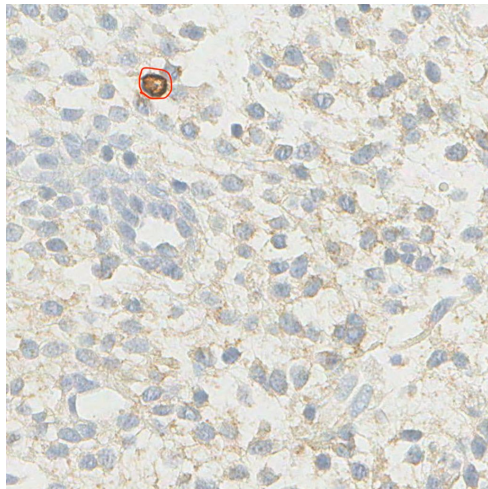


Применение нейросетевой модели детекции к гистологическим изображениям для выявления хронического эндометрита

Определение хронического эндометрита (ХЭ)



Определение ХЭ - подсчет
плазматических клеток (≥ 5
plasma cells per 10 HPF)

Пример плазматической клетки

Постановка задачи

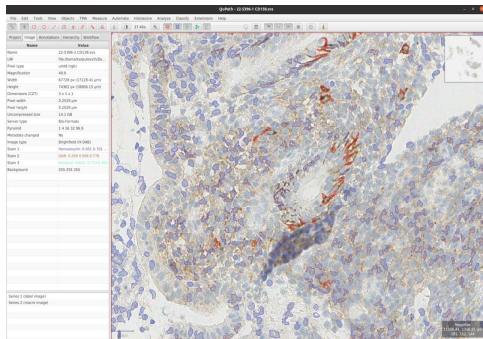
Необходимо подсчитать положительно окрашенные плазматические клетки на гистологических слайдах, **окрашенных** маркером **CD138** (для автоматического определения хронического эндометрита)

Требования к решению:

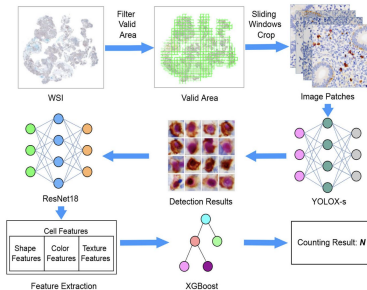
- Минимизировать ошибку второго рода (плазматическая клетка неверно отвергнута, приоритет метрика **recall**). В спорных ситуациях, мы больше хотим ставить диагноз, чем не ставить
- Использовать нейросетевую модель детекции **EndoNet**

Обзор существующих решений

- Встроенные методы QuPath



- Статья с окрашиванием двумя маркерами CD138 и MUM1 (21.12.2022)



Отличие от нашего решения:

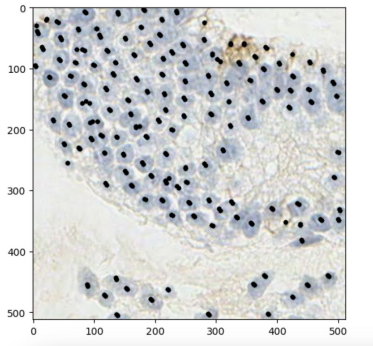
Наше решение окрашивает с помощью **CD138 single-staining** и пытается избавиться от существующей **проблемы**: CD138 может быть выражен не только в плазматических клетках, но также в структурах тканей, которые могут быть похожи на плазматические клетки, таких как эндометриальные эпителиальные клетки и стромальные клетки

Применение нейросетевой модели детекции EndoNet

EndoNet - детектирует клетки и классифицирует их на строму и эпителий на гистологических изображениях (**mAP 0,76**)

