

most effective approaches is based on the deep convolutional neural networks. Nonetheless, it is necessary to use echocardiogram frames of the cardiac muscle, which show the boundaries of the cardiac structures annotated by experts to train it. However, the number of databases containing the necessary information is relatively small. Therefore, generated echocardiogram frames are used to increase the amount of training samples. The article proposes an improved method for generating echocardiograms using a generative adversarial neural network (GAN) with a patch-based conditional discriminator. It has been demonstrated that it is possible to improve the quality of generated echocardiogram frames in both two and four chamber views (AP4C, AP2C) using the masks of cardiac segmentation with sub-pixel convolution layer (pixel shuffle). It is expected that this method will improve the accuracy of solving the direct problem of automatic segmentation of the left ventricle.

Key words: echocardiography, neural networks, generative adversarial networks, image generation, left ventricle.

ВВЕДЕНИЕ

Эхокардиография (ЭхоКГ) является основным способом оценки состояния сердца и диагностики сердечных заболеваний. При использовании ЭхоКГ получают ультразвуковые двумерные изображения сердца (УЗИ-изображения), на которых далее идентифицируют границы внутренних отделов сердца и получают оценки их геометрических размеров. Знание размеров областей сердца позволяет оценить количественные характеристики, используемые для диагностики состояния сердечной мышцы. Например, на основе геометрических размеров и формы левого желудочка (ЛЖ) оценивают: фракцию выброса (ФВ), ударный объем, конечно-систолический и конечно-диастолический объемы ЛЖ. Эти показатели играют важнейшую роль при диагностике заболеваний сердца.

Необходимо отметить, что в большинстве современных программных средств используются ручные и полуавтоматические алгоритмы построения контуров внутренних камер сердца. Сначала эксперт наносит опорные точки, и далее по ним строится контур. Данные методы являются трудоемким процессом и требуют от врачей-кардиологов соответствующего опыта. При этом, определенных правил построения контуров камер на ЭхоКГ снимках, несмотря на имеющиеся руководства Американского общества эхокардиографии [1], нет. Это обусловлено низким качеством снимков, наличием шумов и особенностями распространения ультразвука во внутренних тканях. В этой связи, задача генерации снимков по известным маскам внутренних камер является актуальной.

В работе приводятся результаты генерации изображений с использованием методов глубокого обучения на основе генеративно-сопоставительной нейронной сети с условным дискриминатором (сGAN) с применением набора данных the Cardiac Acquisitions for Multi-structure Ultrasound Segmentation (CAMUS), содержащего изображения сердца 500 пациентов в двухкамерных (2ch) и четырехкамерных (4ch) проекциях по длинной оси ЛЖ.

Существующие методы генерации изображений

Генеративно-сопоставительных сетей (GAN) традиционно используются для решения для различных задач компьютерного зрения, в том числе:

- дорисовки недостающих частей изображения [2-4];
- удаления шумов [4, 5];
- раскрашивания изображений (image colorization) [6, 7];
- 8
- 1
- 0 повышения разрешения изображения [12].

Также известен опыт использования генеративно-сопоставительных сетей для генерации УЗИ-изображений на основе известных масок внутренних областей сердца.

В [13] предлагается использовать две сети, одна из которых сегментирует левый желудочек сердца, а вторая - генеративная учится обратному отображению из эхокардиограммы в сегментационную маску. Данный метод позволяет использовать неразмеченные данные, что в свою очередь, улучшает точность сегментации.

В [14] предложена архитектура генеративно-сопоставительной сети с условным дискриминатором на основе патчей для генерации эхокардиограмм набора данных CAMUS [15] в соответствии с заданными масками сегментации, используемыми в качестве условия синтеза изображения. Данный подход позволяет генерировать изображения размером 256x256 пикселей, однако на синтезированных изображениях отсутствует спекл-шум, присутствующий на соответствующих УЗИ-кадрах, и возникает эффект “шахматной доски”.

Таким образом, ранее полученные результаты свидетельствуют о возможности использования GAN для генерации УЗИ-изображений. Однако необходимо модифицировать GAN с целью устранения отмеченных выше недостатков.

В данной работе, как и в [14], используется набор данных CAMUS, состоящий из ультразвуковых двумерных изображений сердца 500 пациентов.

