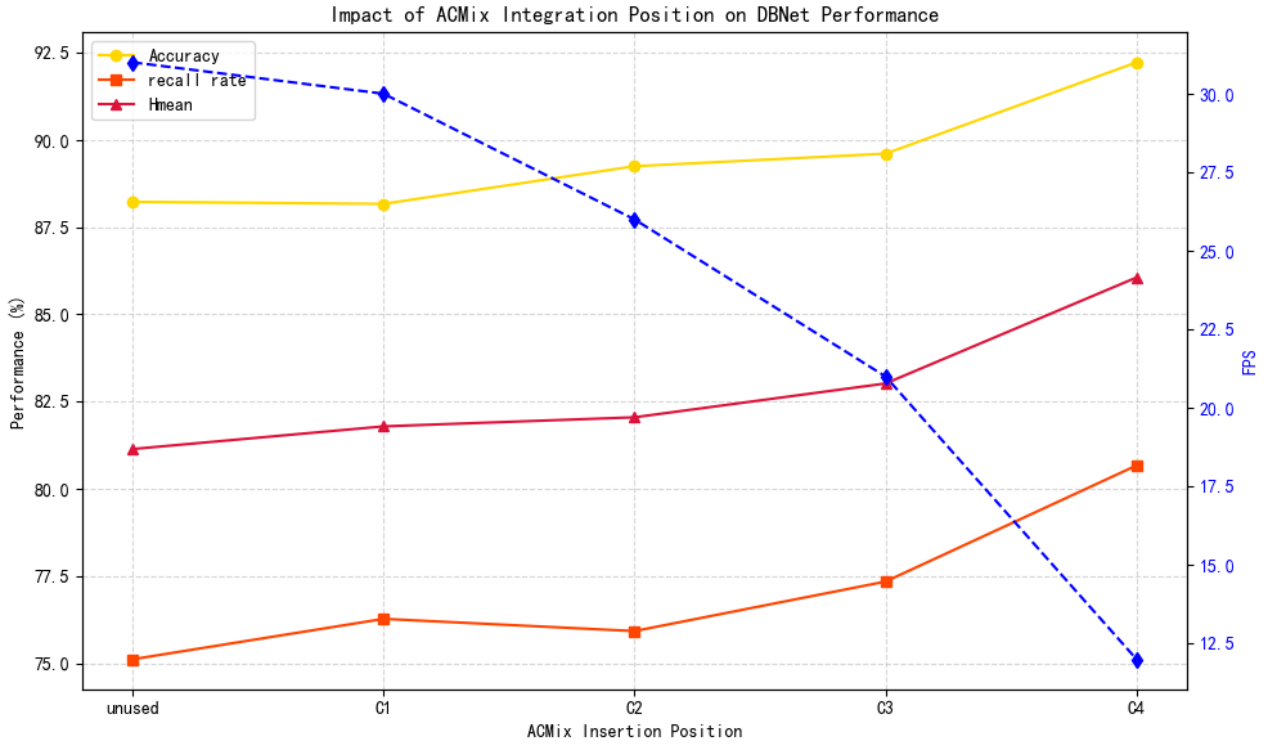


Figure 6: Impact of ACMix insertion stage on detection performance and inference efficiency

The experimental results show that as the ACMix module is inserted from the shallow to the deep layers, the overall detection precision of the model increases, but the inference speed decreases accordingly. Specifically, inserting ACMix at the C4 stage can achieve the highest detection precision, but the frame rate drops significantly, making it difficult to meet the requirements of actual deployment for inference efficiency. In contrast, when ACMix is inserted at the C3 stage, the model achieves stable improvements in precision, recall, and Hmean, while still maintaining a relatively high inference speed, demonstrating a better performance-efficiency balance. Therefore, considering both detection performance and inference efficiency, this paper ultimately opts to insert the ACMix module at the C3 stage of the DBNet backbone network as the default configuration for subsequent experiments and system construction.

