Data Science and Engineering (DSE) Record, Volume 6, issue 1.

Video Sharing Platform Data Extraction: Transforming Images into Structured Data

Xiaofan Zhou [1] and Jakramate Bootkrajang [1,2]

[1] Data Science Consortium, Chiang Mai University,
[2] Department of Computer Science, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand
xiaofan_zhou@cmu.ac.th, jakramate.b@cmu.ac.th

Abstract. This study presents a robust framework for automated extraction and performance evaluation of video interaction metrics across major Chinese social media platforms (Bilibili, Douyin, Xiaohongshu) characterized by heterogeneous interface designs. Leveraging a synergistic combination of YOLOv8 object detection and Optical Character Recognition (OCR), the proposed system addresses platform-specific challenges in identifying engagement indicators (likes, comments, shares, views etc.) through icon localization and numerical extraction. A dataset of 250 annotated screenshots encompassing diverse interface variations was utilized to train and validate the deep learning model, achieving mean average precision (mAP@50) of 99.5% across all interaction categories. The extracted metrics were standardized and validated against third-party Key Performance Indicators (KPIs) from commercial analytics platforms (Pugongying, Huahuo and Xingtu), demonstrating 98% alignment in performance classification. Hyperparameter optimization and spatial pyramid pooling enhancements enabled cross-platform generalization, with error analysis revealing OCR misinterpretations (e.g., unit omission in "万" (10k) as the primary accuracy limitation. The framework advances social media analytics by enabling scalable, platform-agnostic performance benchmarking, offering practical value for content optimization, advertising compliance verification, and engagement trend analysis in the evolving short video ecosystem.

Keywords: Image Recognition, Deep Learning, Social Media Analytics.

1 Introduction

The exponential growth of short video platforms like Douyin, Xiaohongshu , and Bilibili has revolutionized digital content consumption, with combined monthly active users exceeding 1.5 billion in China. These platforms drive social interaction, e-commerce, and digital advertising, generating over 254 billion RMB in ad revenue in 2022. Key performance indicators (KPIs) such as likes, comments, shares, and views serve as critical metrics for evaluating content engagement and advertising efficacy. However, the heterogeneous interface designs and iconography across platforms pose significant challenges for automated extraction and standardization of interaction metrics, hindering cross-platform performance analysis.

This study addresses these challenges by developing a unified deep learning framework integrating YOLOv8 object detection and Optical Character Recognition (OCR) to automate the extraction of interaction metrics from platform-specific screenshots. Leveraging a dataset of 250 annotated screenshots capturing interface variations, the system localizes engagement icons (e.g., likes, shares) and extracts associated numerical values, converting them into structured data. Extracted metrics are validated against third-party KPIs derived from commercial analytics platforms, achieving 98% alignment in performance classification. By resolving platform-specific design disparities through spatial pyramid pooling and hyperparameter optimization, the framework demonstrates robust generalizability across Douyin, Xiaohongshu, and Bilibili.

This work advances social media analytics by enabling scalable, real-time evaluation of video performance, offering actionable insights for advertisers, content creators, and platform operators to optimize strategies and verify advertising compliance in China's dynamic short video ecosystem.

2 Literature Review

2.1 Applications of Object Detection

Recent advancements in deep learning have propelled object detection technologies into diverse domains. In healthcare, YOLO-based systems demonstrate exceptional utility: Qian Xue et al. [1] achieved real-time pharmaceutical defect detection, while Ju Gongwei et al. [2] automated surgical glove orientation identification, reducing manual errors. Pan Xiwen [3] enhanced tumor diagnostics through medical image segmentation, and Wang Feng [4] developed pandemic-compliant PPE monitoring systems.

Transportation systems benefit from improved safety and efficiency. Li Shan [5] optimized road damage classification for maintenance prioritization, while Li Lei [6] integrated vehicle detection with traffic flow prediction. Jiang Jinhong et al. [7] refined traffic sign recognition, balancing accuracy (94.2%) and real-time performance. Engineering applications include Gao Zhao's [8] remote sensing solution for large-scale object detection and Li Xiaoxuan's [9] fire safety equipment recognition system, achieving 97.3% precision in architectural plan analysis.

Agricultural automation has seen transformative impacts. Kuznetsova et al. [10] enabled robotic apple harvesting with 98% detection accuracy, while Mathew et al. [11] mitigated crop losses through early plant disease identification. Sozzi et al. [12] revolutionized viticulture by automating grape variety classification. Industrial advancements include Mahendrakar et al. [13], who reduced satellite collision risks through component tracking, and Darma et al. [14], who digitized cultural artifacts via 3D carving recognition.

2.2 Applications of OCR Recognition

Stilianos Fountas and Don Brendin.[15] used quarterly data for the UK from 1978 to 1998, with exports as the explanatory variable and relative prices, real income and exchange rates as explanatory variables, it is concluded that exchange rates do not affect exports in the short run, but have a significant effect in the long run.

OCR technology has evolved from foundational text recognition to domain-specific solutions. Early innovations by Cao Erqiang [16] in handwritten character recognition (92.1% accuracy) and Shi Suxia et al. [17] on UCI datasets laid groundwork for modern systems. Financial applications stand out: He Xiaolin [18] automated credit authorization document processing at the People's Bank of China, reducing processing time by 68%. Xiao Minghan et al. [19] enhanced logistics efficiency through address recognition in parcel sorting (96.4% success rate).

