Hemalatha S, **International Journal of Advanced Trends in Computer Applications (IJATCA)**
Volume 10, Number 1, December - 2024, pp. 23-28
ISSN: 2395-3519

Polyps Detection in Endoscopy Live Feeds

Hemalatha S[1], Santhiya A K[2], Dharshanashri G[2], Prabin R[2], Rahul Shriram S G[2]
[1]HOD, Dept. of AIML
Sri Shakthi Institute of Engineering and Technology, Coimbatore
[2]Dept. of AIML
Sri Shakthi Institute of Engineering and Technology, Coimbatore

Abstract: *Endoscopy procedures are essential in detecting and preventing colorectal cancer, primarily through identifying and removing polyps. Despite their critical role, polyps are often missed during routine colonoscopies due to factors such as physician fatigue or suboptimal viewing angles. Automated, real-time computer-aided detection (CAD) systems offer a solution to enhance polyp detection rates and accuracy during endoscopic procedures. However, real-time implementation in live feeds presents unique challenges, requiring a balance between computational speed and detection precision to minimize disruption to clinical workflows. This study addresses these challenges by introducing an optimized detection framework specifically designed to detect polyps in real-time in live endoscopic feeds. Using advanced detection techniques tailored to the endoscopy setting, we achieve both high precision and speed, critical for supporting endoscopists in real-time. Our results highlight the potential of real-time polyp detection systems to aid early colorectal cancer prevention by enhancing polyp detection rates without impeding clinical performance.*

Keywords: Colorectal cancer prevention, polyp detection, real-time detection, endoscopy, clinical computer-aided detection systems, deep learning.

I. Introduction

Colorectal cancer (CRC) remains a leading cause of cancer-related deaths worldwide, with high mortality rates primarily due to late diagnosis. Polyps, or abnormal tissue growths within the gastrointestinal tract, serve as a precursor to CRC, making their detection and removal crucial in early-stage cancer prevention. Among various screening methods, colonoscopy is considered the most effective for polyp detection, as it allows direct visualization of the gastrointestinal tract. However, the success of colonoscopy relies heavily on the skill of the endoscopist, with studies reporting that up to 20% of polyps are missed during routine procedures. These limitations highlight the need for automated, real-time polyp detection systems capable of assisting endoscopists in live settings, reducing human error, and ensuring comprehensive examinations.

In recent years, computer-aided detection (CAD) systems have gained traction in the medical imaging field, particularly in tasks requiring speed and accuracy. Recent advancements in deep learning have made it possible to develop CAD systems that support real-time detection. However, in the case of live endoscopic feeds, balancing processing speed with detection accuracy remains a significant challenge. In this study, we introduce a real-time polyp detection system specifically tailored for endoscopic feeds, demonstrating its ability to enhance detection rates without compromising the efficiency required in clinical applications.

II. Literature Survey

Automated polyp detection has been an active topic for research over the last two decades and considerable work has been done to develop efficient methods and algorithms. Earlier works were especially focused on polyp color and texture, using handcrafted descriptors-based feature learning [1], [2]. More recently, methods based on Convolutional Neural Networks (CNNs) have received significant attention [3], [4], and have been the go to approach for those competing in public challenges [5], [6].

Wang et al. [7] designed algorithms and developed software modules for fast polyp edge detection and polyp shot detection, including a polyp alert software system. Shin et al. [8] have used region-based CNN for automatic polyp detection in colonoscopy videos and images. They used Inception ResNet as a transfer learning approach and post-processing techniques for reliable polyp detection in colonoscopy. Later on, Shin et al. [9] used generative adversarial network [10], where they showed that the generated polyp images are not qualitatively realistic; however, they can help to improve the detection performance. Lee et al. [11] used YOLO-v2 [12], [13] for the development of polyp detection and localization algorithm. The algorithm produced high sensitivity and near real-time performance. Yamada et al. [14] developed an artificial intelligence system that can automatically detect the

sign of CRC during colonoscopy with high sensitivity and specificity. They claimed that their system could aid endoscopists in real-time 40498 detection to avoid abnormalities and enable early disease detection.