To further validate the effectiveness of the enhanced text detection model, the original DBNet and the DBNet+ACMix models are evaluated on the ICDAR2015 dataset, with the results illustrated in Figure 7. In addition to the comparison with the DBNet baseline, the proposed method is also compared with several representative text detection approaches discussed earlier, including PSE- and FCE-based models. These methods are selected as they are widely adopted and representative segmentation-based text detectors for handling complex text shapes and layouts. The results demonstrate that incorporating the ACMix module into the DBNet backbone leads to a clear improvement

in text detection performance and achieves competitive results compared with other commonly used detection models.

To verify the performance of the improved text recognition model, both the trained ABINet model and the ABINet+InceptionNeXt model were tested on the ICDAR2015 dataset, with the results shown in Figure 8. In addition to the comparison with the ABINet baseline, the proposed method is also compared with representative text recognition models discussed earlier, including CRNN and RARE. The results demonstrate that integrating the InceptionNeXt module enhances the recognition performance of ABINet.

To further analyze the computational efficiency of the proposed modification, a comparison between the standard ABINet and the InceptionNeXt-based model is conducted in terms of parameter count and computational complexity. As shown in Table 2, the introduction of InceptionNeXt results in only a marginal increase in model parameters and FLOPs, while achieving notable performance gains, indicating a favorable trade-off between recognition accuracy and computational overhead.

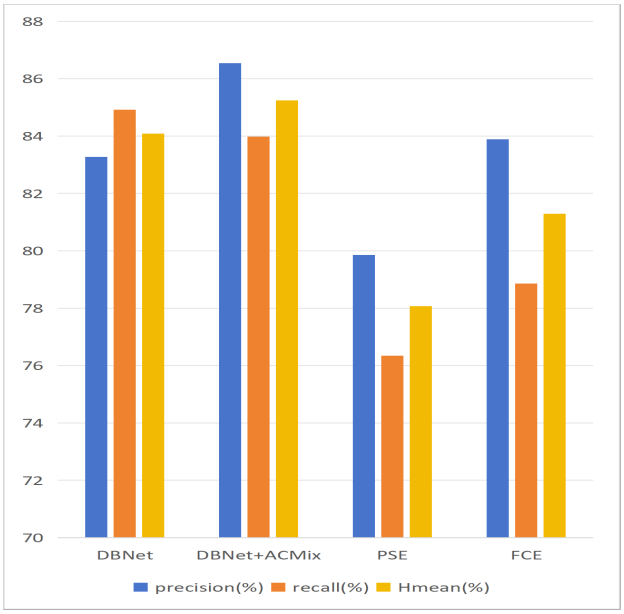


Figure 7: Comparison of DBNet and DBNet+ACMix with representative text detection methods

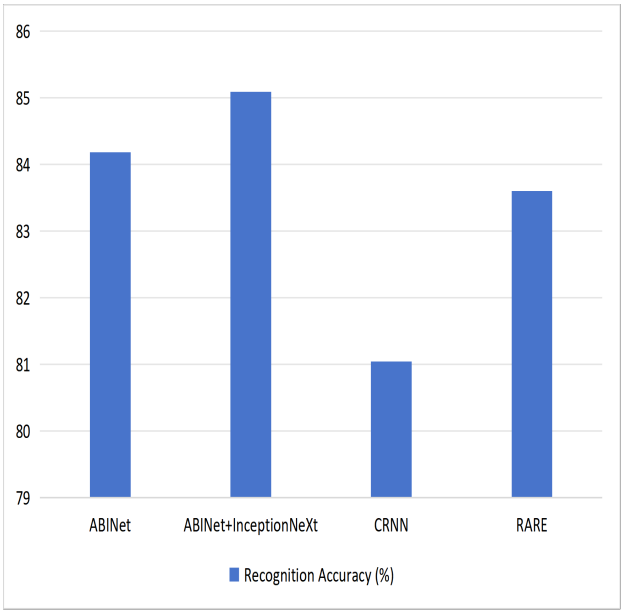


Figure 8: Comparison of ABINet, ABINet+InceptionNeXt, CRNN, and RARE in text recognition

