Enhanced Cellular Detection Using Convolutional Neural Networks and Sliding Window Super-Resolution Inference

Iván García-Aguilar[1,2], Rostyslav Zavoiko, Jose David Fernández-Rodríguez[1,2], Rafael Marcos Luque-Baena[1,2], and Ezequiel López-Rubio[1,2]

[1] ITIS Software. University of Málaga. C/ Arquitecto Francisco Peñalosa, 18. 29010 Málaga, Spain.
[2] Biomedical Research Institute of Málaga (IBIMA), C/ Doctor Miguel Díaz Recio, 28, Málaga, 29010, Spain.

Abstract. Histopathology currently serves as the standard for breast cancer diagnosis, but its manual execution demands time and expertise from pathologists. Artificial intelligence, particularly in digital pathology, has made significant strides, offering new opportunities for precision and efficiency in disease diagnosis. This study presents a methodology to enhance cell nuclei detection in breast cancer histopathological images using convolutional neural network models to apply super-resolution and object detection. Several model architectures are explored, and their performance is evaluated regarding accuracy and sensitivity. The results affirm the potential of the proposed approach for automated cell nuclei identification. These AI advancements in digital pathology open avenues for early and precise cancer detection, influencing clinical practices and patient well-being and improving diagnostic efficiency.

Keywords: Convolutional Neural Networks · Super-resolution · Object Detection · Nuclei · Cancer Diagnosis.

1 Introduction

Histopathology examines disease presence, extent, and progression through microscopic tissue analysis. Thin sections are obtained using a microtome after fixing the sample in formaldehyde and embedding it in paraffin or resin. These cells are then stained to highlight cellular components before microscopic observation. In clinical medicine, a pathologist analyzes histological sections to formulate a pathology report, which is crucial for cancer diagnosis and treatment decisions. However, the reliance on visual analysis, particularly in specific areas of interest, can lead to errors in nucleus counting and classification. This may result in a false diagnosis or misjudgment, which presents a risk to patients by potentially overlooking harmful cells and compromising treatment efficacy. Nowadays, the increase in data generation and collection and the improvement in computational capacity have led to the application of Deep Learning in many fields such as video surveillance [1–4] or autonomous drive [5, 6], improving computer vision compared to classical techniques.

In cancer cell detection, the application of neural networks has emerged as an essential tool to improve accuracy and efficiency in the early diagnosis and monitoring of cancer progression. The collection and use of medical images, such as those obtained by microscopy, have enabled applying techniques based on deep learning to get an early and effective diagnosis. Although conventional methods such as microscopy and biochemical tests have been instrumental in detecting cancer cells, they have significant limitations in accuracy and efficiency, especially for those with biological variability. Recent research has demonstrated significant advances. Hou et al. propose a

sparse Convolutional Autoencoder [7], which efficiently detects and encodes nuclei in histopathology tissue images. Other works, such as NucDETR [8], introduce a transformer-based approach for automated nucleus detection in histopathology images, addressing the challenges of manual detection. This leverages the success of transformer models in object detection tasks and offers simplicity and reduced post-processing requirements. Tian et al. [9] present a novel semi-supervised learning framework addressing the challenge of acquiring extensive labeled data. This approach utilizes unlabeled images for reconstruction, enhancing the detection network by promoting spatial consistency between original and reconstructed images. Additionally, [10] proposes an end-to-end graph-based nuclei feature alignment (GNFA) method to enhance cross-domain nuclei detection in histopathology images. This method addresses challenges such as limited nuclei features and the presence of background pixels by leveraging a nuclei graph convolutional network (NGCN) to generate sufficient discriminative features for successful alignment. Despite the advances, this field still has outstanding challenges, such as needing more annotated data or identifying small elements.

