Journal of Theory and Practice of Engineering Science ISSN: 2790-1505
DOI: 10.53469/jtpes.2024.04(05).08

High-Precision Neuronal Segmentation: An Ensemble of YOLOX, Mask R-CNN, and UPerNet

Zhengning Li[1], Yibo Yin[2], Zibu Wei[3], Yang Luo[4], Guokun Xu[5], Ying Xie[6]

[1]Computer Science, Georgetown University, Washington, D.C. USA
[2]Computer Science, Contemporary Amperex Technology USA Inc, Auburn Hills, USA
[3]Computer Science, University of California, Los Angeles, Los Angeles, USA
[4]Computer Science, China CITIC Bank Software Development Center, Beijing, China
[5]Computer Science, Beijing Foreign Studies University, Beijing, China
[6]Computer Science, San Francisco Bay University, Fremont, USA
[1]zl132@georgetown.edu, [2]epark00007@gmail.com, [3]zibuwei@ucla.edu, [4]luoyangdxx@163.com, [5]nicolashsu@hotmail.com, [6]floraxlr999@gmail.com

Abstract: ***In the realm of neuron cell segmentation from microscopic images, computer vision technologies have shown potential in accelerating drug discovery processes for neurological disorders. This paper presents an innovative approach that combines YOLOX for object detection, Mask R-CNN for instance segmentation, and UPerNet for semantic segmentation to precisely delineate individual cells. Our methodology emphasizes advanced data preprocessing techniques and model ensembling to improve detection and segmentation of neuronal cell types. Extensive experiments conducted on the Sartorius Cell Instance Segmentation dataset demonstrate the superiority of our approach, achieving state-of-the-art results.***

Keywords: Cell Segmentation; YOLOX; Mask R-CNN; UPerNet; Deep Learning; Microscopic Image Analysis.

1. INTRODUCTION

Neuronal cell instance segmentation from microscopic images is a cornerstone in understanding the intricate structure and function of the nervous system. This process, crucial for the advancement of neurological disorder research, faces significant challenges due to the complex morphology of neuronal cells and the inherent difficulties in visualizing and distinguishing individual cells within dense cellular networks. Traditionally, the segmentation of such cells has been performed manually, a procedure that is not only laborious and prone to human error but also impractical at a larger scale needed for significant biomedical research and therapeutic development.

We present an integrated approach that synergizes the strengths of three powerful models: YOLOX, Mask R-CNN, and UPerNet. Each of these models is well-regarded in the domains of object detection, instance segmentation, and semantic segmentation respectively. By weaving together their capabilities, we aim to not only identify the presence of individual neuronal cells within the complex tapestry of microscopic images but also delineate their precise boundaries and understand their contextual relationships within the tissue.

Our approach begins with the meticulous preprocessing of the data, an often overlooked yet critical step in the modeling process. Through innovative techniques of segmentation, we enhance the quality of the training data, ensuring that the models are primed with accurate and representative information. Subsequently, we employ YOLOX for its robustness in object detection, capturing the varied forms of neuronal cells across different images. Mask R-CNN is then applied to carve out the exact contours of each cell, providing detailed instance segmentation. Finally, UPerNet brings its semantic segmentation prowess to refine the contextual understanding of the segmented instances.

The methodology proposed herein is not merely an amalgamation of models; it is a sophisticated ensemble that emphasizes the collaborative strength of its constituents. By harmonizing these models, we tackle the prevalent issues of small object detection, overlapping cell structures, and the nuances of varying cell morphologies. This ensemble not only achieves higher precision in cell detection and segmentation but also ensures robustness across different cell types and image qualities.

2. RELATED WORK

The advent of deep learning has revolutionized the field of medical image analysis, especially in neuron instance segmentation from microscopic images. Wu et al. [1] introduced a framework based on Efficient UNet coupled with morphological post-processing, showcasing the potential of automated feature identification and reducing the manual effort in biomedical image segmentation. The ability of deep learning to produce high-accuracy segmentation with minimal human intervention post-training has been well documented [2].

