LIVE CELL IMAGING ANALYSIS USING MACHINE LEARNING AND SYNTHETIC FOOD IMAGE GENERATION

Yue Han

Major Advisor: Edward J. Delp

Video and Image Processing Laboratory (*VIPER*)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana USA

Acknowledgements

- **Professor Edward J. Delp – Advisor, Committee Chair**
- **Professor Fengqing M. Zhu – Committee Member**
- **Professor Jan P. Allebach – Committee Member**
- **Professor Mary L. Comer – Committee Member**

- **HP Lab Microfluidics team for their help**

- **VIPER Lab members for their support and friendship**

- **My family and my wife for their love and support**

Acknowledgements

Chapter 1 to Chapter 3 of this work was supported by HP.

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of HP.

Contributions of This Work

- **We modify the Kalman filter in DeepSORT to use prior knowledge of the object's velocity to assist the motion estimation and improve the tracking result while outperforming other approaches in the literature**

- **We propose the Weighted Mask Fusion (WMF), combining the results of multiple segmentation models, to improve the final nuclei segmentation mask**

- **We propose an edge-aware Mask RCNN method, and introduced a hybrid architecture, an ensemble of CNNs and Swin-Transformer Edge Mask R-CNNs, to accurately segment irregularly shaped nuclei in microscopy images**

- **We propose an effective clustering-based training framework, named ClusDiff, for generating high-quality and representative food images**

Publications

Publications Resulting From This Thesis:

- **Y. Han**, Y. Lei, V. Shkolnikov, D. Xin, S. Barcelo, J. Allebach, E. J. Delp, "Label-Free Mammalian Cell Tracking Enhanced by Precomputed Velocity Fields", *Proceedings of the IEEE International Conference on Machine Learning and Application (ICMLA)*, December 2022, Nassau, Bahamas

- **Y. Han**, Y. Lei, V. Shkolnikov, D. Xin, A. Auduong, S. Barcelo, J. Allebach, E. J. Delp, "Ensemble Processing and Synthetic Image Generation for Abnormally Shaped Nuclei Segmentation", Feburary 2023, *bioRxiv*

- **Y. Han**, Y. Lei, V. Shkolnikov, D. Xin, A. Auduong, S. Barcelo, J. Allebach, E. J. Delp, "Ensemble Processing and Convexity Measure for Abnormally Shaped Nuclei Segmentation", SPIE Medical Imaging, February 2024, San Diego, CA

- **Y. Han**, Y. Lei, V. Shkolnikov, D. Xin, A. Auduong, S. Barcelo, J. Allebach, E. J. Delp, "An Ensemble Method With Edge Awareness for Abnormally Shaped Nuclei Segmentation", IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshop (CVPR), June 2023, Vancouver, Canada

- **Y. Han,** J. He, M. Gupta, E. J. Delp and F. Zhu "Diffusion Model with Clustering-based Conditioning for Food Image Generation", 8th International Workshop on Multimedia Assisted Dietary Management (MADiMa), October 2023, Ottawa, Canada

Publications

Publications Not Resulting From This Thesis:

- G. Tonmoy, **Y. Han**, V. Raju, D. Hossain, M. A. McCrory, J. Higgins, C. Boushey, E. J. Delp, and E. Sazonov. "Integrated image and sensor-based food intake detection in free-living", Scientific Reports 14, no. 1: 1665, January 2024

- W. Fu, **Y. Han**, J. He, S. Baireddy, M. Gupta, and F. Zhu, "Conditional Synthetic Food Image Generation", Proceedings of the Imaging and Multimedia Analytics in a Web and Mobile World Conference, IS&T Electronic Imaging Symposium, February 2023, Burlingame, CA

- Z. Shao, **Y. Han**, J. He, R. Mao, J. Wright, D. Kerr, C. J. Boushey, and F. Zhu, "An Integrated System for Mobile Image-Based Dietary Assessment", Proceedings of the 3rd Workshop on AIxFood, ACM Multimedia, October 2021, Chengdu, China (Hybrid)

- **Y. Han**, S. K. Yarlagadda, T. Ghosh, F. Zhu, E. Sazonov, and E. J. Delp, "Improving Food Detection For Images From a Wearable Egocentric Camera", Proceedings of the Imaging and Multimedia Analytics in a Web and Mobile World Conference, IS&T Electronic Imaging Symposium, February 2021, Burlingame, CA (Virtual)

Outline

- **Label-Free Mammalian Cell Tracking Enhanced by Precomputed Velocity Fields**
- **Ensemble Processing and Convexity Measure for Abnormally Shaped Nuclei Segmentation**
- **An Ensemble Method with Edge Awareness for Abnormally Shaped Nuclei Segmentation**
- **Diffusion Model with Clustering-based Conditioning for Food Image Generation**

Label-Free Mammalian Cell Tracking Enhanced by Precomputed Velocity Fields

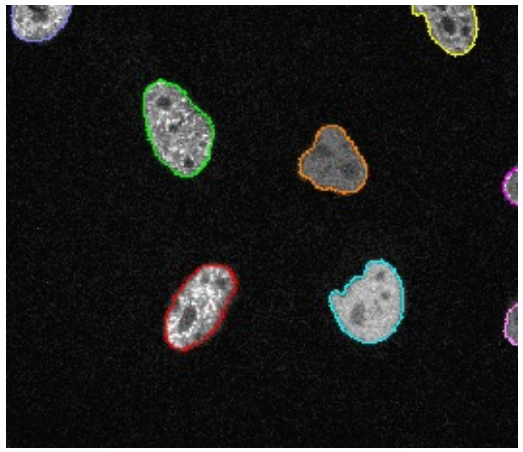
Motivation

- **Cell sorting is the process through which a particular cell type is separated from others on the basis of its physical or biological properties**
- **Cell sorting using flow cytometry has become indispensable in all cellular biology research**
- **Sorted cells are very useful for numerous downstream studies such as functional assays and cell line development for therapeutic research and development**
- **Label-free cell imaging is increasingly gaining interest in biomedical research, as chemical labeling processes risk altering the cells' properties**

Our Goal

- **Label-free imaging provides few human-interpretable image features (HIFs)**
  - **HIFs include cell sizes and shapes, nucleus sizes and shapes, textures, and morphologies**
- **Counting cell sorting result is very time-consuming and labor intensive**
- **Can we use machine learning approaches to obtain more image features from label-free image?**
- **Can we automate visual inspection by tracking the cells during sorting process?**

Existing Work in Image-Based Cell Tracking



ISBI Cell Tracking Challenge

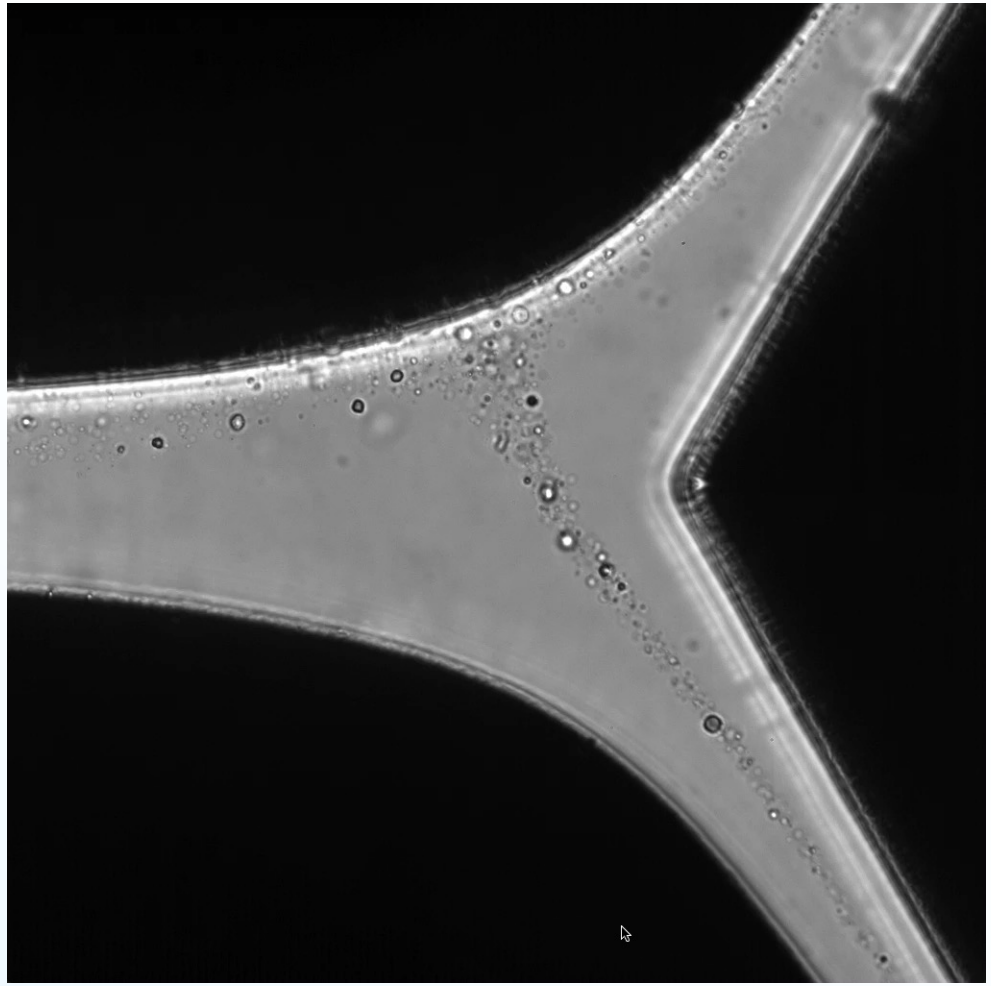
CTMC-v1

- **There are two related public cell tracking challenges using machine learning:**
  - **ISBI Cell Tracking Challenge***
  - **CTMC-v1****
- **The differences, such as having distinct motion patterns and whether the cells are contained in parent cells, are significant**

***Ulman, Vladimír, et al . "An objective comparison of cell-tracking algorithms," *Nature methods* 2017**

**** Anjum, Samreen et al. "CTMC: Cell Tracking With Mitosis Detection Dataset Challenge," *CVPR* 2020**

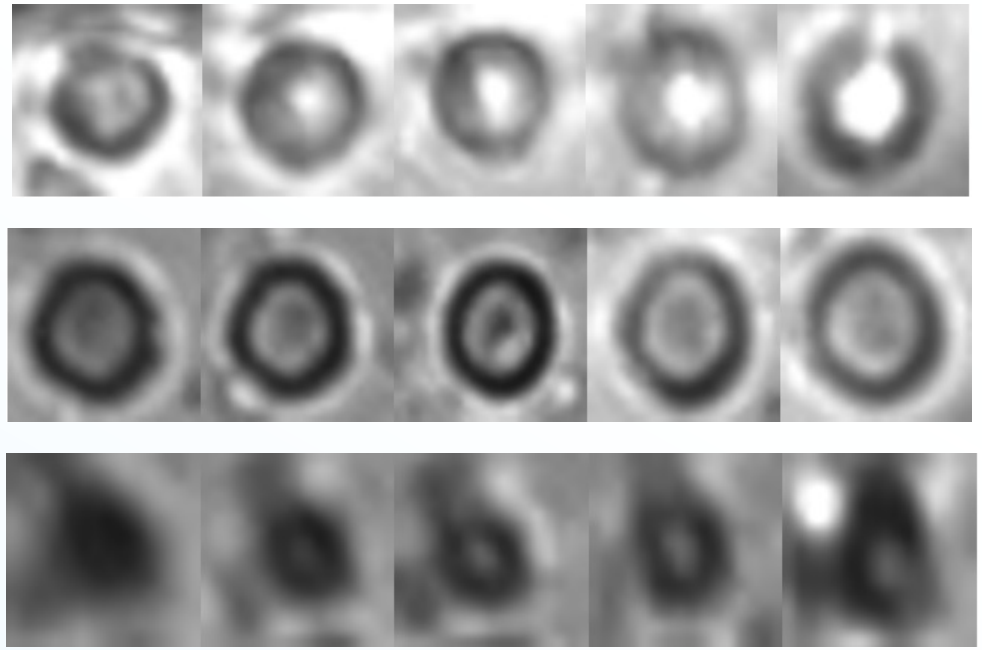
Tracking Cells in the Chip



A Clip of Cell Sorting Using the “chip”

- Our cell sorting device is a physical cell classifier – referred as the “chip”
- The device focuses input cells into streamlines, then uses a uniform dielectrophoretic force field to physically classify the cells into their respective outlet channels
- Our unique challenges include:
  - Large displacement between frames
  - Unique motion pattern
  - The appearance between cells are hard to distinguish

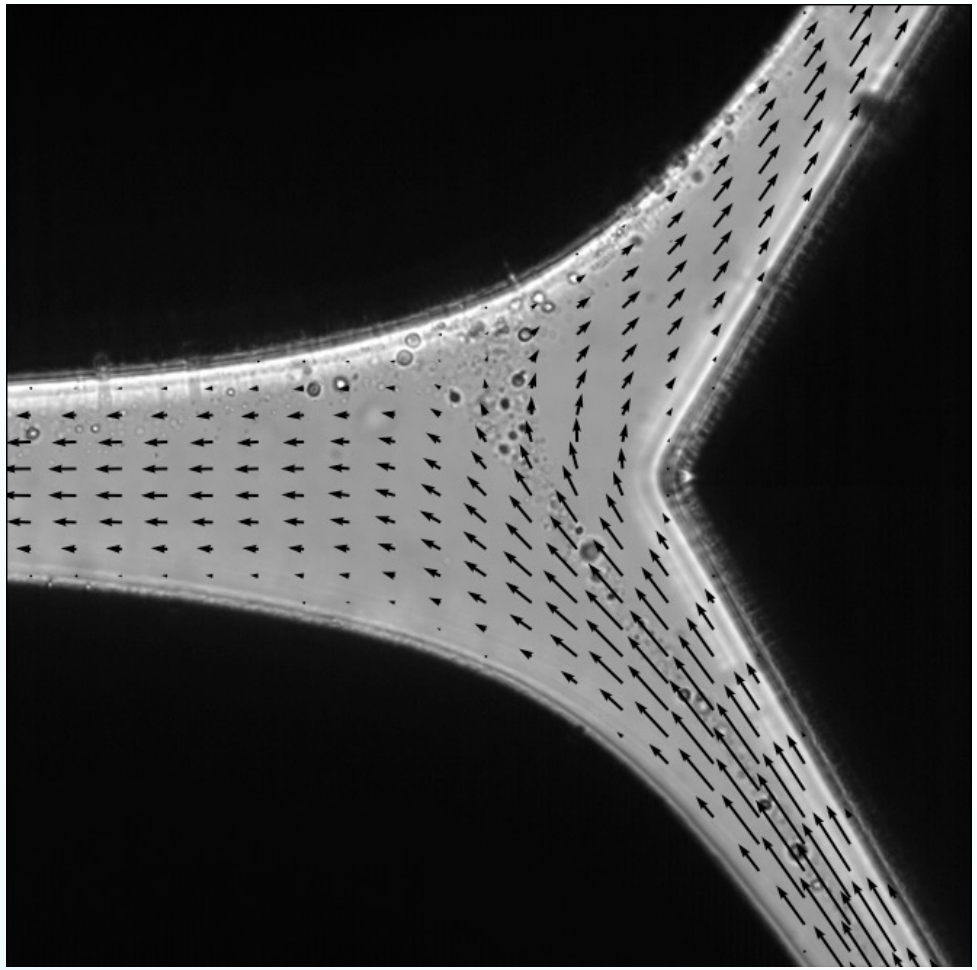
Tracking Cells in the Chip



Appearance of 3 Cells in 5 Different Frames

- Our cell sorting device is a physical cell classifier – referred as the “chip”
- The device focuses input cells into streamlines, then uses a uniform dielectrophoretic force field to physically classify the cells into their respective outlet channels
- Our unique challenges include:
  - Large displacement between frames
  - Unique motion pattern
  - The appearance between cells are hard to distinguish

