

НЕЭТАЛОННАЯ ОЦЕНКА КАЧЕСТВА ИЗОБРАЖЕНИЙ КОЛОНОСКОПИЧЕСКИХ ИССЛЕДОВАНИЙ

Хрящев В.В., к.т.н., доцент кафедры цифровых технологий и машинного обучения Ярославского государственного университета им. П.Г. Демидова, e-mail: v.khryashchev@uniyar.ac.ru

Седов А.Г., старший преподаватель кафедры дискретного анализа Ярославского государственного университета им. П.Г. Демидова, e-mail: a.sedov@uniyar.ac.ru

Приоров А.Л., д.т.н., профессор кафедры цифровых технологий и машинного обучения Ярославского государственного университета им. П.Г. Демидова, e-mail: pri@uniyar.ac.ru

NON-STANDARD ASSESSMENT OF THE IMAGE QUALITY OF COLONOSCOPIC EXAMINATIONS

Khryashchev V.V., Sedov A.G., Priorov A.L.

The article investigates the possibility of using classical and deep machine learning methods for non-reference assessment of the quality of images obtained during colonoscopic examination. A test set of colonoscopic images was created, its expert assessment was carried out, the correlation of a number of popular non-reference algorithms (BRISQUE, NIQE, TOPIQ, PaQ-2-PiQ) with expert values was estimated, and two proprietary algorithms using machine learning methods were proposed. Typical artifacts and distortions in images of this type are considered: blur, glare, interlacing effects. The results show that both proposed algorithms based on machine learning methods generally cope with image quality assessment, while slightly underestimating the average expert assessment. The findings and recommendations can be used in developing a video stream analysis system in an endoscopic system operating in real time during colonoscopic examinations.

Key words: image quality assessment, non-reference algorithms, endoscopy, colonoscopic images, machine learning, neural networks.

Ключевые слова: оценка качества изображений, неэталонные алгоритмы, эндоскопия, колоноскопические изображения, машинное обучение, нейронные сети.

Введение

В современных системах медицинской диагностики все чаще используются различные визуальные технологии, основанные на методах и алгоритмах цифровой обработки статических и динамических изображений (видеопоследовательностей). Традиционно для этих целей используются также методы как классического, так и глубокого машинного обучения, повышающие уровень интеллектуализации таких систем. В свою очередь, это дает возможность частичной автоматизации как диагностики, так и контроля качества соответствующих медицинских процедур [1-4], повышающих общий уровень медицинского обслуживания населения.

Примером успешного применения указанных визуальных технологий и систем искусственного интеллекта является эндоскопия желудочно-кишечного тракта, в том числе колоноскопические исследования кишечника [1, 2].

Технические особенности современных эндоскопических систем таковы, что они имеют некоторые недостатки, в определенной степени осложняющие анализ полученных с их помощью колоноскопических изображений, причем это характерно как для работы врачей-экспертов, так и при использовании алгоритмов обработки информации программно-аппаратными комплексами [1, 5-7]. К ним можно отнести:

Исследована возможность использования методов классического и глубокого машинного обучения для неэталонной оценки качества изображений, полученных при колоноскопическом обследовании. Создан тестовый набор колоноскопических изображений, проведена его экспертная оценка, оценена корреляция ряда популярных неэталонных алгоритмов (BRISQUE, NIQE, TOPIQ, PaQ-2-PiQ) с экспертными значениями, предложено 2 собственных алгоритма, использующих методы машинного обучения. Рассмотрены типовые артефакты и искажения на изображениях рассматриваемого типа: размытие, блики, эффекты чересстрочности. Результаты показывают, что оба предложенных алгоритма, базирующихся на методах машинного обучения, в целом справляются с оцениванием качества изображений, при этом несколько занижая среднюю экспертную оценку. Полученные выводы и рекомендации могут использоваться при разработке системы анализа видеопотока в эндоскопической системе, работающей в режиме реального времени, при проведении колоноскопических исследований.

