

УДК 621.391

## РАЗРАБОТКА АЛГОРИТМА ДЕТЕКТИРОВАНИЯ ПОЛИПОВ НА ЭНДОСКОПИЧЕСКИХ ИЗОБРАЖЕНИЯХ С ИСПОЛЬЗОВАНИЕМ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

*Лебедев А.А., аспирант Ярославского государственного университета им. П.Г. Демидова,  
e-mail: lebedevdes@gmail.com;*

*Хрящев В.В., к.т.н., доцент Ярославского государственного университета им. П.Г. Демидова,  
e-mail: v.khryashchev@uniyar.ac.ru;*

*Среднякова А.С., м.н.с. Ярославского государственного университета им. П.Г. Демидова,  
e-mail: a.srednyakova2@uniyar.ac.ru;*

*Казина Е.М., м.н.с. Ярославского государственного университета им. П.Г. Демидова,  
e-mail: e.kazina2@uniyar.ac.ru*

## DEVELOPMENT OF AN ALGORITHM FOR DETECTING POLYPS ON ENDOSCOPIC IMAGES USING CONVOLUTIONAL NEURAL NETWORKS

*Lebedev A.A., Khryashev V.V., Srednyakova A.S., Kazina E.M.*

*A neural network algorithm for polyps detection in endoscopic images based on the SSD convolutional network architecture is proposed. The developed algorithm has a performance sufficient for analyzing a video data in real time, and also has a relatively high value of quality assessment metrics. The highest metric value  $F1 = 81,53$  on the ETIS-LaribPolypDB images database was obtained for the SSD algorithm with the VGG-512 core network. To train the neural network algorithm, a database of 6430 images of polyps with a resolution of 626x532 pixels was collected on the basis of the endoscopic department of the Yaroslavl Regional Clinical Oncological Hospital. A modified algorithm for input data augmentation is proposed.*

**Key words:** deep learning, convolutional neural network, augmentation, neural network algorithm, basic network, endoscopic image analysis, colonoscopy, polyp detection, interlacing, deinterlacing, polyp video database, post-processing of results, algorithm quality assessment metric.

**Ключевые слова:** глубокое обучение, сверточная нейронная сеть, аугментация, нейросетевой алгоритм, базовая сеть, анализ эндоскопических изображений, колоноскопия, детектирование полипов, интерлейсинг, деинтерлейсинг, база видеоданных полипов, постобработка результатов, метрика оценки качества работы алгоритма.

### Введение

В настоящее время актуальной научно-технической задачей является применение методов цифровой обработки изображений и машинного обучения для создания систем поддержки принятия врачебного решения при проведении эндоскопических исследований желудочно-кишечного тракта [1, 2]. Анализ таких исследований отличается высокой трудоемкостью, что связано с высокой вариативностью изображений слизистых оболочек желудочно-кишечного тракта, многочисленными артефактами на изображениях, получаемых с эндоскопа, а также необходимостью осмотра тканей в различных режимах, в том числе в режиме с оптическим увеличением, достигающим 115 крат. Применение систем поддержки принятия врачебного решения в эндоскопии позволит повысить точность диагностики, уменьшить влияние человеческого фактора на качество исследований, снизить стоимость и временные затраты на их проведение [3-5].


Рак толстой кишки является одним из наиболее рас-

*Предложен нейросетевой алгоритм детектирования полипов на эндоскопических изображениях на основе архитектуры сверточной сети SSD. Он обладает производительностью, достаточной для анализа видеопотока в режиме реального времени, а также имеет сравнительно высокие значения метрик оценки качества. Наибольшее значение метрики  $F1 = 81,53$  на тестовой базе изображений ETIS-LaribPolypDB получено для алгоритма SSD с базовой сетью VGG-512. Для обучения нейросетевого алгоритма собрана база из 6430 изображений полипов, разрешением 626x532 пикселей, на базе эндоскопического отделения Ярославской областной клинической онкологической больницы. Предложен модифицированный алгоритм аугментации входных данных.*

пространенных видов онкологических заболеваний, который в 95% случаев развивается из полипов [6-8]. Эффективной стратегией повышения выживаемости пациентов является регулярное проведение эндоскопического исследования толстого кишечника – колоноскопии и своевременное удаление опасных полипов. Задача автоматического детектирования и классификации полипов методами цифровой обработки эндоскопических видеоданных представляет огромный интерес как для производителей, так и для пользователей современной эндоскопической техники.