Advancements such as the YOLO2U-Net framework have pushed the envelope in detection-guided 3D instance segmentation for microscopy, highlighting the synergy between deep networks and traditional segmentation methods [3]. Additionally, addressing the challenge of cluttered cells, Peng et al. [4] proposed a multiclass weighted loss function to refine instance segmentation tasks. In the field of neurophysiology, the DISCo framework presented by Kirschbaum et al. [5] leverages deep learning for instance segmentation and analysis of calcium imaging, providing insights into the intricate workings of neuronal cells.

Furthering the goal of reliable cell segmentation in crowded environments, the NuSeT tool has been developed, integrating U-Net with a modified Region Proposal Network (RPN) for the segmentation of fluorescently labeled nuclei [6]. Meanwhile, efforts in neuronal cell type classification have employed deep learning to categorize cells based on morphological and electrophysiological properties [7].

Yan et al. [8] have contributed significantly to the neuronal soma segmentation field by introducing a 3D multi-task learning approach that employs a U-shaped Fully Convolutional Neural Network (FCNN), ensuring accuracy and efficiency. Peng et al. [9] addressed domain generalization in 3D human pose estimation with a dual-augmentor framework that innovatively employs weak and strong data augmentation strategies. This approach allows for effective generalization across different domains without prior knowledge of the target domain, overcoming some of the limitations inherent in the single-augmentor models previously used in the field. Despite these advances, the field still grapples with several challenges, including the segmentation of cells with irregular shapes, low-contrast boundaries, and the need for extensive annotated datasets. Our work proposes an integrated approach that leverages the strengths of YOLOX, Mask R-CNN, and UPerNet to address these limitations, significantly enhancing segmentation quality and robustness across various neuronal cell types and imaging conditions.

3. ALGORITHM AND MODEL

The segmentation of cells from microscopic images is not only a fundamental task in computational biology but also a significant challenge due to the high variability in cell shapes, sizes, and clustering. Traditional methods often fail to accurately segment all cells in a given image, particularly when cells are closely packed or when their intensities are not uniform. To overcome these obstacles, we employ a combination of three advanced models: YOLOX for object detection, Mask R-CNN for instance segmentation, and UPerNet for semantic segmentation. By integrating these models, we aim to create a robust system capable of delivering high-precision cell segmentation, which is critical for subsequent biological analysis and applications. The following sections detail the architecture and functionality of each component within our methodology, illustrating their roles and synergies in tackling this demanding task.

3.1 YOLOX - Object Detection

YOLOX serves as the cornerstone of our object detection module, chosen for its exceptional speed and accuracy, crucial for handling the vast datasets typical in medical image analysis. Unlike traditional detectors that utilize a dense sampling of anchor boxes, YOLOX adopts an anchor-free approach, streamlining the prediction process and significantly boosting inference speed.

(1) **Anchor-Free Detection:** YOLOX eliminates the need for pre-defined anchors, using a grid-based mechanism to predict object locations directly. This simplification leads to a reduction in model complexity and an increase in speed, making it well-suited for real-time applications.

(2) **Decoupled Head:** The model separates the tasks of object classification and bounding box prediction into distinct heads. This decoupling allows for specialized learning, where each head is optimized for its specific task, leading to improved accuracy and efficiency.

(3) **MOSAIC Data Augmentation:** A powerful augmentation technique that combines four different training images into one composite image. This method enhances the model's exposure to various object scales and layouts, improving its robustness against diverse transformations and lighting conditions.

(4) **SiLU Activation:** YOLOX incorporates the Sigmoid Linear Unit (SiLU) as its activation function, which has been shown to converge faster than the commonly used ReLU. This choice supports quicker training and potentially better performance on complex segmentation tasks.

(5) **Progressive Learning:** The model is equipped with a progressive learning strategy where the size of the input images is gradually increased during the initial phase of training. This technique helps the model to stabilize its predictions early in training and adapt more effectively to various image resolutions.

3.2 Mask R-CNN - Instance Segmentation

Mask R-CNN is an advanced model for instance segmentation that extends the capabilities of Faster R-CNN by incorporating a dedicated branch for mask prediction, alongside the existing branches for object classification and bounding box regression. This model is particularly valued for its ability to generate precise pixel-level masks and is instrumental in our cell segmentation tasks.