– наличие размытия на полученных изображениях, вызванного отсутствием автофокусировки на современных эндоскопах, что обусловлено их конструктивными особенностями;

– дополнительное размытие изображений, возникающее из-за турбулентного движения эндоскопа в полости исследуемого органа;

– блики на слизистой оболочке кишечника, обусловленные отражением света источника, находящегося на конце вводимого в него световода;

– яркость, контрастность и некоторые другие характеристики получаемого эндоскопического изображения

могут меняться в значительных пределах, поскольку различна геометрия анализируемой области;

– видеопроцессор эндоскопической системы выдает видеопоследовательности с эффектом чересстрочной развертки, что приводит к необходимости дополнительной обработки.

Указанные недостатки приводят к необходимости дополнительной процедуры контроля качества сохраняемых изображений и видеопоследовательностей, которые в дальнейшем используются как входные данные для систем машинного обучения. В большинстве практических случаев после проведения обследования врач-эндоскопист должен сохранить полученный набор изображений согласно соответствующему регламенту в медико-информационную систему (МИС). С большой вероятностью произвольно взятый из видеопотока колоноскопический кадр не будет достаточно информативным. Следовательно, на данном этапе развития систем анализа изображений в эндоскопии актуален вопрос о контроле качества сохраненных кадров в МИС с использованием алгоритмических критериев оценивания качества изображений.

Оценка качества – важнейшая часть процесса исследования при разработке систем обработки и анализа изображений [8, 9]. Как известно, проводить ее можно как путем усреднения экспертных оценок, так и путем разработки соответствующих алгоритмов. Кажется естественным использовать для указанных целей эталонные алгоритмы оценки качества изображений, дающие наилучшие результаты. Однако во многих реальных задачах эталонное изображение, необходимое для этих алгоритмов, отсутствует, поэтому приходится использовать неэталонные алгоритмы. Они подразделяются на два больших класса по признаку использования либо неиспользования априорной информации о виде искажения анализируемого изображения (рис. 1).

К неэталонным алгоритмам оценки качества изображений, учитывающим априорную информацию, относятся те, в которых определенным образом измеряется уровень соответствующего типа искажения. Примерами могут служить различные виды шумов, размытие, артефакты при сжатии по соответствующим стандартам (чаще всего – это стандарты сжатия статических изображений JPEG и JPEG2000).

На современном этапе развития визуальных технологий значительное внимание разработчиков уделяется

созданию и исследованию таких неэталонных алгоритмов оценки качества, работа которых уже не ограничивается только одним типом искажения изображений. К таким методам и алгоритмам, например, относятся метрики, базирующиеся на статистике естественных изображений (NSS – Natural Scene Statistics). Важно отметить, что указанная статистика может рассчитываться и в пространственной области [10, 11], и в области трансформант вейвлет преобразования, а также дискретного косинусного преобразования [12, 13], которые традиционно широко используются в цифровой обработке изображений. Кроме того, для неэталонной оценки качества изображений могут применяться различные методы как классического, так и глубокого машинного обучения [14, 15].

Обучение алгоритмов традиционно осуществляется на предварительно собранной базе изображений с различными типами искажений и усредненными оценками экспертов MOS (Mean Opinion Score). В практических приложениях часто используют такие общедоступные наборы изображений, как LIVE [16], TID2013 [17] и др. Указанный этап обучения алгоритмов нужен для того, чтобы определить функцию, связывающую значения соответствующих признаков и оценку качества изображений на выходе системы обработки. Фактически поиск такой функциональной зависимости часто сводится к решению задачи регрессии, которая обычно выполняется при помощи различных методов машинного обучения, включая и глубокое обучение. При использовании для решения указанной задачи алгоритмов неэталонной оценки качества часто применяют рандомизированные деревья, машину опорных векторов, а также различные нейросетевые архитектуры.

Использование методов глубокого машинного обучения в таких задачах также возможно, но требует сбора и оценивания большой базы изображений, релевантных решаемой практической задаче.

Рассмотрим особенности построения неэталонных алгоритмов оценивания качества изображений на наборе изображений с колоноскопического исследования.