Compared to other emerging methods, our approach proposes a solution based on the re-inference of super-resolved areas to improve detection without re-training or modifying the model. The term super-resolved areas refers to regions of the image that undergo super-resolution techniques, resulting in greater clarity and detail than the original resolution. In the framework of our proposal, these enhanced areas play a crucial role in object detection by providing richer and more detailed information. Importantly, super-resolution improves the visual quality of the relevant areas and contributes significantly to the accuracy of the detection process. Allowing the model to benefit from more detailed information without re-training avoids this procedure's computational and time costs. Furthermore, by not modifying the original model architecture, the integrity and consistency of its performance are maintained, which is particularly valuable in environments where stability and efficiency are paramount.

The following sections are structured as follows: section 2 shows the proposed methodology, section 3 on page 4 explains the experiments supporting our proposal, and section 4 on page 9 explains our conclusions.

2 Methodology

The proposed methodology for enhancing image detection is shown in Figure 1. The approach involves a Deep Learning neural network designed for object detection, denoted as $\mathcal{M}$ to process an input image $\mathbf{X}_{LR}$ to produce a set of detections $\mathcal{S}$.

$$S = \mathcal{M}\left(\mathbf{X}_{LR}\right) \tag{1}$$

$$S = \{(\alpha_i, \beta_i, \gamma_i, \delta_i, \epsilon_i, \omega_i) \mid i \in \{1, ..., N\}\} \tag{2}$$

Here, N represents the number of detections, $(\alpha_i, \beta_i) \in \mathbb{R}^2$ denotes the coordinates of the upper left corner of the i^{th} detection within the input $\mathbf{X}$, $(\gamma_i, \delta_i) \in \mathbb{R}^2$ represents the coordinates of the lower right corner of the i^{th} detection within $\mathbf{X}$, ϵ_i is the class label of the detection, and $\omega_i \in \mathbb{R}$ signifies the class score of the detection. For this proposal, consideration of the $epsilon_i$ class label of the detection will be omitted. The main goal is to increase the total number of detections, regardless of their specific class. This strategic decision is driven by the objective of increasing the exhaustiveness of the detection model. In this context, specific focus is placed on detections characterized by a class score $\omega_i \in \mathbb{R}$ that exceeds a predefined threshold set at 50%. This threshold

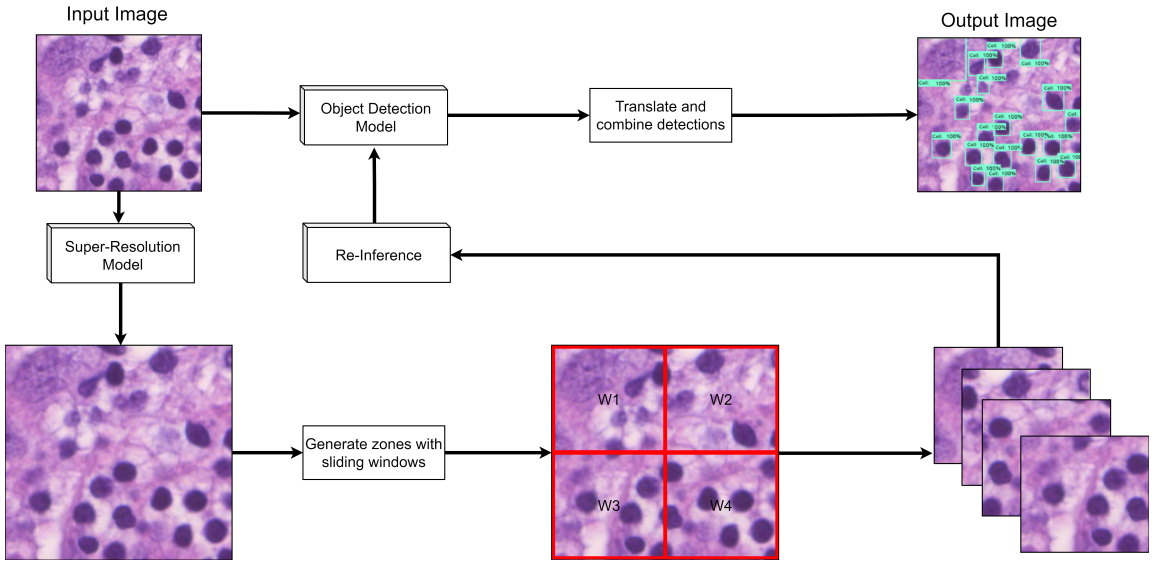


Fig. 1: Workflow of the proposed technique.

serves as a confidence criterion, ensuring that only detections with a class score above 50% are valid.