(1) **Architecture Overview:** At its core, Mask R-CNN uses a convolutional neural network (CNN) to scan the image in a single pass and generates proposals about regions where objects might be located (Region Proposals). Each region proposal is then branched out to perform both classification and bounding box regression, and importantly, to predict a segmentation mask in a pixel-to-pixel alignment.

(2) **ROIAlign:** This feature corrects the misalignment issues found in previous ROI pooling layers by applying bilinear interpolation. This method precisely extracts a small feature map for each Region of Interest, preserving exact spatial locations, crucial for high-quality segmentation. ROIAlign significantly enhances the model's accuracy by ensuring that the extracted features are perfectly aligned with the input, making it extremely effective for medical imaging where precision is paramount.

(3) **Multi-Task Loss:** The multi-task loss function in Mask R-CNN is a composite function that combines the losses for classification, bounding box prediction, and mask prediction. This integration allows Mask R-CNN to optimize for several objectives simultaneously, improving learning efficiency and performance:

$$L = L_{cls} + L_{box} + L_{mask} \tag{1}$$

where is the classification loss, is the bounding box loss, and is the mask loss.

(4) **Heuristic Features:** The mask branch uses a small fully convolutional network on top of each ROI, predicting segmentation masks at a pixel level. This branch operates independently of the bounding box prediction, allowing for more precise segmentation that is not constrained by the limits of the bounding box dimensions.

By implementing these sophisticated architectural and functional enhancements, Mask R-CNN provides a powerful solution for instance segmentation tasks. Its ability to accurately segment individual objects at the pixel level makes it an ideal choice for medical applications where detailed morphological analysis is required.

3.3 UPerNet - Semantic Segmentation

UPerNet stands out as a versatile and robust framework for semantic segmentation, integrating an array of advanced features from various backbone networks. This model excels in processing images at multiple scales through its sophisticated use of a Feature Pyramid Network (FPN) and a Pyramid Pooling Module, making it highly effective for detailed and complex image segmentation tasks like those needed in cellular morphology studies.

(1) **Feature Pyramid Network (FPN)**: UPerNet's FPN uses a multi-scale pyramid hierarchy to process spatial hierarchies within an image. Each level of the pyramid corresponds to a different scale of features, which are then upsampled and combined to preserve both high-resolution details and high-level semantic information. This structure is particularly beneficial for segmenting images with varied object sizes and complex backgrounds.

(2) **Pyramid Pooling Module**: This module further enhances the capability of the network to handle multi-scale information by aggregating the global context, which is crucial for accurate semantic segmentation. It pools features at different levels, from global to local, to capture contexture nuances that might be missed at a single scale.

(3) **Feature Fusion**: The fusion process integrates these multi-level features to maintain a balance between resolution and semantic richness. This is crucial for accurately segmenting small and detailed structures like cells, where precision is paramount.

(4) **Dilated Convolutions**: To increase the receptive field without losing resolution, UPerNet employs dilated convolutions. This allows the network to encompass more contextual information without compromising the sharpness of the image. Dilated convolutions are essential for detailed texture analysis in medical images, where understanding the broader context can be as important as capturing minute details.

(5) **Auxiliary Losses**: In an innovative approach to training, UPerNet incorporates auxiliary losses during the intermediate stages of the network. This technique helps in refining the feature extraction capabilities by penalizing the loss at multiple scales, thereby ensuring that the network learns robust and invariant features throughout its depth.

The architectural advancements and meticulous design choices made in developing UPerNet make it exceptionally capable for the task of semantic segmentation. Its ability to integrate and process multi-scale features effectively allows it to perform well even in the challenging scenarios presented by microscopic cellular images.

3.4 Model Integration and Ensemble Strategy

Our ensemble strategy effectively combines the distinct advantages of three state-of-the-art models to achieve unparalleled accuracy and robustness in cell segmentation tasks. This strategy is designed to harness the unique strengths of YOLOX, Mask R-CNN, and UPerNet by using a sophisticated algorithm that integrates their outputs into a single, cohesive segmentation map.