Существуют различные подходы к решению задачи детектирования полипов, основанные как на классических методах машинного обучения, так и на основе методов глубокого обучения, при котором формирование признаков осуществляется автоматически [9-12]. Анализ

---



научно-технической литературы показывает, что подходы, основанные на сверточных нейронных сетях (СНС), показывают в таких задачах наилучший результат. Например, в [7] для решения рассматриваемой задачи применяется модель, осуществляющая сегментацию изображений с помощью СНС с архитектурой U-Net. В [8] рассматривается применение нейронных сетей с генерацией областей интереса, при которой происходит последовательный выбор областей-кандидатов из изображения и их дальнейшая классификация. Несмотря на достигнутые на сегодняшний день определенные результаты в детектировании полипов, данная задача еще не является решенной с точки зрения практического использования в ежедневной медицинской практике. Главная проблема внедрения подобных автоматизированных систем во врачебную практику – наблюдаемое большое количество ложных срабатываний нейросетевого детектора. Проблемы с ложными срабатываниями усугубляются при обработке видеоданных в реальном времени, что также характерно для эндоскопической практики.

**Целью работы** является разработка алгоритма детектирования полипов на эндоскопических изображениях толстой кишки с помощью СНС на основе архитектуры SSD [13]. При таком подходе желаемые объекты интереса детектируются на изображении за один проход, что дает выигрыш в производительности по сравнению с описанными выше нейросетевыми подходами.

Для обучения СНС собрана собственная база изображений полипов, которая описана в следующем разделе. В последующих разделах приведены описание алгоритма и использованной архитектуры нейронной сети, а также описаны сравнительные результаты тестирования предложенного нейросетевого подхода с имеющимися аналогами на унифицированной базе эндоскопических изображений.

### **Создание базы изображений полипов**

Эндоскопические изображения являются сложными для автоматического анализа из-за разнообразия текстур исследуемой поверхности, ее существенной неоднородности, широкого диапазона масштабов обрабатываемых изображений, применения красителей в процессе исследования слизистой, а также вследствие использования узкоспектральной эндоскопии. Кроме того, существуют дополнительные факторы, усложняющие анализ изображения – множество видеок кадров бывают размытыми из-за резкого движения эндоскопа, присутствуют блики на слизистой, из-за использования чересстрочной развертки наблюдается интерлейсинг, надписи, выводимые на изображение эндоскопом, могут перекрывать собой искомый объект интереса.

В исследовании использованы открытые международные базы эндоскопических изображений для процедуры колоноскопии:

- открытая база изображений CVC-ClinicDB, которая является набором из 612 кадров, разрешением 384×288 пикселей, с полипами, полученных из 29 эндоскопических видеопоследовательностей, разметка к которым дана в виде бинарных масок [14];

- открытая база изображений CVC-ColonDB, которая является набором из 380 кадров, разрешением 574×500 пикселей, с полипами, полученных из 15 эндоскопических видеопоследовательностей, разметка к которым дана в виде бинарных масок [15];

- открытый тестовый набор ETIS-LaribPolypDB, состоящий из 196 изображений полипов, разрешением 1225×966 пикселей, с разметкой в виде бинарных масок [16].