The second step is to process the input image $\mathbf{X}_{LR}$, which has a low resolution, with a convolutional neural network $\mathcal{F}$ to enhance the resolution. The model generates a high-resolution image, denoted as $\mathbf{X}_{HR}$, through a specific upscaling factor (e.g., upscaling factor = 2).

$$\mathbf{X}_{HR} = \mathcal{F}\left(\mathbf{X}_{LR}\right) \tag{3}$$

Then, a sliding windows approach is applied to the super-resolved image X_{HR}, incorporating user-defined displacement parameters d:

$$x_{\text{windows}} = i \times d \qquad y_{\text{windows}} = j \times d \tag{4}$$

Where the coordinates of the sliding window are represented by $(x_{\text{windows}}, y_{\text{windows}})$. The indices representing the window's position are i and j. For each window $\mathcal{I}_{\text{window}}$, perform re-inference using the object detection model to obtain an additional set of bounding boxes denoted as S'.

$$S' = \left\{ \left(\tilde{\alpha}_i, \tilde{\beta}_i, \tilde{\gamma}_i, \tilde{\delta}_i, \epsilon_i, \omega_i \right) \mid i \in \{1, ..., I\} \right\} \tag{5}$$

Where I represents every super-resolved window extracted from X_{HR}. After the re-inference process, coordinates $(\tilde{\alpha}_i, \tilde{\beta}_i, \tilde{\gamma}_i, \tilde{\delta}_i)$ are obtained in the super-resolved image. It is necessary to undo the up-scaling and the displacement to translate these coordinates to the original image.

$$x_{\text{original}} = \frac{x'}{f} + x_{\text{window}}, \quad y_{\text{original}} = \frac{y'}{f} + y_{\text{window}} \tag{6}$$

Detections belonging to the same entity are grouped using the Intersection over Union (IOU). The formula for IOU between two bounding boxes is:

$$IOU(A, B) = \frac{\text{Area of Intersection}}{\text{Area of Union}} \tag{7}$$

Bounding boxes with an Intersection over Union (IOU) greater than a predefined threshold are grouped. After grouping coincident detections of the same element, selecting the most accurate detection becomes necessary. A weighting is applied to the detections obtained within the generated sliding windows to facilitate this. The decision to assign five times more importance to bounding boxes obtained in the initial inference stage on the normal-resolution image without super-resolution is motivated by the need to handle scenarios where, despite accurate detection of a cell in the first stage, re-inference through sliding windows may only yield partial detections of the same cell.

$$w_{\text{initial}} = \text{selected weigth} \times w_{\text{initial reinference}} \tag{8}$$

$$w_{\text{window}} = 1 \times w_{\text{initial reinference}} \tag{9}$$

The reason for this weighting stems from recognizing that the initial inference, conducted on the original low-resolution image, holds paramount significance. It should exert a more substantial influence in determining the precise position and shape of the cell. Re-inference, occurring after super-resolution and sliding window scanning, may face limitations due to potential partial cell detections.

After clustering bounding boxes for the same element, the average of all established bounding boxes is determined. Subsequently, the Euclidean distance between the calculated average and each bounding box is computed:

$$D(\text{bbox}_{\text{average}}, \text{bbox}_i) = \sqrt{\sum_{j=1}^{n} (\text{bbox}_{\text{average}}^{(j)} - \text{bbox}_i^{(j)})^2} \tag{10}$$

Where n represents the dimensionality of the bounding boxes. Following this, the index of the bounding box with the smallest distance to the average is identified:

$$\text{index}_{\min} = \arg\min_i \left(\sum_{j=1}^{n} (\text{bbox}_{\text{average}}^{(j)} - \text{bbox}_i^{(j)})^2 \right) \tag{11}$$

Finally, the closest bounding box to the calculated average is obtained as:

$$\text{closest}_{\text{bbox}} = \text{bbox}_{\text{index}_{\min}} \tag{12}$$

A unified class label is obtained by majority voting among the class labels in the clustering operation for each element, and a unified class score is computed as the maximum of the class scores.

