



Московский государственный университет имени М.В. Ломоносова

Факультет вычислительной математики и кибернетики

Кафедра математической физики

Багамаев Мурад Аммаевич

**Метод автоматического выбора оптимального  
масштаба изображения для использования  
нейросетевых моделей**

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

**Научный руководитель:**

к.ф-м.н., м.н.с

А.В.Хвостиков

Москва, 2022

# Оглавление

<b>1 Введение</b>	<b>3</b>
<b>2 Цель работы</b>	<b>4</b>
<b>3 Обзор существующих методов</b>	<b>5</b>
<b>4 Описание модельной задачи</b>	<b>6</b>
4.1 Описание набора данных для обучения и тестирования нейронной сети . . . . .	6
4.2 Обучение нейронной сети на синтетических данных . . . . .	6
4.3 Описание архитектуры нейронной сети . . . . .	7
<b>5 Алгоритм оценки оптимального маштаба на основе анализа откликов нейронной сети</b>	<b>8</b>
5.1 Извлечение пирамид масштаба . . . . .	8
5.2 Извлечение откликов нейронной сети . . . . .	8
5.3 Сглаживание откликов . . . . .	9
5.4 Вычисление веса для каналов нейронной сети на основе градиента . . . . .	9
5.5 Вычисление производной откликов . . . . .	12
5.6 Оценка оптимального масштаба для каждой пирамиды масштаба . . . . .	12
5.7 Оценка оптимального масштаба для каждого класса изображений . . . . .	13
5.8 Подбор параметров метода . . . . .	13
<b>6 Результаты</b>	<b>15</b>
<b>7 Заключение</b>	<b>17</b>
7.1 Программная реализация метода . . . . .	17
7.2 Дальнейшее развитие . . . . .	17

# 1. Введение

При применении нейросетевых методов обработки и анализа изображений часто возникает проблема выбора масштаба изображения, подаваемого на вход нейронной сети. Особенно актуально это в задачах сегментации, детекции и трекинга.

Можно выделить два основных сценария, когда необходимо подобрать правильный масштаб изображения:

1. Имеется уже обученная нейросетевая модель, и необходимо применить её для обработки (сегментации, детекции, анализа и т.п.) серии изображений, снятых в другом масштабе, отличным от того, на чем обучалась нейронная сеть.
2. И имеется модель, обученная на изображениях одного масштаба, и набор изображений снятых в другом масштабе. При этом требуется использовать этот набор для дообучения модели. Вариант дообучения нейронной сети напрямую, не подбирая масштаб, не рассматривается, так как это будет не эффективно.

Касательно областей применения, в которых нужен алгоритм автоматического анализа масштаба, можно выделить задачу сегментации полнокадровых гистологических изображений. Большое оптическое увеличение позволяет получить изображение на различных масштабах, что можно увидеть на Рис. 1.

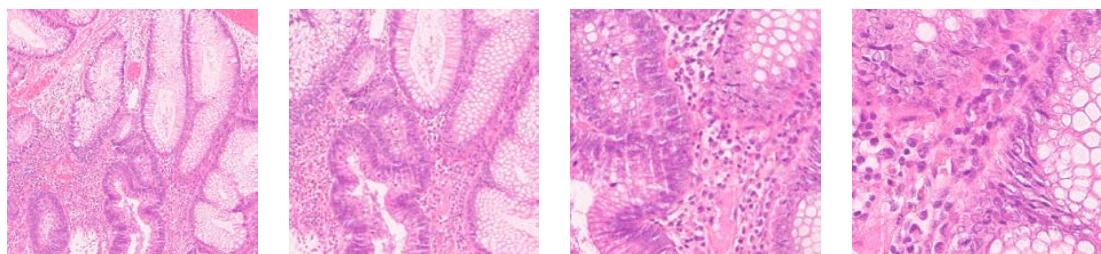


Рис. 1: Одно и то же полнослайдовое гистологическое изображение, взятое на различных масштабах.

Предлагаемый в данной работе метод оценки масштаба изображения разрабатывается для модельной задачи классификации синтетических изображений посредством сверточной нейронной сети. Нейронная сеть обучается на изображениях, взятых на определенном масштабе, а метод, основываясь только на откликах нейронной сети, оценивает этот масштаб. Предлагаемый подход является общим и потенциально может быть применен для произвольной сверточной нейронной сети для сегментации и классификации изображений.

## **2. Цель работы**

Целью работы является разработка метода автоматического определения оптимального масштаба изображения для применения обученной сверточной нейронной сети для классификации изображений к новым изображениям того же или схожего типов.