Однако размер данных баз изображений слишком мал, чтобы использовать их для обучения современных СНС без проведения процедуры аугментации (размножения) данных. Кроме того, в указанных базах отсутствуют сложные изображения с описанными выше искажениями, которые встречаются на практике. Например, там нет изображений, полученных с применением узкоспектральной эндоскопии и красителей. Поэтому, для проведения исследований собрана собственная база изображений колоноскопии. Работа выполнялась совместно с врачами-эндоскопистами Ярославской областной клинической онкологической больницы. Эта база на сегодняшний день больше приближена к реальной практической ситуации. Видеофрагменты и изображения получены с моделей видеокколоноскопов Olympus CF-Q180AL и Olympus CF-Q190AL, которые использовались совместно с системными видеоцентрами Olympus EVIS EXERA II и EXERA III. В ходе выполнения работы получены и обработаны 67 видеофрагментов колоноскопии 64-х пациентов. Средняя продолжительность одного видеофрагмента составила 15 минут. Из видеофрагментов получено 6430 изображений полипов разрешением 626×532 пикселей.

Для всех изображений с помощью врачей создана экспертная разметка – выделены прямоугольниками области изображений, на которых присутствуют полипы. Примеры изображений с аннотацией приведены на рис. 1.

Также из вышеперечисленных баз изображений составлены наборы для обучения, валидации и тестирования. В обучающий набор вошли базы CVC-ClinicDB, CVC-ColonDB и часть из собранной собственной базы изображений – 6330 изображений. Таким образом, итоговый размер набора для обучения составил 7322 изображения. Для валидации результатов во время процедуры обучения использовались 100 изображений из собранной базы. Для тестирования использовалась база изображений ETIS-LaribPolypDB, что позволило провести сравнение разработанного нейросетевого алгоритма с имеющимися аналогами.

### **Разработка нейросетевого алгоритма детектирования полипов**

#### *Выбор архитектуры сверточной нейронной сети*

В работе исследуется применение архитектуры SSD как основы для алгоритма обнаружения полипов на колоноскопических изображениях. Такая сеть использует мультимасштабные карты признаков для детектирования, что важно для решения поставленной задачи, так как полипы на видеоизображениях представлены множеством масштабов.

Данная архитектура представляет собой СНС прямого распространения, которая генерирует набор обрамля-