III.Methodology

We have used the Kvasir-SEG [17] for detection, localisation, and segmentation tasks. This dataset is the outcome of an initiative for open and reproducible results. It contains 1000 polyp images acquired by high-resolution electromagnetic imaging system, i.e., ScopeGuide, Olympus Europe, their corresponding masks and bounding box information. The images and their ground truths can be used for the segmentation task, whereas the bounding box information provides an opportunity for the detection task.

Detection methods aim to predict the object class and regress bounding boxes for localisation, while segmentation methods aim to classify the object class for each pixel in an image. Ground truth masks for segmentation task are shown in 2nd column while corresponding bounding boxes for the detection task are in 3rd column. This section describes the baseline methods for detection, localisation and segmentation methods used for the automated detection and segmentation of polyp in the Kvasir-SEG dataset.

A. DETECTION AND LOCALISATION BASELINE METHODS

Detection methods consist of input, backbone, neck, and head. The input can be images, patches, or image pyramids. The backbone can be different CNN architectures such as VGG16, ResNet50, ResNext-101, and Darknet. The neck is the subset of the backbone network, which could consist of FPN, PANet, and Bi-FPN. The head is used to handle the pre- diction boxes that can be one stage detector for dense predic- tion (e.g., YOLO, RPN, and RetinaNet [50]), and two-stage detector with the sparse prediction (e.g., Faster R-CNN [51] and RFCN [52]). Recently, one stage methods have attracted much attention due to their speed and ability to obtain optima accuracy. This has been possible because recent networks utilise feature pyramid networks or spatial-pyramid pool- ing layers to predict candidate bounding boxes which are regressed by optimising loss functions.

In this paper, we use EfficientDet [53] which uses Effi-cientNet [54], as the backbone architecture, bi-directional feature pyramid network (BiFPN) as the feature network, and shared class/box prediction network. Additionally, we also use Faster R-CNN [51], which uses region proposal net- work (RPN), as the proposal network and Fast R-CNN [55] as the detector network. Moreover, we use YOLOv3 [56] that utilises multi-class logistic loss (*binary cross-entropy* for classification loss and *mean square error* for regres-sion loss) modeled with regularizers such as objectness pre- diction scores. Furthermore, we also used YOLOv4 [57], which utilises an additional bounding box regressor based on the Intersection over Union (IoU) and a cross-stage par- tial connections in their backbone architecture. Additionally, YOLOv4 allows on fly data augmentation, such as mosaic and cut-mix.

RetinaNet [50] takes into account the data driven prop-erty that allows the network to focus on ‘‘hard’’ samples for improved accuracy. The easy to adapt backbones for feature extraction at the beginning of the network provides the opportunity to experiment with deeper and varied archi- tectures such as ResNet50, and ResNet101 for RetinaNet and 53 layered Darknet53 backbone for YOLOv3 and YOLOv4 architecture.

B. SEGMENTATION BASELINE METHODS

In the past years, data-driven approaches using CNNs have changed the paradigm of computer vision methods, includ- ing segmentation. An input image can be directly be fed to convolution layers to obtain feature maps, which can be later upsampled to predict pixel-wise classification provid- ing object segmentation. Such networks learn from available ground truth labels and can be used to predict labels from other similar data. A Fully Convolutional Network (FCN) based segmentation was first proposed by Long *et al.* [58] that can be trained end-to-end. Ronneberger *et al.* [59] modified and extended the FCN architecture to a UNet architecture. The UNet consist of an analysis (*encoder*) and a synthesis (*decoder*) path. In the analysis path of the network, deep fea- tures are learnt, whereas in the synthesis path segmentation is performed on the basis of the learnt features.

Pyramid Scene Parsing Network (PSPNet) [60] introduced a pyramid pooling module aimed at aggregating global con- text information from different regions which are upsampled and concatenated to form the final feature representation. A final per-pixel prediction is obtained after a convolution layer. For feature extraction, we have used the ResNet50 architecture pretrained on imageNet. Similar to the UNet architecture, DeepLabV3+ [61] is an encoder-decoder network. However, it utilizes separable convolutions and spatial pyramid pooling for fast inference and improved accuracy. Atrous convolution controls the resolution of features com- puted and adjust the receptive field to effectively capture multi-

scale information. In this paper, we have used an output stride of 16 for both encoder and decoder networks of DeepLabV3+ and have experimented on both ResNet50 and ResNet101 backbones.