Задача состоит из следующих подзадач:

1. Создание набора изображений для обучения и тестирования нейронной сети,
2. Обучение нейронной сети на синтетических данных,
3. Разработка алгоритма оценки оптимального масштаба изображения на основе анализа откликов нейронной сети,
4. Оценка применимости и качества работы предлагаемого метода автоматического выбора оптимального масштаба изображения.

### 3. Обзор существующих методов

Очень часто проблема оценки масштаба возникает в задаче трекинга объектов, а конкретно в задаче оценки местоположения объекта в каждом кадре последовательности изображений, так как в каждом из них объект может иметь различный масштаб [1]. Одним из самых популярных подходов к решению этой проблемы является обучение дискриминационных корреляционных фильтров (англ. Discriminative Correlation Filters, DCF) [2] на основе представления пирамиды масштабов [3]. Подход является общим, поскольку он может быть включен в любой метод трекинга без оценки масштаба.

Другим решением является применение фильтров перевода и оценки масштаба для трекинга объектов [4]. Сначала метод извлекает в качестве признаков гистограмму направленных градиентов (англ. Histogram of Oriented Gradients, HOG) [5] из нескольких фрагментов данного объекта на разных разрешениях. Далее извлеченные признаки приводятся к одномерному дескриптору, чтобы иметь возможность получить одномерную функцию Гаусса в качестве желаемого результата. После этого одномерные дескрипторы объединяются для построения двумерного DCF, который и используется для точной и устойчивой оценки масштаба.

Помимо этого, существуют также методы, основанные на применении легковесных сверточных нейронных сетей: метод оценки масштаба на основе целостного представления (англ. Holistic Representation-Based Scale Estimation Method, HRSEM) и метод оценки масштаба на основе частичного представления (англ. Region Representation-Based Scale Estimation Method, RRSEM) [6]. В HRSEM на вход предобученной нейронной сети для трекинга объектов подается вся сцена. Далее извлекаются признаки откликов глубоких сверточных слоев, которые используется для устойчивой оценки масштаба объекта. Как и HRSEM, RRSEM использует для оценки масштаба извлеченные из нейронной сети признаки, только на вход нейронной сети в нем подается набор областей с объектом на разных масштабах.

## 4. Описание модельной задачи

В этом разделе дается описание архитектуры сверточной нейронной сети и набора данных, на котором нейронная сеть была обучена и протестирована.

### 4.1. Описание набора данных для обучения и тестирования нейронной сети

Для разработки и тестирования алгоритма определения оптимального масштаба изображения был создан искусственный набор изображений, ориентированный на задачу классификации текстур. Набор включает в себя 300 цветных изображений (200 для обучения, 100 для тестирования) размера  $5000 \times 5000$  пикселей. Каждое изображение представляет собой набор случайно расположенных небольших фигурок определенного типа (квадрат, круг, треугольник, крестик) и цвета (красный, зеленый, синий, желтый, фиолетовый). Таким образом, рассматриваемый синтетический набор данных содержит изображения  $4 \cdot 5 = 20$  классов изображений.

### 4.2. Обучение нейронной сети на синтетических данных

С целью моделирования реальной задачи нейронная сеть для классификации изображений обучается на фрагментах изображений созданного набора данных размера  $224 \times 224$  пикселей, взятых на масштабе 5. Под «взятием фрагмента на масштабе  $a$ » понимается следующее: из изображения вырезается блок размера  $(a \cdot 224) \times (a \cdot 224)$ , который кубической интерполяцией приводится к разрешению  $224 \times 224$ . На Рис. 2 приведены примеры фрагментов, на которых нейронная сеть обучается и тестируется.

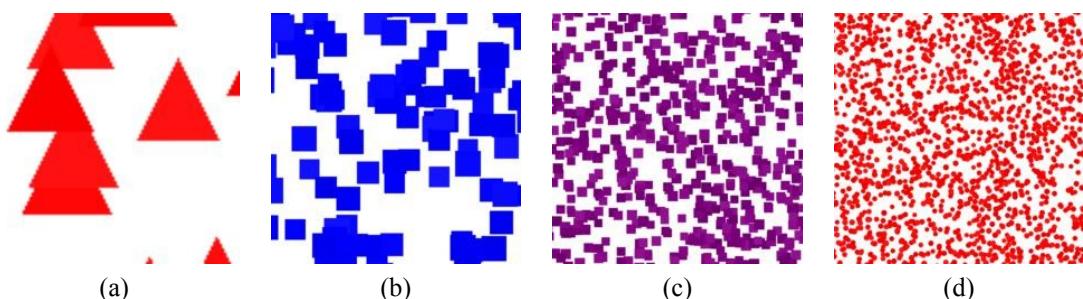


Рис. 2: Примеры фрагментов изображений: а - фрагмент изображения класса 3 масштаба 0.5; б - фрагмент изображения класса 8 масштаба 2; в - фрагмент изображения класса 16 масштабе 5; г - фрагмент изображения класса 1 масштаба 8.8.