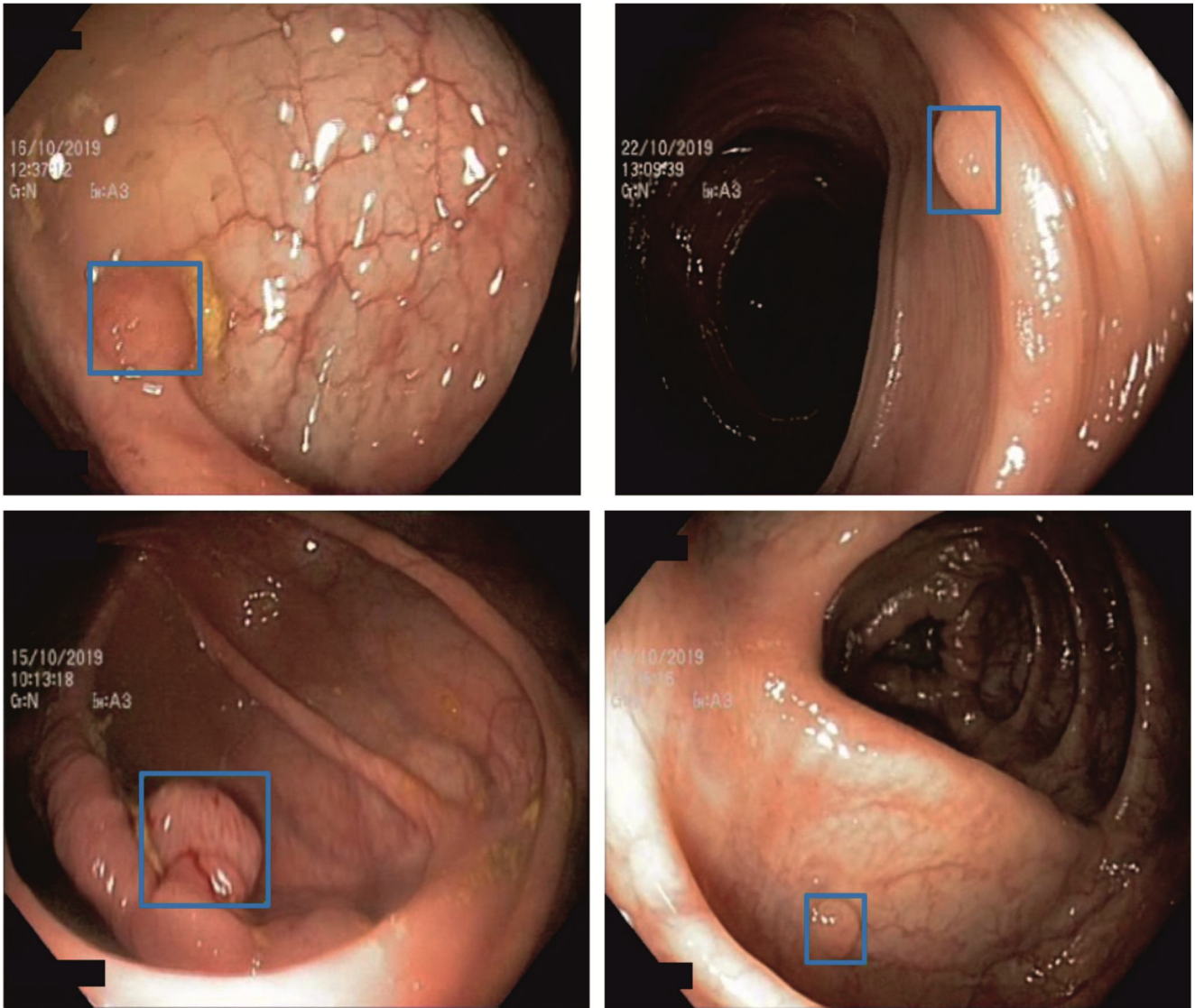


Рис. 1. Примеры изображений полипов из созданной базы изображений с аннотацией

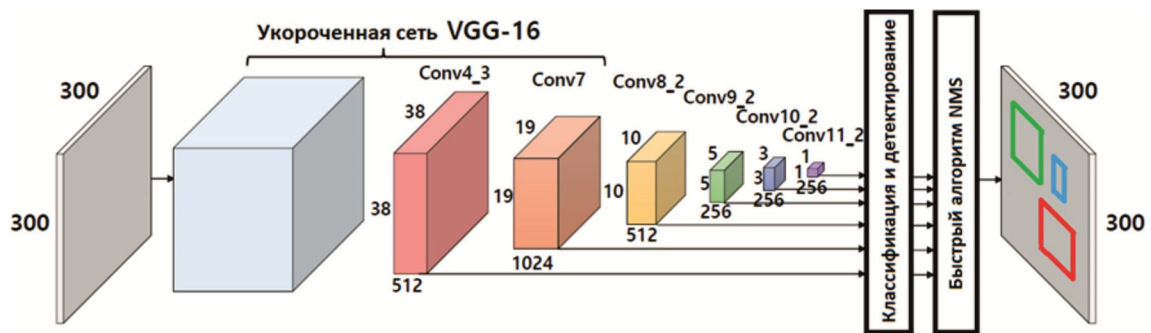


Рис. 2. Архитектура сверточной нейронной сети SSD

ющих рамок фиксированных размеров, а также выдает вероятность наличия объекта в рамке. В дальнейшем эти данные подаются на вход алгоритма подавления немаксимумов для окончательного обнаружения объекта в кадре. Ряд начальных слоев сети SSD (так называемая базовая сеть) представляет собой стандартную архитектуру сети VGG [13], предназначенную для анализа изображений высокого качества, при этом отсекаются слои классификации. Далее к базовой сети добав-

ляется вспомогательная структура, производящая детектирование по вычисленным картам признаков в базовой части, как показано на рис. 2.