Набор колоноскопических изображений

Собранная совместно с врачами Ярославской областной клинической онкологической больницы база состоит из 1000 статических изображений, взятых из 10 колоноскопических видеопоследовательностей раз-

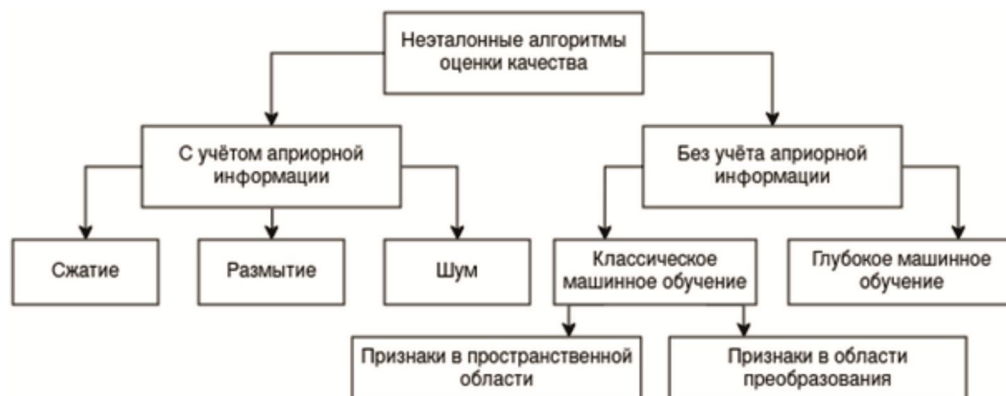
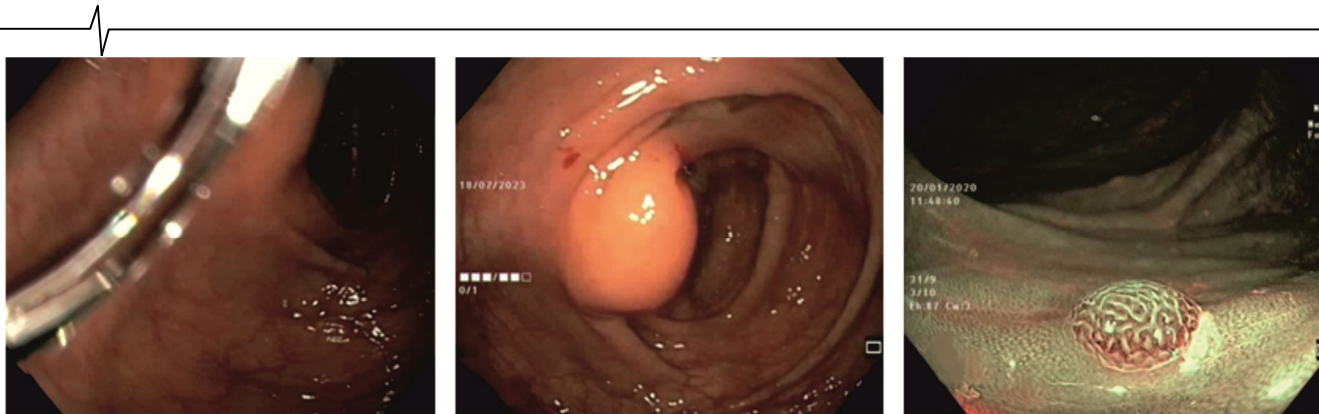