Industrial digitization efforts include Cai Jun et al. [20], whose construction archive metadata system improved retrieval speed by 40%, and Liu Chao et al. [21], who replaced error-prone manual steel slab identification with 99.1%-accurate OCR. Energy sector innovations feature Qiu Hao et al. [22], whose power permit verification system decreased risk assessment errors by 32%. Cultural preservation benefits from Wang Yuzhen's [23] offline handwriting recognition for exam grading (93.7% accuracy).

Cross-industry advancements demonstrate OCR's versatility. Sato et al. [24] enabled rapid news archive indexing via subtitle extraction, while Gupta [25] accelerated document digitization workflows by 55%. Modern systems address complex challenges: Plamondon et al. [26] bridged online/offline handwriting recognition gaps, and Stoliński [27] solved historical document preservation issues through adaptive binarization techniques.

3 Methodology and Experiment

This chapter presents the technical framework and experimental implementation of the proposed system for automated extraction and validation of video interaction metrics. The methodology integrates YOLO-based object detection with OCR processing, supported by systematic data preparation and model optimization strategies. Experimental validation demonstrates the effectiveness of the approach across three major Chinese social media platforms.

3.1 Technical Framework

The system architecture comprises three core components:

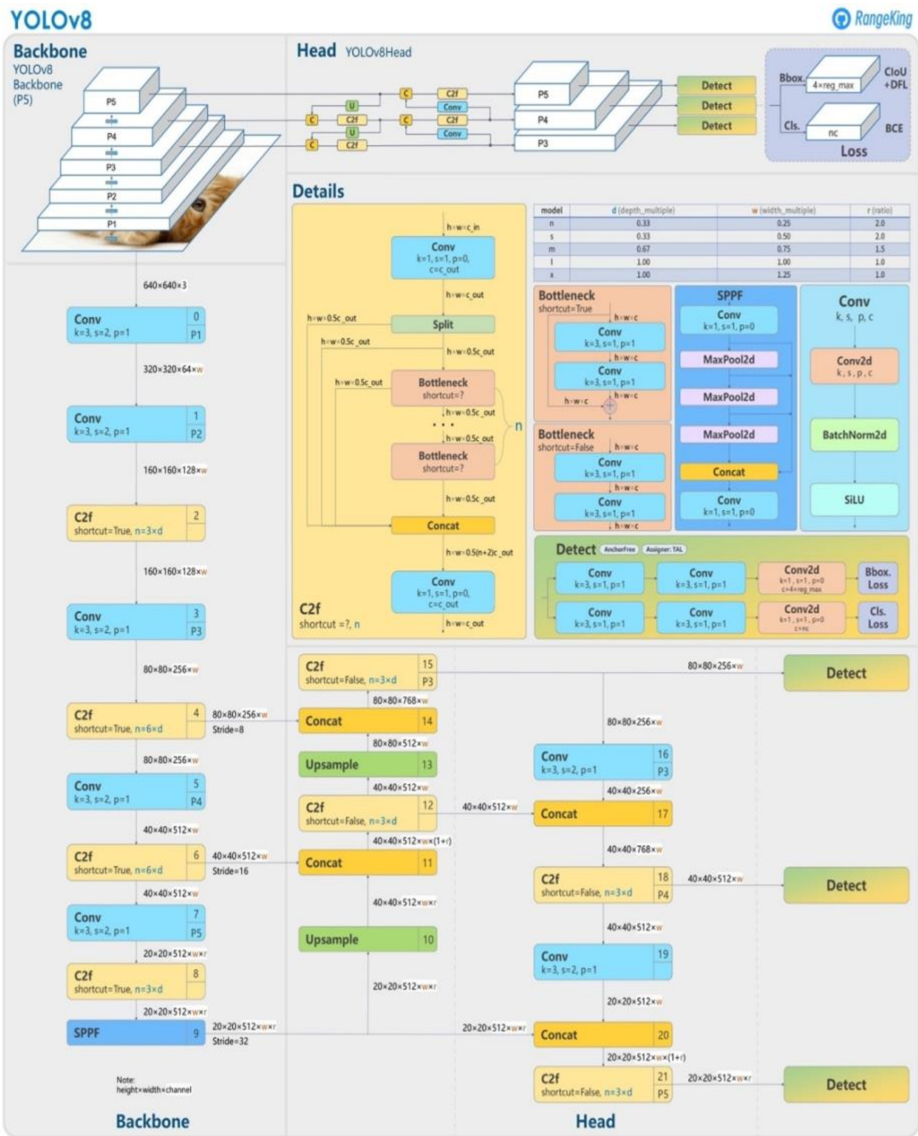

Fig. 1. YOLOv8 Network Structure Diagram [28]

1) Data Acquisition Module: Captures platform screenshots containing interaction metrics
2) Object Detection Module: YOLOv8 model for icon localization
3) OCR Processing Module: Text extraction from detected regions

The YOLOv8 model introduces several structural improvements over previous versions:

1) SPPF Module: Accelerated spatial pyramid pooling using successive 5×5 max-pool layers.