(1) **Consensus Voting Mechanism:** At the core of our ensemble strategy is the consensus voting mechanism. This approach utilizes the confidence scores from each of the three models to determine the most probable segmentation outcome for each pixel in the image. By weighting these predictions based on their respective confidence levels, the mechanism ensures that the final output is not only accurate but also minimizes potential errors inherent in any single model.

(2) **Feature Level Fusion:** Before the final consensus voting, an intermediate fusion stage amalgamates features extracted by each model. YOLOX contributes with its rapid detection capabilities, capturing quick and efficient bounding boxes. Mask R-CNN adds precise instance segmentation details within these boxes, and UPerNet provides a broad semantic context to the mix. By blending these features at a data level, the model can leverage a richer set of information than any individual component could offer on its own.

(3) **Adaptive Weighting:** The ensemble model dynamically adjusts the influence of each individual model's prediction based on its performance during validation phases. This adaptive weighting scheme allows for flexible model tuning that can accommodate variations in image types and segmentation tasks, optimizing performance across a diverse set of imaging conditions.

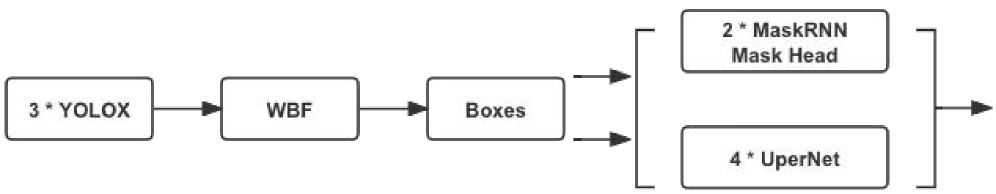

Figure 1: Detailed schematic of the ensemble integration strategy, illustrating the convergence of outputs from YOLOX, Mask R-CNN, and UPerNet.

The integration of these advanced methodologies enables our ensemble to effectively combine the detection precision of YOLOX, the detailed segmentation accuracy of Mask R-CNN, and the contextual sensitivity of UPerNet, resulting in superior overall performance. The ensemble strategy not only enhances the reliability and accuracy of the segmentation but also significantly reduces the likelihood of errors that could occur if relying on a single model approach.

3.5 Model Predictions Visualization

The effectiveness of our integrated ensemble model is demonstrated through visualizations of segmentation results. These visualizations show the model's capability to accurately segment complex cellular structures.

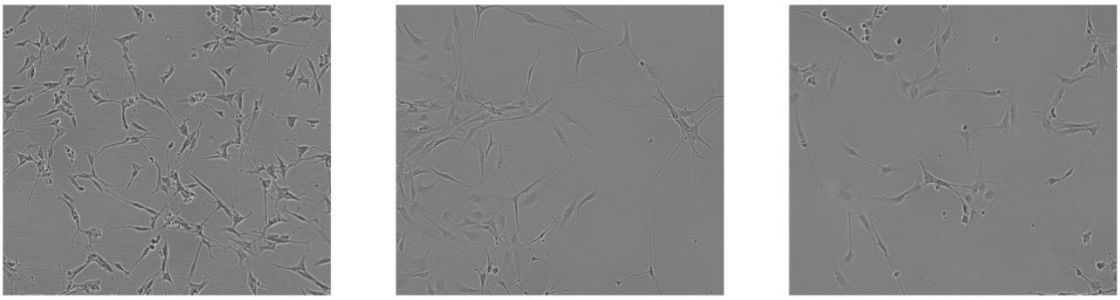

Figure 2: Original microscopic image used for testing.