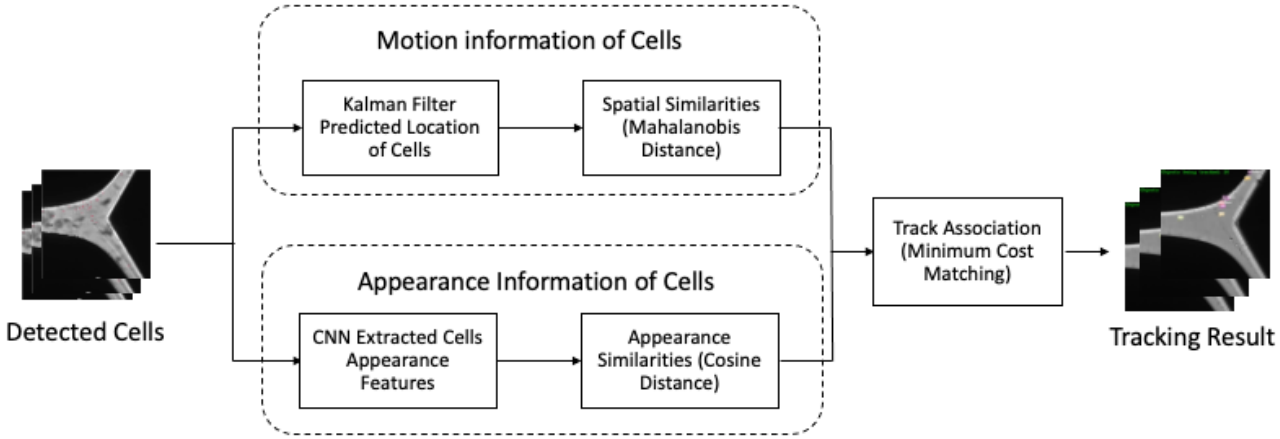
Representative Velocity Field



The Representative Velocity Field in the Chip

- A cell's velocity is mainly determined by the hydrodynamic drag force from the local flow velocity
- We estimate the flow velocity and use as the representative velocity field to indicate the velocity of the cell in the chip
- We assign a velocity to each pixel of the image using bilinear interpolation

Label-Free Mammalian Cell Tracking



The Block Diagram of DeepSORT

- **Ideal cell tracking here should utilize both cell motion and cell appearance features for better track association**
- **DeepSORT uses**
  - **Kalman Filter to obtain motion information**
  - **A pre-trained CNN to obtain appearance features**
  - **Minimum cost matching to associate/combine detected cells to tracks**
- **The DeepSORT assigns track-ids to detected cells in each frame**

Kalman Filter Modification

- **In DeepSORT's Kalman Filter, the measurement of a cell only contains the bounding box center position *(x, y)*, aspect ratio *a*, height *h* of the detected object**

- **Using the representative velocity field information, we can estimate a velocity of the cell based on its location**

- **We modify the Kalman Filter to add new measurement of a cell *(x , y, a, h, $\dot{x}$, $\dot{y}$, $\dot{a}$, $\dot{h}$)***
  - **$\dot{x}$ and $\dot{y}$ are the velocity on the x and y direction, we use the representative velocity field as input**
  - **$\dot{a}$ and $\dot{h}$ are the rate of change of the aspect ratio and the height of the bounding box, we initialize them to 0 since the aspect ratio and the height of the cell change slightly**

Kalman Filter Modification

- **To adapt non-constant velocity in the chip, we modify the Kalman filter state vector X to a 12-dimensional vector *(x , y, a, h, $\dot{x}$, $\dot{y}$, $\dot{a}$, $\dot{h}$, $\ddot{x}$, $\ddot{y}$, $\ddot{a}$, $\ddot{h}$)***

- **With changes in the Kalman filter state vector, the state transition matrix F in the state extrapolation equation can be adjusted to a linear constant acceleration mode**

$$x_{n+1} = x_n + \dot{x}_n \Delta t + 0.5\ddot{x}_n \Delta t^2$$
$$\dot{x}_{n+1} = \dot{x}_n + \ddot{x}_n \Delta t$$
$$\ddot{x}_{n+1} = \ddot{x}_n$$

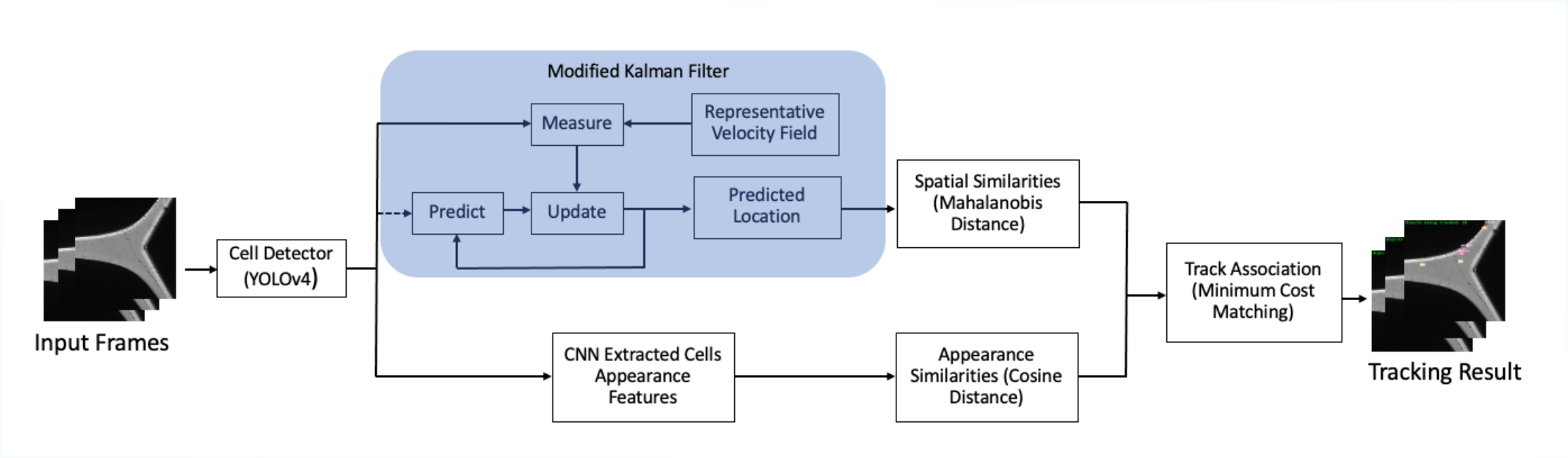
(The rest of the parameters in the system state vector are updated in the same way as x)

- **With the new system state and measurement vectors, we adjust the system observation matrix H**

$$\begin{aligned}\mathbf{z_n} &= \mathbf{H}\mathbf{y_n} + \mathbf{v_n} \\ &= \mathbf{H}\left[x, y, a, h, \dot{x}, \dot{y}, \dot{a}, \dot{h}, \ddot{x}, \ddot{y}, \ddot{a}, \ddot{h}\right]^T + \mathbf{v_n} \\ &= \left[x, y, a, h, \dot{x}, \dot{y}, \dot{a}, \dot{h}\right]^T + \mathbf{v_n}\end{aligned}$$

z is the measurement vector,
y is the newly measured true system state vector,
v is a Gaussian random noise vector

Our Proposed Cell Tracking System



The Block Diagram of the Modified DeepSORT with YOLOv4* using the representative velocity field

Alexey, Bochkovskiy∗, et al . "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv, 2020

Dataset

- **Our dataset consists of phase contrast images captured during the cell sorting process in the chip**
  - **The input is a stream of melanoma cancer cells and white blood cells**
  - **The dataset contains more than 40,000 images captured at 500 frames per second**
- **We randomly sample 50 images and annotate them as the training set for the cell detector**
- **We randomly select the starting frame and annotated three 50-frame sequences**
  - **None of these image sequences have any temporal overlap with each other or with training dataset for detector**
- **We are only interested in tracking in-focus cells, thus only in-focus cells were annotated**

Experiments

- **We train a YOLOv4 to detect the cells in each frames - one-stage object detection and classification system**

- **We choose three tracking methods for comparing with our proposed method**
  - **original DeepSORT**
  - **FairMOT* (Appearance information only)**
  - **CenterTrack** (Motion information only)**

- **We train FairMOT and CenterTrack with one of the three image sequences, the other two are used for evaluation of all four methods**

***Yifu, Zhang, et al . "FairMOT: On the Fairness of Detection and Re-Identification in Multiple Object Tracking," *International Journal of Computer Vision*, 2021**

**** Xingyi, Zhouet al. "Tracking Objects as Points," *ECCV* 2020**

Evaluation Metric

$$MOTA = 1 - \frac{\sum_t (FN_t + FP_t + IDSW_t))}{\sum_t GT_t}$$

Where:
FP: Total number of bboxes not covering an GT bbox
FN: Total number of GT bboxes not covering by and bbox
IDSW: Identity Switch
GT: Total number of GT bboxes

$$IDF1 = \frac{2IDTP}{2IDTP + IDFP + IDFN}$$

Where:
IDTP: Total correctly identified detections
IDFP: False Positive identification
IDFN: False Negative identification

- **We evaluate tracking methods using two metrics: Multiple Object Tracking Accuracy (MOTA) and ID F1 Score (IDF1)**
  - **MOTA represents object coverage**
  - **IDF1 quantifies the objects' identity across frames in the sequence**
  - **MOTA scores can range between negative values and 100, with better performance indicated by values lying closer to a 100.**
  - **IDF1 can range between 0 and 100%, with better performance indicated by values lying closer to 100%**
  - **For both MOTA and IDF1 the larger the metric indicates better performance**

Quantitative Results

Method	MOTA	IDF1
DeepSORT*	17.9	24.6
FairMOT**	11.4	12.1
CenterTrack***	17.5	16.3
Proposed Method (modified DeepSORT)	**26.4**	**34.7**

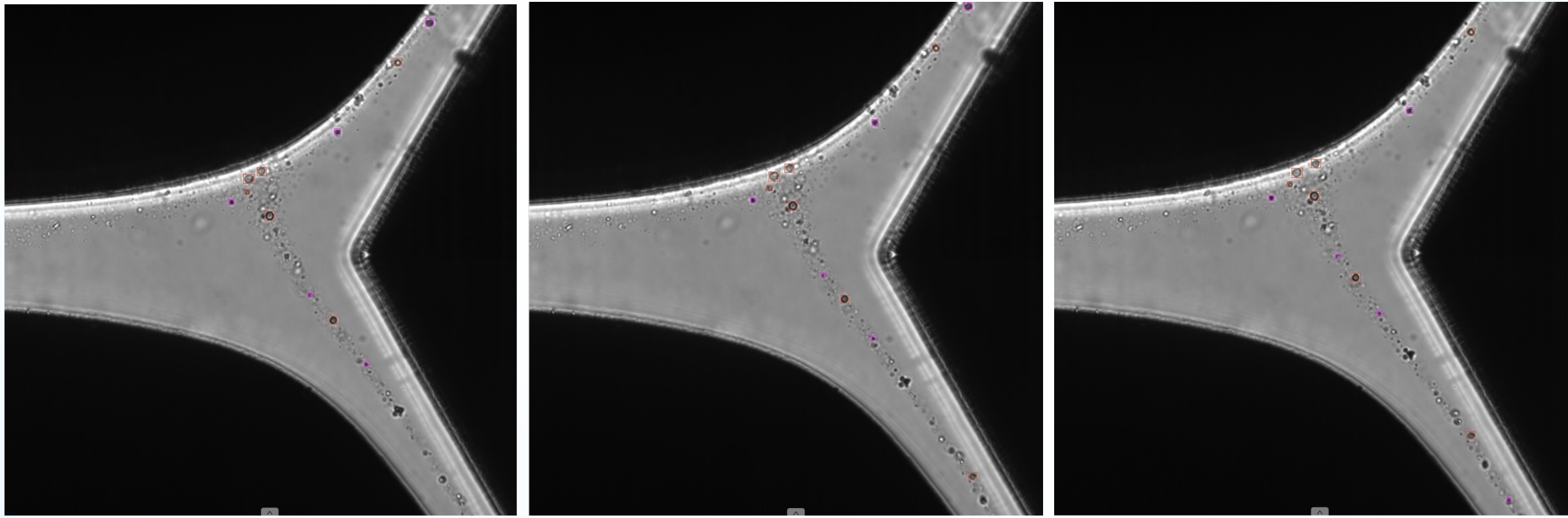
The Cell Tracking Results of Different Methods

***Nicolai, Wojke, et al . "SIMPLE ONLINE AND REALTIME TRACKING WITH A DEEP ASSOCIATION METRIC," *ICIP* 2017**

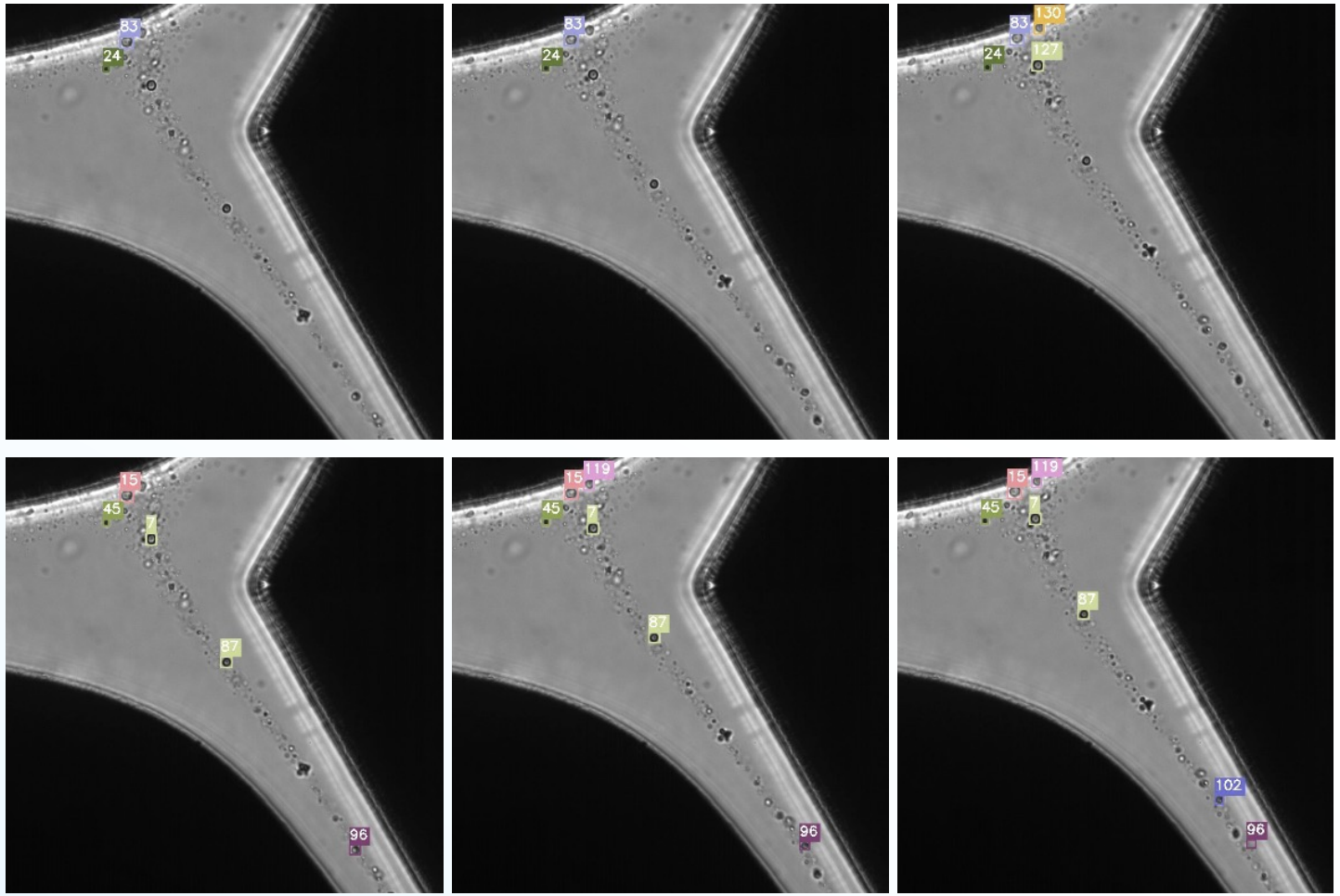
****Yifu, Zhang, et al . "FairMOT: On the Fairness of Detection and Re-Identification in Multiple Object Tracking," *International Journal of Computer Vision*, 2021**