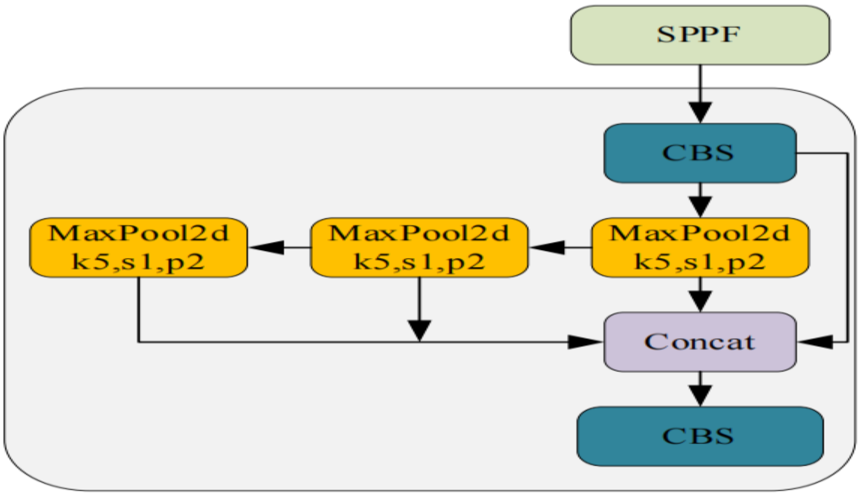

Fig. 2 SPPF Module Sructure Diagram [29]

2) Decoupled Head: Separate branches for classification and regression tasks.

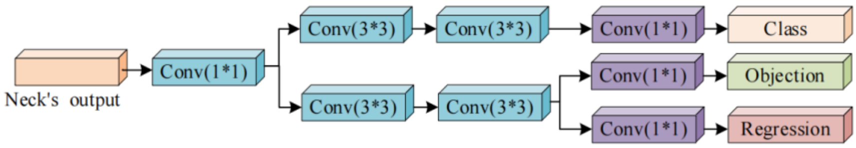

Fig. 3 Decoupled Head Structure Diagram [30]

3) Anchor-Free Detection: Direct coordinate prediction eliminating preset anchor boxes.

3.2 Data Preparation and Annotation

3.2.1 Dataset Composition

The experimental dataset contains 250 annotated screenshots from three platforms:

Table 1. Distribution of Labeled Instances Across Interaction Categories

Category	Number of Labeled instances
Likes	250
Comments	250
Shares	200

Collections	250
Playbacks	100
Coins	100

3.2.2 Annotation Process

Three-stage quality control ensured annotation accuracy:

1) Primary Annotation: Initial bounding boxes for interaction elements
2) Consistency Check: Cross-validation by multiple annotators
3) Error Correction: Resolution of discrepancies through group discussion

Platform-specific UI characteristics are shown in annotation examples:

Douyin: Dense arrangement

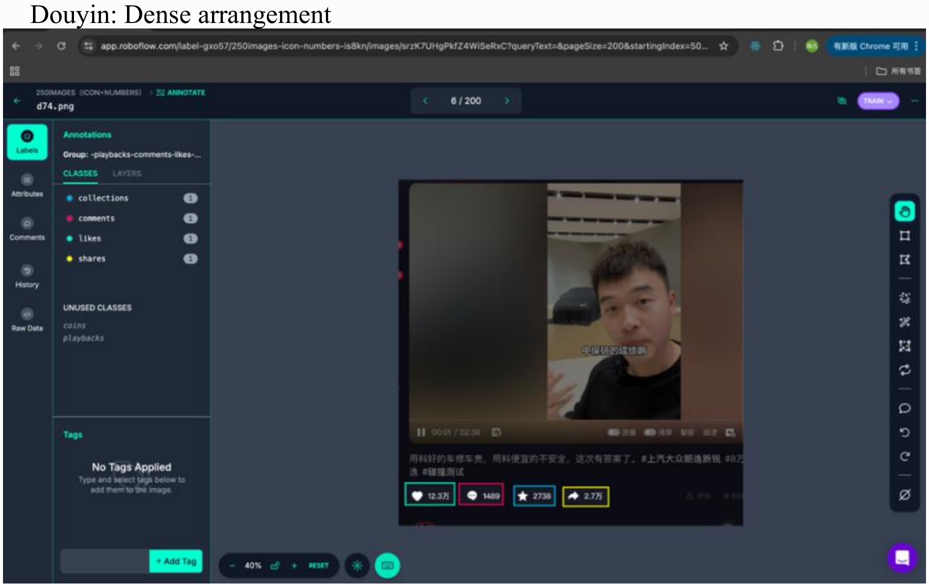

Fig. 4 .Annotation Example Image of Douyin

Bilibili: Decentralized layout

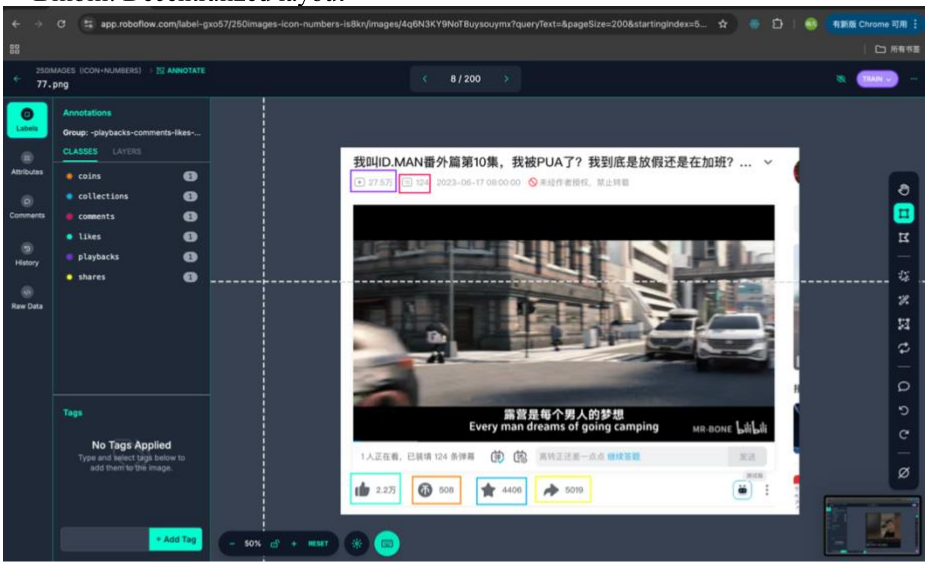

Fig. 5. Annotation Example Image of Bilibili

Xiaohongshu: Different location layout

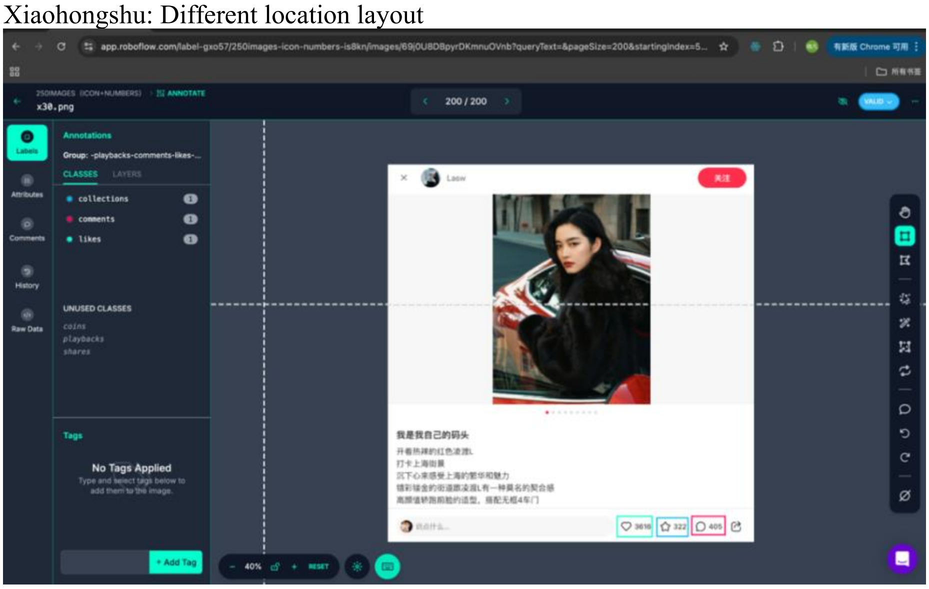

Fig. 6. Annotation Example Image of Xiaohongshu

3.3 Model Training and Optimization

3.3.1 Hyperparameter Configuration

Training parameters were optimized through grid search:

Table 2. Parameter table

Parameter	Value
Input Resolution	640×640
Batch Size	16
Optimizer	SGD+NAG
Epochs	100

The learning rate scheduler implemented cosine decay with warmup.

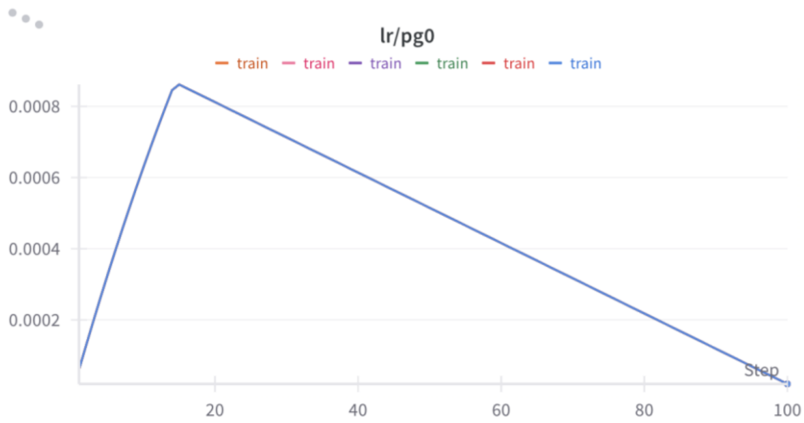

Fig. 7. Learning Rate Scheduler Curve

During the initial phase of training, the learning rate gradually increases to accelerate model convergence, reaching its peak around epoch 20. After that, it progressively decreases to prevent oscillations in the later stages of training, ensuring more stable optimization of model weights. This learning rate scheduling strategy significantly enhances the stability of the training process, allowing YOLOv8 to achieve optimal convergence during training.

3.3.2 Training Process Analysis

Training metrics demonstrated stable convergence:

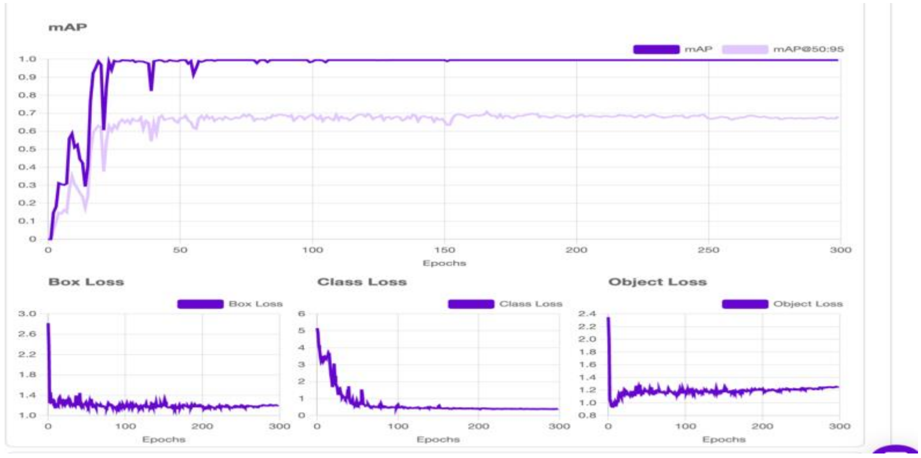

Fig. 8. Training Process and Loss Curves

The Box Loss was initially high at approximately 2.8, but it rapidly decreased as training progressed and finally converged at 1.0. This indicates that the model successfully optimized the bounding box predictions, ensuring that the detected bounding boxes more accurately cover interaction icons. Similarly, the Class Loss started at 5.5 but steadily declined to nearly 0, suggesting that the model achieved minimal classification errors. This demonstrates its ability to accurately distinguish different categories of interaction icons. The Object Loss exhibited significant fluctuations in the early stages of training but gradually stabilized, ultimately converging at 1.2. This loss metric reflects the model's confidence in determining whether an object exists in the detected region. While the overall convergence was satisfactory, slight variations remained, possibly due to the challenges of detecting small-scale objects. This issue is particularly relevant in scenarios involving low-contrast elements or complex backgrounds, where the model might occasionally struggle with icon detection.

To further assess the balance between Precision and Recall, Precision-Recall curves and Recall-Confidence curves were generated.

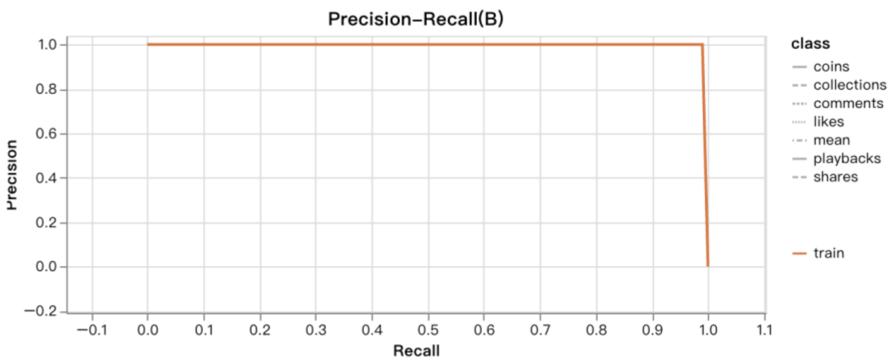

Fig.9. Precision-Recall

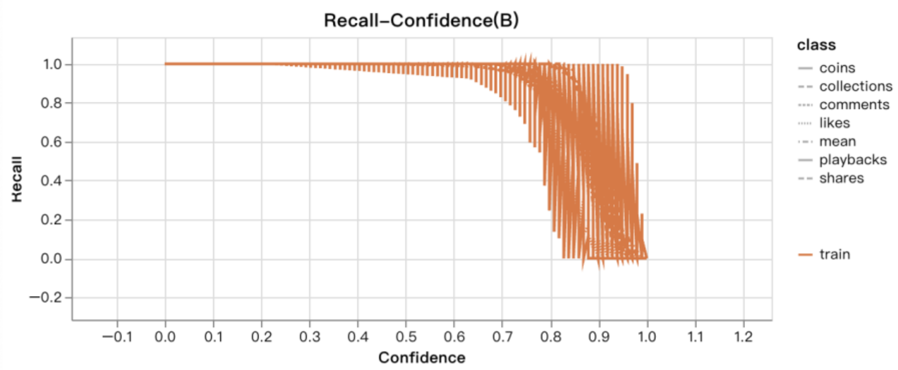

Fig.10. Recall-Confidence Curves

The Precision-Recall curve demonstrates that Precision remains consistently high across all Recall thresholds, indicating that the model effectively minimizes false positives while maintaining a strong recall rate. This suggests that even when Recall increases, Precision does not significantly decline, highlighting the model's robustness in various detection scenarios. The Recall-Confidence curve analysis reveals a slight decline in Recall at high confidence thresholds, which may be attributed to challenges in detecting small interaction icons such as coins or view counts. These elements, due to their small size, are more susceptible to background noise and resolution constraints. Additionally, certain interaction icons may possess less distinguishable visual features, leading to lower confidence scores and occasional detection failures. Overall, the balance between Precision and Recall confirms that YOLOv8 exhibits strong detection stability, making it well-suited for detecting various types of interaction icons across different social media platforms.