Рис. 1. Классификация неэталонных алгоритмов оценки качества изображений



а) MOS = 1,6

б) MOS = 2,8

в) MOS = 4

Рис. 2. Колоноскопические изображения с усредненными экспертными оценками

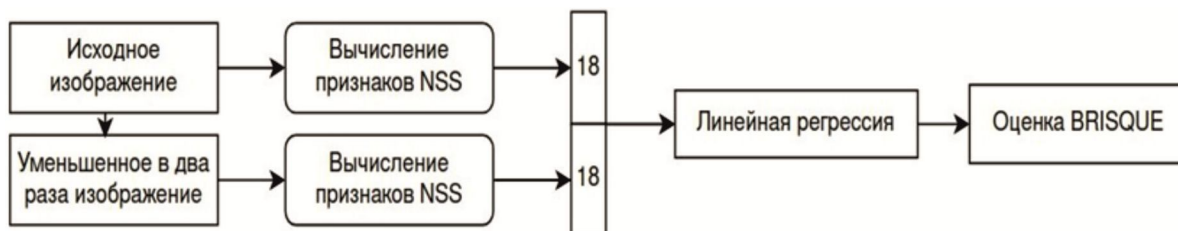


Рис. 3. Структурная схема вычисления метрики BRISQUE

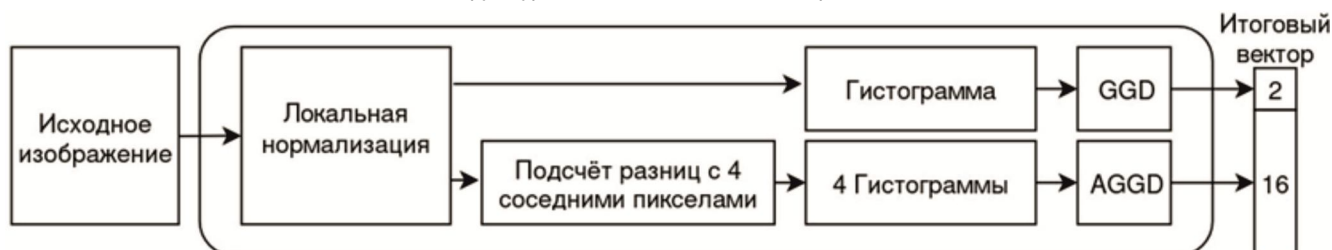


Рис. 4. Структурная схема вычисления признаков NSS

личных исследований. Четыре видеопоследовательно-сти получены на оборудовании с чересстрочным форматом изображения, на остальных видеоданных эффект чересстрочности (ЭЧ) отсутствует. Статические изображения нарезались из видеоданных путем сохранения из них каждого десятого кадра. Затем вручную отбрасывались последовательные кадры, не имеющие заметных видимых различий. Из оставшихся кадров случайным образом выбрано 1000 изображений. Изображения с ЭЧ в итоге составили около половины от всех собранных в наборе [18].

Далее происходила процедура субъективного (экспертного) оценивания качества собранного набора изображений. Группа из пяти экспертов поставила каждому изображению оценки от «1» до «5». Оценка «1» ставилась, если изображение являлось полностью не информативным. Оценка «5» ставилась для идеального четкого изображения. Для изображений с оценкой «3» допускалось наличие зашумленных или размытых областей, при условии наличия полезных областей. Промежуточные оценки «2» и «4» ставились, если изображения логически сложно было соотносить с установленными критериями для оценок «1», «3» или «5».

На рис. 2 показаны примеры колоноскопических изображений с усредненными экспертными оценками качества MOS.

На этом этапе дополнительно рассчитывался также ряд статистических показателей. Установлено, что ко-

эффициент линейной корреляции оценок экспертов друг с другом составил от 0,6 до 0,68. Экспертам дополнительно предлагалось повторно оценить 50 изображений. Корреляция собственных оценок в двух случаях составила от 0,75 до 0,84. Рассчитывалось также среднее арифметическое пяти экспертных оценок MOS. Среднее квадратичное отклонение полученной оценки составило 0,72.

Анализ известных незатонных алгоритмов

Алгоритм расчета метрики BRISQUE, приведенный на рис. 3, сопоставляет каждому изображению вектор, состоящий из 36 признаков [11]. Из них 18 признаков вычисляется для исходного изображения, а другие 18 – для изображения, уменьшенного в 2 раза. Итоговое значение метрики вычисляется с использованием предварительно обученного на базе экспертных оценок алгоритма машинного обучения. Для расчета признаков алгоритм BRISQUE вычисляет распределение коэффициентов, отвечающих за распределение статистики в естественных изображениях – NSS (рис. 4) оцениваемого изображения. Признаки представляют собой коэффициенты симметричного (GGD) и асимметричного (AGGD) нормальных распределений, наилучшим образом аппроксимирующих наблюдаемую статистику данных.