### 4.3. Описание архитектуры нейронной сети

В работе рассматривается AlexNet-подобная [7] сверточная нейронная сеть для классификации изображений. На Рис. 3 приведена архитектура этой нейронной сети. Для обучения используются оптимизатор Adam[8] и многоклассовая перекрестная энтропия (англ. Categorical Cross-Entropy) в качестве функции потерь. Количество эпох при обучении равно 7, а размер пакета равен 32.

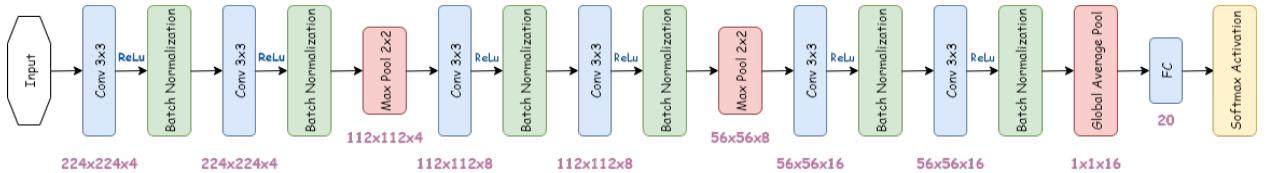


Рис. 3: Архитектура сверточной нейронной сети. Снизу фиолетовым цветом указаны размеры выходных тензоров после каждого слоя. Input - исходный фрагмент изображения размера  $224 \times 224 \times 3$ ; Conv  $3 \times 3$  - сверточный слой с фильтром размера  $3 \times 3$ ; ReLu - функция нелинейной активации сверточного слоя; Batch Normalization - слой пакетной нормализации [9]; Max Pool  $2 \times 2$  - слой субдискретизации, поникающий разрешение тензора по пространственным измерениям вычислением максимального значения в окне размера  $2 \times 2$ ; Global Average Pool - слой субдискретизации с функцией усреднения; FC - полносвязный слой, вычисляющий оценку для каждого из 20 классов. Нейронная сеть имеет 5192 параметра.

## **5. Алгоритм оценки оптимального маштаба на основе анализа откликов нейронной сети**

В данной работе предлагается метод автоматического определения наиболее оптимального масштаба изображения для использования обученной сверточной нейронной сети. Метод основан на анализе откликов и градиентов Global Average Pool слоя этой нейронной сети. Выбор такого слоя в качестве анализируемого обусловлен тем, что отклик последнего сверточного слоя содержит в себе преимущественно семантическую информацию об изображении, которая необходима для оценки масштаба подаваемого на вход нейронной сети фрагмента [1].

Предлагаемый подход состоит из следующих шагов:

1. Извлечение пирамид масштаба из набора изображений,
2. Извлечение откликов нейронной сети,
3. Сглаживание извлеченных откликов,
4. Вычисление веса для каждого из  $k = 16$  каналов Global Average Pool слоя,
5. Вычисление производной откликов,
6. Оценка оптимального масштаба для каждой пирамиды масштаба,
7. Оценка оптимального масштаба для каждого класса изображений,
8. Подбор параметров метода.

### **5.1. Извлечение пирамид масштаба**

Для анализа откликов и градиентов для каждого класса формируются  $m = 50$  пирамид масштаба, каждая из которых представляет собой набор из 95 блоков изображения размера  $224 \times 224$  пикселей, где каждый блок взят на одном из масштабов в интервале  $[0.5, 9.9]$  из одного и того же места изображения. На Рис. 4 показан пример создания пирамиды масштаба.