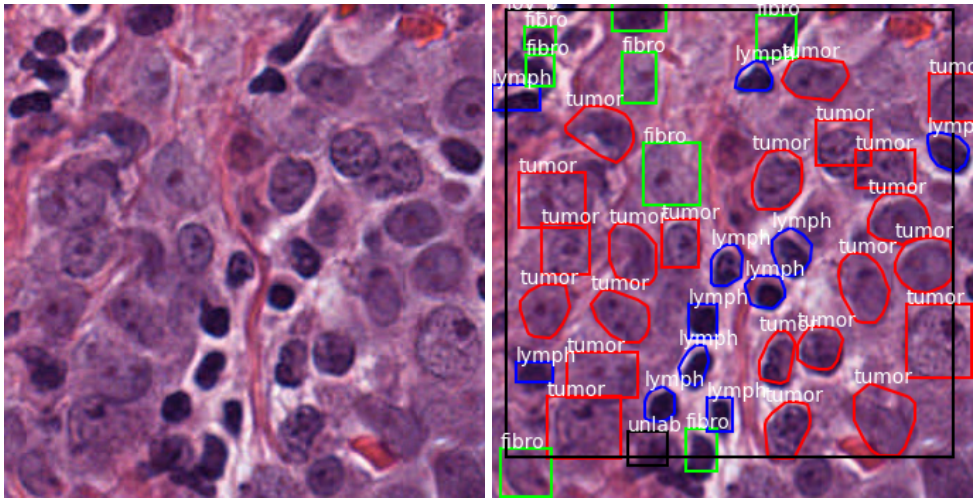
3 Experiments

This section is composed of a series of subsections. First, the dataset used is described to provide a context for the type of elements to be identified and to discuss the attributes of the data. Subsequently, several models focused on object detection are stated, which have been subsequently

evaluated with well-defined metrics. Finally, the results are provided, comprising a comparative analysis between the raw object detection models and the results obtained after applying the described methodology.

3.1 Dataset

In the field of breast pathology, a series of datasets have been published that open the possibility of applying algorithms and deep learning models. In this case, a group of datasets called NuCLS [11] has been chosen. Non-pathological individuals (NPs) were employed to generate these datasets, following two strategies to acquire nuclei labels.



(a) Original Image (b) Annotated Image

Fig. 2: Example of the annotated images that compose the NuCLS Dataset.

The first strategy (single-rater) aimed for broad data collection across numerous fields of view, with NPs receiving feedback and coordinators standardizing annotations. The second strategy (multi-rater) measured reliability using annotations from multiple NPs on a smaller set of shared fields alongside pathologist annotations for comparison. The project seeks to obtain a considerable amount of high-quality data, which is why the reviewed single-rater dataset was used to fine-tune object detection models in this project. The NuCLS dataset contains breast cancer slides from 125 patients obtained from TCGA, a US National Cancer Institute project. The NuCLS dataset creators concentrated on triple-negative breast cancer images, marked by the absence of estrogen, progesterone receptors, and HER2 protein. This aggressive cancer type exhibits poorer survival rates. The complete dataset consists of 1744 images. An advantageous characteristic of this dataset is the substantial variability in cell types and patterns, enabling the training and evaluation of pathology-based object detection models using Convolutional Neural Networks. The objective of this study is to detect the maximum number of cells possible, making the specific cell types (classes) irrelevant to this proposal. Figure 2 shows an example of a labeled frame.

