2024 9th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM)

2024 9th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM) | 979-8-3315-0448-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/SEEDA-CECNSM63478.2024.00039

Yolov9 models for cancer detection in biopsy images: segmentation on pan-cancer data

Panagiotis N. Smyrlis, Aimilia Ntetska, Thomai Karamitsou, Foteini Dimaraki,
Katerina D. Tzimourta, Pantelis Angelidis, Markos G. Tsipouras
Laboratory of Biomedical Technology and Digital Health
School of Electrical and Computer Engineering
University of Western Macedonia
Kozani, Greece
pan.smyrlis@gmail.com, antetska@uowm.gr, tkaramitsou@uowm.gr, fodim96@gmail.com,
ktzimourta@uowm.gr, paggelidis@uowm.gr, mtsipouras@uowm.gr

***Abstract*—This work examines the use of recent Yolo models on segmenting cancer biopsy images to highlight the tissue classes of interest. The experiments include recent YoloV9 and V8 architectures, trained on a large pan-cancer dataset, which contains examples for 19 distinct cancer cases. The models' predictive capability was measured and comparatively studied, and detailed Panoptic Quality and precision metrics are discussed. The results obtained indicate the fine performance of recent Yolo models on cancer segmentation task, while all version 9 algorithms did surpass preceeding architectures. This study also examined the effect of the Yolo framework data augmentation methods for performance enhancement, by conducting experiments for two image sizes, 256 (original size) and 512 pixels.**

Keywords—Biopsy, Cancer, Convolutional Neural Network, YOLO-V9, Segmentation, Medical Imaging

I. Introduction

Cancer diagnosis is yet performed with considering varying imaging modalities, namely CT, ultrasound, biopsy, magnetic resonance images. The various exam evaluation, facilitates accurate diagnosis and enstaging estimation, and contributes to patients' condition monitoring for cancer treatment.

Sophisticated machine learning and Artificial Intelligence algorithms are nowadays tested to decision support systems for accurate cancer detection in medical images. To train and evaluate these algorithms, various datasets including digital images with tissue samples are presented in literature [1]. They include precisely labelled images for a variety of cases, in order to achieve the accurate identification and recognition of cancer cells, while initially, specialized datasets that focus on distinct cancer type were studied, including data such as mammograms or lymph node images, focusing on specific cancers type detection.

For instance, a wide range of primary tumors with varying cellular differentiation can develop from the aorta. Due to their rare occurrence, complex clinical and imaging features that often mimic thrombosis, it is a challenge to precisely distinguish a primary tumor to a large vessel [2].

Furthermore, wider and more complex data collections are seen in literature [3], to facilitate generic pan-cancer detection efforts. Such data collections provide a comprehensive benchmark to developed models and a wide example set for cancer detection applications. They are thoroughly descriptive of real-world conditions and shall facilitate robust nuclei segmentation and cancer detection in the wild, as they include a wide range of tissue instances and conditions, and may improve precise diagnosis efforts.

Automated cancer detection in biopsy images includes image segmentation efforts [4], to precisely detect and delineate the maligant tissue, and to find the neural network architecture and parameters to optimally model the data of interest. Until today, the researchers have tried to employ various Deep Learning segmentation algorithms, including state-of-the-art architectures like YOLO[5], Mask-RCNN[6], or other application-wise and custom made works[7].

Works in literature have focused on specific cancer types[8], and earlier efforts have tried to work with multi-tissue data, presenting sophisticated architectures. Pan-cancer data, like Pannuke nowadays allow generalized modelling on cancer delineation, with sophisticated taxonomy for them. In this paper, the authors use the Pannuke pan-cancer dataset to train various YOLO flavors of versions 8 and 9[9].

The authors discuss the use of these algorithms for cancerous / maligant tissue segmentation in microscopy biopsy images, and reside detailed metrics and results. YOLO is nowadays a rapidly evolving algorithm, which yet has been widely adopted by the research community, as having presented state-of-the-art results to common object recognition tasks and applications. Major benefits to consider using YOLO is the small computational overhead needed to train and predict, as also the minimal size that tries to preserve. This work examines all the basic v8 flavors, as well as the v9 flavors v9-c, and GELAN-C, which is a part of v9 algorithm and the authors distribute it as of the YOLOv9 release.