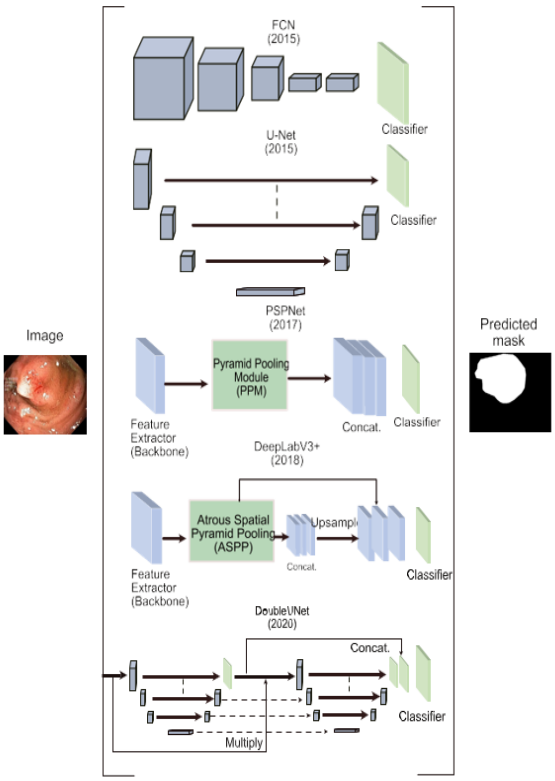

Figure: CNN Model Architecture

ResUNet [62] combines the capabilities of residual neural networks and UNet. The updated ResUNet architecture is called ResUNet++ [3]. Atrous Spatial Pyramid Pooling (ASPP), attentiveness block, and squeeze-and-excite block are some of its extra layers. These extra layers aid in the learning of deep features that can better predict pixels for tasks involving object segmentation. DoubleU-Net [43] is made up of two UNet architecture modifications. The first encoder it employs is VGG-19 pretrained on ImageNet [63]. VGG-19, which is comparable to UNet [64], was chosen primarily because it is a lightweight model. The ASPP block and the squeeze and-excite block are the extra parts of the DoubleUNet. High-Resolution Network (HRNet) [65] continually exchanges information across the resolution while maintaining high-resolution representation convolution in parallel. One of the newest and most often used approaches in the literature is this one. Additionally, we trained the model using UNet with ResNet34 as the backbone network so that it could be compared to other cutting-edge semantic segmentation networks.

The hyperparameters for each of the benchmark techniques based on semantic segmentation are displayed in Table 4. The table shows that the baseline approaches have a significant number of trainable parameters. The network becomes complex due to a large number of trainable parameters, which lowers the frame rate. Therefore, creating a lightweight, effective architecture that can offer improved performance and a faster frame rate is crucial. In light of this, we suggest a novel architecture called ColonSegNet that can reduce training and inference time by requiring a small number of training parameters. The section below contains more information on the architecture.

C. COLONSEGNET

This encoder-decoder's primary components are the squeezing and excitation network [67] and residual block [66]. Compared to other baseline networks like U-Net [59], PSPNet [60], DeepLabV3+ [61], and others, the network is made to have a very small number of trainable parameters. The suggested architecture is a highly lightweight network that achieves real-time performance by using fewer trainable parameters.

Two encoder blocks and two decoder blocks make up the network. The input image is transferred to the decoder when the encoder network has learnt to extract all of the required information from it. Two skip connections from the encoder make up each decoder block. Simple concatenation is used in the first, and a transpose convolution is used in the second skip connection to add multi-scale features to the decoder. In the form of a segmentation mask, these multi-scale characteristics assist the decoder in producing more semantic and significant information.