### **5.2. Извлечение откликов нейронной сети**

Для каждой пирамиды масштаба извлекаются отклики на слое Global Average Pool, которые для каждого фрагмента пирамиды представляют собой вектор из  $k$  элементов, где  $k$

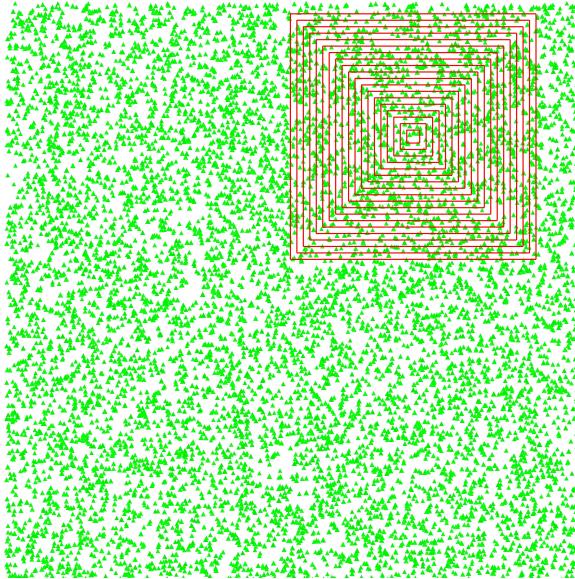


Рис. 4: Пример извлечения пирамиды масштаба из изображения класса 6 (зеленые крестики). Красными квадратами на рисунке отмечены фрагменты, которые извлекаются из изображения, а затем приводятся к размеру  $224 \times 224$  пикселей, таким образом формируя пирамиду масштаба.

- число нейронов на слое Global Average Pool. У используемой в данной работе сверточной нейронной сети  $k = 16$ . Каждый элемент этого вектора представляет собой среднее значение соответствующего канала выходного тензора последнего сверточного слоя. На рис. 5 приведено  $k$  графиков зависимости каждого из  $k$  значений Global Average Pool слоя от масштаба, на котором был взят фрагмент.

### 5.3. Сглаживание откликов

Для того, чтобы эффективно проанализировать полученные на предыдущем шаге отклики, их нужно предварительно сгладить. Для этого в работе предлагается вычислить свертку функций на графиках с одномерным фильтром Гаусса:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}},$$

с некоторой  $\sigma$ , которая будет одним из параметров предлагаемого метода. На рисунке Рис. 6 представлен пример сглаживания.

### 5.4. Вычисление веса для каналов нейронной сети на основе градиента

Глядя на графики можно заметить, что отклики на одних нейронах имеют экстремум в районе искомого масштаба 5, на фрагментах которого нейронная сеть обучалась, а отклики

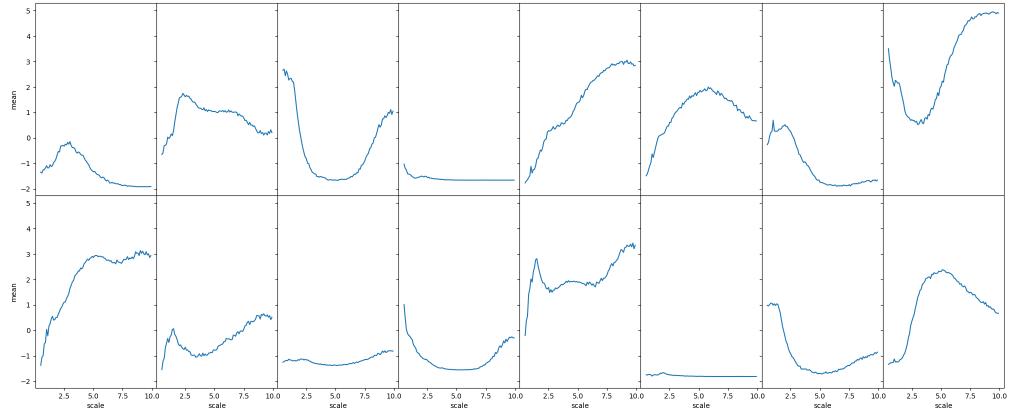


Рис. 5: Отклики для пирамиды масштаба класса 6, взятые с Global Average Pool слоя обученной сверточной нейронной сети. Каждый отклик соответствует одному из 16 нейронов этого слоя. У каждого графика на оси абсцисс расположен масштаб, на котором был взят соответствующий фрагмент пирамиды масштаба, на оси ординат - среднее значение соответствующего канала выходного тензора последнего сверточного слоя.