3.2 Super-resolution Model

The use of super-resolution involves the application of advanced models and techniques to increase the number of pixels that compose the input. A series of models are available for super-resolution applications such as [12, 13]. One of the most effective models is FastSRCNN (Fast Super-Resolution Convolutional Neural Network) [14]. This convolutional neural network was designed to address the super-resolution task efficiently and quickly. It has many advantages, such as preserving fine details during the resolution enhancement process. Compared to other models, it is particularly fast, making it suitable for applications that require results in a short time. It is important to note that the proposed methodology allows flexibility in switching to other super-resolution techniques according to the specific needs of the problem to be solved, thus ensuring the method's adaptability.

3.3 Object Detection Models

Object detection models capable of identifying cell nuclei are required to apply the presented methodology. The dataset was divided into training, validation, and test subsets with proportions of 70%, 20%, and 10%, respectively. Following data splitting, fine-tuning was performed on various object detection models from the Tensorflow Object Detection Model Zoo[3]. The following models were selected:

- **SSD MobileNetV2 FPNLite** [15]: SSD (Single Shot Detector) is a single-shot object detection model using MobileNetV2 for feature extraction and lightweight FPNLite.
- **EfficientDet D0 and D2** [16]: High-performance models with BiFPN architecture, utilizing EfficientNet for scalability. D0 has a 3-layer, 64-filter BiFPN, and D2 features a scaled EfficientNet and a 5-layer, 112-filter BiFPN.
- **CenterNet Hourglass 104 512x512** [17]: CenterNet adopts a unique keypoint network to detect central points, leveraging an Hourglass backbone for deep feature extraction from stacked convolutional and max-pooling modules.

The choice of these models is based on the diversity of architectures to determine the improvement after applying the proposed methodology. The evaluation of SSD addresses the need for efficient single-stage models, while EfficientDet D0 and D2 represent variants of an efficient and scalable family. On the other hand, CenterNet provides a unique perspective. The generic design allows versatile adaptation to explore and evaluate new architectures.

3.4 Results

Metrics like Mean Average Precision (MAP) were utilized to assess the proposed approach. The COCO evaluator[4] was employed for mAP, considering multiple confidence thresholds across IoU values from 0.5 to 0.95. This range enables a detailed evaluation of model performance regarding overlap with ground truth bounding boxes. The results are shown from the exhaustive evaluation of several object detection models, both in their original state (RAW) and after applying the proposed methodology (OURS). The main metric used to quantify the performance of the models was the Mean Average Precision (mAP), considering different Intersection over Union (IoU) configurations.

[3] https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tf2_detection_zoo.md
[4] `https://github.com/cocodataset/cocoapi/blob/master/PythonAPI/pycocotools/cocoeval.py`

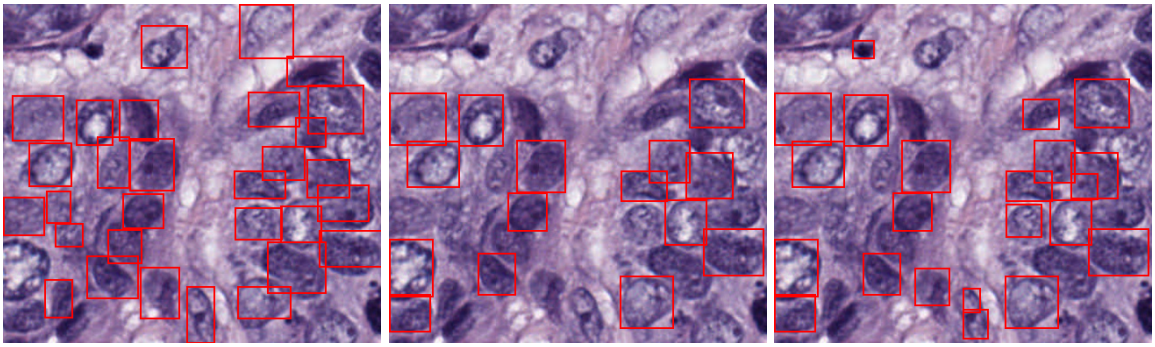
Model	Method	Area All IoU 0.50:0.95	Area All IoU 0.50	Area All IoU 0.75	Area Small IoU 0.50:0.95	Area Medium IoU 0.50:0.95
EfficientDet D0	RAW	0.244	0.462	**0.226**	0.207	0.280
	OURS	**0.253**	**0.502**	0.216	**0.213**	**0.291**
EfficientDet D2	RAW	0.195	0.356	0.196	0.175	0.212
	OURS	**0.239**	**0.459**	**0.221**	**0.235**	**0.242**
SSD MobileNetV2	RAW	0.262	0.502	0.241	0.258	0.271
	OURS	**0.296**	**0.593**	**0.261**	**0.266**	**0.319**
CenterNet Hourglass 104	RAW	0.097	0.161	**0.103**	0.078	0.114
	OURS	**0.098**	**0.171**	0.102	**0.078**	**0.117**