The first encoder receives the input image and uses a 3 x 3 strided convolution between two residual blocks. After this layer, there is a 2x2 max-pooling. In this case, the spatial dimensions of the output feature map are lowered to one-fourth of the original image. A 3 × 3 strided convolution sits between two residual blocks in the second encoder.

The first decoder employs a stride value of 4, which expands the feature map's spatial dimensions by 4, and begins with a transpose convolution. Likewise, the second decoder increases the spatial dimensions by two by using a stride value of two. The network then proceeds to a residual block and a straightforward concatenation. The second skip connection is then concatenated with it, and a residual block follows once more. The binary segmentation mask is produced by passing the output of the last decoder block through a sigmoid activation function and a 1 × 1 convolution.

i) DATA AUGMENTATION

Supervised learning techniques are data-hungry and need a lot of data to produce dependable and effective models. Such training data must be obtained manually through data collection, curation, and annotation, which requires a large investment of time and money from computational scientists as well as clinical specialists.

One popular method for computationally increasing a dataset's training sample count is data augmentation. We employ fundamental augmentation methods for our DL models, including random rotation, random scale, random cropping, vertical and horizontal flipping, and random rotation. The photos used in each experiment are shrunk to a fixed 512 × 512 size after being normalised. The image is normalised by dividing it by the standard deviation and subtracting it from the mean.

IV. Results

i) EVALUATION METRICS

This section, defines the evaluation metrics used to measure the performance of the CNN model.

True Positive (TP): The number of positive samples that are identified correctly by the classifier means that sample falls in polyp class and classified as such.

False Positive (FP): The number of negative sample that are wrongly identified in a positive category, means that sample falls in non-polyp class, but classified as polyp class.

True Negative (TN): The number of negative sample that are identified correctly in its category. Samples are non-polyp class and classified as such.

False Negative (FN): The number of positive samples that are wrongly identified in another category means that the sample falls in polyp class, but classified as non-polyp class.

Confusion Matrix is a table utilized to describe the overall performance of the classification model on test data whose actual values are known. The relation between true positive, false positive, true negative and false negative are shown in Table 3.2.

Table: Confusion Matrix

		Actual Class	
		Polyp	Non-polyp
Predicated Class	Polyp	True	False
	Non-	False	True

Recall (REC): Calculates the proportion of all true positive samples from cases that are actually positive. Also it referred to as sensitivity and true positive rate.

Precision (PREC): Calculates the proportion of all true positive samples from cases that are predicated as positive.

F1 score (F1): Another accuracy measure, also referred F-measure, utilized to seek the relation between precision and recall by counting the weighted average.

ROC curve: The receiver operating characteristics is a two-dimensional graph in which created by plotting the false positive rate FPR on the x-axis against true positive rate TPR represents the y-axis at various threshold settings.

Specificity (SPEC): Calculates the proportion of all true negative samples from cases that are actually negative, also referred to false positive rate.

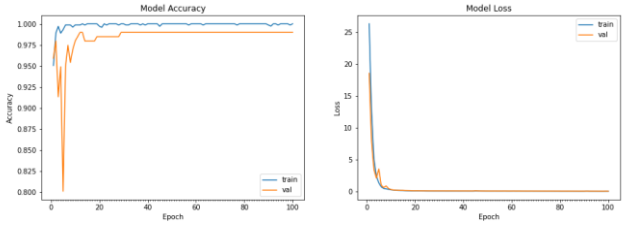

Figure: Evaluation of Model

ii) EXPERIMENTAL RESULTS

The experimental results shown in the image illustrate the performance of the proposed model in detecting polyps during colonoscopy. The outputs are divided into three parts: the **original image**, the **predicted mask**, and the **polyp detection visualization**.

1. **Original Image**: This is the raw colonoscopy frame input to the model. It showcases the gastrointestinal tract, with a visible polyp present in the center of the frame.
2. **Predicted Mask**The mask highlights the region of interest (ROI) corresponding to the detected polyp. The clear and accurate segmentation of the polyp validates the model's ability to isolate relevant structures effectively.
3. **Polyp Detection Visualization**: This frame overlays the bounding box and depth estimation onto the original image. The bounding box tightly encloses the polyp, demonstrating the localization accuracy of the model. The depth annotation indicates the model's potential in providing additional diagnostic insights, such as size or proximity.