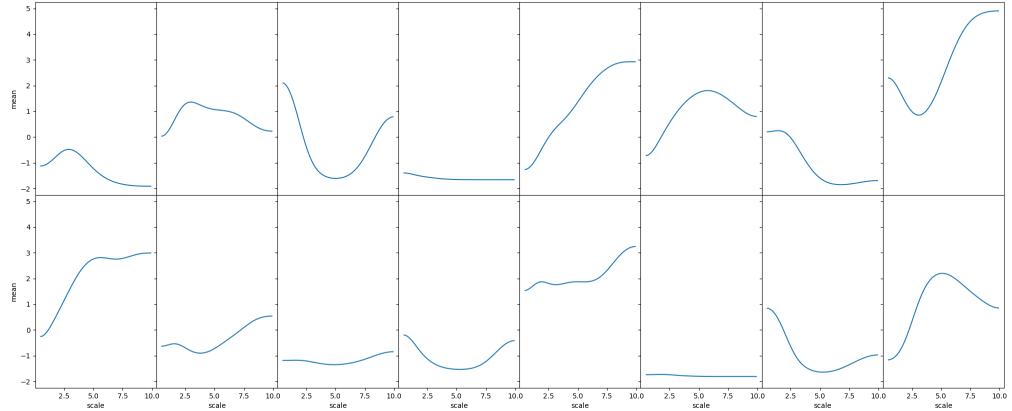


Рис. 6: Результат сглаживания откликов с Рис.5 с одномерным фильтром Гаусса с  $\sigma = 8$ .

на других нейронах не дают информации, коррелирующей с поиском оптимального масштаба, тем самым внося лишь шум. Поэтому необходима числовая характеристика, позволяющая оценить, насколько отклик является полезным для нахождения оптимального масштаба для использования нейронной сети.

В качестве такой характеристики, называемой далее в работе *весом*, предлагается использовать градиент функции предсказания нейронной сети для соответствующего класса  $c$ ,  $y^c$  (до применения функции softmax), по переменным активации  $a^k$  Global Average Pool слоя, усредненного по всем масштабам:

$$\alpha_k^c = \frac{\partial y^c}{\partial a^k}.$$

Таким образом, метод будет использовать градиентную информацию, проходящую через последний сверточный слой нейронной сети, для оценки важности каждого нейрона для конкретного целевого класса изображения [10].

В Табл. 1 приведен результат вычисления весов предлагаемым способом.

0.00002	0.00011	0.00018	0.00025	0.00008	0.00011	0.00009	0.00010
0.00009	0.00008	0.00025	0.00034	0.00012	0.00024	0.00020	0.00001

Таблица 1: Модули весов для каждого нейрона Global Average Pool слоя нейронной сети для пирамиды масштаба класса 6 из Рис. 4.

Вычислив вес для каждого нейрона Global Average Pool слоя, можно оставить только  $P\%$  нейронов с наибольшими по модулю весами, а остальные пометить как шумовые. На Рис. 7 показано, что при таком действии остаются наиболее полезные для метода отклики, имеющие экстремум в районе масштаба 5, на фрагментах которого нейронная сеть обучалась.

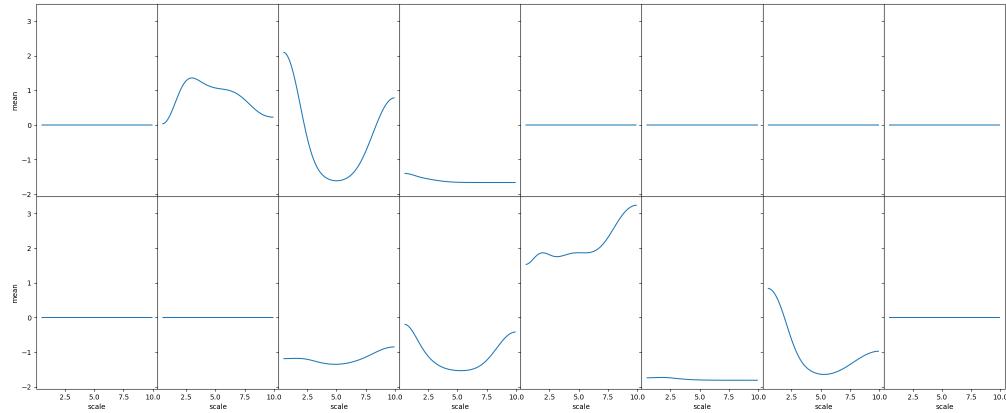


Рис. 7: Результат отбрасывания  $P = 50\%$  откликов с Рис.6 с наименьшими по модулю весами.

## 5.5. Вычисление производной откликов

Для улучшения метода предлагается анализировать помимо самих откликов еще и производные от них. Так как все отклики представляют собою сеточные функции, определенные на отрезке  $[0.5, 9.9]$  на равномерной сетке с шагом  $h = 0.1$ , то для приближения производной используется центральная разностная производная

$$y_{\dot{x},i} = \frac{y_{i+1} - y_{i-1}}{2h}, \quad i = 1, 2, \dots, N - 1$$

где  $y_i = y(x_i)$ ,  $x_i = 0.5 + ih$ ,  $i = 0, 1, \dots, N$ ,  $N = l/h$ ,  $l$  - длина отрезка. На левой и правой границах производная вычисляется как разностная производная вперед и назад соответственно. На Рис. 8 приведен пример вычисления производной.