Table 1: Results obtained after using different object detection models comparing the Raw model and the presented methodology. The best results are marked in **bold**.

The obtained analysis reveals overall improvements using the presented methodology. A sustained elevation in Mean Average Precision (mAP) is discernible across various Intersection over Union (IoU) configurations, underscoring the effectiveness of this approach in different object detection scenarios. Notably, deploying the methodology generates overall performance improvements and demonstrates a uniformly positive impact regardless of the base model chosen. This finding underscores the overall robustness and applicability of the proposed method, suggesting that its implementation can consistently produce positive results across a spectrum of object detection environments. Table 1 compares the mean average precision (mAP). The results obtained by each proposal are shown.

Beyond the overall performance improvement, the performance optimization of our methodology is particularly striking in the total area under different IoU values, where a more pronounced increase in mAP is observed. This specific result underlines the methodology's effectiveness in scenarios requiring a high degree of overlap between predictions and actual annotations.

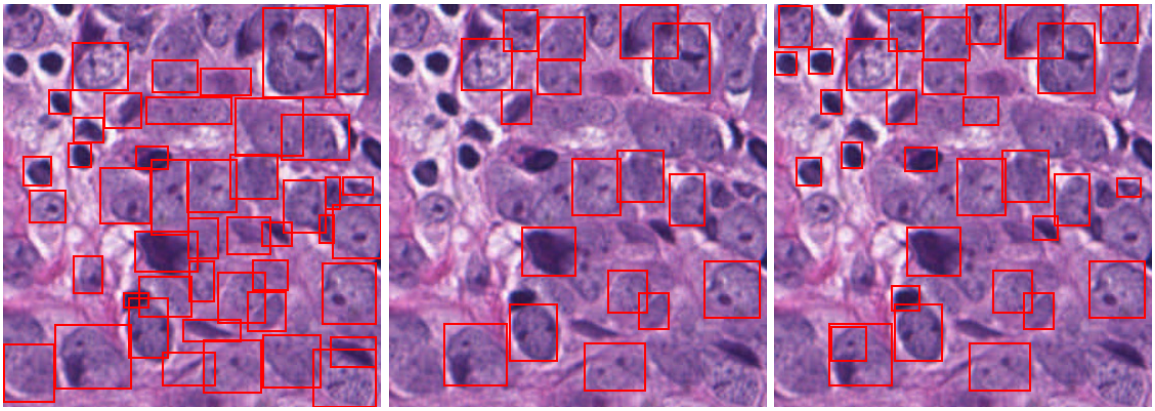
While the RAW model occasionally exhibits a marginally higher Mean Average Precision (MAP), providing quantitative insights into the improvement is crucial. Our methodology, labeled as OURS, consistently outperforms the RAW model across various evaluation metrics. For example, considering the Area All with IoU 0.50:0.95, OURS achieves a notable improvement, with an increase of approximately 3.5% compared to RAW. This emphasizes the effectiveness of our methodology in enhancing object detection accuracy in scenarios requiring a high degree of overlap. Examining specific areas, such as Area Small and Medium with IoU 0.50:0.95, OURS demonstrates significant advancements. Area Small has an improvement of around 3.1%, while Area Medium shows an enhancement of approximately 4.9%. These results underscore the ability of OURS to address challenges associated with object detection in contexts with varying dimensions, especially in smaller and medium-sized regions. Table 1 presents a comprehensive overview of the quantitative results obtained from different object detection models, comparing the RAW model with OURS. The values in **bold** highlight the superior performance of OURS across diverse model configurations and evaluation criteria.