With respect to the hyperparameters set, which we tried to minimally alter from the original paper proposals, we discuss some pros and cons of these models, providing various precision tests for them with respect to their size. We examine their usage on biopsy image segmentation for cancer, and also we discuss the YOLOv9 advancements comparing to the preceeding v8 architecture. We consider that these same-parameter testing delineates a baseline capability measurement

979-8-3315-0448-9/24/$31.00 ©2024 IEEE
DOI 10.1109/SEEDA-CECNSM63478.2024.00039

for the architecture and different flavors potential, as it can be further optimized model-specifically and enhance the obtained results.

II. Materials and Methods

A. PanNuke Dataset

The Pannuke dataset is a biopsy samples' image collection, containing microscopy image examples for 19 distinct cancer types. It is considered a pan-cancer dataset, and it is distributed as 7901 image segments (patches) of 256x256 pixels. The original Whole-Slide Images were of x20 and x40 magnification and were all transformed into x40 scale by resampling. The dataset introduces a 5-class nuclei segmentation taxonomy to detect maligant cases.

This dataset may be used to train cancer detection and precise localization models (cancerous tissue segmentation) and the collection includes pixel-wise annotation for all patches, which was achieved using a semi-automatic procedure involving expert doctors. Specifically, Pannuke contains $\sim$190000 pixel-wise annotated nuclei examples whose distribution is not balanced for the classes of interest.

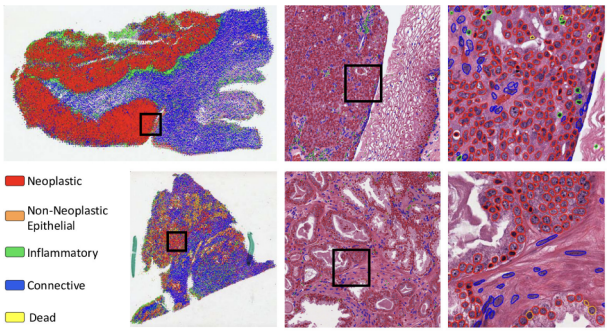

Fig. 1: Visualisation of Pannuke dataset, together with the respective class annotation highlight. The images was found in [3]

The authors propose a 5-class categorization scheme for cell segmentation: Neoplastic, Non-Neoplastic Epithelial, Inflammatory, Connective and Dead cells. The different tissue samples supported include: Adrenal Gland, Bile Duct, Bladder, Breast, Cervix, Colon, Esophagus, Head and Neck, Kidney, Liver, Lung, Ovarian, Pancreatic, Prostate, Skin, Thyroid, Uterus. The images were retrieved from The Cancer Genome Atlas [10], while some were given by a local hospital.

The annotation procedure involved pretrained convolutional neural networks to perform cell detections, which were then evaluated by pathologists. The pretraining was performed using 4 widely used datasets seen in literature. A relabelling step was included to end up in the proposed 5-class taxonomy. The method contained back-and-forth evaluation and retrain for accurate neural net predictions. Nuclick tool was used to produce nuclei pixel-wise masks.

B. The Yolo Framework

The Yolo algorithm state-of-the-art performance relies on the relevant framework ecosystem, which in every distribution includes the respective Deep Learning architecture, together with a set of data augmentation methods to enhance the generalized learning ability. The Deep Learning architecture is yet based on Convolutional Neural Networks.

Also, traditional image transformation functions, such as scaling and cropping are employed in order to additionally feed the training process. The recent framework includes HSV transformation, rotation, translation, perspective manipulation, scaling, shearing, flipping, mosaicing, image mixing (MixUp) and partial image mixing (CutMix) and partial image masking (CutOut).

C. The Yolov8 Model

The YOLOv8 architecture uses a Darknet based network (CSPDarknet53) as backbone, which employs 53 convolutional along pooling layers with cross-stage connections. A feature pyramid network is used to facilitate generalized detection of different object sizes, providing features on different predefined scales.

A number of detection heads make predictions for the available feature representations, while the predictions are filtered using Non-Maximum Suppression to end up with the optimal final prediction removing the overlapping detections. In order to perform the segmentation task, the YOLO head is modified to use segmentation heads, to end up with the final pixel-wise predictions.