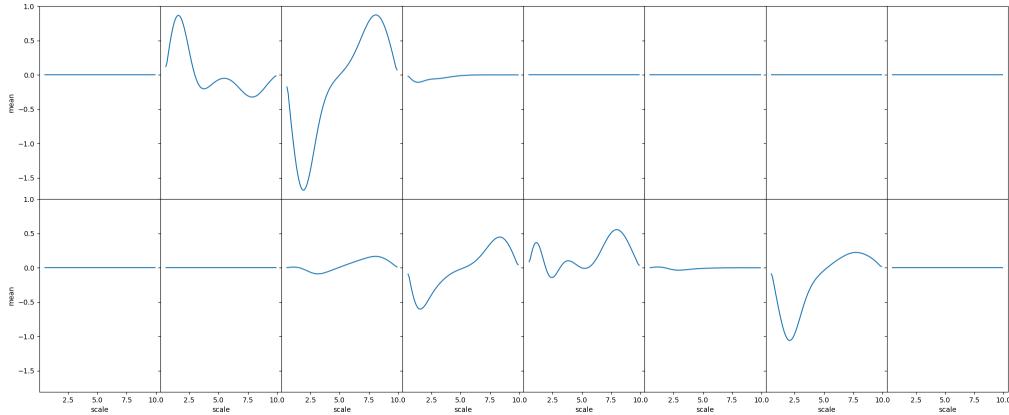


Рис. 8: Производные откликов с Рис.7.

## 5.6. Оценка оптимального масштаба для каждой пирамиды масштаба

Для каждого из  $P\%$  нейронов с наибольшими весами находятся все масштабы, на которых достигаются экстремумы их откликов и производных этих откликов. В качестве оценки оптимального масштаба для каждой пирамиды масштаба пробовались два способа: медиана полученных масштабов и их среднее значение. Медиана показала лучшие результаты, поэтому она и используется в качестве оценки оптимального масштаба для каждой пирамиды масштабов в предлагаемом методе.

## 5.7. Оценка оптимального масштаба для каждого класса изображений

Оценка оптимального масштаба для каждого класса изображений вычисляется как медиана предсказаний для всех 50 пирамид масштаба, взятых из изображений этого класса. Для оценки можно было бы использовать и всего одну пирамиду масштаба, но для устойчивости метода предлагается использовать большее количество. На Рис. 9 показан пример извлечения нескольких пирамид масштаба из исходного изображения.

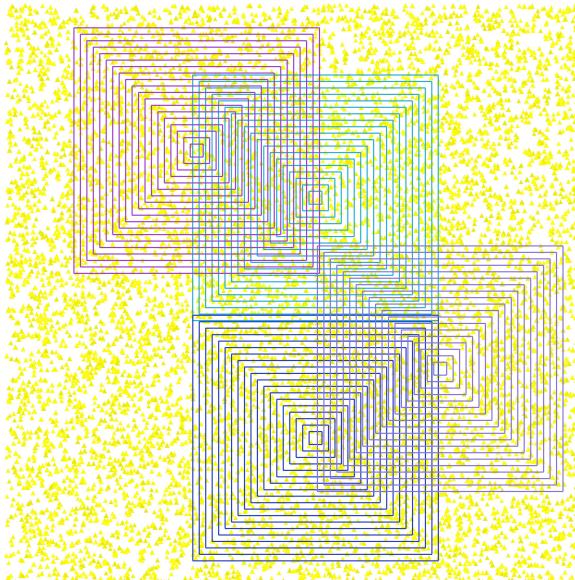


Рис. 9: Пример извлечения 4-х пирамид масштаба из изображения класса 15.

## 5.8. Подбор параметров метода

У предлагаемого метода есть два параметра:  $\sigma$  - параметр фильтра Гаусса для сглаживания откликов и  $P$  - процент нейронов слоя Global Average Pool с наибольшими по модулю весами. Результат естественно зависит от выбора этих параметров. В Табл. 2, 3, 4 приведены различные метрики для результатов оценки оптимального масштаба для всех классов. Из них видно, что лучшие результаты метод показывает либо при  $\sigma = 8, P = 50$ , либо при  $\sigma = 10, P = 40$ .

$P \backslash \sigma$	7	8	9	10	11	12	13	14	15
40	15.570	12.810	11.055	<b>10.068</b>	10.933	10.823	12.368	12.553	13.840
50	12.565	<b>10.675</b>	11.415	11.998	12.230	11.883	11.860	14.948	15.610
60	11.108	10.065	9.445	12.850	10.433	9.480	10.325	11.190	12.153

Таблица 2: Среднеквадратичная отклонение оценок оптимального масштаба для всех классов от оптимального масштаба 5.