Для того чтобы обеспечить работу с различными разрешениями входных видеоданных, получаемый кадр видеопоследовательности перед подачей на вход копируется, полученная копия масштабируется до разрешения, необходимого для работы нейронной сети, например 300×300 или 512×512 пикселей в зависимости от

выбранной модели базовой сети. После этого выполняется такая же предобработка изображения, что применялась к обучающей базе изображений – вычитание среднего. Выходные координаты масштабируются до размеров, соответствующих исходному разрешению видеокadra.

*Процедура обучения нейронной сети с модификацией процесса аугментации данных*

Размер базы изображений играет ключевую роль при глубоком обучении СНС. Следует отметить, что в области эндоскопии создание баз изображений, содержащих десятки тысяч аннотированных примеров, затруднено тем, что далеко не все видеоматериалы врачей исследований хранятся и доступны для исследователей. Кроме того, процесс привлечения врачей-эндоскопистов для разметки подобных данных представляет собой сложную организационную задачу. Таким образом, возникает необходимость в организации этапа алгоритмического расширения (аугментации) базы эндоскопических изображений до размеров, достаточных для обучения и тестирования современных моделей СНС. В данной работе рассмотрены два варианта аугментации: стандартная аугментация, предложенная в работе [13], а также ее модифицированный вариант.

При стандартной аугментации, для того чтобы сделать нейросетевую модель более устойчивой к различным размерам и формам входных объектов, к каждому изображению обучающей выборки случайным образом применяется один из следующих вариантов преобразований:

- используется все исходное входное изображение;
- из изображения вырезается фрагмент (патч), для которого минимальное пересечение с объектом согласно мере Жаккара составляет 0,1; 0,3; 0,5; 0,7 или 0,9;
- из изображения вырезается фрагмент (патч) случайным образом.

Размер каждого фрагмента составляет от 10 % до 100 % от размера исходного изображения, и соотношение сторон находится в диапазоне от 1 к 2 до 2 к 1.

После выполнения данного этапа производится изменение размеров каждого фрагмента до некоторого фиксированного значения и осуществляется зеркальное отображение части фрагментов по горизонтали. Также с вероятностью 0,5 выбирается одна из двух последовательностей фотометрических преобразований, аналогичных описанным в [17].

При использовании разработанного модифицированного варианта аугментации дополнительно к описанным используются преобразования, при которых учитывается тот факт, что рассматриваемые в исследовании изображения не имеют «верха» и «низа». В частности, в отличие от стандартной аугментации, в данном варианте рассматривается также зеркальное отображение изображений по вертикали.

При использовании модифицированного варианта аугментации выполняется преобразование изображения из цветового пространства RGB в HSV, рандомизация яркости, контрастности, насыщенности и оттенка в заданных пределах, осуществляется случайный поворот изображения на 90, 180 и 270 градусов с вероятностью

0,5, зеркальное отображение по горизонтали и вертикали, вырезание из изображения патчей случайным образом и изменение их размера с выбором различного соотношения сторон.

Для оценки качества работы алгоритма аугментации использовались кривые зависимости точности от полноты (Precision-Recall), а также следующие метрики [9]:

Precision ( $P$ ) (точность): значение данной метрики рассчитывалось по следующей формуле:

$$P = \frac{TP}{TP + FP},$$

где  $TP$  – истинно-положительные решения модели,  $FP$  – ложно-положительные ответы;

Recall ( $R$ ) (полнота или чувствительность):

$$R = \frac{TP}{TP + FN}, \quad (2)$$

где  $FN$  – ложно-отрицательные решения;

$F1$  мера: данная метрика вычисляется как среднее гармоническое полноты и точности согласно следующему соотношению:

$$F1 = 2 \cdot \frac{P \cdot R}{P + R}, \quad (3)$$

$AP$  (Average Precision): значение данной метрики вычислялось как усреднение значений точности при разных значениях порога:

$$AP = \frac{\sum_{r=1}^N P_r}{N}, \quad (4)$$

где  $P_r$  – значение точности (Precision) при фиксированном значении порога  $r$ , а  $N$  – множество значений порога;

$mAP$  (mean Average Precision): данная метрика является одной из ключевых для анализа качества работы детекторов объектов и представляет собой усреднение значений метрики  $AP$  по всем классам; в данном исследовании число классов равно 1, следовательно, значение метрик  $AP$  и  $mAP$  совпадает.