D. The Yolov9 Model

The YOLOv9 is a novel algorithm approach to the YOLO series baseline concept, which examined a set of modular modifications to enhance the YOLO precise object recognition. The algorithm is proposed for a set of 5 model 'flavors', namely tiny (v9-t), small (v9-s), medium (v9-m), compact (v9-c), extended (v9-e), with the distinct models to differ to the number of trainable parameters. The authors aim to better exploit the raw train information to end up with better feature sets extracted.

In that work, the authors experimented and developed on two novel modules, Generalized ELAN and Programmable Gradient Information, which combination forms the V9 algorithm to optimize the object detection precision. The above novel concepts were thoroughly measured on the common object recognition task to large benchmark data.

On top of the ELAN[11] schema, the authors experiment using CSPNet[12] modules. Its layer-stacking architecture is extended using stronger computational blocks to formulate GELAN. Based on Yolov7[13] outline, with respective modifications for anchor-free functionality, the GELAN is made for gradient path planning training. Complementary, the PGI concept is based on the reversible functions' principle, when some data may be backwards reconstructed losslessly.

PGI employs a 3-branch architecture to optimally extract the relevant features and minimize the semantic loss. The main branch is split into the "Auxiliary Reversible Branch" and "Multi-level Auxiliary Information Branch" to achieve the main branches' optimized train. At testing, the strong main branch formulates the model to perform the predictions.

E. Experiment Setup

To setup the experiment, all yolo models were trained on the initial 3-fold dataset split, as distributed by the dataset authors. To assess cross validated metrics, the results of all splits were considered and Mean Average Precison detailed metrics were calculated, for each architecture. All models were trained setting identical parameterization, except for the image size and the model architecture. The batch size was set to 8, the training strategy followed the Yolo authors recommendations, same as the data augmentation methods and quantity.

Distinct models were trained of all the v8 version flavors, which trainable parameters significally differ, to test the model size impact to the final performance. Also the version 9 flavors v9-C and its module Gelan-c, which is embodied into v9-C architecture combined to the PGI. The model size for all of the experiment models (trainable parameters in Millions) are resided in Table I

TABLE I: Yolo Model Parameters (Millions)

Model	V8-N	V8-S	V8-M	V8-L	V8-X	GELAN-C	V9-C
Params (M)	3.4	11.8	27.3	46	71.80	25.3	27.9

III. Experimental Results

The detection and segmentation algorithm ability were tested measuring mAP50 and mAP50-95 and all models were trained for image sizes 256x256 and 512x512. The detailed results are resided in Table II. Our low resolution models (256 pixels) score from 29.4 to 32.2% for segmentation mAP50-95, while the 512 models from 33.81% to 35.93%. The results are analogous for detection mAP50-95, which ranges from 34.03% to 37.4% in low resolution, while in high resolution, it ranges from 36.6% to 38.7%.

As seen in Average Precision results, YOLOv9-C with image size of 512 did score highest to all object detection and segmentation metrics. The score is 35.93% for segmentation mAP50-95 and 38.7% for the detection mAP50-95. Thus, within the scope of this experiment it shows the best ability both to highlight the nuclei classes presence and precisely localize it pixel-wise. For each separate image resolution, the YOLOv9 models did surpass the v8 architectures and the trained model size does have impact to the model precision.

As observed, the nano models of v8 architecture, using a very small trainable parameters set, presented reduced capability to optimally model the data, when other v8 models did score closer to v8 peak performance. On the other hand, regarding the parameters' increase from v8-S to v8-X versions, we did not observe an analogous precision gain.

The v9 versions, GELAN-C and YOLOv9-C presented increased precision compared to all preceeding architectures. The GELAN schema ebodied in YOLOv9 resulted in increased precision metrics, while the v9-C model which includes it gave us the best scores when trained with 512x512 image resolution. All models when trained for 512 image size recorded a precision increase comparing to the their 256-size train, as a result of data augmentation mechanisms, considering that the Pannuke original image size is 266x256 pixels.