Алгоритм NIQE [19] оценивает различие между распределением коэффициентов NSS для изображения и приближением того же самого распределения с исполь-

Таблица 1. Значения коэффициента линейной корреляции

Алгоритм	Все изображения	Изображения с ЭЧ	Изображения без ЭЧ
BRISQUE	0,46	0,71	0,08
NIQE	0,5	0,67	0,37
TOPIQ	0,29	0,18	0,36
PaQ-2-PiQ	0,18	0,38	0,36

Таблица 2. Значения среднеквадратичной ошибки

Алгоритм	Все изображения	Изображения с ЭЧ	Изображения без ЭЧ
LRG	$MSE = 0,20$	$MSE = 0,10$	$MSE = 0,20$
PR	$MSE = 0,14$	$MSE = 0,11$	$MSE = 0,18$

зованием многомерного нормального распределения. Метрика NIQE для своего расчета не использует методы машинного обучения на экспертных оценках.

Известно, что традиционные метрики оценки качества плохо решают проблему контекста: они не способны анализировать содержимое изображения, хотя оно влияет на субъективную оценку качества. Для решения этой проблемы исследователи прибегают к созданию алгоритмов, использующих сверточные нейронные сети и другие алгоритмы глубокого машинного обучения. К ним, например, относятся алгоритмы TOPIQ [20] и PaQ-2-PiQ [21]. Область их применения на практике ограничена наличием большого (желательно от 10000 штук) обучающего набора изображений с экспертными оценками MOS.

В ходе проведенного эксперимента рассчитан коэффициент линейной корреляции между экспертной оценкой MOS и описанными выше известными алгоритмическими метриками неэталонной оценки качества:

$$\rho_{x,y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}.$$

Здесь x_i, y_i – экспертные и алгоритмические оценки i -го изображения, а \bar{x}, \bar{y} – средние значения оценок на всем множестве изображений.

Исследование проводилось: на всем наборе изображений, только на изображениях с ЭЧ, только на изображениях без ЭЧ. Результаты приведены в табл. 1. Вычисление значений метрик произведено с использованием библиотеки IQA-PyTorch [22].

Алгоритмы BRISQUE, NIQE, TOPIQ оценивают степень зашумленности, поэтому обратно коррелируют с качеством изображения, поэтому в табл. 1 приведены абсолютные значения коэффициента корреляции. Установлено, что наилучшие результаты на всем наборе демонстрируют метрики BRISQUE и NIQE, использующие признаки NSS и два уровня пространственной пирамиды (используется как исходное, так и уменьшенное в два раза изображение, см. рис. 3, [11]).

Одним из этапов вычисления признаков NSS является нахождение распределений разностей значений между соседними пикселями. В частности, вычисляются разности с соседями по горизонтали и вертикали. У изображений с ЭЧ эти распределения будут отличаться, но только на первом уровне пространственной пирамиды. На ее втором уровне изображение уменьшается в

два раза, и характерные ЭЧ пропадают. Эти особенности могут объяснять высокую корреляцию оценок алгоритмов BRISQUE и NIQE для изображений с ЭЧ.

Разработка алгоритмов на основе признаков NSS

Необходимо было обучить алгоритм машинного обучения на наборе колоноскопических изображений, не изменяя способов вычисления коэффициентов NSS. При этом рассматривались методы линейной регрессии (LRG) и использование персептрона (PR) с одним скрытым слоем (100 нейронов в скрытом слое, функция активации ReLU). В качестве функции потерь использована среднеквадратичная ошибка (MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^n (z_i - \bar{z}_i)^2,$$

где z_i и \bar{z}_i – исходная и предсказанные оценки для i -го изображения. Набор изображений делился на обучающее и тестовое множество в соотношении 7:3. В табл. 2 приведены результаты, полученные на тестовом множестве.

Разработка алгоритма на основе сверточной нейронной сети

На субъективную оценку качества изображения среди прочих факторов отрицательно влияют низкая освещенность и наличие бликов отраженного света. Кроме того, движение эндоскопа в органах желудочно-кишечного тракта, в том числе и кишечника, приводит к появлению артефактов размытия на изображениях. Напротив, если на изображении отчетливо видны стенки внутренних органов – это положительно влияет на субъективную оценку. Выявление перечисленных выше признаков должно быть под силу алгоритму, базирующемуся на небольшой СНС. Одна из первых попыток использования СНС для оценки качества изображений приведена в работе [14].