На рис. 3 приведены кривые зависимости значений точности от полноты, построенные на стандартном наборе ETIS-LaribPolypDB для моделей, обученных с использованием стандартной и модифицированной аугментации. Как следует из приведенных данных, при использовании обоих типов аугментации наблюдаются достаточно высокие значения метрики  $mAP$ . Однако максимальное ее значение получено при использовании модифицированного варианта аугментации ( $mAP = 0,80$ ). Наибольшее значение метрики  $F1$  для стандартной аугментации составило 0,76, а для модифицированного алгоритма аугментации – 0,77.

Таким образом, наилучшие результаты работы нейронной сети SSD получены при использовании разработанного модифицированного варианта аугментации.

#### Тестирование нейросетевого алгоритма

В стандартном варианте модели на основе архитектуры SSD размер входного слоя нейронной сети составляет 300×300, однако согласно работе [13] в качестве базовой сети в данной архитектуре могут успешно использоваться и другие варианты СНС.

Для создания базы изображений использовалось оборудование для колоноскопии Olympus с разрешением

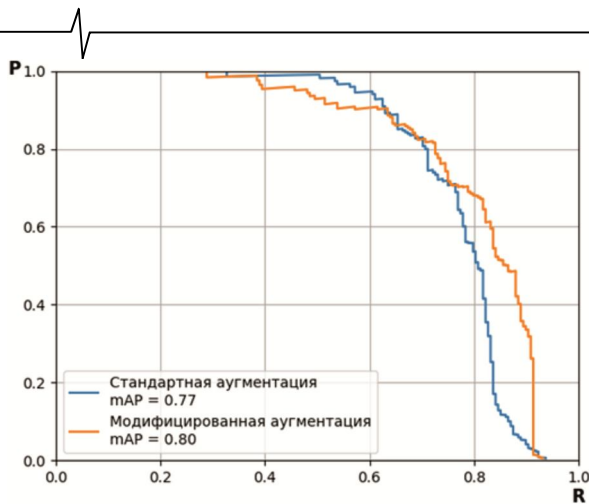


Рис. 3. Сравнение качества работы нейросетевых моделей, обученных с использованием стандартной и модифицированной аугментации

ем кадра видеоданных, равным  $626 \times 532$  пикселей. В связи с этим, принято решение использовать в исследовании помимо стандартной архитектуры SSD-300 также архитектуру SSD с базовой сетью VGG, размер входного слоя которой составляет  $512 \times 512$ , при этом используется еще один дополнительный масштаб карт признаков для детектирования.

Кроме того, в связи с высокими требованиями разрабатываемой системы к скорости работы алгоритма детектирования, связанной с необходимостью обработки данных в режиме реального времени, реализована архитектура сети SSD, в которой в качестве базовой сети использовалась часть архитектуры сети MobileNet [18]. Особенностью архитектуры MobileNet является ее «легковесность» относительно других сетей, связанная с отсутствием слоев пулинга, вместо которых используются свертки с шагом 2.

Таким образом, для дальнейших исследований выбраны следующие архитектуры СНС:

- 1) SSD с базовой сетью VGG, размер входного слоя  $300 \times 300$  (SSD VGG-300);
- 2) SSD с базовой сетью VGG, размер входного слоя  $512 \times 512$  (SSD VGG-512);
- 3) SSD с базовой сетью MobileNet, размер входного слоя  $300 \times 300$  (SSD MobileNet-300).

Кривые зависимости значений точности от полноты, построенные на наборе изображений ETIS-LaribPolypDB для моделей на основе сетей SSD VGG300, SSD

VGG512 и SSD MobileNet300 приведены на рис. 4. Результаты показывают преимущество нейросетевого алгоритма, построенного на основе базовой сети VGG-512, настроенного под характеристики входных колоноскопических видеоданных.