(a) Ground Truth (b) Single-model detections (c) Enhanced detections

Fig. 3: Comparison of the detections obtained. The left-hand side is the Ground Truth, followed by the Raw model, and on the right-hand side is the presented methodology using the EfficientDet D2 model.

Figures 3 and 4 provide a visual comparison of the detections obtained by applying the model in its original state (RAW) and the proposed methodology (OURS). These figures highlight how our methodology's application has significantly influenced the cell detection process.



(a) Ground Truth (b) Single-model detections (c) Enhanced detections

Fig. 4: Comparison of the detections obtained. The left-hand side is the Ground Truth, followed by the Raw model, and on the right-hand side is the presented methodology using the SSD model.

As seen in the images, the number of detections increases when applying the OURS methodology compared to the RAW model. This observation underlines our approach's effectiveness in improving the system's ability to accurately identify and delimit the cells present in the analyzed images.

4 Conclusions and Future lines

In conclusion, the results obtained in this study support the effectiveness of the presented methodology in improving the performance of object detection models compared to their original versions. The consistency in the observed improvements, regardless of the base model and variations in IoU configurations, highlights our approach's overall relevance and applicability in computer vision. In a quantitative analysis, the positive impact of the methodology is especially highlighted by observing the significant increase in Mean Average Precision (mAP), being most notably evident in the case of SSD MobileNet V2, where the best result was achieved with a mAP of 0.296 in the total area with IoU of 0.50:0.95. As illustrated in Figures 1 and 2, visual comparisons further support the quantitative improvement, showing a substantial increase in the number of detections.

Considering the achievements obtained, several lines of research are open for future development. One promising direction is integrating more advanced deep learning techniques and considering emerging model architectures, which could contribute to the evolution and continuous improvement of the proposed methodology. In addition, incorporating specialized data augmentation techniques could enhance the methodology's ability to address specific challenges in complex environments.

Acknowledgements

This work is partially supported by the Ministry of Science and Innovation of Spain, grant number PID2022-136764OA-I00, project name Automated Detection of Non Lesional Focal Epilepsy by Probabilistic Diffusion Deep Neural Models. It is also partially supported by the University of Málaga (Spain) under grants B1-2021_20, project name Detection of coronary stenosis using deep learning applied to coronary angiography; B4-2022, project name Intelligent Clinical Decision Support System for Non-Obstructive Coronary Artery Disease in Coronarographies; B1-2022_14, project name Detección de trayectorias anómalas de vehículos en cámaras de tráfico; and, by the Fundación Unicaja under project PUNI-003_2023, project name Intelligent System to Help the Clinical Diagnosis of Non-Obstructive Coronary Artery Disease in Coronary Angiography. The authors thankfully acknowledge the computer resources, technical expertise and assistance provided by the SCBI (Supercomputing and Bioinformatics) center of the University of Málaga. They also gratefully acknowledge the support of NVIDIA Corporation with the donation of a RTX A6000 GPU with 48Gb. The authors also thankfully acknowledge the grant of the Universidad de Málaga and the Instituto de Investigación Biomédica de Málaga y Plataforma en Nanomedicina-IBIMA Plataforma BIONAND.

References

1. Iván García-Aguilar, Jorge García-González, Rafael Marcos Luque-Baena, and Ezequiel López-Rubio. Object detection in traffic videos: an optimized approach using super-resolution and maximal clique algorithm. *Neural Computing and Applications*, 35(26):18999–19013, June 2023.
2. S. Ushasukhanya, T. Y. J. Naga Malleswari, M. Karthikeyan, and C. Jayavarthini. An intelligent deep learning based capsule network model for human detection in indoor surveillance videos. *Soft Computing*, 28(1):737–747, December 2023.
3. Iván García-Aguilar, Jorge García-González, Rafael Marcos Luque-Baena, and Ezequiel López-Rubio. Automated labeling of training data for improved object detection in traffic videos by fine-tuned deep convolutional neural networks. *Pattern Recognition Letters*, 167:45–52, 2023.