*****Xingyi, Zhouet al. "Tracking Objects as Points," *ECCV* 2020**

Example of Ground Truth Image Sequence



Tracking Result Comparison



Tracking Result Comparison: DeepSORT (TOP), Proposed Method (Bottom)

- With the supplement of the modified Kalman filter and the precomputed velocity field, our method:
  - Associates the tracks to the cells in high velocities regions
  - Picks up the cells for tracking sooner
  - Adapts better to the unique motion pattern in the chip

Example: Cell Tracking Result

Future Work

- **Examine other machine learning methods to extract appearance features to better distinguish and re-identify the individual cells**

- **Explore motion estimation methods and graph network track association methods to better predict the cell's complex motion pattern in the chip as more annotated data becomes available**

Ensemble Processing and Convexity Measure for Abnormally Shaped Nuclei Segmentation

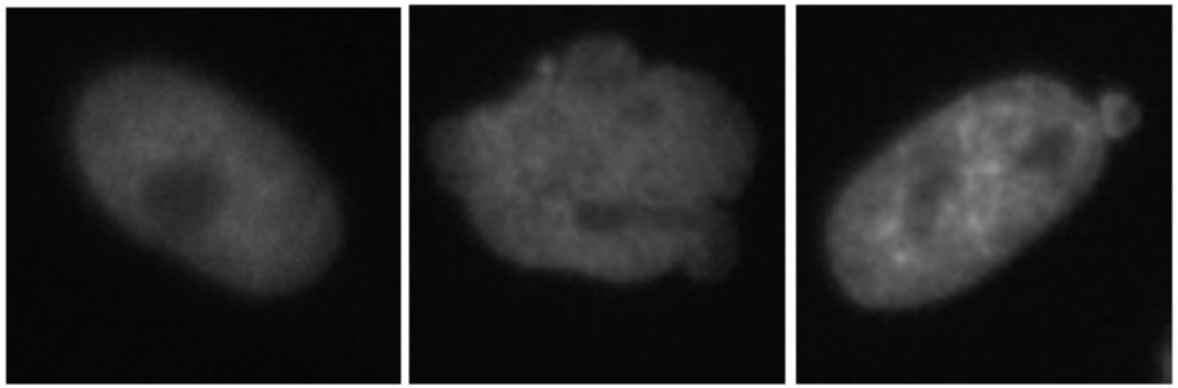
Motivation

- **While most nuclei typically have an elliptical shape, deviations from this shape can arise in certain stages of the cell cycle, due to external stress, or in certain disease state**
- **Characterization of nuclei shapes yields important information for many applications such as determining cell cycle stage, measuring cellular response, indicating genetic instability, and cancer diagnostics**
- **Methods have been developed for elliptical nuclei segmentation while approaches for segmenting non-elliptical nuclei are lacking**
- **We want to design a method to segment abnormally shaped nuclei**

Our Goal

- **Can we generate synthetic nuclei images with annotation as ground truth data for training the nuclei segmentation model?**
- **Can we design a method to segment abnormally shaped nuclei and reduce the false positive of segmenting the elliptical nuclei?**
- **Can we avoid the nuclei segmentation model learning the bias from synthetic images?**

Introduction



The images of
(a): A typical elliptical nucleus
(b): An abnormally shaped nucleus with large differences in shape
(c): An abnormally shaped nucleus which looks similar to elliptical shape but with a small bump on its surface

- **There is a wide range of differences between typical elliptical nuclei and abnormally shaped nuclei of the cells in our experimental dataset**
- **While most of the abnormally shaped nuclei have significant differences from the typical elliptical nuclei, the most difficult abnormally shaped nuclei are the ones that are elliptical in shape with additional bumps and protrusions**

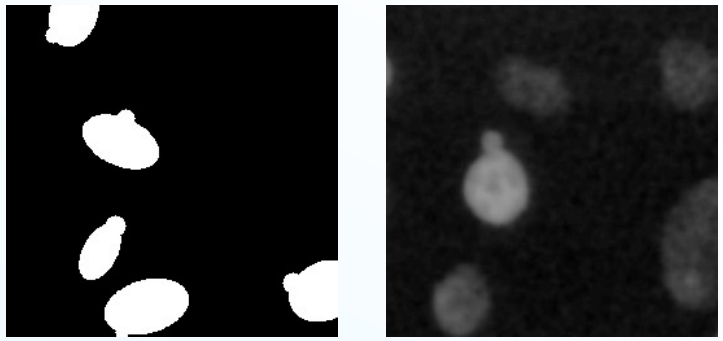
Synthetic Ground Truth Images Generation

- **Since a specific shape of abnormal nuclei is challenging to segment and ground truth images for abnormal nuclei are hard to obtain, we choose to synthetically generate ground truth images**

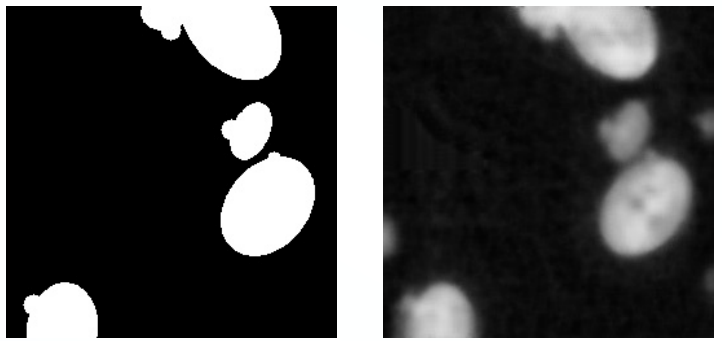
- **SpCycleGAN is an extension of CycleGAN with spatial constraints added to the loss function to maintain the location between the synthetic ground truth mask and the generated nuclei image**

$$\mathcal{L}(G, F, H, D_2, D_1) = \mathcal{L}_{\text{GAN}}\left(G, D_2, I^{\text{bi}}, I^{\text{orig}}\right) + \mathcal{L}_{\text{GAN}}\left(F, D_1, I^{\text{orig}}, I^{\text{bi}}\right) + \lambda_1 \mathcal{L}_{\text{cycle}}\left(G, F, I^{\text{orig}}, I^{\text{bi}}\right) + \lambda_2 \mathcal{L}_{\text{spatial}}\left(G, H, I^{\text{orig}}, I^{\text{bi}}\right)$$

Where I^{bi} is the binary mask and I^{orig} is the original microscopy images, and $\lambda 1$ and $\lambda 2$ are weight coefficients



(a) Training Pair Example

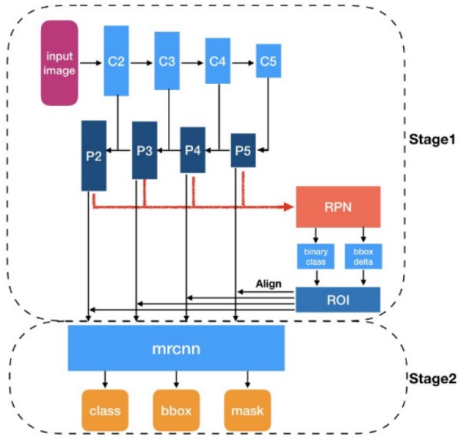


(b) Generated Pair Example

Abnormally Shaped Nuclei Segmentation

- **We train Mask R-CNN for abnormally shaped nuclei segmentation**

- **Mask R-CNN is a two-stage instance segmentation network**
  - **First stage: a region proposal network (RPN) to generate bounding boxes indicating the potential objects (ROIs)**
  - **Second stage: classify and further refine the bounding box of the object, then provide the segmentation mask**
  - **Architecture: feature pyramid network (FPN)**



The block diagram of Mask R-CNN

Using Ensembles

- **Our annotated real dataset is very limited – segmentation model may learn the errors from synthetic images if they overwhelm the training dataset**
- **We decide to divide the synthetic images into small batches and mix them with real image to train multiple Mask R-CNNs, and use ensemble to process the Mask R-CNNs' segmentation**
- **It has been shown that employing ensembles leads to better and more stable performance**
- **False positives are averaged out, meaning the true positives are allowed to show more often**

Weighted Mask Fusion

- **We propose a Weighted Mask Fusion to fuse the output from each Mask R-CNN**
- **We first define three lists containing the segmentation mask:**
  - **A potential mask list B: contains segmentation masks Seg_i^n from each segmentation model and is sorted based on confidence $Prob_i^n$**
  - **A matched mask list L: contains multiple masks Seg_i^n for each object**
  - **A refined mask list F: contains one single mask Seg^n for each object**

- **Iterate through masks in B and try to find a matching in F (IoU matching)**
  - **If not found, add to L and F**
  - **If found, add this box to L at the position corresponding to the matching in list F**

$$Seg^n = \frac{1}{|C_n|} \sum_{i \in C_n} Seg_i^n \cdot Prob_i^n > \tau$$

$$Prob^n = \frac{1}{|C_n|} \sum_{i \in C_n} Prob_i^n$$

C_n : cluster of masks of object n that matched
$|C_n|$: number of masks in this cluster

Convexity Measure

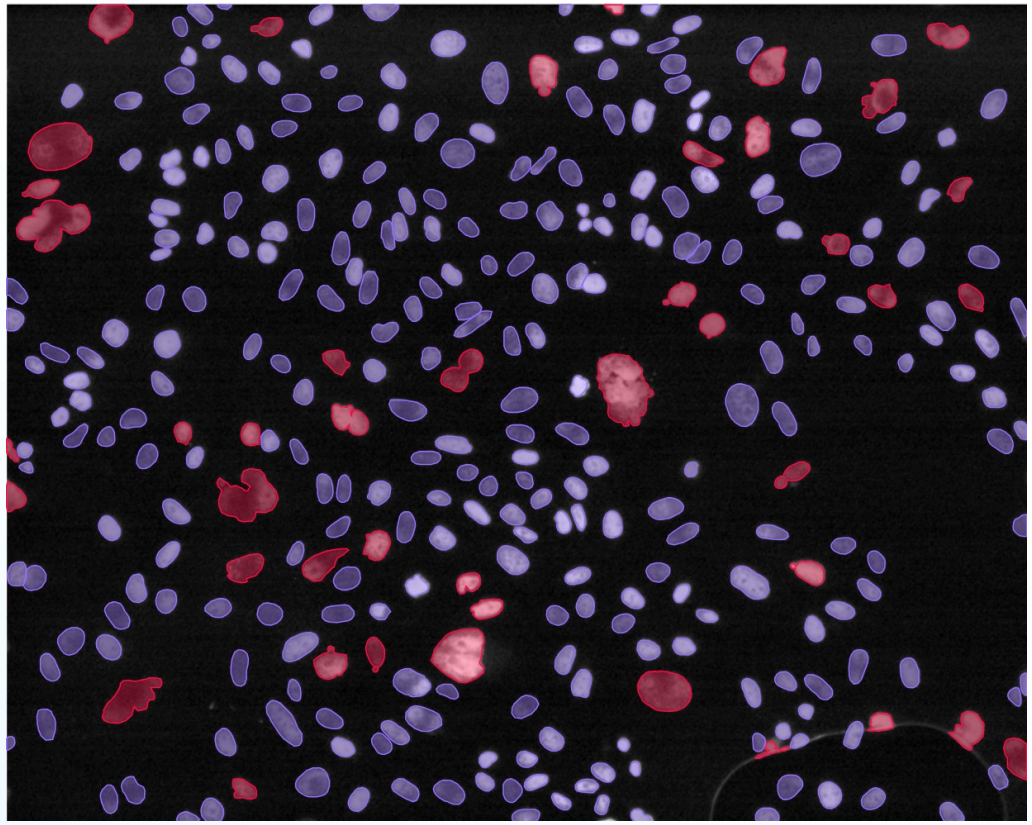
- **Some abnormally shaped nuclei are very similar to typical elliptical nuclei – challenging to distinguished**
- **We use a Convex hull-based convexity measure for shape analysis as a post-processing stage**
  - **Without adding more parameters to the network structure**
  - **Removing the incorrectly segmented elliptical nuclei**
- **We denote the convexity measure C_A by equation:**

$$C_A = \frac{area(S)}{area(CH(S))}$$

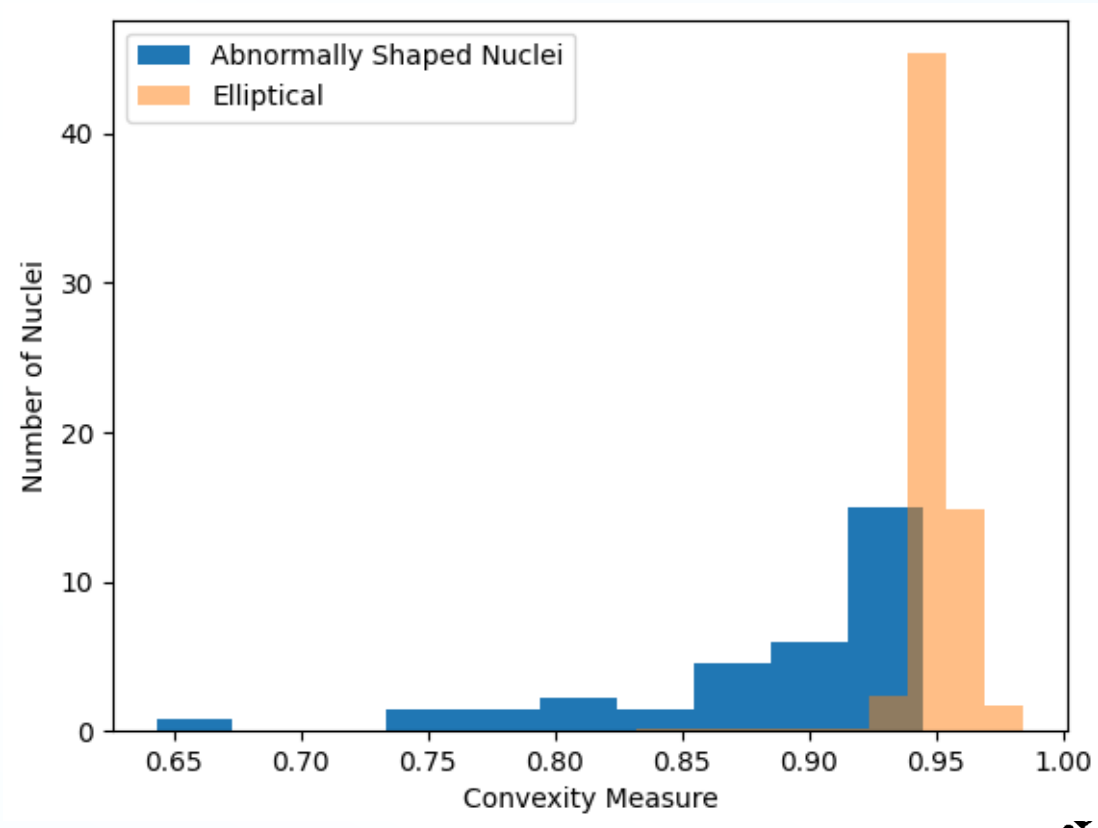
where area(S) is the area of the polygon shape S, and CH(S) is the convex hull of polygon shape S. The range of the convexity measure CA is [0,1] with a higher value meaning that a shape is more convex and closely matches its convex hull

Convexity Measure

- **An example of an annotated microscopy image of nuclei and the distribution of the convexity measure of the nuclei**