Table 2: Computational cost comparison of ABINet and its improved variant

Model	Parameters (M)	FLOPs (G)
ABINet	25.6	0.32
ABINet + InceptionNeXt	28.1	0.33

To evaluate the overall effectiveness of the proposed OCR framework for cross-border logistics documents, we conduct comprehensive end-to-end experiments on our self-built cross-border logistics document dataset. Comparative experiments are carried out with several representative end-to-end OCR approaches, including Mask TextSpotter v3, PSE+CRNN, FCE+RARE, and DBNet+ABINet. The evaluation metrics include end-to-end precision, end-to-end recall, and end-to-end Hmean.

Table 3 presents the quantitative comparison results among different models.

As shown in Table 3, the proposed model achieves the highest end-to-end precision and Hmean among all compared methods. Although the end-to-end recall is slightly lower than that of some competing approaches, the proposed method maintains the best overall end-to-end Hmean, indicating a more balanced performance between detection and recognition.

Table 3: End-to-end OCR performance comparison on the cross-border logistics document dataset

Models	e2e_precision(%)	e2e_recall(%)	e2e_Hmean
PSE[8]+CRNN[13]	73.15	67.46	70.18
FCE[11]+RARE[17]	74.32	75.24	74.78
DBNet[12]+ABINet[24]	80.57	73.78	77.03
Mask TextSpotter v3[10]	82.31	75.22	78.61
DBNet_ACMix+ABINet_InceptionNeXt (Our model)	**84.72**	**74.95**	**79.54**

Notably, all end-to-end evaluation metrics of the proposed model outperform those of the baseline DBNet+ABINet framework, further verifying the effectiveness of the proposed improvement strategy.

To visually demonstrate the actual performance of the improved model, we present the recognition results of typical samples. As shown in Figure 9, the improved model successfully identified all the key information in the cross-border logistics documents.