TABLE II: All models - Mean Average Precision results

Image size (px)	Model	mAP50 (box)	mAP50-95 (box)	mAP50 (seg)	mAP50-95 (seg)
256	**YOLOv8-N**	58.30	34.03	57.28	29.40
	YOLOv8-S	60.68	36.11	59.66	31.13
	YOLOv8-M	61.18	36.59	59.98	31.55
	YOLOv8-L	60.95	36.61	60.01	31.50
	YOLOv8-X	60.93	36.57	60.02	31.53
	GELAN-C	62.53	37.57	61.43	32.36
	YOLOv9-C	62.13	37.40	61.07	32.20
512	**YOLOv8-N**	60.56	36.60	59.79	33.81
	YOLOv8-S	62.01	37.78	61.36	35.00
	YOLOv8-M	62.63	38.29	62.07	35.48
	YOLOv8-L	61.82	37.60	61.28	34.92
	YOLOv8-X	61.21	36.98	60.68	34.41
	GELAN-C	63.00	38.67	62.37	35.87
	YOLOv9-C	**63.07**	**38.70**	**62.50**	**35.93**

In table III, we reside the Panoptic Quality metrics obtained, where we observe the per class Panoptic qualtiy, the mean Panoptic Quality and the Binary Panoptic Quality for all classes and tissues. For the Panoptic Quality testing, we additionally defined the classes' score threshold to 0.15 for all classes of interest. The v9 architectures did score the best Panoptic Quality metrics for all tests, too.

As seen, all architectures show significantly reduced capability to model the "Dead" class, as a consequence of the very small number of examples included in dataset. For this category, all models scored from 8.81% to 13.31% and the best model was Yolov8-M for image size of 512 pixel.

As of the other classes, the Neoplastic and Non-Neoplastic classes were more successfully modelled, scoring from 43.6% to 48.56% and from 38.49% to 46.56% respectively. The Inflammatory and Connective cells scored from 35.98% to 39.45% and from 33% to 37.45%. The mean Panoptic Quality top was by GELAN model, same as mean Binary Panoptic Quality (at resolution 256)

The 512-resolution v9 models did score better for the Neoplastic (v9-C), Non-Neoplastic (v9-C), mean Panoptic Quality (GELAN), while competed by the 256-resolution models for Inflammatory (GELAN), Connective (GELAN), mean Panoptic Quality (GELAN) and mean Binary Panoptic Quality (GELAN). The best models to all tests achieved a 3% to 5% gain to quality metrics compared to the minimal size YOLOv8-N model at 256 image size, which we consider to delineate the V8 version base capability on these data.

IV. Conclusion

For a comprehensive evaluation, we conducted a comparative analysis of all five architectures within the YOLOv8 model series (YOLOv8-N, YOLOv8-S, YOLOv8-M, YOLOv8-L, YOLOv8-X) alongside the GELAN-C model and the YOLOv9-C architecture. The models were trained on the Pannuke pan-cancer dataset, which encompasses samples from 19 distinct cancer types. Additionally, this study examined the impact of data augmentation techniques within the YOLO

TABLE III: Panoptic Quality results obtained for all models

Resolution	Model	Neoplastic	Inflammatory	Connective	Dead	Non Neoplastic	mean PQ	mean PQ bin
256	**YOLOv8-N**	43.60	35.98	33.00	8.81	38.49	36.80	47.24
	YOLOv8-S	46.10	37.80	35.23	10.59	41.19	39.11	49.62
	YOLOv8-M	46.59	37.61	36.55	10.89	43.34	39.80	50.85
	YOLOv8-L	47.21	38.59	37.20	10.28	44.22	40.72	51.59
	YOLOv8-X	47.51	38.11	36.90	11.21	44.23	40.40	51.16
	GELAN-C	48.13	**39.45**	**37.45**	11.40	44.43	**41.16**	**52.48**
	YOLOv9-C	48.13	38.46	36.82	11.93	44.66	40.66	51.93
512	**YOLOv8-N**	43.87	35.61	33.53	10.90	42.57	37.38	46.60
	YOLOv8-S	45.94	36.92	35.43	12.79	44.30	39.54	49.32
	YOLOv8-M	47.18	37.63	36.03	**13.31**	45.69	40.12	50.92
	YOLOv8-L	47.36	37.59	36.99	12.55	46.41	40.91	51.98
	YOLOv8-X	46.87	38.20	36.62	13.25	45.89	40.68	51.18
	GELAN-C	48.48	38.36	36.62	13.18	46.53	**41.16**	52.18
	YOLOv9-C	**48.56**	38.34	36.55	13.01	**46.56**	41.11	52.13

framework on performance enhancement, utilizing two image resolutions: 512 pixels and 256 pixels (original size). The metrics employed for model evaluation included Mean Average Precision at a 50 threshold (mAP50) and Mean Average Precision across 50-95 thresholds (mAP50-95) in both bounding box and segmentation tasks, as illustrated in Table II. Furthermore, Panoptic Quality (PQ) and Binary Panoptic Quality (bPQ) for each nuclei category within the Pannuke dataset are detailed in Table III.