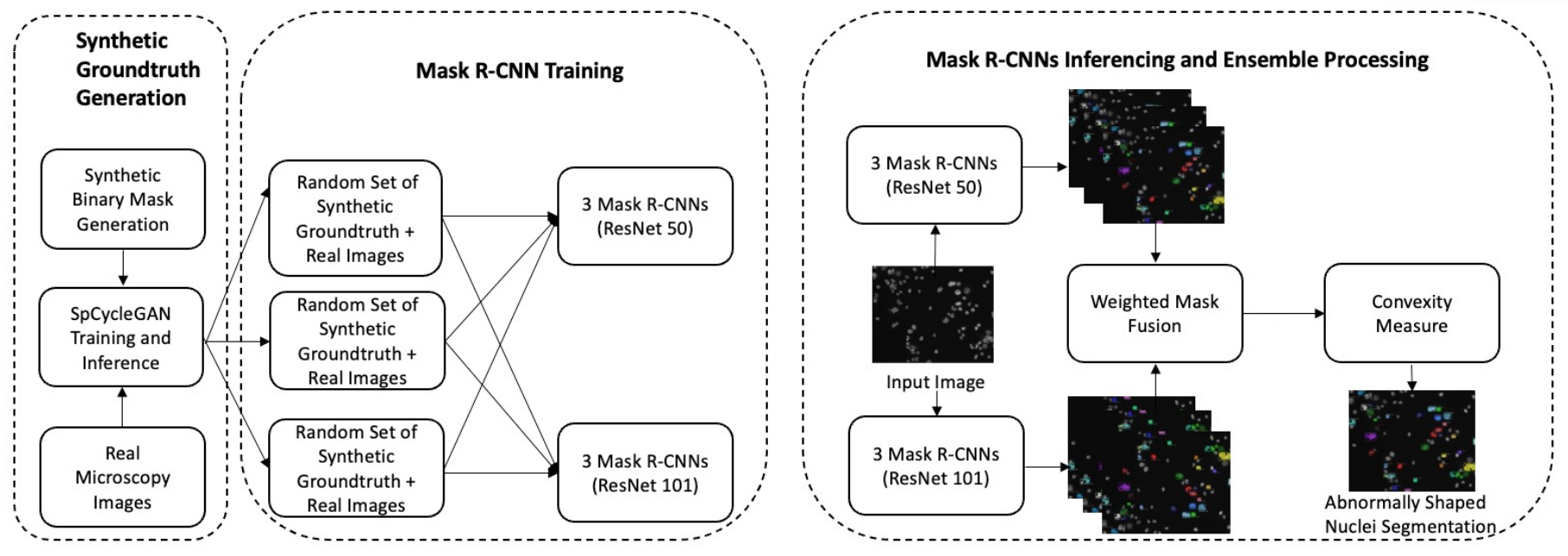
An example of annotated microscopy image



The distribution of nuclei convexity measure

Blue: Elliptical nuclei
Red: Abnormally shaped nuclei

Our Proposed System



Block diagram of our proposed abnormally shaped nuclei segmentation system

Dataset and Experiment

- **We generated 200 128×128 synthetic ground truth binary maps and along with 200 real image patches to form the unpaired training set to train a SpCycleGAN**
- **We generated 600 synthetic nuclei ground truth images**
- **The synthetic ground truth image are randomly split into 3 sets evenly and combined with the same 30 real training images to form 3 training sets**
- **The 3 training sets are used for training 3 Mask R-CNNs with ResNet-50 and 3 Mask R-CNNs with RestNet-101**

Evaluation Metric

- The Precision (P), Recall (R), and F_1 Scores are defined as

$$P = \frac{TP}{TP + FN}, R = \frac{TP}{FP + TN}, F_1 = \frac{2PR}{P + R}$$

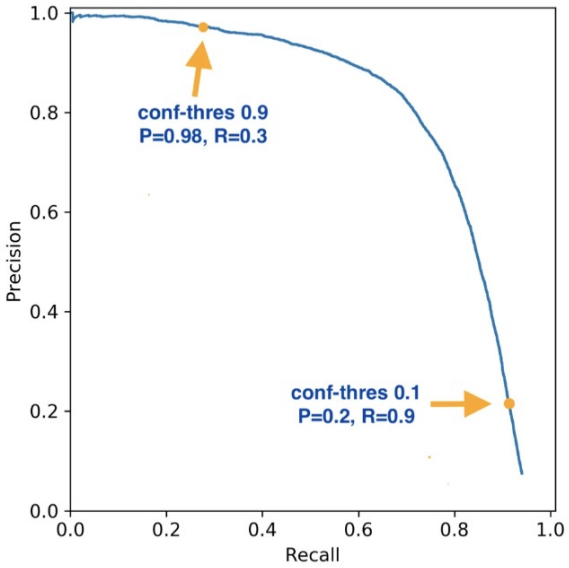
- Average precision (AP) is defined as the Area Under Curve (AUC) of the Precision-Recall Curve (11-points [0, 1, 0.1])

$$AP = AUC = \int_0^1 p(r)dr$$

- $p(r)$ is the Precision given a Recall r
- The mean Average Precision (mAP) is the mean of AP for all IoU thresholds T_{IoU}

$$mAP = \frac{1}{|T_{IoU}|} \sum_{t \in T_{IoU}} AP(t)$$

  - Three commonly used mAP: mAP@[0.5:0.95,0.05], mAP@50, mAP@75



Precision and Recall Curve

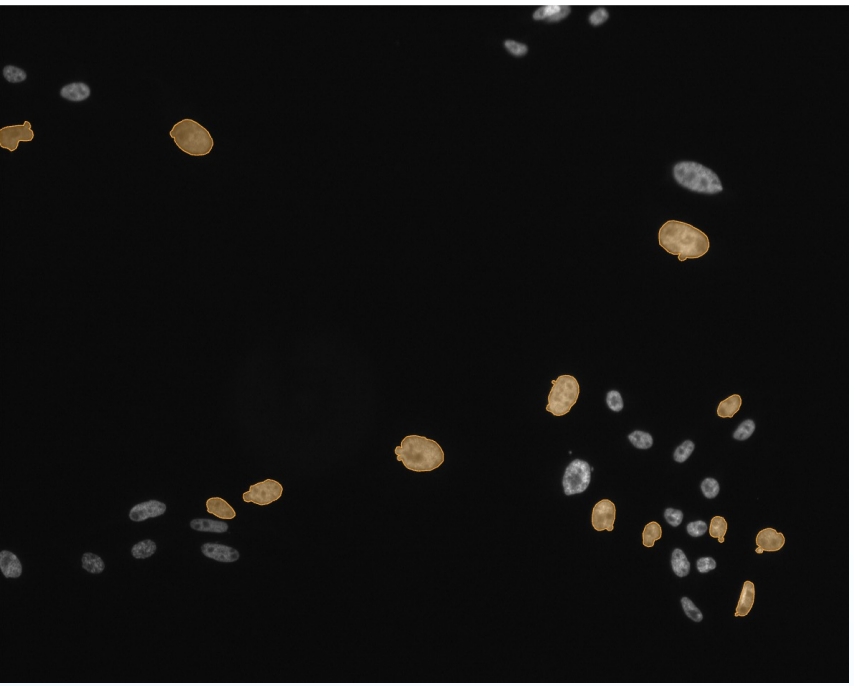
Quantitative Results

Method	AP50	AP75	mAP
U-Net[13]	47.17	36.22	29.21
StarDist[17]	47.64	41.34	32.42
CellPose[18]	63.33	55.97	44.67
MRCNN[12]	65.27	63.41	51.38
MRCNN[12] + Syn	65.98	64.45	52.58
MRCNNs[12] + Syn + WMF	67.15	65.69	53.24
Proposed System	**73.81**	**67.69**	**57.17**

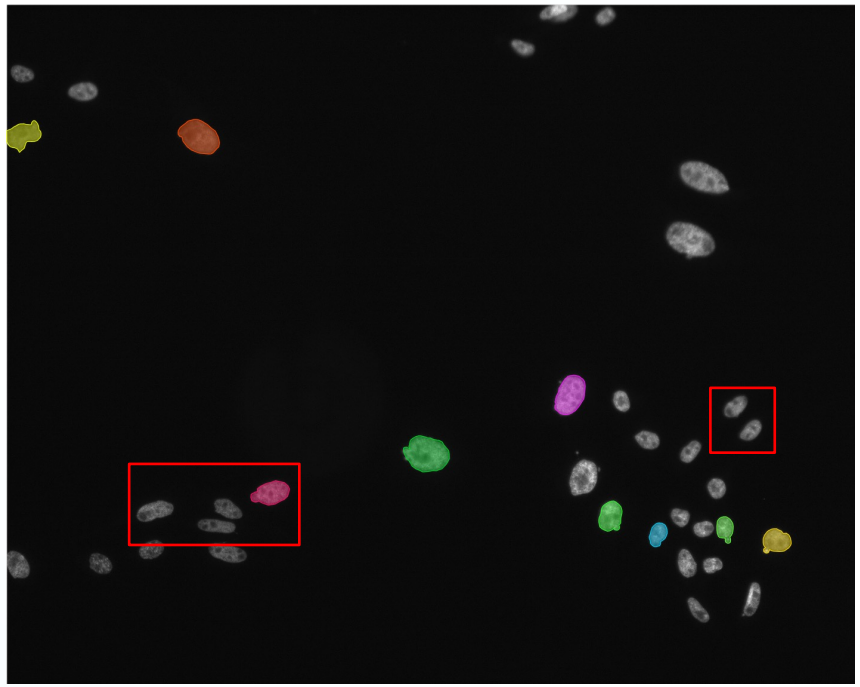
The performance metrics of abnormally shaped nuclei segmentation using AP50, AP75 and mAP (The higher number indicates better performance)

- **For simplicity, Mask R-CNN is written as "MRCNN", methods trained with the extra synthetic ground truth images are denoted as "+syn", and Weighted Mask Fusion is denoted as "WMF"**

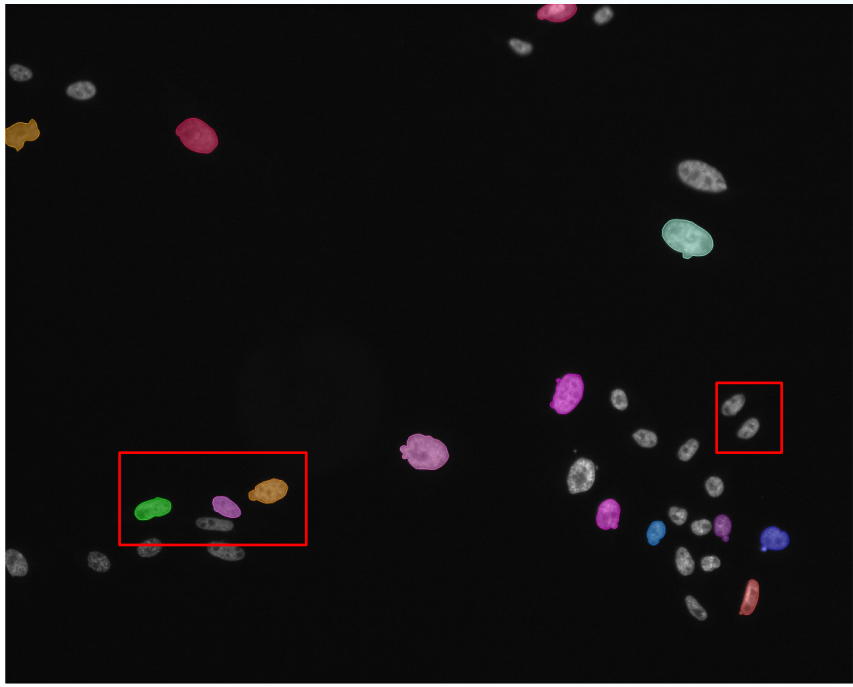
Example: Abnormally Shaped Nuclei Segmentation



Ground Truth Annotation

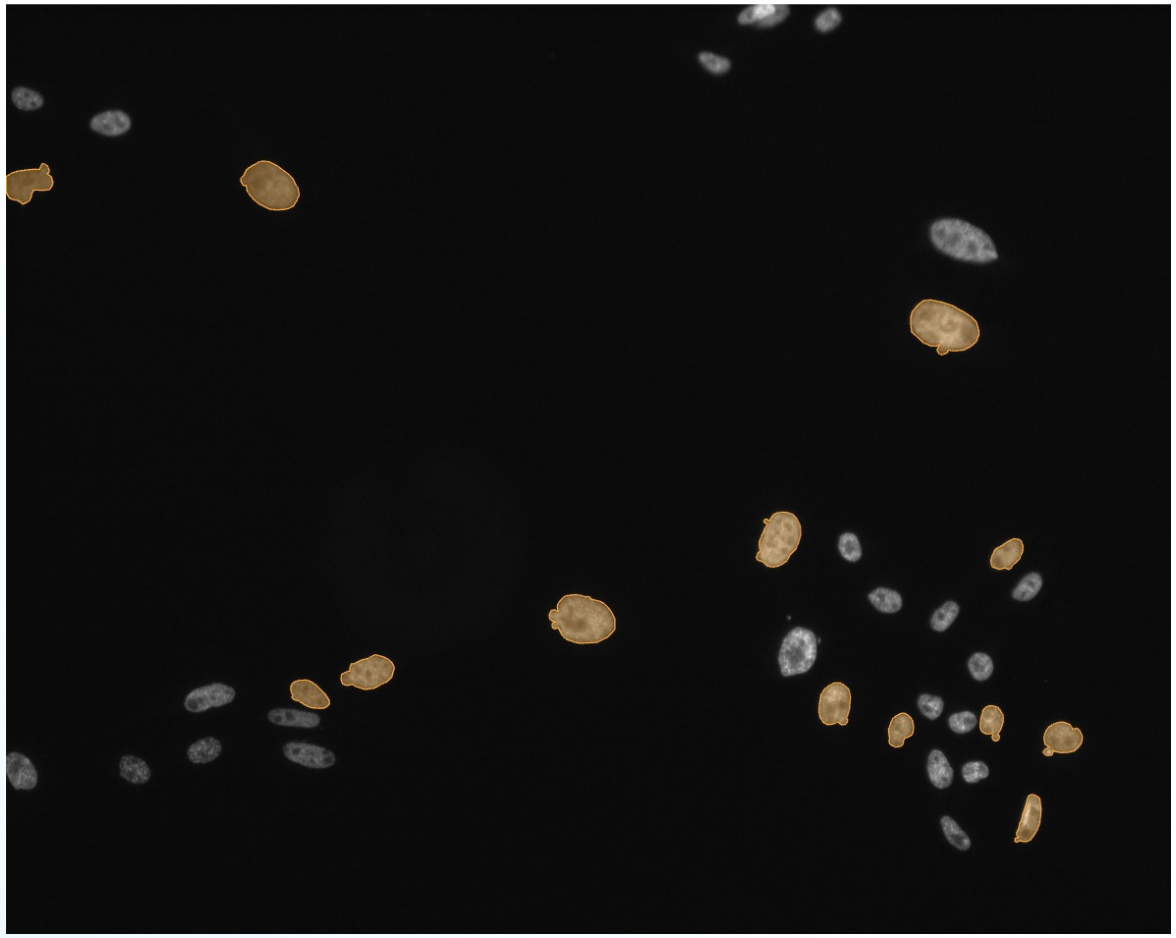


MRCNN+ syn

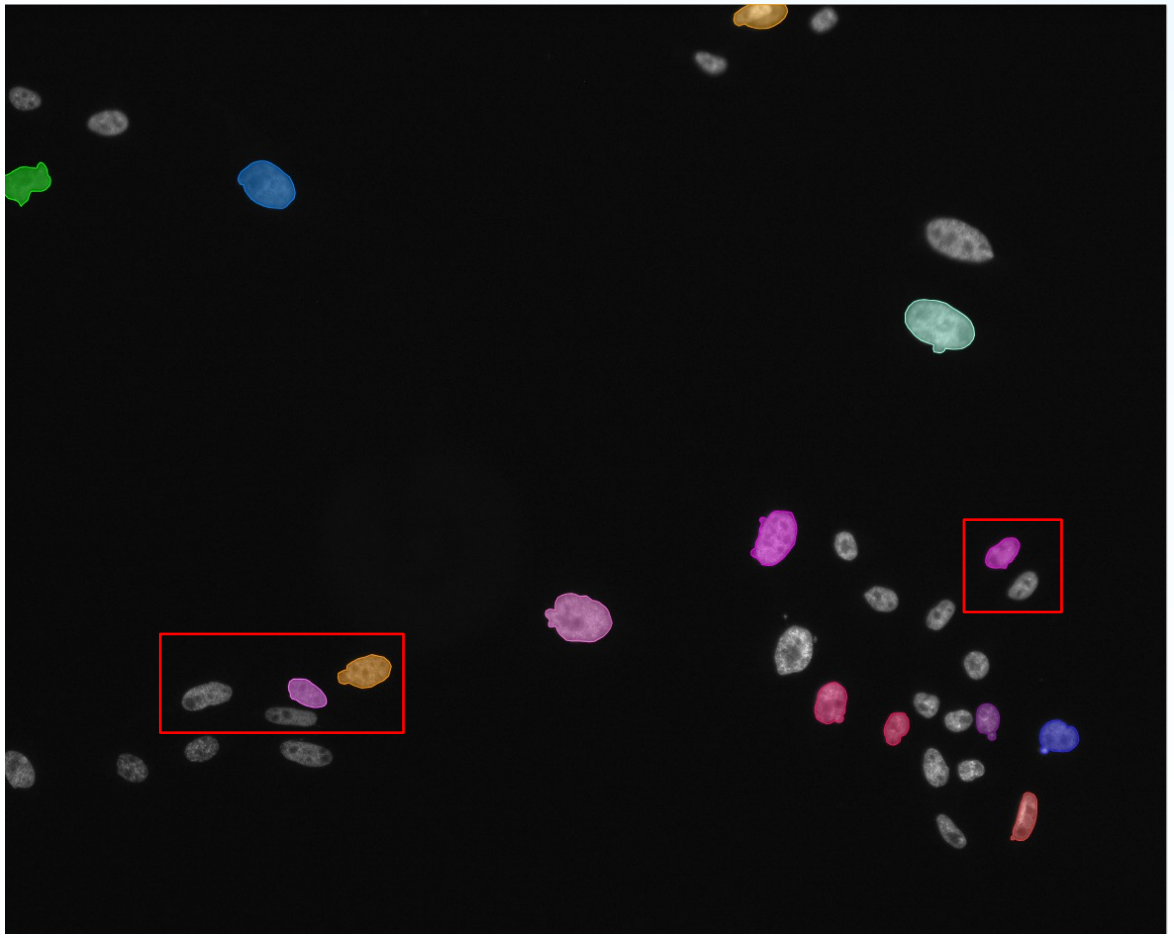


MRCNNs + WMF

Example: Abnormally Shaped Nuclei Segmentation



Ground Truth Annotation



Proposed System

Future Works

- **Investigate different loss functions which could better model the shape of abnormal nuclei**
- **Reduce the use of multiple networks and maintain the robustness of the segmentation result**
- **Explore semi-supervised and weakly-supervised learning approaches, which would allow one to learn from unlabeled image**