3.4 OCR Integration and Validation

The OCR module achieved 99.22%-character recognition accuracy through:
1) Text Detection: Easy OCR for precise text localization
2) Text Recognition: CRNN with CTC decoding

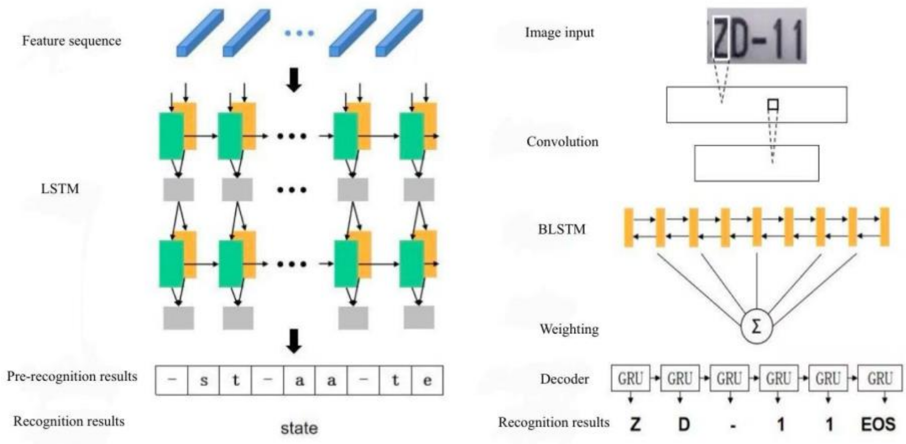

Fig. 11. Convolutional Recurrent Neural Network [31]

3.5 KPI Validation Mechanism

After a video is published, advertising agencies extract the actual view count and engagement rate from the screenshots and compare them against the contractual KPI thresholds:
1) If actual performance ≥ KPI, the content creator is eligible for full payment.
2) If actual performance < KPI, the agency may impose a penalty or renegotiate the contract terms

Sample validation results demonstrate effective KPI compliance prediction:

Table 3. KPI Prediction Results Example

Order	Blogger ID	Link	Qualification Status
1	swing_窝	https://www.xiaohongshu.com/explore/6593faac000000000f0110eb	Pred: Qualified | Actual: Qualified
2	草莓菌🍓	https://www.xiaohongshu.com/explore/65953c13000000001d037d4d	Pred: Qualified |

			Actual: Qualified
3	An jallalla	https://www.xiaohongshu.com/explore/65951a74000000001a001d76	Pred: Qualified | Actual: Qualified
4	白洋-Aries	https://www.xiaohongshu.com/explore/659623b4000000001c013d29	Pred: Qualified | Actual: Qualified
5	跟滕老师去旅行	https://www.xiaohongshu.com/explore/65955ac0000000001e007186	Pred: Qualified | Actual: Qualified
6	游川	https://www.xiaohongshu.com/explore/6596733d00000000120091e3	Pred: Qualified | Actual: Qualified
7	爱笑的小虾米	https://www.xiaohongshu.com/explore/6597beb60000000011030978	Pred: Qualified | Actual: Qualified
8	洋芋咸	https://www.xiaohongshu.com/explore/659cce96000000001500117b	Pred: Qualified | Actual: Qualified
9	皮晶姐姐	https://www.xiaohongshu.com/explore/65a0ed18000000000f01249a	Pred: Qualified | Actual: Qualified
10	杨柳依	https://www.xiaohongshu.com/explore/65a0d382000000001e0085fb	Pred: Qualified | Actual: Qualified

Order	Interaction Numbers KPI	likes	comments	collections	interactions
1	767	70000	269	867	71136

2	414	87000	325	5718	93043
3	550	26000	42	27	26069
4	511	74000	876	14000	88876
5	691	49000	367	517	49884
6	697	19000	79	2027	21106
7	526	49000	551	3610	53161
8	600	83000	151	1239	84390
9	3602	117000	1180	2577	120757
10	1766	28000	197	758	28955

3.6 KPI Prediction Accuracy and Model Performance

In this study, the KPI prediction accuracy reached 98%, demonstrating the high feasibility of the YOLO-OCR-based data extraction approach for automated social media data analysis. Among the 250 test samples, OCR recognition errors were minimal, with only 2% of KPI predictions being affected by misrecognized numerical values. This indicates that the proposed method can accurately compute KPIs and reliably determine content qualification in most cases. Although OCR still exhibits minor recognition errors, particularly in identifying the "万"(10k) character, the overall accuracy and stability of the YOLO-OCR combination remain exceptionally high. Future optimization efforts could focus on enhancing the recognition of the "万" (10k) character, such as leveraging a custom OCR dictionary or improving character contrast to reduce misclassification. Additionally, implementing numerical range constraints could help automatically detect unreasonable value deviations in recognition results, further minimizing the impact of OCR errors on KPI computation.

4 Discussion

This section elaborates on the strengths, limitations, and future development opportunities of the proposed YOLOv8-OCR-based KPI extraction framework based on experimental findings. Particular attention is given to the system's cross-platform generalizability, OCR-related errors, and enhancement strategies.

4.1 Cross-Platform Generalizability

The proposed framework demonstrated strong generalization capability across Bilibili, Douyin, and Xiaohongshu, despite significant visual and stylistic differences among these platforms.

1) Bilibili: Icons and numbers are relatively large and uniformly spaced, resulting in the highest detection and OCR accuracy.
2) Douyin: Layout variations are frequent, including changes in icon order and placement, slightly lowering OCR performance.
3) Xiaohongshu: Interface is stable and minimalist, leading to consistently high extraction accuracy even with a smaller dataset.

These results highlight YOLOv8's robust feature extraction ability and the importance of selecting adaptive OCR parameters when dealing with different text densities and font styles.

4.2 OCR Error Impact and Analysis

Although object detection achieved near-perfect precision and recall, OCR errors remained a primary source of end-to-end KPI extraction inaccuracies. Major types of OCR failures included:

1) Character Omissions: Especially the omission of '万' (10k), critically affecting high-value KPIs.
2) Numeric Distortion: Misrecognition of characters under compression or low resolution.
3) Single-Digit Instability: Misinterpretation of isolated digits.