According to the results in Table II, when evaluating Mean Average Precision at threshold 50 and across 50-95 thresholds, the GELAN-C architecture demonstrates superior overall performance for images of 256 pixels compared to the other architectures. Conversely, for 512-pixel images, the YOLOv9-C architecture exhibits the highest performance among the models. In Table III all average PQ results are presented, for all Pannuke dataset nuclei categories, including all YOLOv9 architectures and GELAN. The obtained results indicate that the proposed GELAN architecture outperforms all other approaches.

This study introduces an innovative approach to the challenge of nuclei recognition across a diverse microscope scans of biopsy images, encompassing samples from 19 different types of malignant tissues, with image resolutions of 256 and 512 pixels. We employ a novel and effective instance segmentation architecture inspired by the YOLOv9 framework. Through the utilization of this lightweight model, we have achieved high performance metrics to approach the existing state-of-the-art results in this field.

ACKNOWLEDGMENT

This research has been financed by the Operational Program "Flagship actions in interdisciplinary scientific fields with a special focus on the productive fabric", under the call Recovery & Resilience Facility - RRF (project code: TAEDR-0535983).

REFERENCES

[1] J. Gamper, N. Alemi Koohbanani, K. Benet, A. Khuram, and N. Rajpoot, "Pannuke: an open pan-cancer histology dataset for nuclei instance segmentation and classification," in *Digital Pathology: 15th European Congress, ECDP 2019, Warwick, UK, April 10–13, 2019, Proceedings 15*. Springer, 2019, pp. 11–19.

[2] C. S. Restrepo, S. L. Betancourt, S. Martinez-Jimenez, and F. R. Gutierrez, "Aortic tumors," in *Seminars in Ultrasound, CT and MRI*, vol. 33, no. 3. Elsevier, 2012, pp. 265–272.

[3] J. Gamper, N. A. Koohbanani, K. Benes, S. Graham, M. Jahanifar, S. A. Khurram, A. Azam, K. Hewitt, and N. Rajpoot, "Pannuke dataset extension, insights and baselines," *arXiv:2003.10778*, 2020.

[4] P. N. Smyrlis, K. Vogklis, O. Tsakai, A. Tzallas, N. Giannakeas, and M. G. Tsipouras, "A novel approach for cancer image segmentation," in *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2023, pp. 4951–4954.

[5] P. N. Smyrlis, O. Tsakai, K. Vogklis, N. Giannakeas, A. Tzallas, G. F. Fragulis, and M. G. Tsipouras, "A comparative testing of yolov8-based algorithm for cancer biopsy images: Case study on pannuke pan-cancer dataset," in *2023 8th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM)*. IEEE, 2023, pp. 1–5.

[6] H. Huang, X. Feng, J. Jiang, P. Chen, and S. Zhou, "Mask rcnn algorithm for nuclei detection on breast cancer histopathological images," *International Journal of Imaging Systems and Technology*, vol. 32, no. 1, pp. 209–217, 2022.

[7] S. Graham, Q. D. Vu, S. E. A. Raza, A. Azam, Y. W. Tsang, J. T. Kwak, and N. Rajpoot, "Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images," *Medical image analysis*, vol. 58, p. 101563, 2019.

[8] A. Kanavos, E. Kolovos, O. Papadimitriou, and M. Maragoudakis, "Breast cancer classification of histopathological images using deep convolutional neural networks," in *2022 7th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM)*. IEEE, 2022, pp. 1–6.

[9] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "Yolov9: Learning what you want to learn using programmable gradient information," *arXiv preprint arXiv:2402.13616*, 2024.

[10] K. Tomczak, P. Czerwińska, and M. Wiznerowicz, "Review the cancer genome atlas (tcga): an immeasurable source of knowledge," *Contemporary Oncology/Współczesna Onkologia*, vol. 2015, no. 1, pp. 68–77, 2015.

[11] C.-Y. Wang, H.-Y. M. Liao, and I.-H. Yeh, "Designing network design strategies through gradient path analysis," *arXiv preprint arXiv:2211.04800*, 2022.

[12] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "Cspnet: A new backbone that can enhance learning capability of cnn," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 390–391.

[13] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 7464–7475.