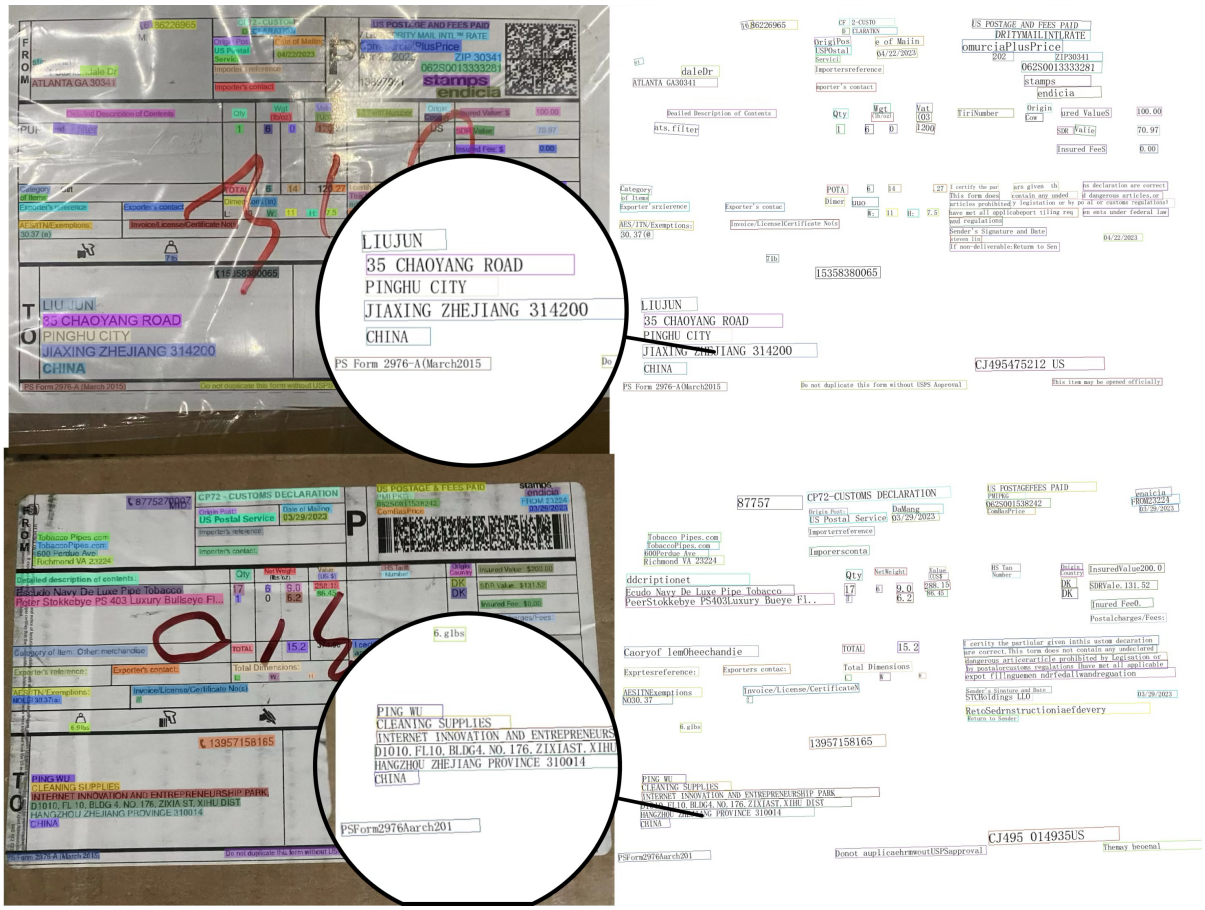


Figure 9: Qualitative end-to-end OCR results of the proposed model

5 Conclusion

To address the challenges posed by diverse formats, complex layouts, and degraded image quality in cross-border logistics documents, this study proposes enhancements to both DBNet and ABINet. By embedding the ACMix module at the C3 stage of the DBNet backbone, the text feature extraction capability is effectively strengthened while maintaining a favorable balance between detection perfor-

mance and inference efficiency. In addition, incorporating InceptionNeXt into ABINet enables more robust visual feature modeling in complex scenarios, leading to improved text recognition performance.

The effectiveness of the proposed improvements is validated on the cross-border logistics document dataset used in this study. The enhanced end-to-end OCR system achieves an end-to-end precision of 84.72% and an end-to-end recall of 74.95%, demonstrating strong overall performance and practical applicability for automated information extraction in cross-border logistics scenarios.

It should be noted that the dataset used in this study is relatively limited in size and originates from a single institutional source. Although the proposed method demonstrates strong performance on this dataset, further evaluation on larger and more diverse datasets would be beneficial to comprehensively validate its generalization capability. In addition, while the reported results consistently show improvements over baseline methods, future work will include multiple training runs and statistical analysis to further assess the robustness of the proposed approach.

From a practical perspective, the proposed framework can be applied to real-world cross-border logistics document processing tasks. However, its performance may still be affected by extreme image degradation, unseen document layouts, or multilingual variations. These limitations will be further investigated in future work, including evaluation on datasets from different logistics providers and multilingual scenarios.

6 Declaration of AI-assisted Technologies

During the preparation of this work, the authors used AI-assisted tools only for Chinese-English translation, English language polishing, and improving readability. These tools were not used to generate scientific ideas, design the methodology, analyze data, interpret results, or draw conclusions.

References

[1] Wei Shen et al. “End-to-End Information Extraction from Courier Order Images Using a Neural Network Model with Feature Enhancement”. In: *Applied Sciences* 15.2 (2025), p. 698.

[2] Xinyu Zhou et al. *EAST: An Efficient and Accurate Scene Text Detector*. 2017. arXiv: 1704.03155 [cs.CV].

[3] Jonathan Long, Evan Shelhamer, and Trevor Darrell. “Fully Convolutional Networks for Semantic Segmentation”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015, pp. 3431–3440.

[4] Yuliang Liu et al. “Single-Shot Arbitrarily-Oriented Scene Text Detection”. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2019, pp. 455–471.