Для 450 пациентов имеется в открытый доступ выложены кадры с нанесенной на них разметкой: область эндокарда ЛЖ, область эпикарда ЛЖ и область левого предсердия. Тестовый набор состоит из неразмеченных УЗИ-изображений 50 пациентов. Для каждого пациента в наборе данных содержится по 2 изображения в двухкамерной и четырехкамерной проекции сердца для диастолической и систолической фаз сердечного сокращения. Общее количество кадров с экспертной разметкой составляет 1800.

Примеры типичных кадров представлены на рисунке 1.

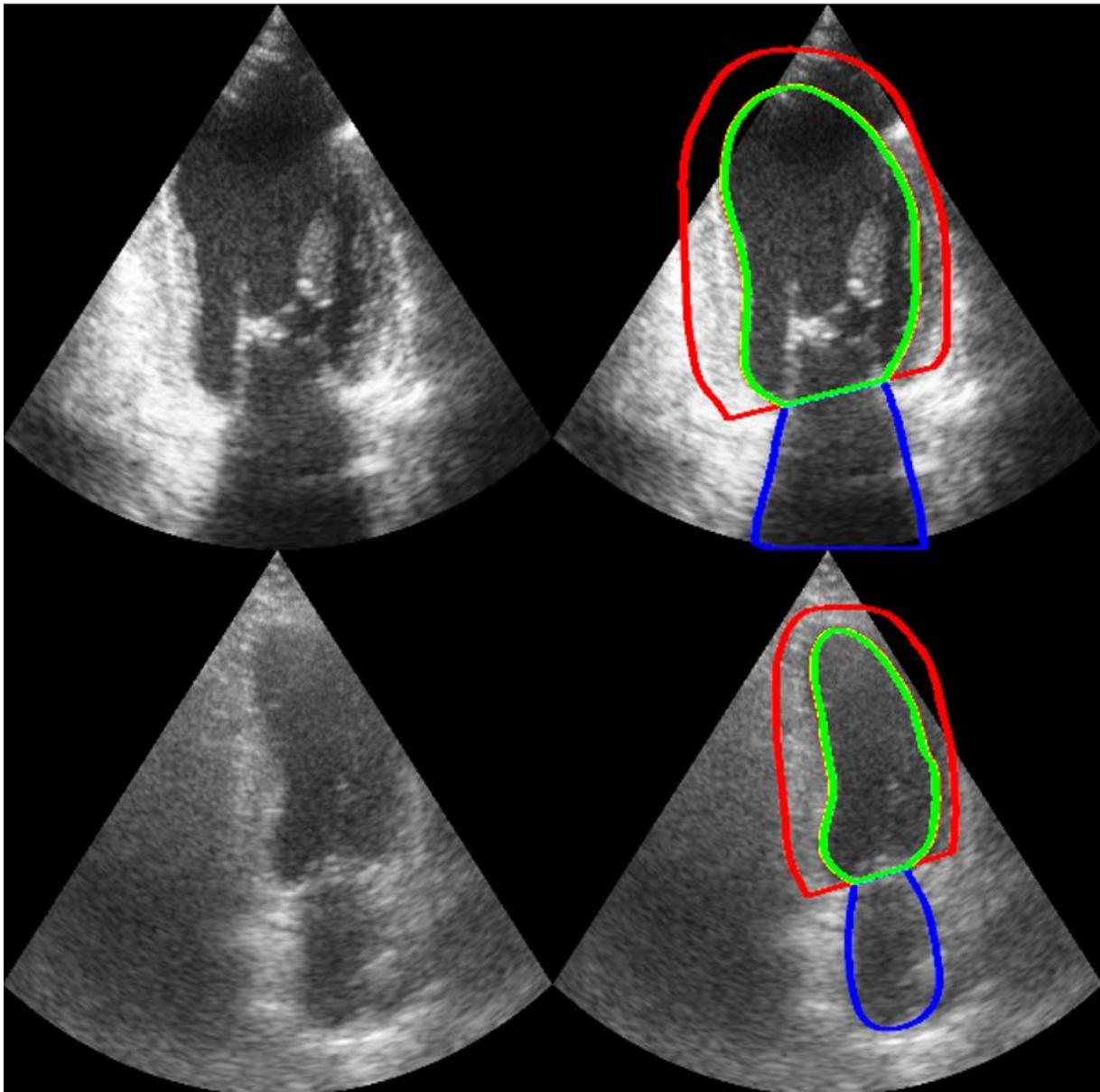


Рисунок 1 – Типичные кадры с экспертной разметкой камер сердца

Модификация архитектуры нейронной сети

В [14] использован процесс обучения, блок-схема которого представлена на рисунке 2.