В используемой архитектуре (рис. 5) несколько сверточных слоев возвращают на выходе тензор размерности $17 \times 17 \times 50$. Затем, как и в оригинальной работе, в каждом из 50 каналов извлекаются максимальное и минимальное значение. Кроме того, добавлено вычисление среднего значения в каждом канале. После конкатенации получается вектор из 150 элементов. С использованием нескольких полносвязных слоев вычисляется оценка качества. Далее в работе будем называть эту модель MMA-CNN (min+max+average – convolutional neural network).

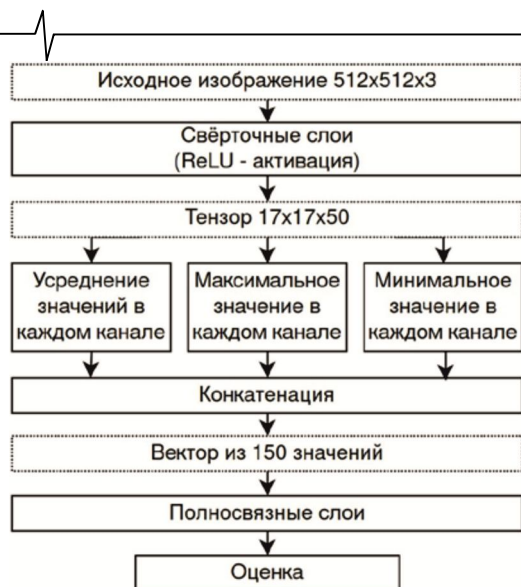


Рис. 5. Алгоритм оценки качества изображения на основе сверточной нейронной сети

В результате обучения среднеквадратичная ошибка на тестовом множестве составила 0,2, как и при использовании линейной регрессии по признакам NSS.

Также визуальный анализ значений в последнем канале сверточного слоя позволяет сделать следующие наблюдения:

Слишком темные и слишком яркие области отчетливо выделились в отдельных каналах. Один канал имеет отклики только для изображений с ЭЧ.

Области «высокого качества» не имеют отдельных каналов.

Примерно на четверти всех изображений виден зонд эндоскопа. Зачастую он неподвижен на подверженном размытию движению изображении или наоборот, размыт или засвечен на переднем плане в целом качественного изображения. Изображение зонда не представляет никакой ценности и не должно влиять на субъективную оценку, однако извлекаемые признаки слишком низкоуровневые, чтобы нейронная сеть научилась его игнорировать.

Сравнительный анализ алгоритмов

На новом тестовом множестве из 300 изображений

вычислен коэффициент линейной корреляции (ρ) с экспертными оценками MOS. Результаты приведены в табл. 3. Они позволяют сделать вывод о преимуществе использования алгоритмов PR и MMA-CNN, обученных на колоноскопических изображениях, по сравнению с универсальными неэталонными алгоритмами типа BRISQUE.

Таблица 3. Сравнительный анализ рассматриваемых алгоритмов

Алгоритм	Коэффициент линейной корреляции
BRISQUE	0,43
PR	0,88
MMA-CNN	0,83

На рис. 6 приведены примеры изображений с экспертными оценками MOS и оценками, полученными рассмотренными выше алгоритмами. Видно, что оба алгоритма, базирующихся на методах машинного обучения (PR и MMA-CNN), в целом справляются с оцениванием качества изображений, при этом несколько занижая среднюю экспертную оценку.

Заключение

В данном исследовании создан тестовый набор колоноскопических изображений, проведена его экспертная оценка, оценена корреляция ряда неэталонных алгоритмов с экспертными значениями, предложено 2 собственных алгоритма, использующих методы машинного обучения.

На данном этапе исследования рассмотрены модели, использующие только низкоуровневые признаки или их статистику распределения. Такие модели менее подвержены переобучению на небольшом наборе изображений, который чаще всего и имеется в распоряжении исследователей.

Дальнейшая работа будет направлена на выявление высокоуровневых признаков. С учетом относительно небольшого количества изображений в обучающих и тестовых наборах потребуется дополнительная разметка с использованием масок или разбиения на патчи, чтобы обучить нейронную сеть отыскивать области, заслуживающие большего или меньшего внимания.