An Ensemble Method with Edge Awareness for Abnormally Shaped Nuclei Segmentation

Motivation

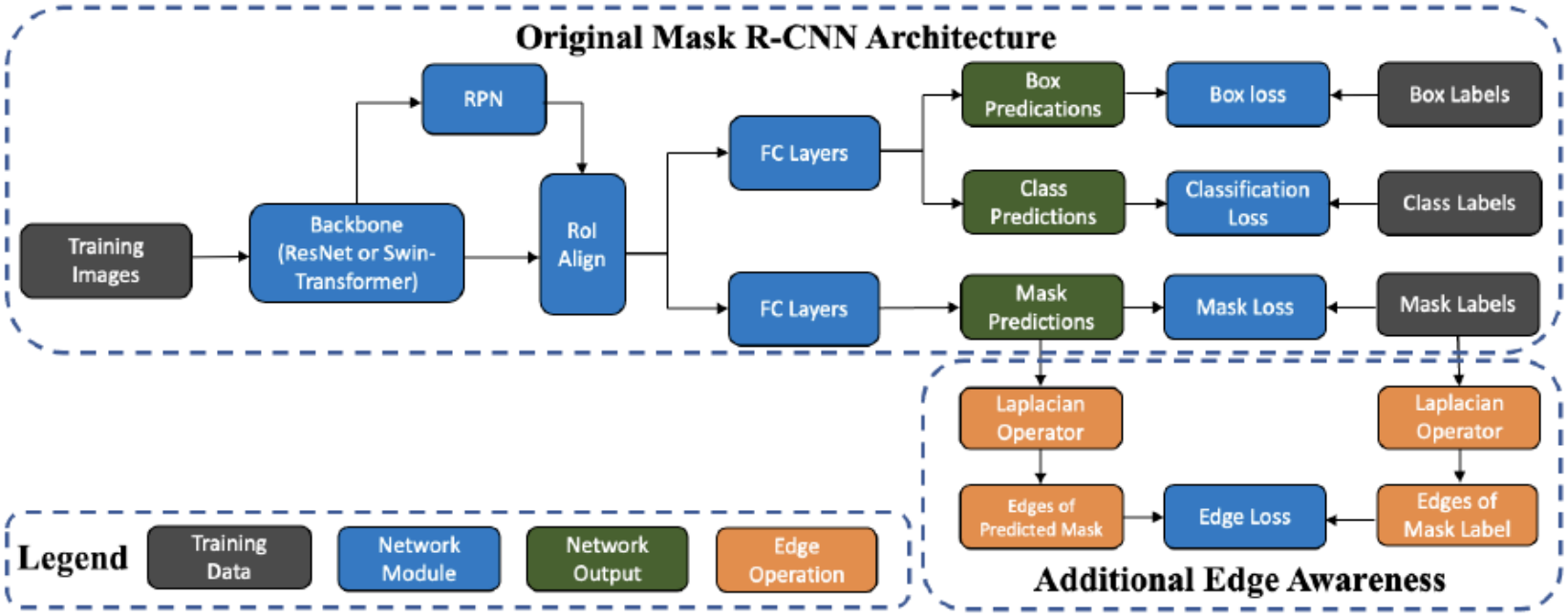
- **In previous work, we improve the ensemble strategy and use the convexity measure to improve the nuclei segmentation result**
- **However, the convexity measure cannot be generalized to other tasks if we want to segment the object other than the elliptical shape.**
- **Transformer-based methods have shown performance, which exceeds that of competing methods, across a wide spectrum in the area of computer vision**
- **We want to design a segmentation method to leverage nuclei boundary information**

Our Goal

- **Can we design a loss function to extract and incorporate the edge information of nuclei into the network to enhance mask prediction**
- **Can we propose a method benefit from both detailed high-resolution spatial information from CNNs and the global context encoded by Transformers**
- **Can we reduce the use of multiple networks and maintain the robustness of the segmentation result?**

Edge-aware Mask R-CNN

- **The main difference between abnormal nuclei and typical elliptical nuclei is the shape of the edge, which sometimes are hard to distinguish**
- **We propose combining an additional edge loss with the existing Mask R-CNN to address this problem – we call it Edge-aware Mask R-CNN**



The block diagram of Edge-aware Mask R-CNN

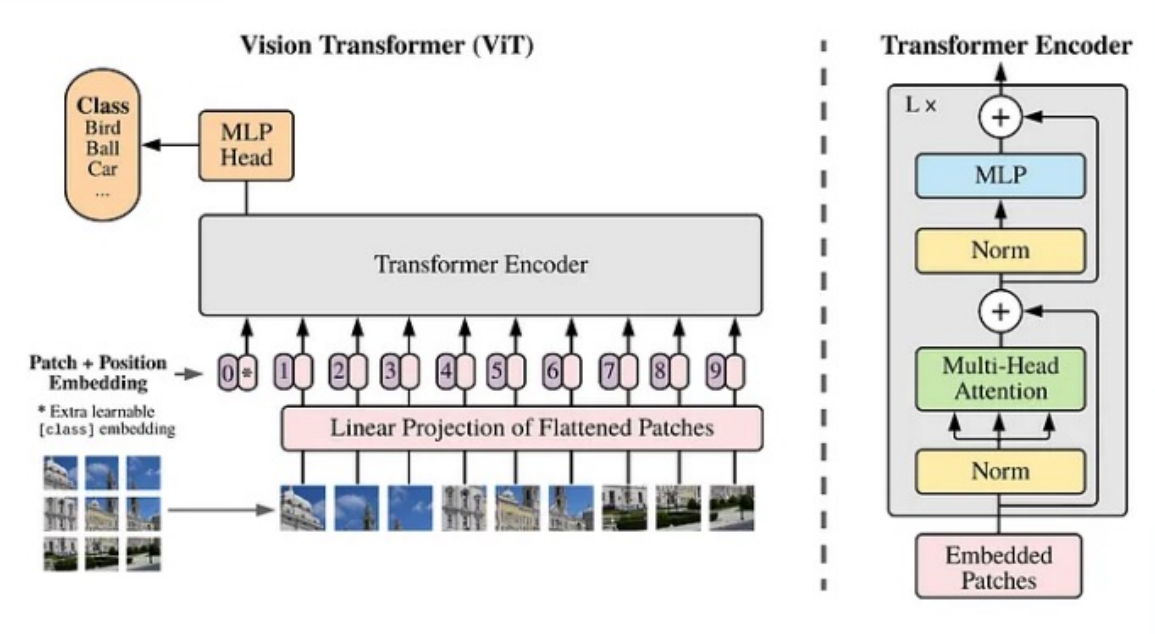
The multi-task loss L is:

$$L = L_{cls} + L_{box} + L_{mask} + \lambda L_{edge}$$

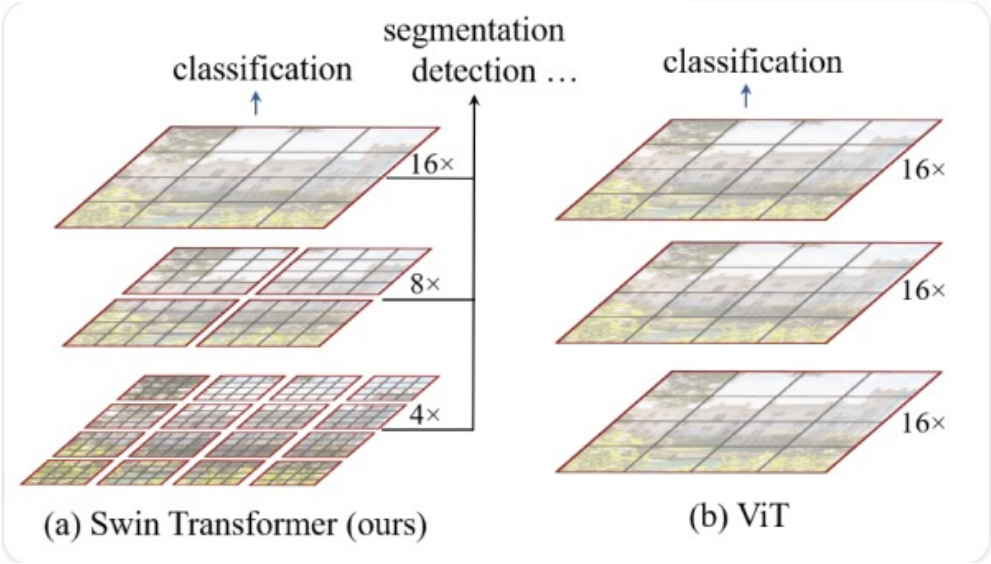
Where L_{cls} is classification loss, L_{box} is bounding box regression loss, and L_{mask} is mask loss L_{edge} is the edge additional loss, and λ is the weight coefficient

Transformer

- **Transformer architectures are based on a self-attention mechanism that learns the relationships between elements of a sequence, hence better at modeling global information**



Network Architecture of a Vision Transformer



Swin Transformer's Hierarchical Feature Maps

Transformer and CNN Ensembles

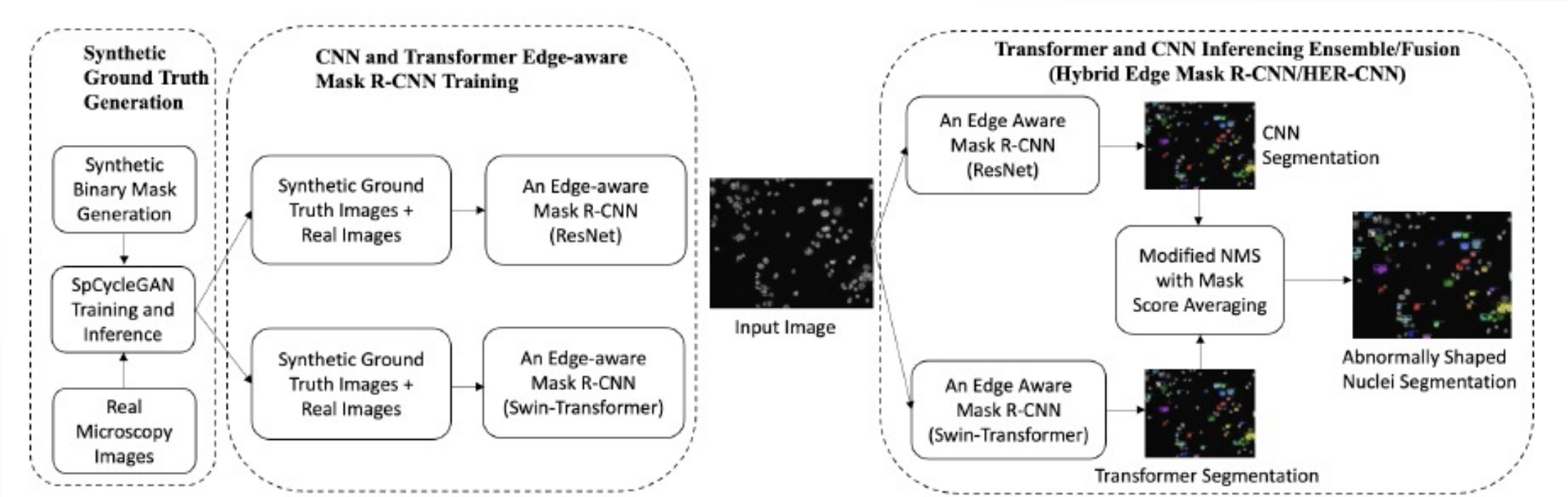
- **However, compared to the CNN-based methods, transformer-based methods sometimes exhibit limitations in extracting detailed localization information**
- **Ensemble processing, where outputs of several segmentation methods are combined or fused, has proven to be useful in improving segmentation performance as well as robustness**
- **We choose to train a ResNet based, and a Swin-Transformer based Edge-aware Mask R-CNN for the segmentation, then fused the result using a modified version of NMS**

$$Seg^n = NMS(Seg_i^n, \tau)$$

$$Prob^n = \frac{1}{M}\sum_{i=0}^{M} Prob_i^n$$

Where M is number of the detector, Seg_i^n is the the segmentation mask of i-th detector of nucleus n, $Prob_i^n$ is its corresponding confidence score. Seg^n and $Prob^n$ are the ensemble mask and confidence score for nucleus n

Hybrid Edge Mask R-CNN (HER-CNN)



The Block Diagram of HER-CNN

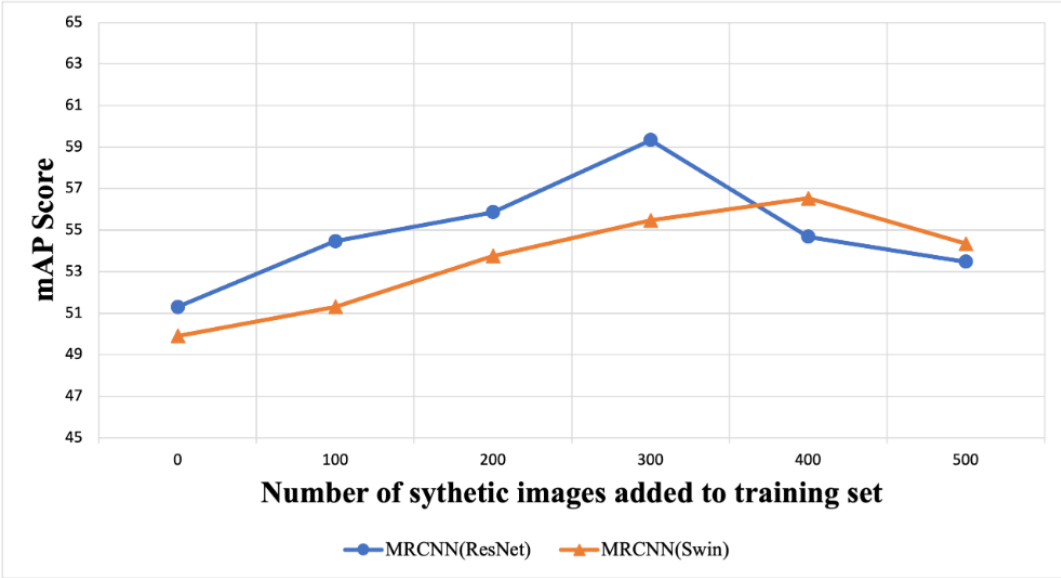
Dataset and Experiment

- We manually annotated abnormally shaped nuclei on 50 fluorescence microscopy images, the nuclei are from CHO-K1 cells stained with ”Hoechst 33342 stain”
  - The size of each image is 2758 × 2208 pixels
  - The images were divided into 30, 5, and 15 images for training, validation, and testing
- We randomly selected 5 images from the training dataset and crop 200 256 × 256 image patches for training the SpCycleGAN to generate synthetic ground truth images
- The training dataset of the Edge-aware Mask R-CNN with ResNet-50 consists of 300 synthetic ground truth images with 30 real images, for the Edge MRCNN with with Swin-S Transformer consists of 400 synthetic ground truth images with the same 30 real images
- We also conducted ablation study on the effect of the use of synthetic ground truth images in trainings and the effect of the L_{edge} in the Edge-aware Mask R-CNN