The results validate the effectiveness of the model initialization, demonstrating robust segmentation and detection capabilities in complex medical imaging scenarios. This performance is critical for assisting clinicians in early detection and accurate diagnosis of

colorectal polyps.

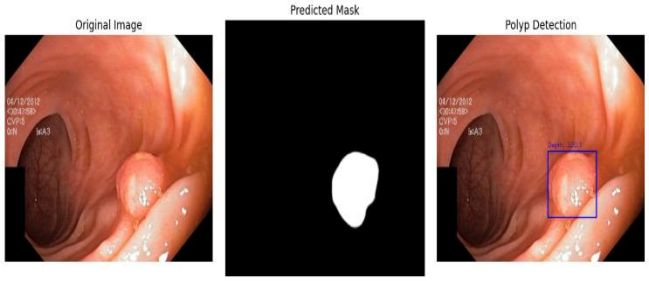

Figure: Detection of Polyp

V. Conclusion

In this project, we developed a deep learning model for the automatic classification of colorectal polyps in colonoscopy images using Convolutional Neural Networks (CNNs). The primary objective was to design a model capable of distinguishing between polyp and non-polyp images, and later classifying various types of polyps based on their features.

The project began with a study of colorectal polyps, their significance in cancer screening, and the existing methods for automatic polyp classification. We explored different machine learning approaches, particularly focusing on CNNs, which are well-suited for image classification tasks. The CNN architecture, with its multiple convolutional layers, allowed the model to automatically learn discriminative features from the colonoscopy images.

Our approach involved a single CNN-based model, which was trained from scratch on the Kvasir-SEG dataset. The image preprocessing phase included patch extraction and data augmentation to increase the robustness of the model. The model was trained to classify images into two categories: polyp and non-polyp. By using the original Kvasir-SEG dataset, which only contained polyp images, we classified them accordingly into polyp and non-polyp categories, focusing on accuracy.

The proposed model achieved impressive results, with an overall classification accuracy of 98.4%, alongside a precision, sensitivity, and F1-score of 98%. This high performance demonstrates the efficacy of using CNNs for this task. Furthermore, the model's architecture is flexible and can be adapted in the future for other medical imaging tasks, providing a foundation for further development.

FUTURE WORK

Although the results of this project are promising, there are several potential improvements and directions for future research:

1. **Classifying Multiple Polyp Types**: The current model distinguishes between polyp and non-polyp images. In future work, we aim to extend the classification to include different types of polyps, such as non-malignant (Type 1), neither malignant nor non-malignant (Type 2A, 2B), and malignant (Type 3). This would broaden the clinical applicability of the model by enabling more precise diagnoses.
2. **Expanding the Dataset**: A larger and more diverse dataset would likely improve the model's performance and generalizability. This would also allow the inclusion of additional polyp types, such as serrated sessile, pedunculated, and tubular polyps. Collecting more colonoscopy images and classifying them accurately would provide a more comprehensive model.
3. **Real-Time Integration with Endoscopy Machines**: One of the most significant future directions is to integrate this model into real-time endoscopy systems. By installing our software in endoscopy machines, the model could automatically classify polyps during the procedure, providing immediate feedback to healthcare professionals and potentially improving patient outcomes.
4. **Model Optimization**: Further optimization of the CNN architecture could improve the model's accuracy and efficiency. This could involve experimenting with different network depths, layer configurations, and hyperparameter tuning to achieve even better performance.

By expanding the dataset, extending the classification to multiple polyp types, and integrating the model into real-world applications such as endoscopy systems, this project has the potential to contribute to improved colorectal cancer screening and diagnosis.

References

[1]. The American Cancer Society medical and editorial content team. (2018, 02 21). What Is Colorectal Cancer? (cancer.org) Retrieved 10 01, 2019, from https://www.cancer.org/cancer/colon-rectal-cancer/about/what-is-colorectal-cancer.html

[2]. Arnold, M., Sierra, M. S., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2016). Global patterns and trends in colorectal cancer incidence and mortality. Gut, 66. doi:10.1136/gutjnl-2015-310912

[3]. Awad, M., & Khanna, R. (2015). Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers. Berkeley, CA: Apress.