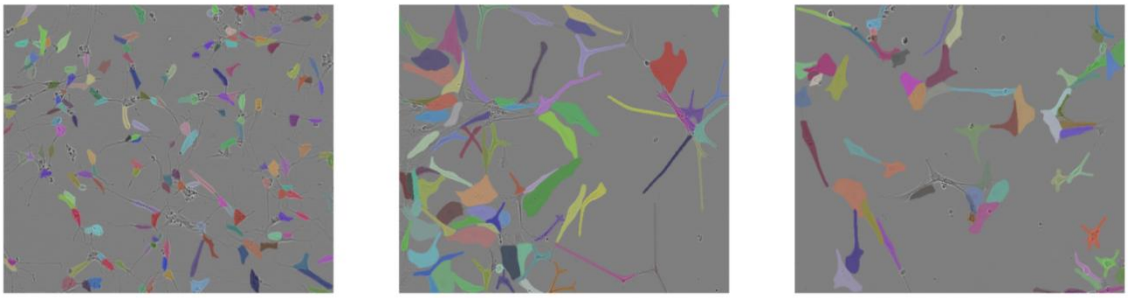

Figure 3: Predicted segmentation visualization showing detailed cell identification.

As illustrated in Figures 2 and 3, the comparison between the original image and the prediction highlights the precision and effectiveness of our segmentation approach.

3.6 Prospects of Large Language Models (LLM)

Large Language Models (LLMs), such as GPT-3 and its successors, have shown remarkable capabilities in natural language processing tasks, but their potential extends beyond text-based applications [10-12]. For instance, IllumiCore explores optimization modeling and implementation for efficient Virtual Network Function (VNF) placement [13]. The use of attention mechanisms has also been employed in news recommendation systems to enhance user experience and engagement [14]. Moreover, models like BERT have been utilized for predicting the graded effect of context in word similarity, showcasing the versatility of advanced language models [15]. Machine learning techniques have further found applications in enhancing network security, providing robust solutions against cyber threats [16]. Additionally, reinforcement learning environments, such as EdgeGYM, have been developed for constraint-aware NFV resource allocation, highlighting the adaptability of AI technologies in cybersecurity [17]. Furthermore, advancements in chatbot technology, like Falcon-7B and 16-bit Full Quantization, have led to significant improvements in e-commerce chatbot performance [18].

Deep learning techniques, including CNN-LSTM hybrid neural networks, have been employed for accurate photovoltaic power generation forecasting [19]. Image captioning has also been explored in the context of news reports, leveraging deep learning methods [20]. Sentiment analysis, particularly in analyzing COVID-19 related tweets, has been enhanced through the fusion of BERT and RCNN models [21]. Optimal resource allocation in SDN/NFV-enabled networks has been achieved using deep reinforcement learning techniques [22]. Additionally, Conv1D-based approaches have been employed for advancing legal citation text classification [23]. Lastly, Particle Filter SLAM methods have been utilized for precise vehicle localization [24].

Advanced network intrusion detection has been realized with TabTransformer [25]. Novel approaches have been developed for the automatic recognition of static phenomena in retouched images [26]. Semi-asynchronous federated learning has been accelerated to improve efficiency [27]. Hybrid machine learning approaches have been proposed for financial time-series forecasting [28]. Insights from customer reviews have been unveiled to enhance e-commerce recommendations using BERTFusionDNN [29]. Image captioning in the news report scenario has been addressed, emphasizing the importance of visual content in news delivery [30]. Equipment health prediction has been enhanced using Enhanced SMOTE-KNN techniques [31]. Furthermore, deep learning has been utilized for crystal system classification in lithium-ion batteries [32].

4. EXPERIMENTS

Our evaluation involved a comparative analysis of different model configurations to assess their effectiveness in segmenting cellular structures on microscopic images. The configurations and their performance measured in Intersection over Union (IoU) are detailed below:

(1) **Mask R-CNN + NMS:** Achieved IoU scores of 0.283 on the public dataset and 0.287 on the private dataset.

(2) **NMS + ResNest + Ensemble:** Recorded IoU scores of 0.326 (public) and 0.341 (private).

(3) **YoloX + Mask R-CNN + UPerNet:** This model configuration outperformed the others, achieving the highest IoU scores of 0.356 (public) and 0.350 (private).

These results clearly indicate that the integration of multiple segmentation methods, particularly the combination of YoloX, Mask R-CNN, and UPerNet, significantly enhances the accuracy and robustness of the segmentation task.