Ablation Study

Method	AP50	AP75	mAP
MRCNN	64.73	63.11	51.46
Edge MRCNN (λ=0.1)	67.10	64.69	52.97
Edge MRCNN (λ=0.2)	**69.37**	**66.61**	**54.48**
Edge MRCNN (λ=0.3)	67.03	66.43	52.94

The abnormally shaped segmentation performance with different weights on edge loss



Mean Average Precision (mAP) as a function of the numbers of generated microscopy images in the training set

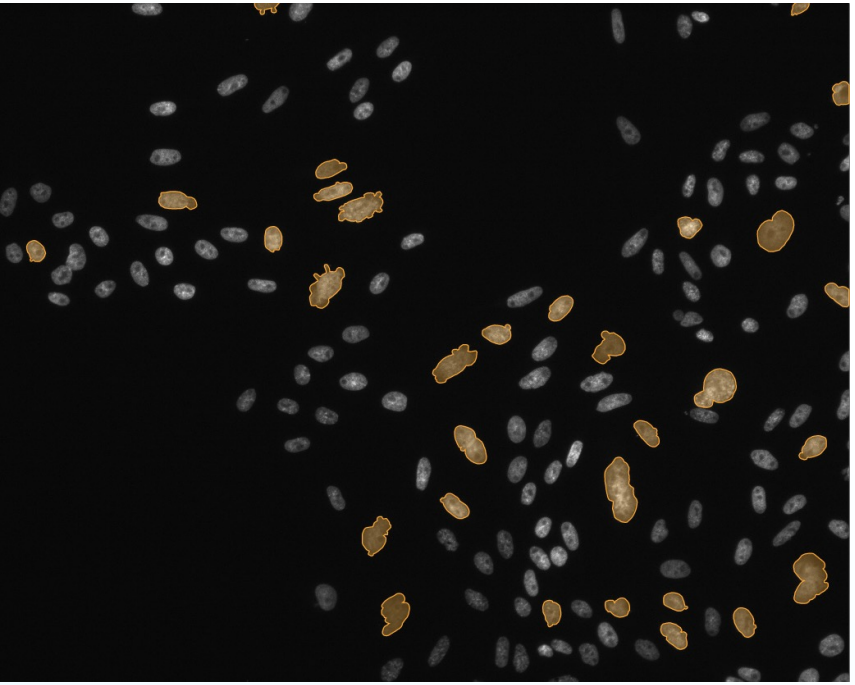
Quantitative Results

Method	AP50	AP75	mAP
U-Net [37]	47.17	36.22	29.21
StarDist [44]	47.64	41.34	32.42
CellPose [41]	63.33	55.97	44.67
MRCNN (ResNet) [21]	65.27	63.41	51.38
Edge MRCNN (ResNet)+syn	73.74	72.87	58.77
Edge MRCNN (Swin)+syn	73.83	71.36	57.42
HER-CNN	**76.07**	**75.04**	**61.35**

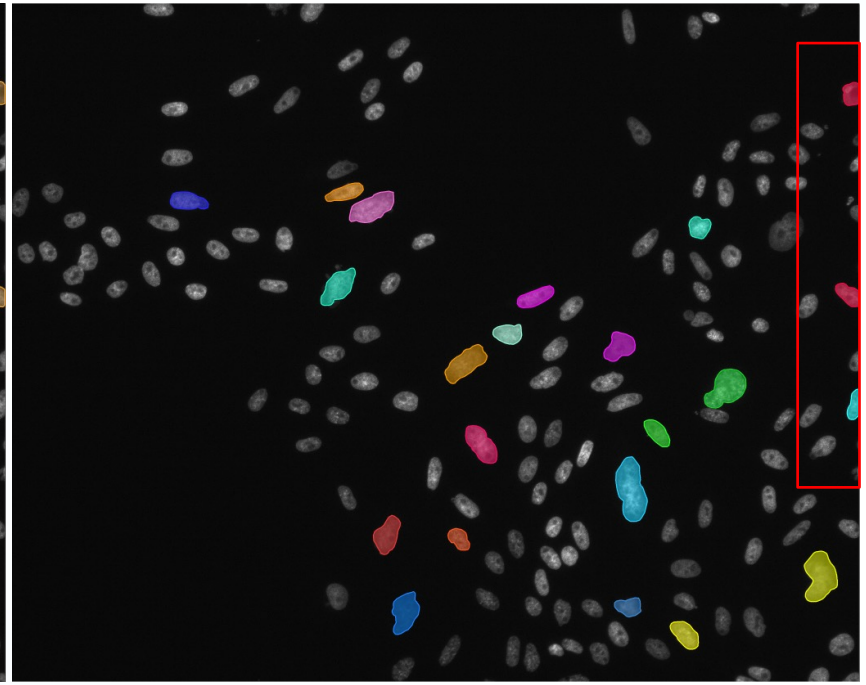
The performance metrics of abnormally shaped nuclei segmentation using AP50, AP75 and mAP (The higher number indicates better performance)

- **For simplicity, Mask R-CNN is written as "MRCNN", Edge-aware Mask R-CNN is written as "Edge MRCNN", and methods trained with the extra synthetic ground truth images are denoted as "+syn"**

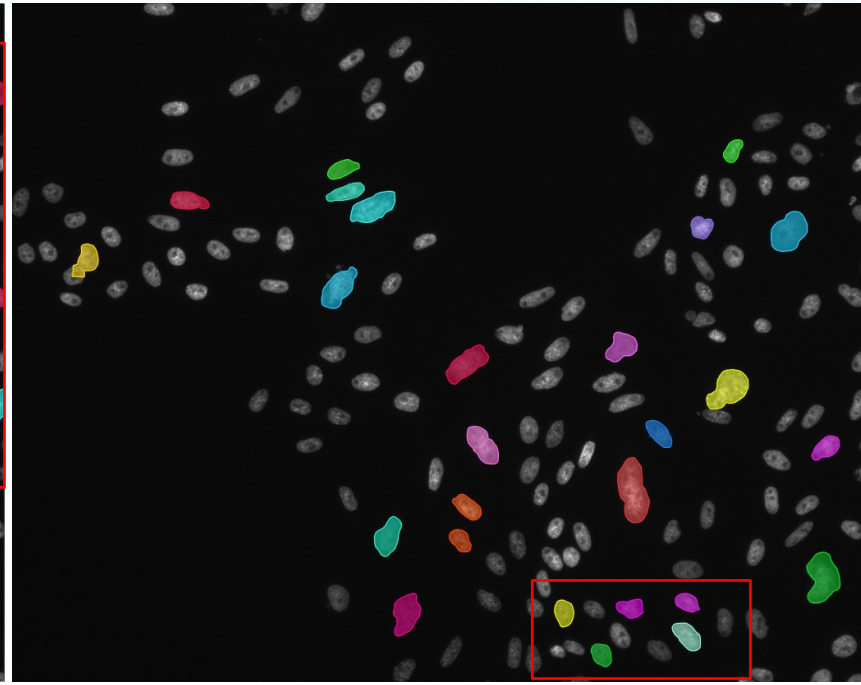
Example: Abnormally Shaped Nuclei Segmentation



Ground Truth Annotation

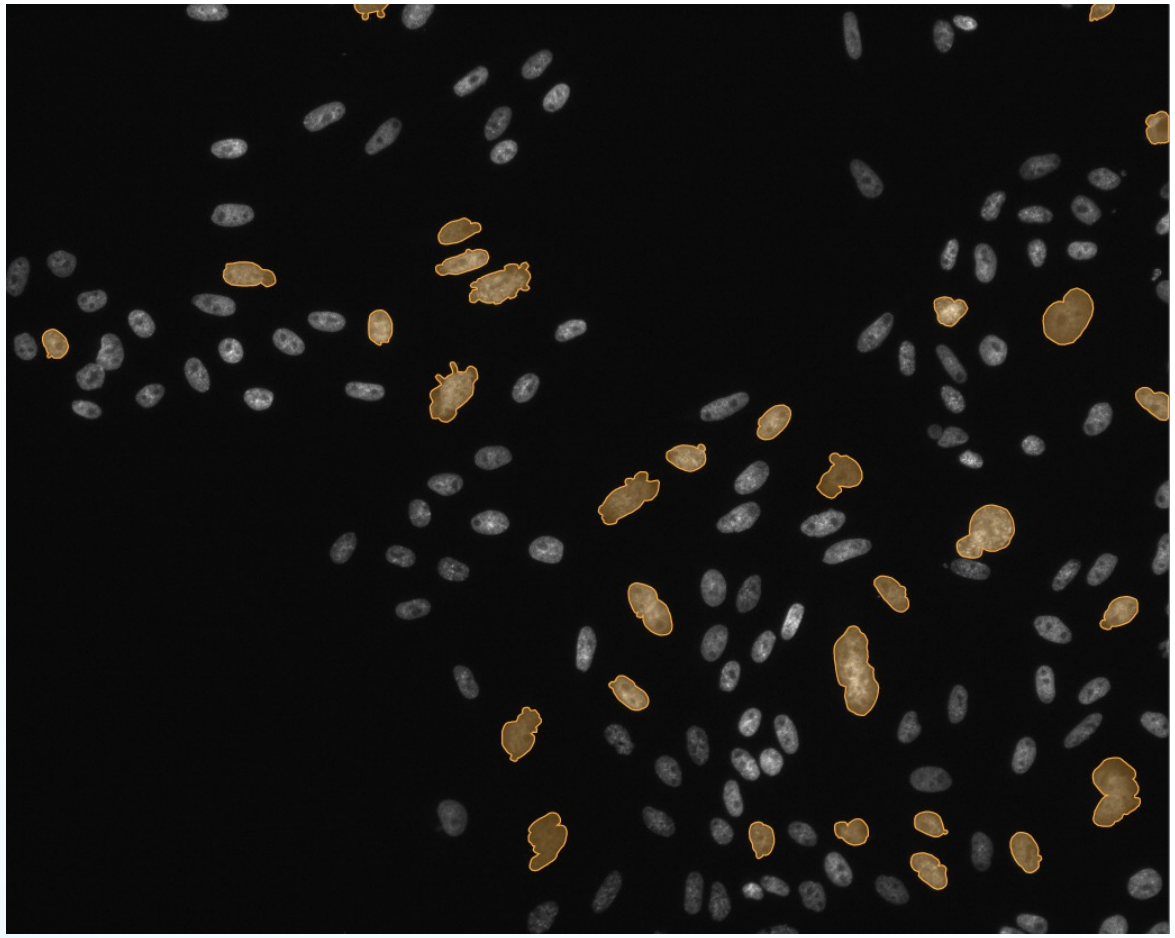


Edge MRCNN (ResNet) + syn

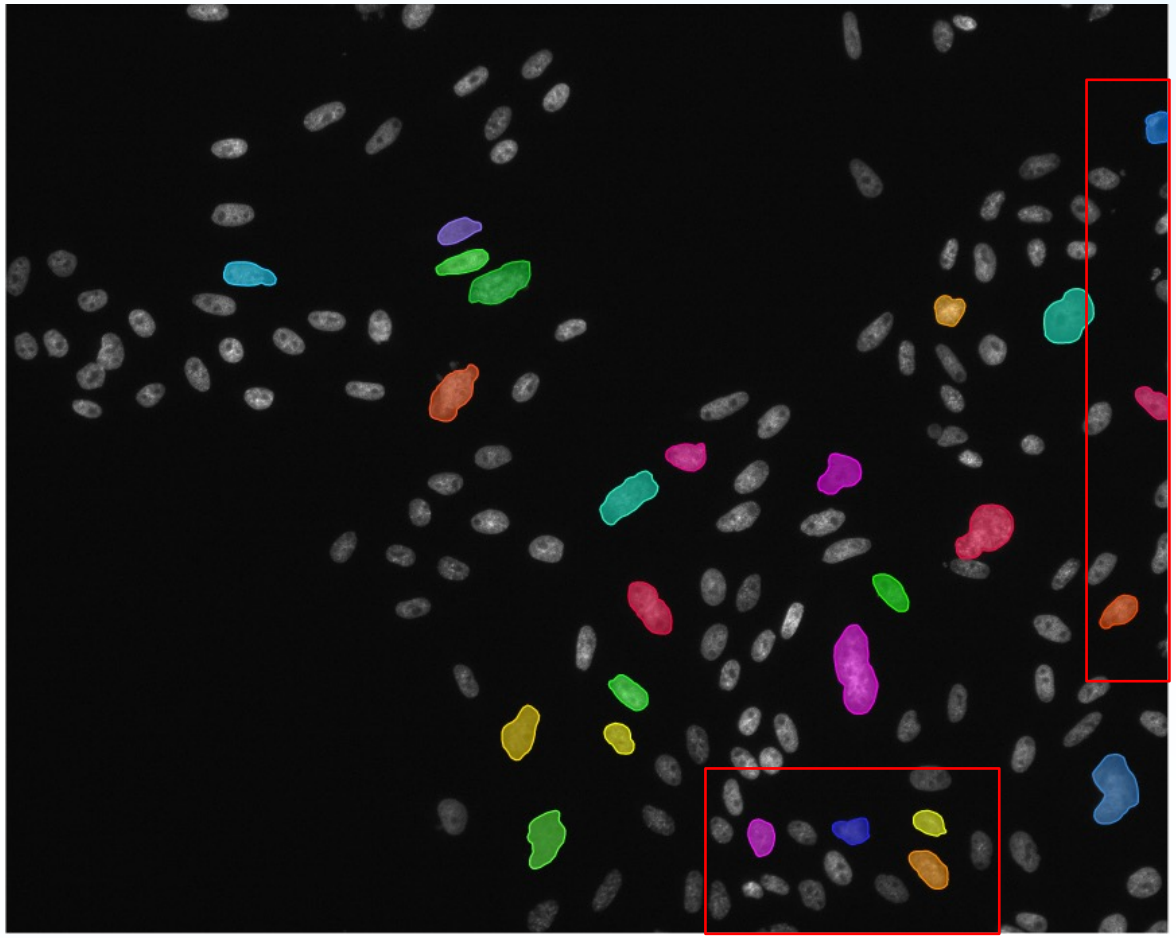


Edge MRCNN (Swin) + syn

Example: Abnormally Shaped Nuclei Segmentation



Ground Truth Annotation



HER-CNN

Future Works

- **Investigate ensemble methods at the network structure level to better incorporate the information**
- **Improve the quality of the synthetically generated ground truth images, which could reduce the need for real annotated image**
- **Explore semi-supervised and weakly-supervised learning approaches, which would allow one to learn from unlabeled image**

Diffusion Model with Clustering-based Conditioning for Food Image Generation

Motivation

- **Dietary intake has a profound impact on personal health and disease. A healthy diet is an essential aspect of keeping lifelong well-being for current and future generations across the lifespan**

- **While image-based dietary assessment methods reduce the time and labor required for nutrient analysis, the major challenge with deep learning-based approaches is that the performance is heavily dependent on the quality and the quantity of the datasets**

- **One of the major issues of food image dataset is that the food data usually suffers from high intra-class variance, another is class imbalance**