Пример работы
(позже будет получено, что **mAP 0.46** на наших данных)

```
1 "model": {  
2   "runner": "HeatmapDetectionRunner",  
3   "model_name": "HeatmapDetector",  
4   "model_kwargs": {  
5     "backbone": "UnetPlusPlus",  
6     "encoder_name": "resnet50",  
7     "encoder_depth": 5,  
8     "encoder_weights": "imagenet",  
9     "in_channels": 3,  
10    "classes": 2,  
11    "activation": null,  
12    "min_peak_value": 0.1,  
13    "pooling_scale": 13,  
14    "out_image_shape": [  
15      512,  
16      512  
17    ],  
18    "supression_range": 7  
19  },  
20  "weights_path": null,  
21  "class_separator": null,  
22  "class_labels_map": {  
23    "0": 0,  
24    "1": 1  
25  },  
26  "input_shape": [  
27    512,  
28    512  
29  ],  
30  "extractor_params": {  
31    "pooling_scale": 23,  
32    "min_peak_value": 0.144,  
33    "supression_range": 14.197  
34  }  
35 }
```

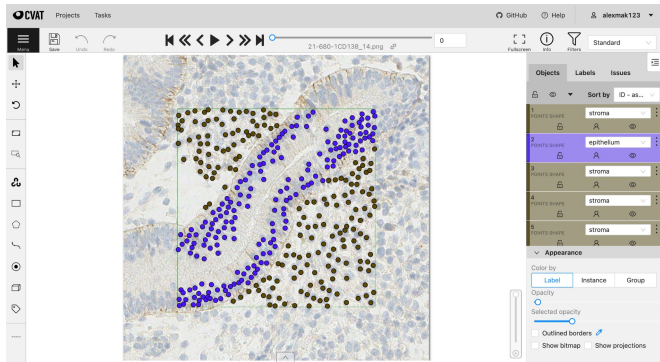


Итог :

- а) работает неточно, т.к. **EndoNet** была обучена на датасете [EndoNuke](#)
- б) нужно дообучить **EndoNet** на наших данных

Датасет “endometrium” :

- по 30 тайлов у каждого эксперта (24 уникальных и 6 совпадающих)
- 200x200 мкм и 790x790 pxl
- к тайлам добавлен контекст (это все, что снаружи зеленой рамки)
- был составлен [протокол разметки](#)
- несогласия в совпадающих тайлах были устранены



Для разметки использовался сервис CVAT

По совпадающим 6 тайлам был посчитан коэффициент согласия Cohen's kappa и он оказался равен 0.81, что свидетельствует о высоком согласии в разметке

Дообучение EndoNet

Было произведено дообучение нейросети **EndoNet** на размеченных данных и сравнение результатов

| <i>Note: endometrium - это 54 размеченных тайла</i> | mAP best_epoch(train) | mAP best_epoch(val) |
|--|-----------------------|---------------------|
| endometrium , предобученные веса imagenet | 0.75 | 0.57 |
| endonuke , предобученные веса imagenet | 0.83 | 0.76 |
| endometrium , предобученные веса endonuke | 0.84 | 0.72 |

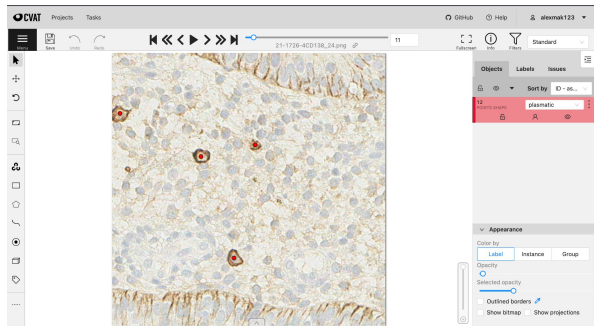
Note: как мы видим, дообучение показало хороший результат

Датасет с плазматическими клетками

Далее был собран и размечен датасет изображений, на которых есть плазматические клетки

Датасет “plasmatic” :

- по 219 тайлов у каждого эксперта (154 уникальных и 65 совпадающих)
- 200x200 мкм и 790x790 pxl
- был составлен протокол разметки
- несогласия в совпадающих тайлах были устранены



Для разметки использовался сервис CVAT

По совпадающим 65 тайлам был посчитан коэффициент согласия Cohen's kappa и он оказался равен 0.89, что свидетельствует о высоком согласии в разметке