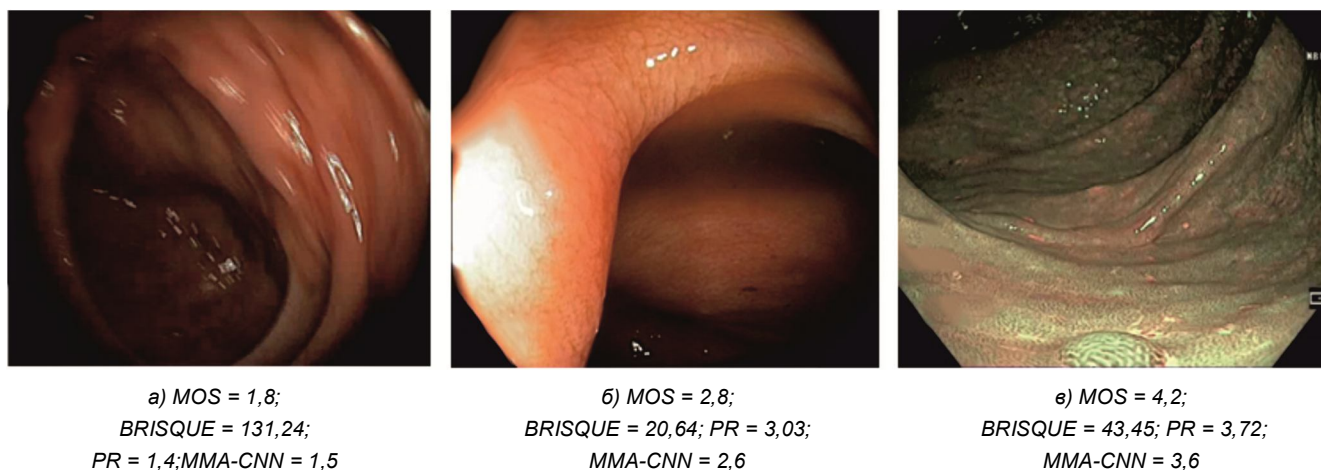


Рис. 6. Изображения с оценками, полученными экспертами (MOS) и рассмотренными алгоритмами