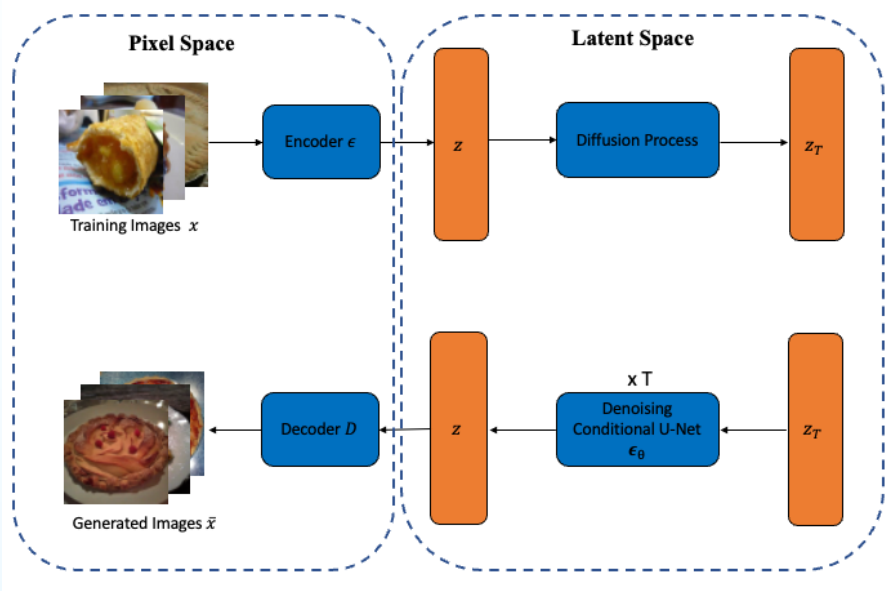
- **Diffusion models have recently shown great capability in image generation, however, there is no existing work utilizing the diffusion model for food image generation**

Our Goal

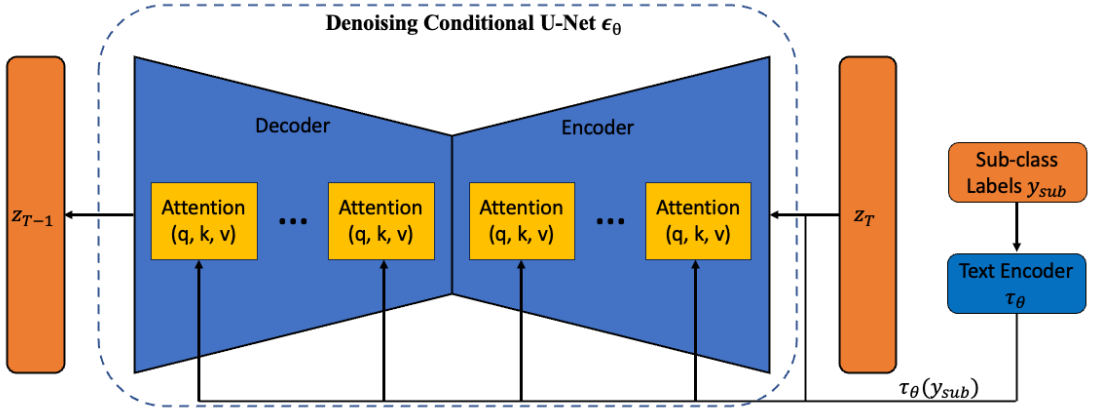
- **Can we explore the performance of latent diffusion methods on food image generation**
- **Can we design a clustering method to reduce the high intra-class variance issue when training the food image generation mode**
- **Can we use synthetic images for data augmentation to address the class-imbalance issue in food dataset**

Latent Diffusion Model

- The diffusion model defines a Markov chain to gradually add random noise to the data, and the model learns to reverse the diffusion process
- As a result, the trained model can construct the desired data samples from the noise



The block diagram of Latent Diffusion Model



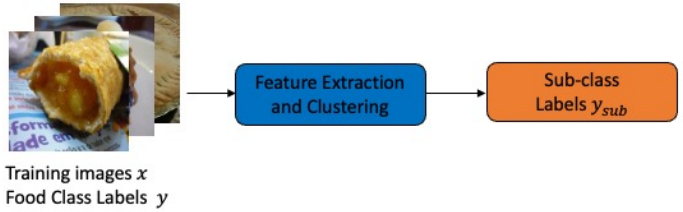
The block diagram of the denoising conditional U-Net architecture with cross-attention mechanism

Image Clustering using Affinity Propagation

- **We utilize Affinity Propagation to cluster the training images from one class into sub-classes to address the high intra-class variance of food dataset**
  - **A pre-trained ResNet-18 on Food-101 dataset is used to map the images into feature vectors**
  - **These feature vectors represent the visual features of the input food images and are used for clustering, cosine distance is used as the similarity between two vectors**
  - **Affinity propagation creates clusters by sending messages between samples until convergence, and it does not require a pre-defined number of clusters**

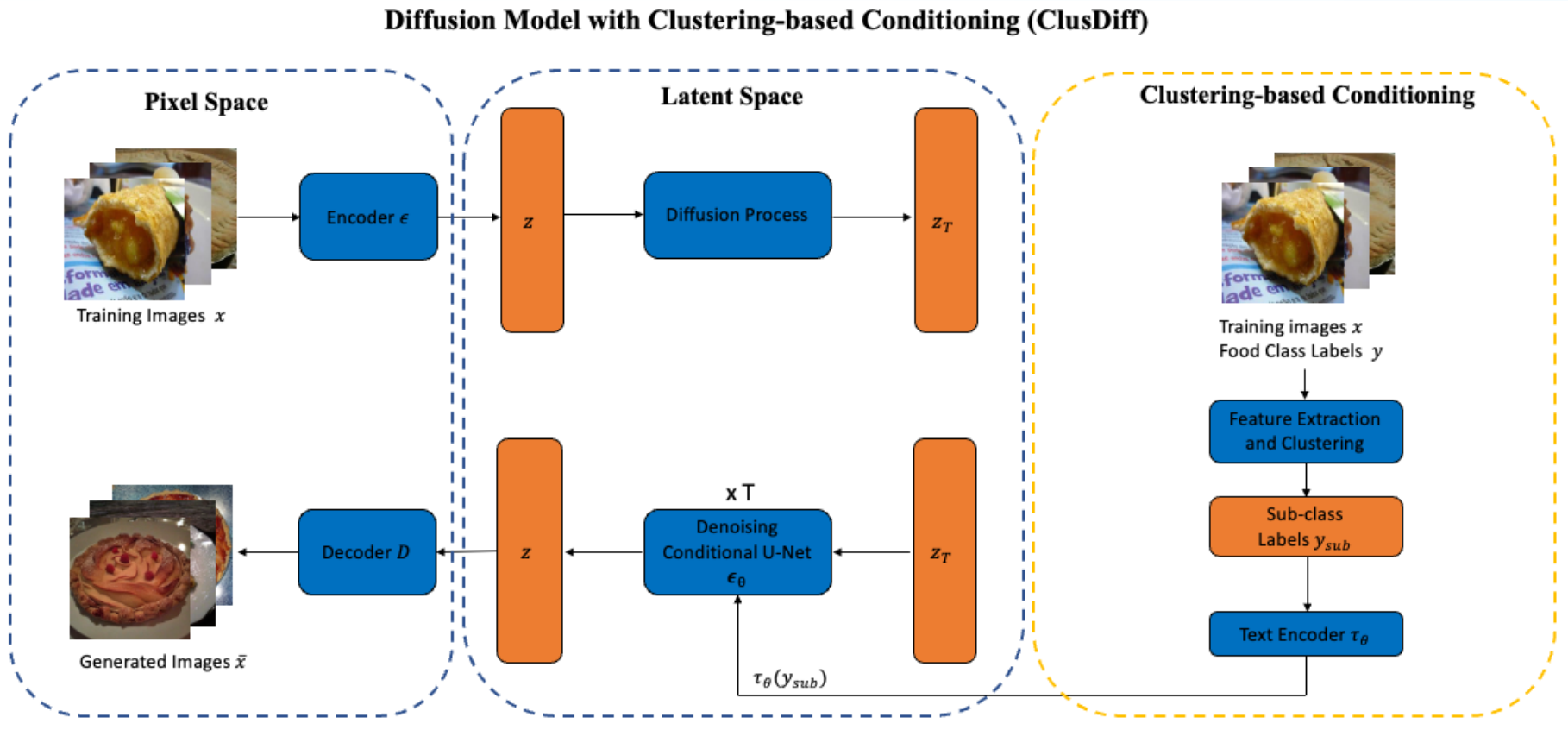
- **In Affinity Propagation, two messages sent between the sample are responsibility r and availability a**
  - **Responsibility:** $r(i,j) = s(i,j) - max[0, a(i,j') + s(i,j')\forall j' \neq j]$
  - **Availability:** $a(i,j) = min[0, a(j,j) + \sum_{i' \notin \{i,k\}} r(i',j)]$
  - **Update functions:**

$$r_{n+1}(i,j) = \lambda \cdot r_n(i,j) + (1-\lambda) \cdot r_{n+1}(i,j)$$
$$a_{n+1}(i,j) = \lambda \cdot a_n(i,j) + (1-\lambda) \cdot a_{n+1}(i,j)$$



Food Image Clustering

Clustering-based conditional training strategy for latent diffusion model (ClusDiff)

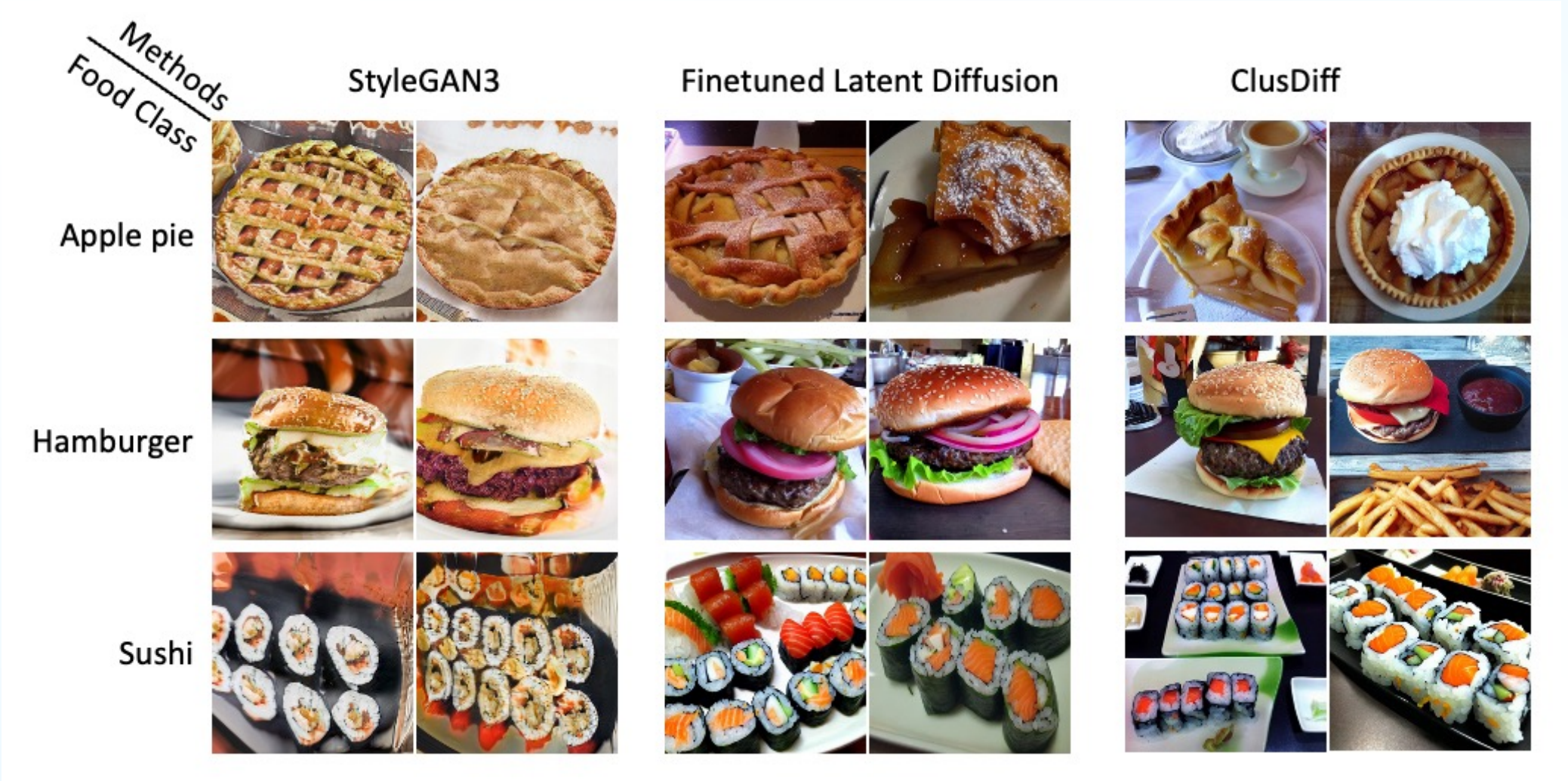


The block diagram of our proposed system, ClusDiff

Dataset and Evaluation Metrics

- **We use Food-101 dataset which consists of 101 food categories and each category contains 1000 images to train the image generation models**
- **Frechet Inception Distance Metric (FID) is main evaluation metric for image generation**
- **We use VFN-LT dataset which containing 2.5K images from 74 common food classes with a maximum of 288 images per class and a minimum of 1 image per class to for long-tailed classification**
  - **Generated food images is used to construct a balanced training set**
- **We use Top-1 image classification accuracy as the evaluation metric. In addition, we provide the performance on both head (instance-rich) and tail (instance-rare) classes**

Food Image Generation Examples



Numerical Evaluations

Datasets	Food-101
Methods	FID
StyleGAN3[41]	39.05
Finetuned Latent Diffusion [65]	30.39
ClusDiff	**27.73**

Food image generation results in terms of FID. The lower FID score indicates better diversity and fidelity of generated image.

Datasets	VFN-LT		
Methods	Head (%)	Tail (%)	Overall (%)
Baseline	62.3	24.4	35.8
ROS[76]	61.7	24.9	35.9
RUS[5]	54.6	26.3	34.8
CMO[58]	60.8	33.6	42.1
LDAM[6]	60.4	29.7	38.9
BS[63]	61.3	32.9	41.9
IB[59]	60.2	30.8	39.6
Focal[49]	60.1	28.3	37.8
Food2stage[25]	61.9	37.8	45.1
Baseline + ClusDiff	**68.7**	**42.4**	**49.5**

Top-1 accuracy on VFN-LT with tail class (Tail) and head class (Head) accuracy

Future Works

- **Utilize the latent vectors extracted by the diffusion encoder to share mutual information for clustering**
- **Integrate cluster information into the loss function of the diffusion model would better capture the variance of food images**
- **Explore better evaluation metrics to accurately assess the variance of the generated food images**

Contributions of This Work

- **We modify the Kalman filter in DeepSORT to use prior knowledge of the object's velocity to assist the motion estimation and improve the tracking result while outperforming other approaches in the literature**

- **We propose the Weighted Mask Fusion (WMF), combining the results of multiple segmentation model, to improve the final nuclei segmentation mask**

- **We propose an edge-aware Mask RCNN method, and introduced a hybrid architecture, an ensemble of CNNs and Swin-Transformer Edge Mask R-CNNs, to accurately segment irregularly shaped nuclei in microscopy images**

- **We propose an effective clustering-based training framework, named ClusDiff, for generating high-quality and representative food images**

Publications

Publications Resulting From This Thesis:

- **Y. Han**, Y. Lei, V. Shkolnikov, D. Xin, S. Barcelo, J. Allebach, E. J. Delp, "Label-Free Mammalian Cell Tracking Enhanced by Precomputed Velocity Fields", *Proceedings of the IEEE International Conference on Machine Learning and Application (ICMLA)*, December 2022, Nassau, Bahamas

- **Y. Han**, Y. Lei, V. Shkolnikov, D. Xin, A. Auduong, S. Barcelo, J. Allebach, E. J. Delp, "Ensemble Processing and Synthetic Image Generation for Abnormally Shaped Nuclei Segmentation", Feburary 2023, *bioRxiv*

- **Y. Han**, Y. Lei, V. Shkolnikov, D. Xin, A. Auduong, S. Barcelo, J. Allebach, E. J. Delp, "Ensemble Processing and Convexity Measure for Abnormally Shaped Nuclei Segmentation", SPIE Medical Imaging, February 2024, San Diego, CA