$P \backslash \sigma$	7	8	9	10	11	12	13	14	15
40	10	8	10	<b>11</b>	8	9	8	10	9
50	9	<b>10</b>	10	8	7	7	7	7	8
60	10	9	7	7	7	5	7	6	8

Таблица 3: Количество оценок масштаба для всех классов в интервале [4.5, 5.5].

$P \backslash \sigma$	7	8	9	10	11	12	13	14	15
40	14	15	15	<b>15</b>	15	14	13	14	14
50	15	<b>16</b>	16	16	16	15	14	13	14
60	16	15	16	14	17	18	17	15	16

Таблица 4: Количество оценок масштаба для всех классов в интервале [4, 6].

## 6. Результаты

В Табл. 5 приведены результаты работы предлагаемого метода с  $\sigma = 8$  и  $P = 50$ .

цвет \ тип	квадрат	круг	крестик	треугольник
красный	4.8	3.8	4.2	4.65
зеленый	4.5	4.8	5.15	4.8
синий	3.7	5.0	4.5	4.85
желтый	3.05	4.4	4.4	5.05
фиолетовый	4.7	4.8	3.95	5.6

Таблица 5: Оценки оптимального масштаба для каждого класса изображения при использовании нейронной сети, обученной на фрагментах масштаба 5, с архитектурой, описанной на Рис. 3.

Как можно видеть, реализованный алгоритм автоматического выбора оптимального масштаба изображения при использовании нейронной сети, обученной на фрагментах масштаба 5, в 50% случаев оценивает масштаб в диапазоне [4.5, 5.5] и в 80% - в диапазоне [4, 6]. Такой результат можно считать допустимым для реальных задач. На Рис. 10 приведены примеры сравнения фрагментов, взятых на целевом масштабе 5 и на масштабе, выбранным методов.