5. CONCLUSION

The experiments conducted reveal that using a hybrid ensemble of advanced modeling techniques substantially improves the performance of cell segmentation tasks in microscopic imaging. The combination of YoloX, Mask R-CNN, and UPerNet not only provided the highest IoU scores but also demonstrated potential for further advancements in medical imaging segmentation. Future work will aim to refine these models further and explore new combinations to enhance the precision and reliability of automated cell segmentation processes. This ongoing research will continue to push the boundaries of what is achievable in the field of medical image analysis.

REFERENCES

[1] Wu, H., Souedet, N., Jan, C., Clouchoux, C., & Delzescaux, T. (2022). A general deep learning framework for neuron instance segmentation based on efficient UNet and morphological post-processing. Computers in Biology and Medicine, 150, 106180.

[2] Kar, A., Petit, M., Refahi, Y., Cerutti, G., Godin, C., & Traas, J. (2022). Benchmarking of deep learning algorithms for 3D instance segmentation of confocal image datasets. PLoS computational biology, 18(4), e1009879.

[3] Vadori, V., Peruffo, A., Graïc, J. M., Finos, L., Corain, L., & Grisan, E. (2023, October). NCIS: Deep Color Gradient Maps Regression and Three-Class Pixel Classification for Enhanced Neuronal Cell Instance Segmentation in Nissl-Stained Histological Images. In International Workshop on Machine Learning in Medical Imaging (pp. 457-466). Cham: Springer Nature Switzerland.

[4] Peng, Q., Zheng, C., & Chen, C. (2023). Source-free domain adaptive human pose estimation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 4826-4836).

[5] Kirschbaum, E., Bailoni, A., & Hamprecht, F. A. (2020). DISCo for the CIA: deep learning, instance segmentation, and correlations for calcium imaging analysis. Preprint at https://arxiv. org/abs/1908.07957 v4.

[6] Vadori, V., Graïc, J. M., Finos, L., Corain, L., Peruffo, A., & Grisan, E. (2023, April). MR-NOM: multi-scale resolution of neuronal cells in Nissl-stained histological slices via deliberate over-segmentation and merging. In 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI) (pp. 1-5). IEEE.

[7] Yang, L., Ghosh, R. P., Franklin, J. M., Chen, S., You, C., Narayan, R. R., ... & Liphardt, J. T. (2020). NuSeT: A deep learning tool for reliably separating and analyzing crowded cells. PLoS computational biology, 16(9), e1008193.

[8] Hu, T., Xu, X., Chen, S., & Liu, Q. (2021). Accurate neuronal soma segmentation using 3D multi-task learning U-shaped fully convolutional neural networks. Frontiers in Neuroanatomy, 14, 592806.

[9] Peng, Q., Zheng, C., & Chen, C. (2024). A Dual-Augmentor Framework for Domain Generalization in 3D Human Pose Estimation. arXiv preprint arXiv:2403.11310.

[10] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.

[11] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI blog, 1(8), 9.

[12] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877-1901.

[13] Popokh, L., Su, J., Nair, S., & Olinick, E. (2021, September). IllumiCore: Optimization Modeling and Implementation for Efficient VNF Placement. In 2021 International Conference on Software, Telecommunications and Computer Networks (SoftCOM) (pp. 1-7). IEEE.

[14] Liu, T., Xu, C., Qiao, Y., Jiang, C., & Chen, W. (2024). News recommendation with attention mechanism. arXiv preprint arXiv:2402.07422.

[15] Bao, W., Che, H., & Zhang, J. (2020, December). Will_Go at SemEval-2020 Task 3: An accurate model for predicting the (graded) effect of context in word similarity based on BERT. In Proceedings of the Fourteenth Workshop on Semantic Evaluation (pp. 301-306).

[16] Zhou, Z., Xu, C., Qiao, Y., Ni, F., & Xiong, J. (2024). An Analysis of the Application of Machine Learning in Network Security. Journal of Industrial Engineering and Applied Science, 2(2), 5-12.

[17] Su, J., Nair, S., & Popokh, L. (2023, February). EdgeGYM: a reinforcement learning environment for constraint-aware NFV resource allocation. In 2023 IEEE 2nd International Conference on AI in Cybersecurity (ICAIC) (pp. 1-7). IEEE.