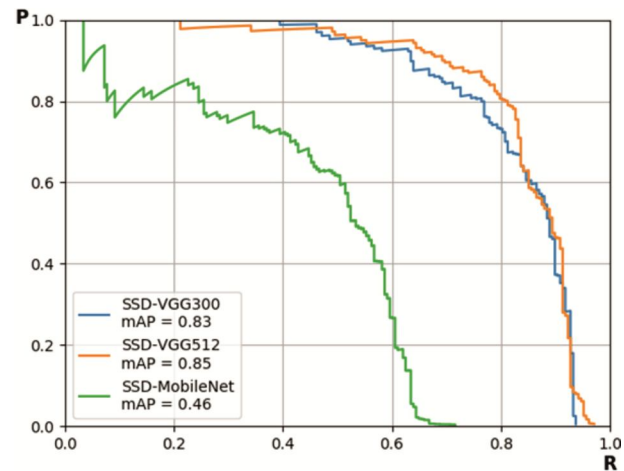


Рис. 4. Сравнение моделей на основе сетей VGG300, VGG512 и MobileNet300

Значение метрик  $P$ ,  $R$  и  $F1$  для исследуемых архитектур в сравнении с алгоритмами, рассмотренными в работах [7, 8], приведены в табл. 1.

Из приведенных результатов следует, что предложенный и обученный нейросетевой алгоритм на базе SSD VGG-512 превосходит другие алгоритмы по метрикам точности и  $F1$ , но уступает некоторым алгоритмам по критерию полноты.

Для оценки производительности обученных нейросетевых моделей проведен тест на персональном компьютере со следующими характеристиками: CPU AMD Ryzen 7 3800X, RAM 64 GB, GPU GeForce GTX 1080 Ti. В тесте измерялось время обработки короткого видеосфрагмента, состоящего из 1249 кадров. Перед обработкой не производилась оптимизация моделей с помощью Nvidia TensorRT, обработка выполнялась с помощью фреймворка машинного обучения TensorFlow.

Для алгоритма на основе SSD VGG-300 скорость обработки составила 62 кадра в секунду, для SSD VGG-512 – 55 кадров в секунду, а для SSD MobileNet-300 – 84 кадра в секунду. Таким образом, предложенные алгоритмы могут применяться для обработки эндоскопического видеопотока в режиме реального времени.

Таблица 1. Сравнение параметров различных моделей на тестовом наборе изображений ETIS-LaribPolypDB

Модель	Точность, $P$	Полнота, $R$	$F1$
SNU	10,20	9,60	9,70
ETIS-LARIB	6,90	49,50	12,20
CVC-Clinic	10,00	49,00	16,50
PLS	15,80	57,20	24,90
UNS+UCLAN	32,70	52,80	40,40
CUMED	72,30	69,20	70,70
OUS	69,70	63,00	66,10
FCN-VGG	73,61	<b>86,31</b>	79,46
Faster R-CNN	72,93	80,29	76,43
U-NET	80,48	81,25	80,86
SSD VGG-300	80,71	76,44	78,51
<b>SSD VGG-512</b>	<b>87,36</b>	76,44	<b>81,53</b>
SSD MobileNet-300	61,76	50,48	55,55



---

## Заключение

В работе предложен алгоритм детектирования полипов на колоноскопических изображениях на основе архитектуры CHC SSD. Данный алгоритм обладает производительностью, достаточной для анализа видеопотока в режиме реального времени, а также имеет сравнительно высокие значения метрик оценки качества алгоритма. Наибольшее значение метрики  $F1 = 81,53$  получено для алгоритма SSD с базовой сетью VGG-512. Для обучения нейросетевого алгоритма совместно с врачами-эндоскопистами Ярославской областной клинической онкологической больницы собрана соответствующая база видеоданных полипов.

Перспективным направлением для дальнейших исследований является улучшение качества работы предложенного алгоритма за счет постобработки результатов с использованием дополнительной информации из соседних кадров видеопотока, а также улучшение качества входного изображения за счет применения алгоритмов деинтерлейсинга.

*Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-37-90153.*

## Литература

1. Rangayyan R.M. Biomedical Image Analysis. – CRC Press, Boca Raton, 2005. – 1306 p.
2. Кирсанова А.В. Современное состояние и перспективы развития экспертных медицинских систем // Новый университет. Серия «Технические науки», 2015, № 11-12. С. 45-46.
3. Куваев Р.О., Никонов Е.Л., Кашин С.В., Капанов В.А., Гвоздев А.А. Контроль качества эндоскопических исследований, перспективы автоматизированного анализа эндоскопических изображений // Кремлевская медицина. Клинический вестник, № 2, 2013. С. 51-56.
4. Khryashchev V., Stepanova O., Lebedev A., Kashin S., Kuvaev R. Deep Learning for Gastric Pathology Detection in Endoscopic Images // ACM International Conference Proceeding Series, 3rd International Conference on Graphics and Signal Processing, ICGSP 2019. Hong Kong, 2019. Pp. 90-94.
5. Хрящев В.В., Ганин А.Н., Лебедев А.А., Степанова О.А., Кашин С.В., Куваев Р.О. Разработка и анализ алгоритма детектирования патологий на эндоскопических изображениях желудка на основе сверточной нейронной сети // Цифровая обработка сигналов. 2018. № 3. С. 70-75.
6. Levin B, Lieberman D, McFarland B, Andrews K, Brooks D, Bond J et al. Screening and Surveillance for the Early Detection of Colorectal Cancer and Adenomatous

Polyps, 2008: A Joint Guideline From the American Cancer Society, the US MultiSociety Task Force on Colorectal Cancer, and the American College of Radiology. Gastroenterology. 2008; 134(5). P. 1570-1595.

7. Sun X., Zhang P., Wang D., Cao Y., Liu B. Colorectal Polyp Segmentation by U-Net with Dilation Convolution. 18th IEEE International Conference On Machine Learning And Applications (ICMLA), 2019.

8. Sornapudi S., Meng F., and Yi S. Region-based automated localization of colonoscopy and wireless capsule endoscopy polyps. Applied Sciences, vol. 9, no. 12, 2019.

9. Николенко С.И., Кадуринов А.А., Архангельская Е.О. Глубокое обучение – СПб.: Питер, 2018. 480 с.

10. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение. ДМК-Пресс, 2017. 652 с.

11. Лебедев А.А., Хрящев В.В., Казина Е.М., Среднякова А.С., Журавлева А.С. Распознавание устья червеобразного отростка на эндоскопических изображениях прямой кишки на основе сверточной нейронной сети // Цифровая обработка сигналов и ее применение (DSPA-2020): докл. 22-й междунар. конф. Москва, 2020. С. 638-642.

12. Хрящев В.В., Среднякова А.С., Ганин А.Н., Кашин С.В. Использование глубоких нейронных сетей для поиска патологий на эндоскопических изображениях желудка // Цифровая обработка сигналов и ее применение (DSPA-2021): докл. 23-й междунар. конф. Москва, 2021. С. 254-258.

13. Liu W., Anguelov D., Erhan D., Szegedy C., and Reed S.E. SSD: Single Shot Multibox Detector. CoRR, abs/1512.02325, 2015.

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

15. Bernal J., Sanchez J. and Vilarico F. Towards automatic polyp detection with a polyp appearance model. Pattern Recognition. Volume 45, Issue 9, September 2012, pp. 3166-3182.

16. Silva J.S., Histace A., Romain O., Dray X., Granda B. Towards embedded detection of polyps in WCE images for early diagnosis of colorectal cancer. International Journal of Computer Assisted Radiology and Surgery, Springer Verlag (Germany), 2014. 9 (2). Pp. 283-293.

17. Howard A.G. Some improvements on deep convolutional neural network based image classification // arXiv preprint arXiv:1312.5402, 2013.

18. Ioffe S., Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Proceedings of ICML, 2015. Pp. 448-456.