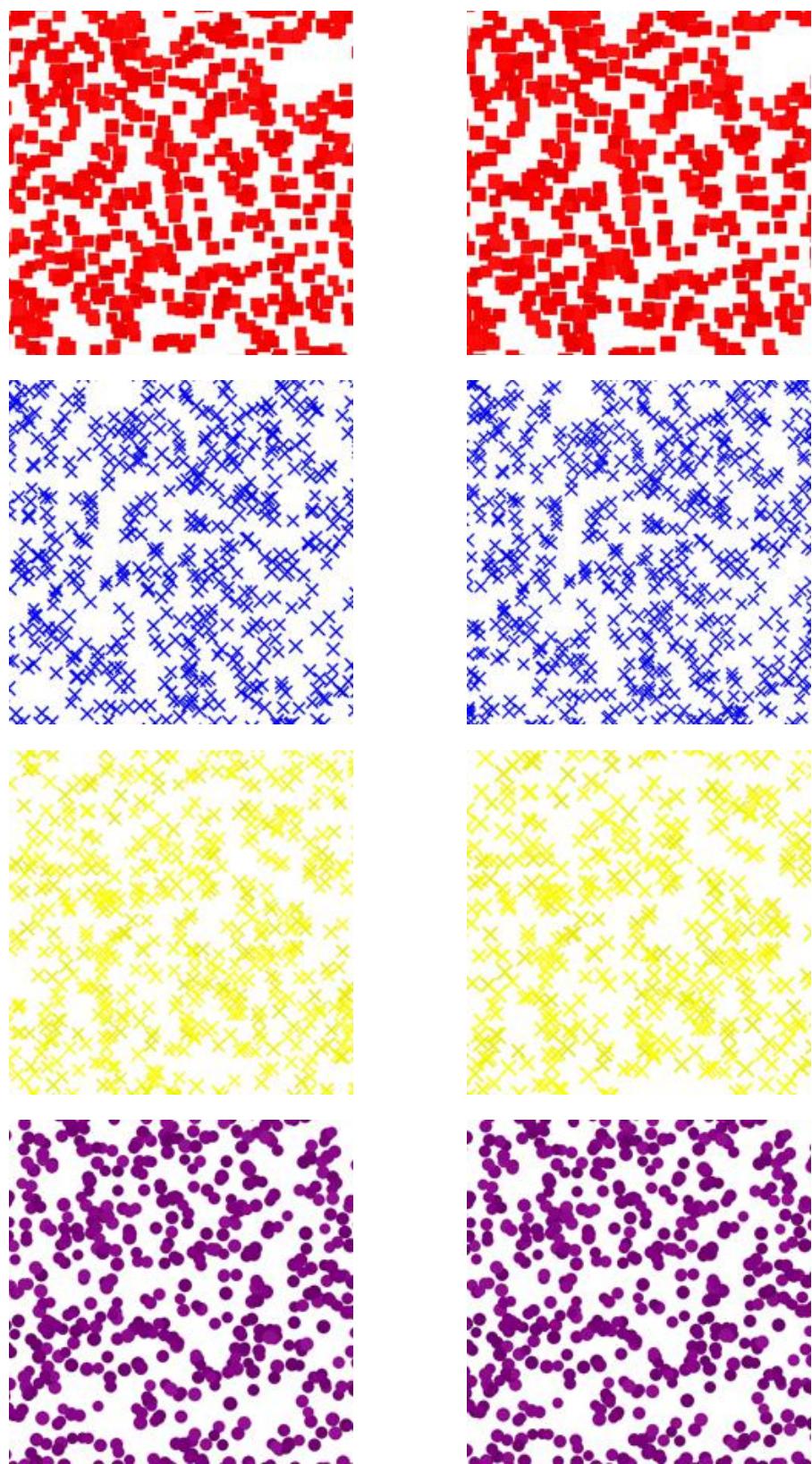


Рис. 10: В первом столбце расположены фрагменты целевого масштаба 5, во втором столбце - соответствующие фрагменты масштаба, выбранного предлагаемым методом.

## **7. Заключение**

Разработан метод автоматического поиска оптимального масштаба изображения для использования AlexNet-подобной сверточной нейронной сети для классификации изображений. Алгоритм был протестирован на сгенерированных данных.

### **7.1. Программная реализация метода**

Предлагаемый метод был программно реализован на языке Python версии 3.7.11 с использованием следующих библиотек:

- NumPy 1.21.2 - для работы с данными,
- TensorFlow 2.1.0 - для работы с нейронной сетью,
- Pillow 8.3.1 - для работы с изображениями,
- SciPy 1.4.1 - для использования одномерного фильтра Гаусса,
- Matplotlib 3.4.3 - для визуализации графиков.

Git-репозиторий с исходным кодом работы: <https://github.com/JokerLord/scale-estimation>.

### **7.2. Дальнейшее развитие**

Дальнейшим развитием работы может стать адаптация описанного в работе алгоритма для задачи сегментации полнокадровых гистологических изображений.

## Список литературы

- [1] Spatial and semantic convolutional features for robust visual object tracking / Zhang J., Jin X., Sun J., Wang J., and Sangaiah A. K. // Multimedia Tools and Applications. — 2020. — Vol. 79, no. 21. — P. 15095–15115.
- [2] Visual object tracking using adaptive correlation filters / Bolme D. S., Beveridge J. R., Draper B. A., and Lui Y. M. // 2010 IEEE computer society conference on computer vision and pattern recognition / IEEE. — 2010. — P. 2544–2550.
- [3] Accurate scale estimation for robust visual tracking / Danelljan M., Häger G., Khan F., and Felsberg M. // British Machine Vision Conference, Nottingham, September 1-5, 2014 / Bmva Press. — 2014.
- [4] Discriminative scale space tracking / Danelljan M., Häger G., Khan F. S., and Felsberg M. // IEEE transactions on pattern analysis and machine intelligence. — 2016. — Vol. 39, no. 8. — P. 1561–1575.
- [5] Object detection with discriminatively trained part-based models / Felzenszwalb P. F., Girshick R. B., McAllester D., and Ramanan D. // IEEE transactions on pattern analysis and machine intelligence. — 2010. — Vol. 32, no. 9. — P. 1627–1645.
- [6] Marvasti-Zadeh S. M., Ghanei-Yakhdan H., Kasaei S. Efficient scale estimation methods using lightweight deep convolutional neural networks for visual tracking // Neural Computing and Applications. — 2021. — Vol. 33, no. 14. — P. 8319–8334.
- [7] Krizhevsky A., Sutskever I., Hinton G. E. Imagenet classification with deep convolutional neural networks // Advances in neural information processing systems. — 2012. — Vol. 25.
- [8] Kingma D. P., Ba J. Adam: A method for stochastic optimization // arXiv preprint arXiv:1412.6980. — 2014.
- [9] Ioffe S., Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift // International conference on machine learning / PMLR. — 2015. — P. 448–456.
- [10] Grad-cam: Visual explanations from deep networks via gradient-based localization / Selvaraju R. R., Cogswell M., Das A., Vedantam R., Parikh D., and Batra D. // Proceedings of the IEEE international conference on computer vision. — 2017. — P. 618–626.