[18] Luo, Y., Wei, Z., Xu, G., Li, Z., Xie, Y., & Yin, Y. (2024). Enhancing E-commerce Chatbots with Falcon-7B and 16-bit Full Quantization. Journal of Theory and Practice of Engineering Science, 4(02), 52-57.

[19] Xu, C., Yu, J., Chen, W., & Xiong, J. (2024, January). Deep learning in photovoltaic power generation forecasting: Cnn-lstm hybrid neural network exploration and research. In The 3rd International scientific and practical conference “Technologies in education in schools and universities” (January 23-26, 2024) Athens, Greece. International Science Group. 2024. 363 p. (p. 295).

[20] Liu, T., Cai, Q., Xu, C., Zhou, Z., Xiong, J., Qiao, Y., & Yang, T. (2024). Image Captioning in news report scenario. arXiv preprint arXiv:2403.16209.

[21] Xiong, J., Feng, M., Wang, X., Jiang, C., Zhang, N., & Zhao, Z. (2024). Decoding sentiments: Enhancing covid-19 tweet analysis through bert-rcnn fusion. Journal of Theory and Practice of Engineering Science, 4(01), 86-93.

[22] Su, J., Nair, S., & Popokh, L. (2022, November). Optimal resource allocation in sdn/nfv-enabled networks via deep reinforcement learning. In 2022 IEEE Ninth International Conference on Communications and Networking (ComNet) (pp. 1-7). IEEE.

[23] Xie, Y., Li, Z., Yin, Y., Wei, Z., Xu, G., & Luo, Y. (2024). Advancing Legal Citation Text Classification A Conv1D-Based Approach for Multi-Class Classification. Journal of Theory and Practice of Engineering Science, 4(02), 15-22.

[24] Liu, T., Xu, C., Qiao, Y., Jiang, C., & Yu, J. (2024). Particle Filter SLAM for Vehicle Localization. arXiv preprint arXiv:2402.07429.

[25] Wang, X., Qiao, Y., Xiong, J., Zhao, Z., Zhang, N., Feng, M., & Jiang, C. (2024). Advanced Network Intrusion Detection with TabTransformer. Journal of Theory and Practice of Engineering Science, 4(03), 191-198.

[26] Qiao, Y., Ni, F., Xia, T., Chen, W., & Xiong, J. (2024, January). Automatic recognition of static phenomena in retouched images: A novel approach. In The 1st International scientific and practical conference “Advanced technologies for the implementation of new ideas” (January 09-12, 2024) Brussels, Belgium. International Science Group. 2024. 349 p. (p. 287).

[27] Xu, C., Qiao, Y., Zhou, Z., Ni, F., & Xiong, J. (2024). Accelerating Semi-Asynchronous Federated Learning. arXiv preprint arXiv:2402.10991.

[28] Liu, S., Wu, K., Jiang, C., Huang, B., & Ma, D. (2023). Financial time-series forecasting: Towards synergizing performance and interpretability within a hybrid machine learning approach. arXiv preprint arXiv:2401.00534.

[29] Zhao, Z., Zhang, N., Xiong, J., Feng, M., Jiang, C., & Wang, X. (2024). Enhancing E-commerce Recommendations: Unveiling Insights from Customer Reviews with BERTFusionDNN. Journal of Theory and Practice of Engineering Science, 4(02), 38-44.

[30] Liu, T., Cai, Q., Xu, C., Zhou, Z., Xiong, J., Qiao, Y., & Yang, T. (2024). Image Captioning in news report scenario. arXiv preprint arXiv:2403.16209.

[31] Zhou, Z., Xu, C., Qiao, Y., Xiong, J., & Yu, J. (2024). Enhancing Equipment Health Prediction with Enhanced SMOTE-KNN. Journal of Industrial Engineering and Applied Science, 2(2), 13-20.

[32] Yin, Y., Xu, G., Xie, Y., Luo, Y., Wei, Z., & Li, Z. (2024). Utilizing Deep Learning for Crystal System Classification in Lithium-Ion Batteries. Journal of Theory and Practice of Engineering Science, 4(03), 199-206.