Модификация EndoNet

Для дообучения **EndoNet** для задачи детекции и классификации **плазматических клеток**, чтобы загрузить предобученные веса с “**endometrium**” и “**endonuke**” были проделаны следующие шаги:

- изменены параметры **аугментации**, чтобы кейпоинты при аугментации не выходили за рассматриваемые рамки
- изменен **config** модели, чтобы модель решала задачу одноклассовой детекции
- заменена **segmentation_head** исходной модели на одноклассовую

```
1 import torch
2 import torch.nn as nn
3
4 class ModifiedModel(nn.Module):
5     def __init__(self, original_model, num_classes):
6         super(ModifiedModel, self).__init__()
7         self.original_model = original_model
8         self.original_model.heatmap_model.segmentation_head[0]
9         = nn.Conv2d(16, num_classes, kernel_size=(3, 3), stride=(1,
10         1), padding=(1, 1))
11
12     def forward(self, x):
13         original_output = self.original_model(x)
14         return original_output
```

train/val/test для задачи детекции и классификации плазматических клеток

Всего 373 уникальных изображения, на которых врачи разместили плазматические клетки
анализ датасета "plasmatic" :

- 826 размеченных плазматических клеток
- в среднем ≈ 2.21 плазматических клеток на тайл
- 23 тайла, на которых нет плазматических клеток

Для разделения на **train/val/test** были выбраны следующие соотношения:

- 35%/15%/50%
- в train 132 тайлов, 7 пустых
- в val 56 тайлов, 4 пустых
- в test 185 тайлов, 12 пустых

Для более корректного определения точности **в test** было **добавлено** еще **160 пустых тайлов**

Итого test : 345 тайлов (173 непустые / 172 пустые)

Дообучение EndoNet для задачи детекции и классификации плазматических клеток

Было произведено дообучение нейросети **EndoNet** на **train/val "plasmatic"** и сравнение результатов

| <i>Note: plasmatic - это 373 размеченных тайла с плазматическими клетками</i> | mAP best_epoch(train) | mAP best_epoch(val) |
|--|-----------------------|---------------------|
| train/val plasmatic , предобученные веса imagenet | 0.73 | 0.54 |
| train/val plasmatic , предобученные веса endonuke | 0.97 | 0.70 |
| train/val plasmatic , предобученные веса endometrium | 0.98 | 0.75 |

*Note: как мы видим, лучший результат дообучения получается, если использовать предобученные веса **endometrium***

Результаты на test

Была взята модель **EndoNet**, решающая задачу детекции и классификации плазматических клеток, предобученные веса **endometrium**

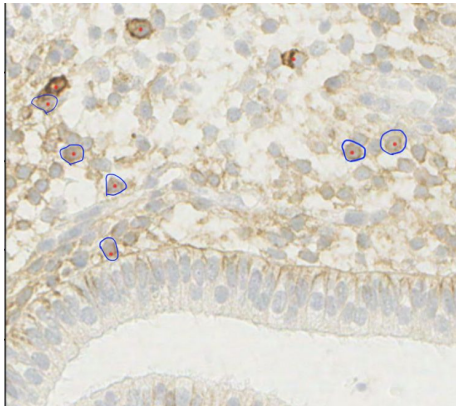
Посчитана стандартная метрика **precision, recall, f1-score** для задачи бинарной классификации : **плазматическая клетка/фон**

Результат на **test**, состоящем из 345 тайлов (**173** непустые / **172** пустые)

- precision = 0.73
- recall = 0.89
- f1-score = 0.80

Отсечение по цвету

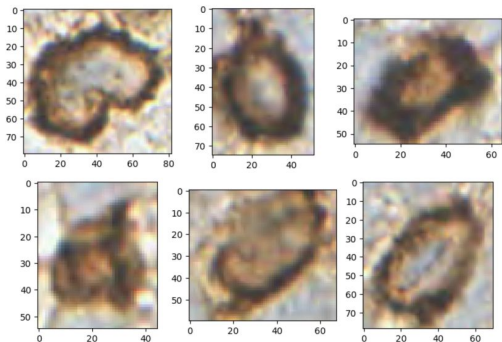
В результате анализа поставленных кейпоинтов, стало понятно, что **EndoNet** ставит кейпоинты на некоторые клетки, которые **не окрашены** маркером **CD-138** и соответственно не являются плазматическими