- **Y. Han**, Y. Lei, V. Shkolnikov, D. Xin, A. Auduong, S. Barcelo, J. Allebach, E. J. Delp, "An Ensemble Method With Edge Awareness for Abnormally Shaped Nuclei Segmentation", IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshop (CVPR), June 2023, Vancouver, Canada

- **Y. Han,** J. He, M. Gupta, E. J. Delp and F. Zhu "Diffusion Model with Clustering-based Conditioning for Food Image Generation", 8th International Workshop on Multimedia Assisted Dietary Management (MADiMa), October 2023, Ottawa, Canada

Publications

Publications Not Resulting From This Thesis:

- G. Tonmoy, **Y. Han**, V. Raju, D. Hossain, M. A. McCrory, J. Higgins, C. Boushey, E. J. Delp, and E. Sazonov. "Integrated image and sensor-based food intake detection in free-living." Scientific Reports 14, no. 1: 1665, January 2024

- W. Fu, **Y. Han**, J. He, S. Baireddy, M. Gupta, and F. Zhu, "Conditional Synthetic Food Image Generation", Proceedings of the Imaging and Multimedia Analytics in a Web and Mobile World Conference, IS&T Electronic Imaging Symposium, February 2023, Burlingame, CA

- Z. Shao, **Y. Han**, J. He, R. Mao, J. Wright, D. Kerr, C. J. Boushey, and F. Zhu, "An Integrated System for Mobile Image-Based Dietary Assessment", Proceedings of the 3rd Workshop on AIxFood, ACM Multimedia, October 2021, Chengdu, China (Hybrid)

- **Y. Han**, S. K. Yarlagadda, T. Ghosh, F. Zhu, E. Sazonov, and E. J. Delp, "Improving Food Detection For Images From a Wearable Egocentric Camera", Proceedings of the Imaging and Multimedia Analytics in a Web and Mobile World Conference, IS&T Electronic Imaging Symposium, February 2021, Burlingame, CA (Virtual)

LIVE CELL IMAGING ANALYSIS USING MACHINE LEARNING AND SYNTHETIC FOOD IMAGE GENERATION

Yue Han

Major Advisor: Edward J. Delp

Video and Image Processing Laboratory (*VIPER*)
School of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana USA

Backup: Representative Velocity Field

- **Since the chip utilizes a dielectrophoretic force field to separate the cells, we can have precomputed velocity fields which are theoretical velocities of the cells in the chip – we shall call this the *representative velocity field***

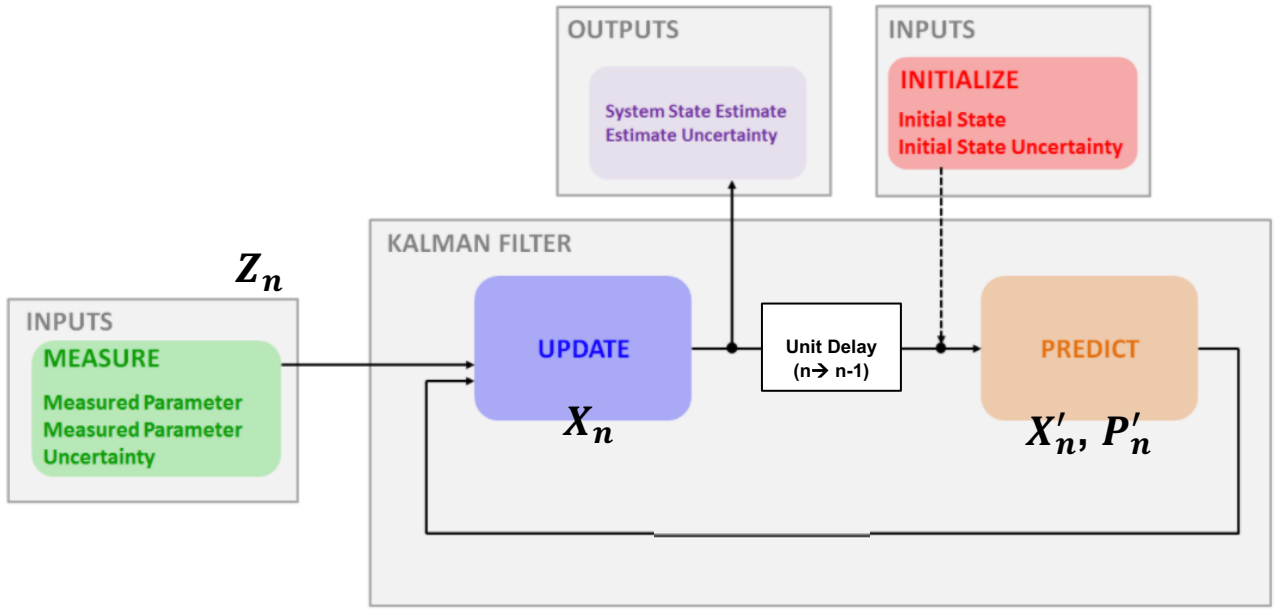
- **The cells' velocities in each experiment may varies from the velocities obtained from the representative velocity filed due to:**
  - **different types of cells**
  - **different flowrates**
  - **interactions between cells**
  - **interactions with the chip boundaries.**

Backup: Label-Free Mammalian Cell Tracking

- **Ideal cell tracking here should utilize both cell motion and cell appearance features for better track association**
- **Take advantage of the representative velocity field to deal with the unique motion pattern in the chip**
- **We modify DeepSORT* and the Kalman Filter** in DeepSORT to incorporate the precomputed velocities**
- **Kalman filter can effectively produce motion estimation to predict the locations of the cells based on inaccurate and uncertain measurements**

***Nicolai, Wojke, et al . "SIMPLE ONLINE AND REALTIME TRACKING WITH A DEEP ASSOCIATION METRIC," *ICIP* 2017**
**** Greg, Welch et al. "An Introduction to the Kalman Filter," *ACM SIGGRAPH Course* 2001**

Backup: Kalman Filter



The Block Diagram of Kalman Filter

X'_n is the predicted system state at time stamp n, P'_n is the uncertainty of the predicted system state (covariance), X_n is the system state estimation, Z_n is the measurement, R is the measurement uncertainty (covariance), F is the transition matrix , H is observation matrix

- Kalman filter can provide a prediction of the future system state, based on the past estimation of the motion and the inaccurate or uncertain measurement of the newly detected cells' motion

- Next State Prediction

$$X'_n = FX_{n-1}$$
$$P'_n = FP_{n-1}F^T$$

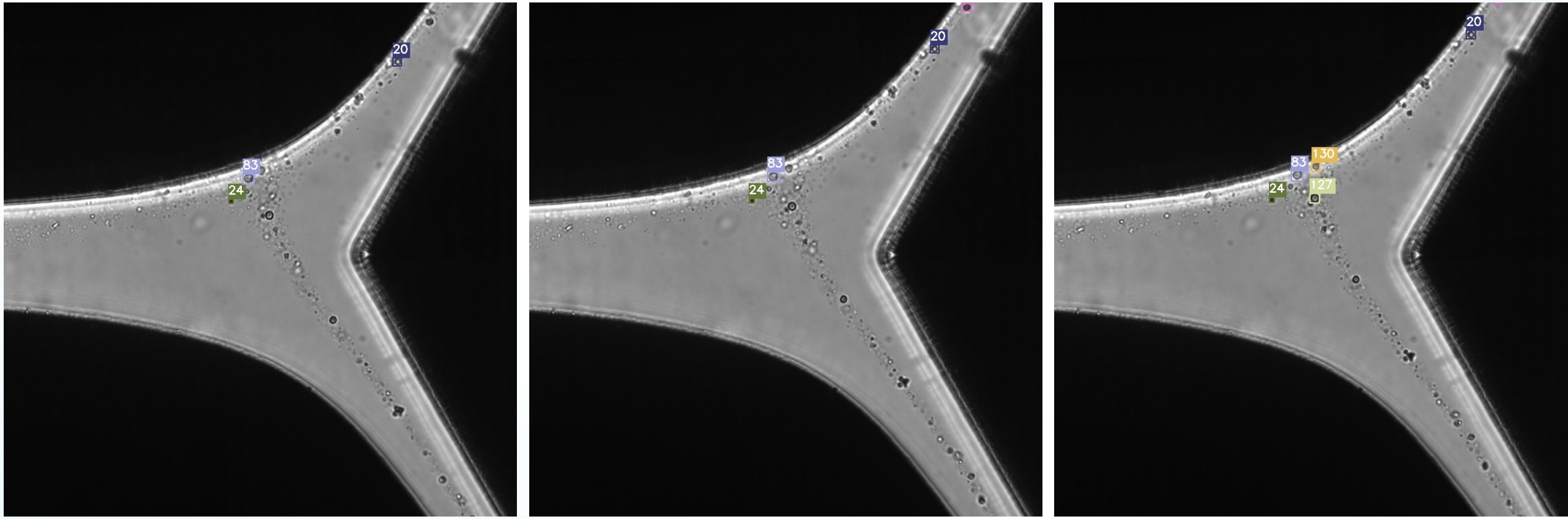
- Kalman Gain Calculation

$$K_n = \frac{P'_n H^T}{HP'_n H^T + R}$$

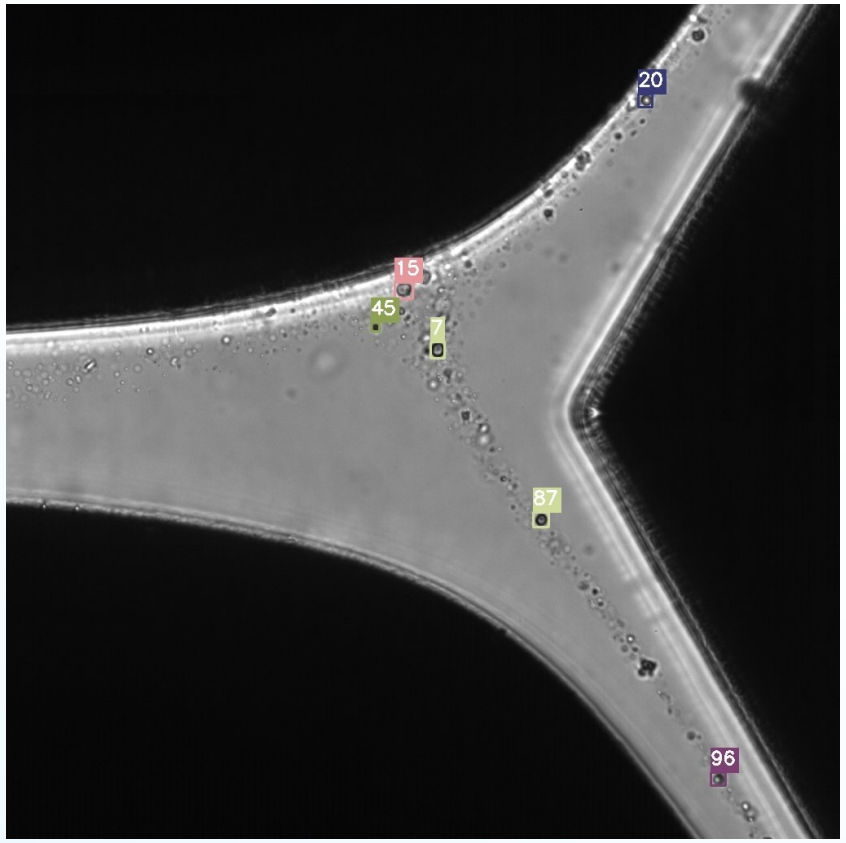
- Current State Update

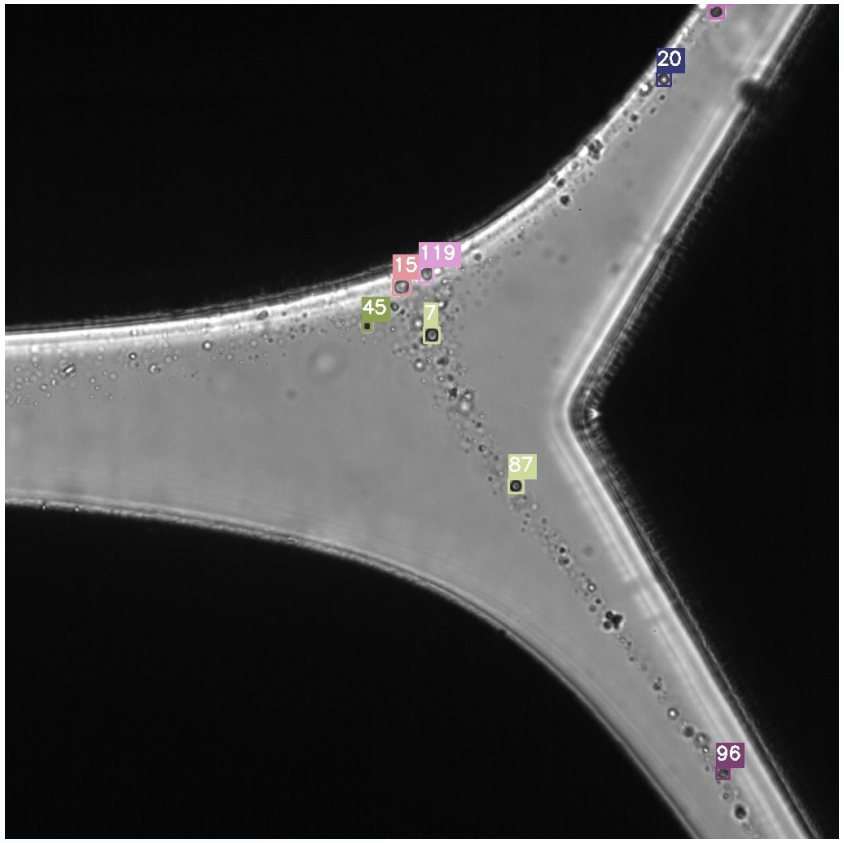
$$X_n = X'_n + K_n[Z_n - HX'_n]$$

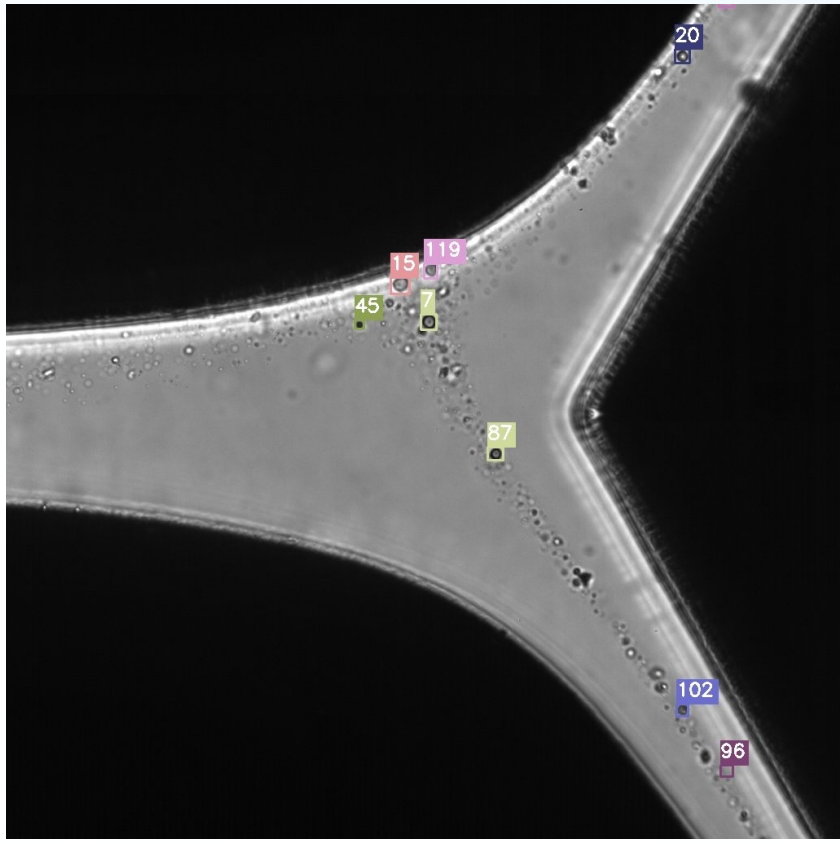
Backup: Example of DeepSORT Tracking



Backup: Example of Proposed Method Tracking







Swin Transformer Block