4.3 Future Directions

Despite multiple rounds of optimization targeting OCR misrecognition issues—including adjustments in OCR preprocessing, detection box refinement, training data enhancement, and contextual correction—certain recognition errors persist. This indicates that even with specialized optimizations, OCR's generalization ability remains limited under specific font styles and UI designs. Factors such as character shape similarity, low resolution of small-font text, and biases in OCR language models toward vocabulary remain challenging for current OCR technologies to fully resolve. These challenges explain why the three main types of misrecognitions observed in this study still occur. Even after implementing all reasonable improvements, OCR remains susceptible to misclassification due to UI design variations, font characteristics, and character similarities. Future research could explore deep-learning-based OCR error correction mechanisms, integrating natural language processing (NLP) or contextual pattern matching to automatically rectify high-risk characters such as “万” (10k) and “9.” Another potential optimization approach is a multi-model fusion strategy, leveraging multiple OCR engines (e.g., combining EasyOCR and Tesseract) to enhance character differentiation. Additionally, neural network-based image super-resolution techniques could be applied to improve the resolution of small-font text, thereby enhancing OCR recognition accuracy and mitigating errors caused by insufficient text resolution.

5 Conclusions

This study proposes a YOLO-based object detection and OCR-based text recognition approach for extracting engagement data from social media platforms, addressing the limitations of traditional Web Scraping in social media data collection. The research focuses on dataset construction, object detection, OCR recognition, KPI computation, and evaluation, leveraging interaction data from three major platforms: Bilibili, Douyin, and Xiaohongshu. The model's detection performance was assessed, and a comprehensive comparison of different data extraction methods was conducted. Experimental results demonstrated that this approach achieved a 99.22% OCR recognition accuracy and a 98% KPI prediction accuracy, offering an efficient and stable automated solution for social media data analysis.

5.1 Key Contributions of the Study

1) Proposing a YOLOv8+ OCR-based social media data extraction method

Integrated YOLO-based object detection to identify engagement icons and EasyOCR for extracting numerical information, enabling the automation of KPI data acquisition. This method circumvents anti-scraping mechanisms, ensuring data collection stability while maintaining cross-platform adaptability and scalability.

2) Constructing a cross-platform dataset and training the model

Developed a dataset covering Bilibili, Douyin, and Xiaohongshu, consisting of 250 screenshots and 1,150 labeled instances. Trained a YOLOv8 model for object detection and optimized the OCR recognition model to enhance the accuracy of numerical data extraction.

3) Systematically evaluating KPI computation accuracy

Extracted engagement data using OCR for KPI calculations and compared the results with manually labeled KPI values. The study found that OCR misrecognition primarily affected the “万” (10k) character and single-digit numbers, but the overall impact on KPI calculations was minimal, with an ultimate KPI prediction accuracy of 98%.

5.2 Limitations of the Study

1) OCR misrecognition issues persist

Primary challenges include: The “万” (10k) character being misrecognized as “仿” or omitted entirely, leading to a 10-fold reduction in numerical values, thereby impacting KPI calculations. Failure to recognize the digit 9, resulting in missing numerical values. Future improvements could leverage custom OCR dictionaries and contrast enhancement techniques to further improve OCR accuracy.

2) Limited dataset size

The dataset used in this study comprises only 250 screenshots, with Xiaohongshu accounting for just 50 images. While the dataset was sufficient for model training, a larger dataset would improve model generalization. Future studies should expand the dataset, incorporating more social media platforms and varied UI designs to enhance the model's adaptability across different environments.

3) Potential for optimizing KPI computation methods

This study relies on a simple numerical summation approach for KPI computation, but KPI calculation logic varies across platforms. Future work could explore more sophisticated KPI computation methods, incorporating time-based factors and user behavior patterns to improve KPI prediction accuracy.

5.3 Future Research Directions

Based on the findings of this study, future research can be expanded in the following areas:

1) Expanding the dataset scale

Incorporate data from additional social media platforms, such as Weibo, Instagram, and Twitter, to improve cross-platform adaptability. Collect screenshots with diverse UI designs to enhance the model’s generalization ability for different interface layouts.

2) Integrating NLP for OCR error correction

Utilize Natural Language Processing (NLP) techniques, such as spell correction models, to improve post-OCR text quality. Train an OCR context correction model to automatically rectify OCR misrecognitions, such as "3.1 万" being misread as "3.仿".

3) Deploying an automated social media data monitoring system

Develop a real-time data collection pipeline, integrating YOLO + OCR + data analysis to monitor social media metrics dynamically and provide business intelligence insights.

5.4 Conclusion

The YOLO + OCR method proposed in this study has demonstrated strong applicability in extracting engagement data from social media platforms, effectively addressing the limitations of Web Scraping, such as anti-scraping restrictions, webpage structure changes, and compliance issues. The experimental results indicate that this method enables efficient and accurate KPI data extraction, maintaining high recognition accuracy across multiple social media platforms. Although OCR misrecognition issues persist, their overall impact on KPI calculations is minimal, with a final KPI prediction accuracy of 98%, confirming the method’s practical value.

Future research should focus on enhancing OCR recognition accuracy, expanding the dataset, integrating NLP-based text correction, and developing automated KPI monitoring systems to further improve the intelligence level of social media data analysis. The findings of this study can be widely applied in social media marketing analysis, KPI monitoring, and influencer performance evaluation, providing an efficient, stable, and scalable technological solution for the social media industry.