[5] Minghui Liao, Baoguang Shi, and Xiang Bai. “TextBoxes++: A Single-Shot Oriented Scene Text Detector”. In: *IEEE Transactions on Image Processing* 27.8 (2018), pp. 3676–3690.

[6] Wei Liu et al. “SSD: Single Shot MultiBox Detector”. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2016, pp. 21–37.

[7] Youngmin Baek et al. *Character Region Awareness for Text Detection*. 2019. arXiv: 1904.01941 [cs.CV].

[8] Wenhai Wang et al. *Shape Robust Text Detection with Progressive Scale Expansion Network*. 2019. arXiv: 1903.12473 [cs.CV].

[9] Wenhai Wang et al. *Efficient and Accurate Arbitrary-Shaped Text Detection with Pixel Aggregation Network*. 2020. arXiv: 1908.05900 [cs.CV].

[10] Pengyuan Lyu et al. *Mask TextSpotter: An End-to-End Trainable Neural Network for Spotting Text with Arbitrary Shapes*. 2018. arXiv: 1807.02242 [cs.CV].

[11] Yiqin Zhu et al. *Fourier Contour Embedding for Arbitrary-Shaped Text Detection*. 2021. arXiv: 2104.10442 [cs.CV].

[12] Minghui Liao et al. *Real-Time Scene Text Detection with Differentiable Binarization*. 2019. arXiv: 1911.08947 [cs.CV].

[13] Baoguang Shi, Xiang Bai, and Cong Yao. *An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition*. 2015. arXiv: 1507.05717 [cs.CV].

[14] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems*. Vol. 25. 2012, pp. 1097–1105.

[15] Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-Term Memory". In: *Neural Computation* 9.8 (1997), pp. 1735–1780.

[16] Alex Graves et al. "Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks". In: *Proceedings of the 23rd International Conference on Machine Learning (ICML)*. 2006, pp. 369–376.

[17] Baoguang Shi et al. *Robust Scene Text Recognition with Automatic Rectification*. 2016. arXiv: 1603.03915 [cs.CV].

[18] Max Jaderberg et al. *Spatial Transformer Networks*. 2016. arXiv: 1506.02025 [cs.CV].

[19] Baoguang Shi et al. "ASTER: An Attentional Scene Text Recognizer with Flexible Rectification". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41.9 (2018), pp. 2035–2048.

[20] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. *Neural Machine Translation by Jointly Learning to Align and Translate*. 2014. arXiv: 1409.0473 [cs.CL].

[21] Hui Li et al. *Show, Attend and Read: A Simple and Strong Baseline for Irregular Text Recognition*. 2019. arXiv: 1811.00751 [cs.CV].

[22] Fenfen Sheng, Zhineng Chen, and Bo Xu. "NRTR: A No-Recurrent Sequence-to-Sequence Model for Scene Text Recognition". In: *arXiv preprint arXiv:1908.08526* (2019).

[23] Zhi Qiao et al. *SEED: Semantics Enhanced Encoder-Decoder Framework for Scene Text Recognition*. 2020. arXiv: 2005.10977 [cs.CV].

[24] Shancheng Fang et al. *ABINet++: Autonomous, Bidirectional and Iterative Language Modeling for Scene Text Spotting*. 2022. arXiv: 2211.10578 [cs.CV].

[25] Xuran Pan et al. *On the Integration of Self-Attention and Convolution*. 2022. arXiv: 2111.14556 [cs.CV].

[26] Weihao Yu et al. *InceptionNeXt: When Inception Meets ConvNeXt*. 2024. arXiv: 2303.16900 [cs.CV].

Journal's webpage: http://univagora.ro/jour/index.php/ijccc/

This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).
https://publicationethics.org/members/international-journal-computers-communications-and-control

Cite this paper as:

Zhang, J.; Shen, W.; Zheng, B. (2026). Research on an OCR-Based Identification Model for Low-Quality Cross-Border Logistics Invoices, *International Journal of Computers Communications & Control*, 21(5), 7060, 2026.

https://doi.org/10.15837/ijccc.2026.5.7060