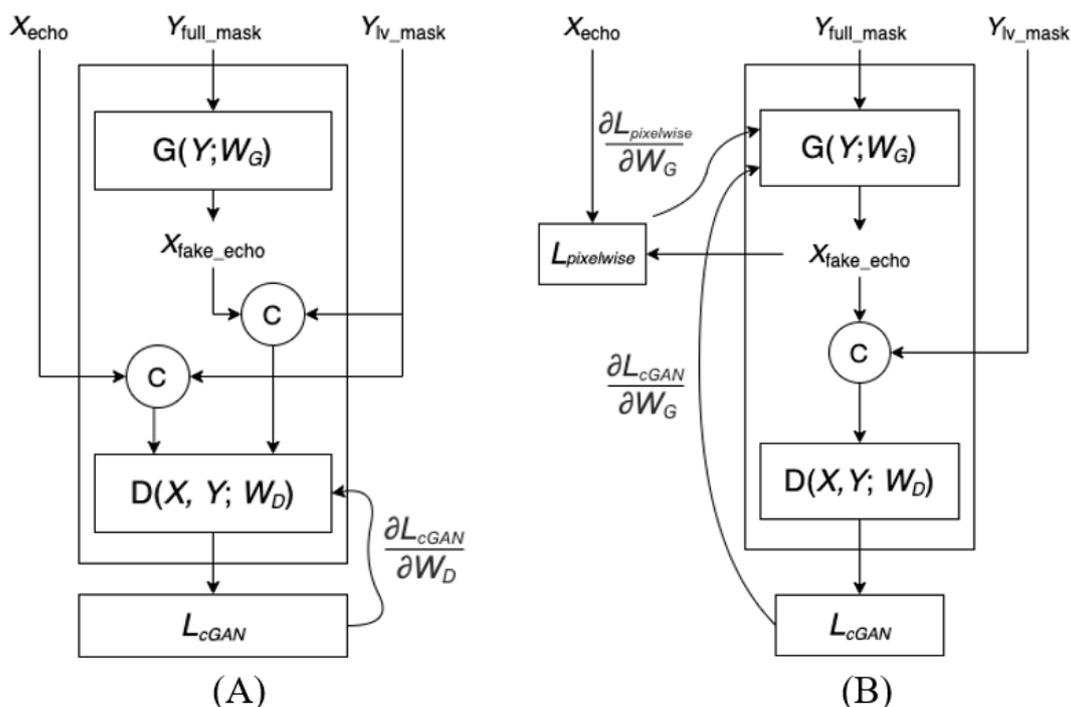


Рисунок 2 – Схема обучения дискриминатора (А) и генератора (В): X_{echo} – УЗИ-изображения, X_{fake_echo} – сгенерированное УЗИ-изображение, Y_{full_mask} – сегментационные маски всех структур сердца, Y_{lv_mask} – сегментационная маска ЛЖ, C – конкатенация изображений, W_G – веса генератора, W_D – веса дискриминатора, $L_{pixelwise}$ – функция потерь для сравнения X_{echo} и X_{fake_echo} , L_{cGAN} – функция потерь для cGAN.

Из рисунка 2 видно, что процесс обучения состоит из двух последовательных этапов:

1. обучение дискриминатора;
2. обучение генератора.

В процессе обучения генеративно-сопоставительной сети с условным дискриминатором на основе патчей (patch-based cGAN) генератор синтезирует на основе экспертных масок УЗИ-изображения, которые далее сравниваются с соответствующими реальными УЗИ-изображениями по критерию средней абсолютной ошибки. Затем генератор обновляет свои веса с учетом значения выбранного критерия. Далее синтезированные кадры и соответствующие маски ЛЖ передаются в дискриминатор, который определяет, являются ли патчи изображения реальными или сгенерированными, используя критерий средней квадратической ошибки.