References

1. Xue, Q., Li, J., & Tang, Q. (2021). Real-time detection method for surface defects of pharmaceuticals based on YOLOv5. *Information Technology and Network Security*, 40(12), 45–50.
2. Ju, G., Jiao, H., & Zhang, J. (2021). Left- and right-hand recognition of medical surgical gloves based on YOLOv5. *Manufacturing Automation*, 43(12), 189–192.
3. Pan, X. (2021). *Research on tumor medical image segmentation and detection methods based on convolutional neural networks* (Master's thesis, Beijing University of Posts and Telecommunications).
4. Wang, F. (2020). Improved YOLOv5-based AI detection and recognition algorithm for mask and safety helmet wearing. *Architecture and Budget*, (11), 67–69.
5. Li, S. (2021). Research on road disease detection and classification based on YOLOv5. *Modern Computer*, 27(35), 75–79.
6. Li, L. (2021). *Research on vehicle detection, tracking, and prediction algorithms based on deep learning* (Master's thesis, East China Jiaotong University).
7. Jiang, J., Bao, S., & Shi, W. (2020). Improved traffic sign recognition algorithm based on YOLOv3. *Computer Applications*, 40(8), 2472–2478.
8. Gao, Z. (2020). *Research on target detection methods for remote sensing images based on deep learning* (Master's thesis, Shandong University of Science and Technology).
9. Li, X. (2020). *Research on automatic recognition of fire-fighting equipment based on deep learning* (Master's thesis, Wuhan Research Institute of Posts and Telecommunications).
10. Kuznetsova, A., Maleva, T., & Soloviev, V. (2020). Detecting apples in orchards using YOLOv3 and YOLOv5 in general and close-up images. In *17th International Symposium on Neural Networks* (pp. 233–243). Springer.
11. Mathew, M. P., & Mahesh, T. Y. (2022). Leaf-based disease detection in bell pepper plant using YOLOv5. *Signal, Image and Video Processing*, 16(3), 841–847.
12. Sozzi, M., Cantalamessa, S., & Cogato, A. (2022). Automatic bunch detection in white grape varieties using YOLOv3, YOLOv4, and YOLOv5 deep learning algorithms. *Agronomy*, 12(2), Article 1–17.
13. Mahendrakar, T., White, R. T., & Wilde, M. (2021). Real-time satellite component recognition with YOLOv5. In *Small Satellite Conference*. Mahendrakar.
14. Krišto, M., Ivašić-Kos, M., & Pobar, M. (2020). Thermal object detection in difficult weather conditions using YOLO. *IEEE Access*, 8, 125459–125476.
15. Sohu. (2024). Short video advertising market growth trends. Retrieved from https://www.sohu.com/a/722321939_121388268
16. Cao, E. (1985). Research on handwritten character recognition using feature recognition method. *Journal of Changchun University of Posts and Telecommunications*, (1), 73–93.

17. Shi, S., Chang, W., & Song, Z. (2022). OCR optical character recognition based on the UCI dataset. *Technological Innovation and Application*, 12(35), 50–53.
18. He, X. (2023). Application analysis of OCR text recognition in personal credit inquiry services. *Financial Technology Era*, 31(1), 60–64.
19. Xiao, M., Deng, D., & Lin, H. (2022). Research on express address information recognition based on OpenCV and Tesseract. *Electronic Testing*, 36(22), 51–54.
20. Cai, J., & Chen, X. (2022). Application of OCR in file-level cataloging information verification for urban construction archives. *Lantai World*, (6), 95–97.
21. Liu, C., Song, H., & Liu, X. (2022). Slab spray number recognition based on Baidu AI open platform. *Software*, 43(10), 166–169.
22. Qiu, H., Zhang, W., & Lin, X. (2022). Work ticket segmentation and information extraction for power operations. *Journal of Electric Power Science and Technology*, 37(6), 198–205.
23. Wang, Y. (2022). Design of grading system for offline handwritten Chinese character recognition. *Information Technology and Informatization*, (9), 207–209.
24. Sato, T., Kanade, T., & Hughes, E. K. (1998). Video OCR for digital news archive. In *IEEE International Workshop on Content-Based Access of Image and Video Database* (pp. 52–60). IEEE.
25. Gupta, G., Niranjan, S., & Shrivastava, A. (2006). Document layout analysis and classification in OCR. In *10th IEEE International EDOC Workshops* (EDOCW'06) (p. 58). IEEE.
26. Plamondon, R., & Srihari, S. N. (2000). Online and offline handwriting recognition: A comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1), 63–84.
27. Stoliński, S., & Bieniecki, W. (2011). Application of OCR systems in document digitization. *Information Systems in Management*, 8, 102–121.
28. Jocher, G. (2023). *Issue #189: [Issue Title]* [GitHub Issue]. Ultralytics YOLOv5 Repository. Retrieved August 2023, from https://github.com/ultralytics/yolov5/issues/189
29. **Ultralytics**. *Comprehensive Guide to Ultralytics YOLOv5*. Ultralytics YOLO Documentation. Retrieved from https://docs.ultralytics.com/yolov5
(Note: Adjust the year if specific publication/update dates are available.)
30. **Carion, N.**, Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-End Object Detection with Transformers. *arXiv preprint* **arXiv:2005.12872**. Retrieved from https://arxiv.org/abs/2005.12872
31. 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. https://doi.org/10.1109/TPAMI.2016.2586371