4. Iván García-Aguilar, Rafael Marcos Luque-Baena, and Ezequiel López-Rubio. Improved detection of small objects in road network sequences using cnn and super resolution. *Expert Systems*, 39(2):e12930, 2022.
5. Iván García-Aguilar, Jorge García-González, Daniel Medina, Rafael Marcos Luque-Baena, Enrique Domínguez, and Ezequiel López-Rubio. Detection of dangerously approaching vehicles over onboard cameras by speed estimation from apparent size. *Neurocomputing*, 567:127057, 2024.
6. Sajjad Ahmad Khan, Hyun Jun Lee, and Huhnkuk Lim. Enhancing object detection in self-driving cars using a hybrid approach. *Electronics*, 12(13), 2023.
7. Le Hou, Vu Nguyen, Ariel B. Kanevsky, Dimitris Samaras, Tahsin M. Kurc, Tianhao Zhao, Rajarsi R. Gupta, Yi Gao, Wenjin Chen, David Foran, and Joel H. Saltz. Sparse autoencoder for unsupervised nucleus detection and representation in histopathology images. *Pattern Recognition*, 86:188–200, 2019.
8. Ahmad Obeid, Taslim Mahbub, Sajid Javed, Jorge Dias, and Naoufel Werghi. Nucdetr: End-to-end transformer for nucleus detection in histopathology images. In Wenjian Qin, Nazar Zaki, Fa Zhang, Jia Wu, and Fan Yang, editors, *Computational Mathematics Modeling in Cancer Analysis*, pages 47–57, Cham, 2022. Springer Nature Switzerland.
9. Chenchen Tian, Lei Su, Zhi Wang, Ao Li, and Minghui Wang. Semi-supervised nuclei detection in histopathology images via location-aware adversarial image reconstruction. *IEEE Access*, 10:42739–42749, 2022.
10. Zhi Wang, Kai Fan, Xiaoya Zhu, Honglei Liu, Gang Meng, Minghui Wang, and Ao Li. Cross-domain nuclei detection in histopathology images using graph-based nuclei feature alignment. *IEEE Journal of Biomedical and Health Informatics*, 28(1):78–88, 2024.
11. Mohamed Amgad, Lamees A Atteya, Hagar Hussein, Kareem Hosny Mohammed, Ehab Hafiz, Maha A T Elsebaie, Ahmed M Alhusseiny, Mohamed Atef AlMoslemany, Abdelmagid M Elmatboly, Philip A Pappalardo, Rokia Adel Sakr, Pooya Mobadersany, Ahmad Rachid, Anas M Saad, Ahmad M Alkashash, Inas A Ruhban, Anas Alrefai, Nada M Elgazar, Ali Abdulkarim, Abo-Alela Farag, Amira Etman, Ahmed G Elsaeed, Yahya Alagha, Yomna A Amer, Ahmed M Raslan, Menatalla K Nadim, Mai A T Elsebaie, Ahmed Ayad, Liza E Hanna, Ahmed Gadallah, Mohamed Elkady, Bradley Drumheller, David Jaye, David Manthey, David A Gutman, Habiba Elfandy, and Lee A D Cooper. NuCLS: A scalable crowdsourcing approach and dataset for nucleus classification and segmentation in breast cancer. *GigaScience*, 11, 05 2022. giac037.
12. Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5835–5843, 2017.
13. Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks, 2016.
14. Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 391–407, Cham, 2016. Springer International Publishing.
15. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. *SSD: Single Shot MultiBox Detector*, page 21–37. Springer International Publishing, 2016.
16. Mingxing Tan, Ruoming Pang, and Quoc V. Le. Efficientdet: Scalable and efficient object detection, 2020.
17. Kaiwen Duan, Song Bai, Lingxi Xie, Honggang Qi, Qingming Huang, and Qi Tian. Centernet: Keypoint triplets for object detection, 2019.