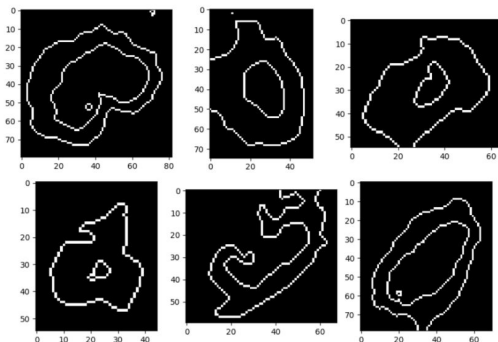
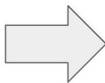
Чтобы устранить эту проблему, было сделано отсечение по цвету, который имеют плазматические клетки

Отсечение по цвету

Был применен **median filter** и **laplacian filter**, благодаря чему мы смогли построить контуры кандидатов на плазматические клетки



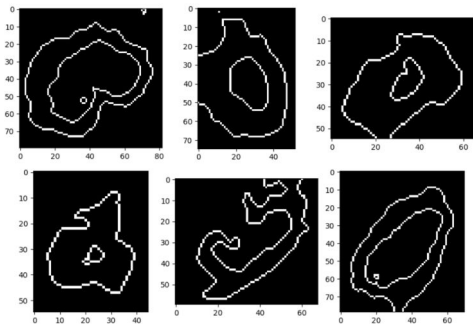
вход – изображение тайлов



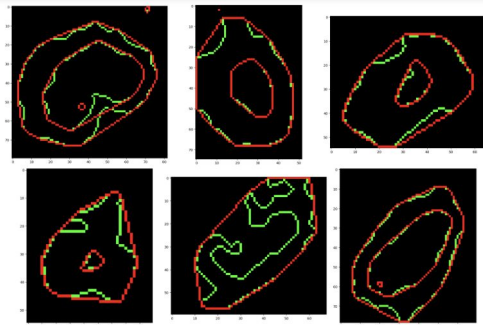
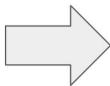
выход – маска с контурами вокруг скоплений коричневого цвета
("насколько коричневый" – регулируется порогом threshold)

Отсечение по цвету

К контурам был применен **ConvexHull**, который сделал из контуров выпуклые объекты, которые можно далее анализировать



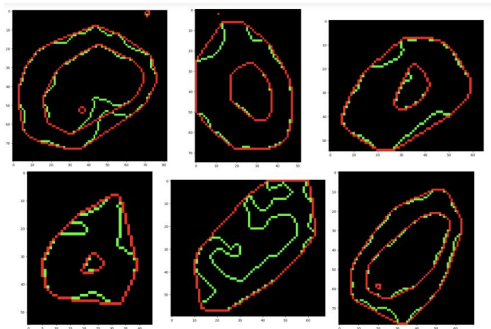
вход – маска с отмеченными на ней контурами кандидатов на плазматические клетки



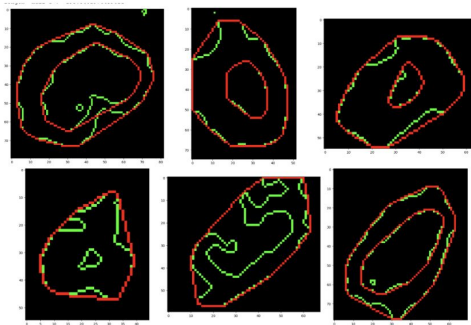
выход – маска с отмеченными на ней контурами выпуклых оболочек кандидатов на плазматические клетки

Отсечение по цвету

Была **произведена предобработка** выпуклых оболочек: отсев оболочек с маленькой длиной, отсев дублирующих оболочек