Литература

1. Палевская С.А., Короткевич А.Г. Эндоскопия желудочно-кишечного тракта. М.: ГЭОТАР-Медиа, 2020. 752 с.
2. Куваев Р.О., Никонов Е.Л., Кашин С.В., Капранов В.А., Гвоздев А.А. Контроль качества эндоскопических исследований, перспективы автоматизированного анализа эндоскопических изображений. Кремлевская медицина. Клинический вестник. 2013. Т. 2. С. 51-56.
3. Munzer B., Schoeffmann K., Boszormenyi L. Content-based processing and analysis of endoscopic images and videos: A survey. *Multimed. Tools Appl.* (2018) 77. pp. 1323-1362.
4. Приоров А.Л., Хрящев В.В., Топников А.И. Обработка и передача мультимедийной информации: учебное пособие. Ярославль: ЯрГУ. 2022. 248 с.
5. Лебедев А.А. Исследование нейросетевых алгоритмов обнаружения объектов на видеоизображениях в медицинских системах прикладного телевидения. Диссертация на соискание ученой степени кандидата технических наук. Ярославль. 2022. 131 с.
6. Лебедев А.А., Хрящев В.В., Среднякова А.С., Казина Е.М. Разработка алгоритма детектирования полипов на эндоскопических изображениях с использованием сверточных нейронных сетей. *Цифровая обработка сигналов*. 2021. № 2. С. 55-60.
7. Lebedev A., Khryashchev V., Stefanidi A., Stepanova O., Kashin S., Kuvaev R. Convolutional neural network for early detection of gastric cancer by endoscopic video analysis. *Proceedings of Twelfth International Conference on Machine Vision (ICMV 2019)*. Amsterdam, Netherlands, 2019, pp. 585-590.
8. Дворкович В.П., Дворкович А.В. Цифровые видеоинформационные системы (теория и практика). Техносфера, 2012. 1008 с.
9. Приоров А.Л., Ненахов И.С., Хрящев В.В. Разработка и анализ алгоритма незатонной оценки качества изображений на основе локальных бинарных шаблонов. *Успехи современной радиоэлектроники*. 2016. № 7. С. 46-52.
10. Ruderman D. The statistics of natural images. *Network: Computation in Neural System*, 1994, vol. 5, no 4, pp. 517-548.
11. Mittal A., Moorthy A., Bovik A. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.* 2012, vol. 7, no 12, pp. 4695-4708.
12. Chandler D., Hemami S. VSNR: a wavelet-based visual signal-to-noise ratio for natural images. *IEEE Trans. Image Process.* 2007, vol. 16, no 9, pp. 2284-2298.
13. Saad M., Bovik A., Charrier C. Blind image quality assessment: A natural scene statistics approach in the DCT domain. *IEEE Trans. Image Process.* 2012, vol. 21, no. 8, pp. 3339-3352.
14. Kang L., Ye P., Li Y., Doermann D. Convolutional Neural Networks for No-reference Image Quality Assessment. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2014, pp. 1733-1740.
15. Николенко С.И., Кадури А.А., Архангельская Е.О. Глубокое обучение. Питер: СПб., 2019. 476 с.
16. Sheikh H., Wang Z., Cormack L., Bovik A. LIVE image quality assessment database release 2, 2006. <http://live.ece.utexas.edu/research/quality/subjective.htm>.
17. Ponomarenko N., Jin L., Ieremeiev O., Lukin V., Egiazarian K., Astola J., Vozel B., Chehdi K., Carli M., Battisti F., Jay Kuo C. Image database TID2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication*. 2015, vol. 30, pp. 57-77.
18. Седов А.Г., Хрящев В.В. Разработка и исследование алгоритмов оценки качества колоноскопических изображений. *Цифровая обработка сигналов и ее применение (DSPA-2024): Докл. 26-й междунар. конф. Москва, 2024. С. 244-247.*
19. Mittal A., Soundararajan R., Bovik A. Making a «completely blind» image quality analyzer. *IEEE Signal processing letters* 20.3 (2012): 209-212.
20. Chaofeng Chen, Jiadi Mo, Jingwen Hou, Haoning Wu, Liang Liao, Wenxiu Sun, Qiong Yan, Weisi Lin. TOPIQ: A Top-Down Approach from Semantics to Distortions for Image Quality Assessment. *IEEE Transactions on Image Processing*. 2024. vol. 33, pp. 2404-2418.
21. Ying Z., Niu H., Gupta P., Mahajan D., Ghadiyaram D., Bovik A. From Patches to Pictures (PaQ-2-PiQ): Mapping the Perceptual Space of Picture Quality. *arXiv:1912.10088 [cs.CV]*.
22. Chaofeng C., Jiadi M. IQA-PyTorch: PyTorch Toolbox for Image Quality Assessment. <https://github.com/chaofengc/IQA-PyTorch> 2022.

НОВЫЕ КНИГИ

Поборчая Н.Е.

Методы и алгоритмы оценивания параметров канала связи в условиях априорной неопределенности в системах с приемником прямого преобразования: Учебное издание для вузов, М.: Изд-во «Горячая линия-Телеком», 2023 г. 240 с.: ил.

Изложены методы и алгоритмы совместного оценивания параметров сигнала (канала связи) в условиях априорной неопределенности относительно статистических характеристик канала связи и законов распределения шумов. Особое внимание уделено системам с приемником прямого преобразования. Рассмотрены вопросы синтеза и анализа процедур оценивания для систем связи с одной передающей и приемной антенной (SISO), с несколькими передающими и приемными антеннами (MIMO), а также для систем с ортогональным частотным мультиплексированием (OFDM).

Предложенные алгоритмы способствуют повышению помехоустойчивости приема информации или понижению вычислительной сложности процедур обработки сигнала.

Для научных работников, инженеров и аспирантов. Может быть полезна студентам старших курсов и магистрантам, обучающимся по направлению подготовки «Радиотехника» и «Инфокоммуникационные технологии и системы связи».

