Paper ID 3

FMIC-AI: Annotation-Free Tumor Cell Detection in Fluorescence Microscopy via Self-Supervised Anomaly Detection

Yinglan Kuang[1], Lin Chen[2], Dongjiang Tang[2], Xing Lu[1,2]
[1] Zhuhai Hengqin Sanmed Aitech Ltd. Zhuhai, Guangdong, China
[2] Zhuhai Sanmed Biotech Inc. Zhuhai, Guangdong, China

Introduction

Accurate identification of tumor cells is crucial for clinical diagnosis and treatment, yet traditional fluorescence microscopy methods rely on manual interpretation, leading to inefficiency and subjectivity. While deep learning offers automation potential, its performance depends heavily on large-scale pixel-level annotations, which are costly and expert-dependent. To overcome these limitations, self-supervised learning (SSL) [1] has emerged as a promising alternative. However, existing SSL approaches face challenges in feature representation, model generalization across modalities, and explainability at the cellular level.

This paper proposes FMIC-AI, an annotation-free framework for tumor cell detection based on self-supervised anomaly detection (SSAD). The framework incorporates a Vision Transformer (ViT) to preserve local cell features and uses Triple Mean-Shift Contrastive Loss (TMSCL) [2] to enhance representation learning of normal regions. It integrates self-supervised fine-tuning (SSFT) with CellPose for universal cell segmentation and combines Grad-CAM++ with cell mask information for multi-level anomaly localization (MAL), enabling precise detection from image patches to individual cells. By reducing annotation dependence and improving accuracy, FMIC-AI offers an efficient tool for clinical pathology analysis.

Method

The FMIC-AI framework integrates three core components: an SSAD-based network for anomaly detection, universal cell segmentation with SSFT, and multi-level anomaly localization (MAL).

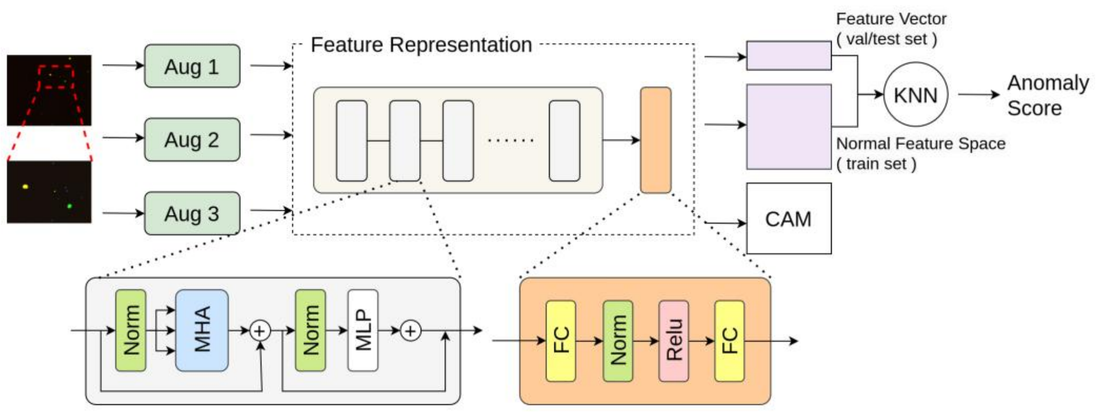

Fig. 1. Method of the SSAD-based Network.

As illustrated in Fig. 1, the SSAD network replaces ViT's fully connected layer with a non-linear projector (comprising linear transformation, batch normalization, and ReLU) to enhance discriminative feature representation. During training, triple-view inputs are generated using RandomResizedCrop and ColorJitter augmentations, while TMSCL constrains feature consistency across perspectives. The anomaly scoring mechanism employs k-nearest neighbors (KNN, k=2) to compute patch-wise distances in the normal feature space, with dynamic thresholding for adaptive abnormality classification across datasets.

For cell segmentation, CellPose [3] —based on an enhanced UNet—generates initial predictions, from which high-quality results with clear boundaries and minimal overlap are manually selected as pseudo-labels for model fine-tuning, ensuring morphological diversity and representation.

The MAL module combines Grad-CAM++ with a sliding window approach (window size: 512×512, stride: 412) to process image patches independently. Results are aggregated using a maximum-value fusion strategy that accounts for overlap and anomaly scores. Cell-level masks from segmentation are integrated to extract per-cell anomaly scores, enabling reconstruction of global tumor cell distribution while optimizing computational efficiency.

Experiments and Results

Backbone comparison (Table 1) shows ViT achieved superior performance (AP: 0.961, ROC-AUC: 0.984, PR-AUC: 0.958), outperforming ResNet, EfficientNet, and RegNet. Comparative SSL analysis (Table 2) demonstrates our method exceeds FPI, MSC, CutPaste3Way, and CutPasteNormal across all metrics. Ablation study (Table 3) validates full framework (AD+Seg+Loc) achieves highest performance (ROC-AUC: 0.934, Recall: 0.89), demonstrating critical component synergy.

Table 1. Comparison of Different Backbones in SSAD-based Network.

Backbone	AP	ROC-AUC	PR-AUC
ResNet	0.954	0.976	0.952
EfficientNet	0.884	0.950	0.881
RegNet	0.946	0.971	0.944
ViT	**0.961**	**0.984**	**0.958**

Table 2. Comparison of Different SSL Methods.

Backbone	AP	ROC-AUC	PR-AUC
FPI	0.774	0.712	0.775
MSC	0.727	0.802	0.723
CutPaste3Way	0.912	0.920	0.911
CutPasteNormal	0.913	0.939	0.917
Ours	**0.961**	**0.984**	**0.958**

Table 3. Ablation Study. "AD" refers to the SSAD-based network, "Seg" to universal cell segmentation with SSFT, and "Loc" to multi-level anomaly localization.

Method	ROC-AUC	Recall
AD	0.881	0.87
AD + Seg + Loc	0.934	0.89

Conclusion

The FMIC-AI framework enables high-precision, annotation-free tumor cell detection using SSL, achieving an AUC of 0.934 and a recall of 0.89 on fluorescence microscopy test data, outperforming conventional methods. By integrating SSFT and MAL, the system accurately locates and segments abnormal cells, demonstrating strong potential for clinical prognosis analysis.

Current limitations include sensitivity to normal cell diversity due to dataset constraints and occasional missed small cells from limited feature resolution. Future work will expand data diversity to include brightfield and phase-contrast microscopy and enhance feature extraction with multi-scale or attention mechanisms. Although currently validated on fluorescence imaging, FMIC-AI's framework is extendable to histopathology and large-scale cancer research, providing a scalable solution for label-scarce scenarios like personalized medicine.

References

1. Qin, J., Gu, C., Yu, J., et al.: Multilevel saliency-guided self-supervised learning for image anomaly detection. In: Signal, Image and Video Processing, vol. 18, pp. 6339–6351. Springer, Heidelberg (2024)
2. Fan, W., Shangguan, W., & Chen, Y.: Transformer-based contrastive learning framework for image anomaly detection. In: International Journal of Machine Learning and Cybernetics, vol. 14, pp. 3413–3426. Springer, Heidelberg (2023)
3. Stringer, C., Wang, T., Michaelos, M., et al.: Cellpose: a generalist algorithm for cellular segmentation. In: Nature Methods, vol. 18, pp. 100–106. Springer, Heidelberg (2021)