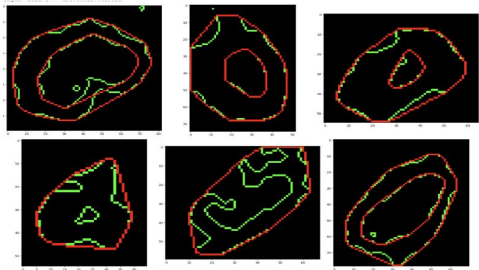
вход – маска с отмеченными на ней контурами
выпуклых оболочек кандидатов на
плазматические клетки



выход – маска с отмеченными на ней контурами
выпуклых оболочек кандидатов на
плазматические клетки без дублирующих
оболочек и оболочек малой длины

Отсечение по цвету

Предобработанные выпуклые **оболочки** были **сгруппированы** во вложенные выпуклые оболочки в виде графа



вход – маска с пред обработанными выпуклыми оболочками



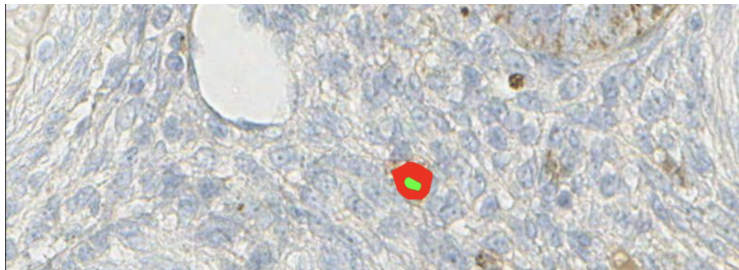
```
amount of convex hulls on picture1 : 2
convex hull index : 0 ; the index of its parent convex hull : -1
convex hull index : 1 ; the index of its parent convex hull : 0
amount of convex hulls on picture2 : 2
convex hull index : 0 ; the index of its parent convex hull : 1
convex hull index : 1 ; the index of its parent convex hull : -1
amount of convex hulls on picture3 : 2
convex hull index : 0 ; the index of its parent convex hull : 1
convex hull index : 1 ; the index of its parent convex hull : -1
amount of convex hulls on picture4 : 1
convex hull index : 0 ; the index of its parent convex hull : -1
amount of convex hulls on picture5 : 1
convex hull index : 0 ; the index of its parent convex hull : -1
amount of convex hulls on picture6 : 2
convex hull index : 0 ; the index of its parent convex hull : 1
convex hull index : 1 ; the index of its parent convex hull : -1
```

выход – массив из вложенных выпуклых оболочек,
которые имеют древовидную структуру:

1. <индекс оболочки, индекс родителя>
2. самые внешние оболочки имеют индекс
родителя -1

Отсечение по цвету

По **10** плазматическим клеткам был подсчитан **порог hsv** цвета, когда мы считаем клетку плазматической



Порог **hsv**:
 $5 \leq \text{hue} \leq 20$

Считался **hsv** цвет пикселей, которые расположены между самой внешней и ближайшей к ней внутренней оболочки:
область на изображении, обозначенная красным

Результаты на test

Была взята модель **EndoNet**, решающая задачу детекции и классификации плазматических клеток, предобученные веса **endometrium**

Добавлено отсечение по пороговому цвету $hsv : 5 \leq hue \leq 20$

Посчитана стандартная метрика **precision, recall, f1-score** для задачи бинарной классификации : **плазматическая клетка/фон**

Результат на **test**, состоящем из 345 тайлов (**173** непустые / **172** пустые)

- precision = 0.84
- recall = 0.76
- f1-score = 0.80

Выводы

Лучший вариант модели **EndoNet** с приоритетом в метрике **recall**:

- предобученные веса **endometrium**
- без отсекация по цвету

Имеет точность на **test**, состоящем из 345 тайлов (173 непустые / 172 пустые):

- precision = 0.73
- recall = 0.89
- f1-score = 0.80

Итог:

такой подход позволяет решить поставленную задачу с приемлемой точностью