[4]. Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017).

SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29, 2481-2495. doi:10.1109/TPAMI.2016.2644615

[5]. Bernal, J., Sánchez, F. J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., & Vilariño, F. (2015). WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Computerized Medical Imaging and Graphics, 43, 99-111.

[6]. Bernal, J., Sánchez, J., & Vilariño, F. (2012). Towards automatic polyp detection with a polyp appearance model. Pattern Recognition, 45, 3166-3182. doi:0.1016/j.patcog.2012.03.002

[7]. Bernal, J., Tajkbaksh, N., Sánchez, F.J., Matuszewski, B., Chen H., Yu, L., Angermann, Q., Romain, O., Rustad, B., Balasingham, I., Pogorelov, K., Choi, S., Debard, Q., Maier-Hein, L., Speidel, S., Stoyanov, D., Brandao, P., Cordova, H., Sánchez-Montes, C. (2017). Comparative Validation of Polyp Detection Methods in Video Colonoscopy: Results from the MICCAI 2015 Endoscopic Vision Challenge. IEEE Transactions on Medical Imaging.

[8]. Burkov, A. (2019). The hundred-page machine learning book. Andriy Burkov.

[9]. Byrne, M. F., Chapados, N., Soudan, F., Oertel, C., Linares Pérez, M., Kelly, R., … Rex, D. K. (2017). Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. Gut, 94–100. doi:0.1136/gutjnl-2017-314547

[10]. Chen, J., Milot, L., Cheung, H. M. C., & Martel, A. L. (2019). Chen, J., Milot, L., Cheung, H. M. C., & Martel, A. L. (2019). Unsupervised Clustering of Quantitative Imaging Phenotypes Using Autoencoder and Gaussian Mixture Model. Lecture Notes in Computer Science Medical Image Computing and Computer Assisted Intervention – MICCAI, 575–582. doi:10.1007/978-3-030-32251-9_63

[11]. Dekker, E., & Rex, D. K. (2018). Advances in CRC prevention: screening and surveillance. Gastroenterology, 154. doi:10.1053/j.gastro.2018.01.069

[12]. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, 248–255.

[13]. Ferlay, J. , Colombet, M. , Soerjomataram, I. , Mathers, C. , Parkin, D. , Piñeros, M. , Znaor, A. and Bray, F. (2019). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. International Journal of Cancer, 144, 1941-1953. doi:10.1002/ijc.31937

[14]. Groff, R. J., Nash, R., & Ahnen, D. J. (2008). Significance of serrated polyps of the colon. Current gastroenterology reports, 490–498. doi:10.1007/s11894-008-0090-z

[15]. Guo, Y., & Matuszewski, B. (2019). GIANA Polyp Segmentation with Fully Convolutional Dilation Neural Networks. Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications.

[16]. He, K., Zhang, X., Ren, S., & Sun, J. . (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[17]. Hilsden RJ, Heitman SJ, Mizrahi B, Narod SA, Goshen R. (2018). Prediction of findings at screening colonoscopy using a machine learning algorithm based on complete blood counts (ColonFlag). PLoS One, 1-9. doi:10.1371

[18]. Horváth, A., Spindler, S., Szalay, M., & Rácz, I. (2016). Preprocessing Endoscopic Images of Colorectal Polyps. Acta Technica Jaurinensis, 9, 65. doi:0.14513/actatechjaur.v9.n1.397

[19]. Huang, Y., Gong, W., Su, B., Zhi, F., Liu, S., & Jiang, B. (2012). Risk and Cause of Interval Colorectal Cancer after Colonoscopic Polypectomy. Digestion, 86, 148-154. doi:10.1159/000338680

Author Profile

Mrs. S. Hemalatha, has over 10 years of experience in academics and machine learning research. **Prabin R, Rahul Shriram S G, Dharshanashri G,** and **Santhiya A K** are final-year AIML students at Sri Shakthi Institute of Engineering and Technology, specializing in computer vision and deep learning applications.