На рисунке 3 представлены типичные результаты, возвращаемые обсуждаемой [14] нейронной сетью.

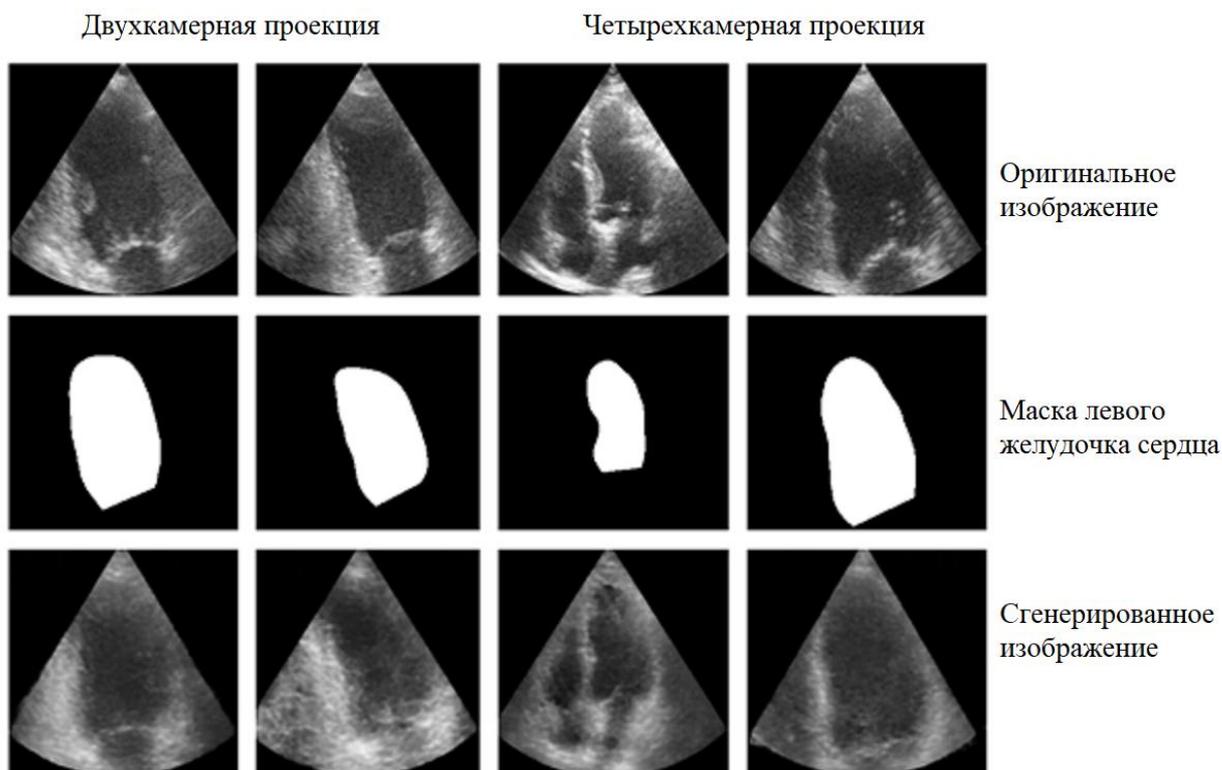


Рисунок 3 – Типичные результаты, возвращаемые обсуждаемой cGAN.

Из рисунка 3 видно, что синтезированным изображениям присущ эффект “шахматной доски” [17], а также на синтезированных УЗИ-изображениях отсутствует спекл-шум, являющийся неотъемлемой частью реальных УЗИ-изображений. Причина возникновения данных проблем связана с использованием в [14] транспонированных сверточных слоев (transposed convolutions) для увеличения размера изображения, реализуемого в генераторе. Здесь использовался генератор на основе стандартной архитектуры Unet [16], состоящий из 7 сверточных и 7 транспонированных сверточных слоев без соединений быстрого доступа (skip-connections). Выходы всех сверточных слоев генератора и дискриминатора, за исключением последних, были нормализованы по батчам. Использованы функции активации LeakyReLU. Необходимо отметить, что при этом применялось ядро свертки 4×4 в каждом блоке генератора. В связи с вышеизложенным была выдвинута гипотеза о возможности избавления от артефактов “шахматной доски” и получения, во-первых, менее размытых изображений с более выраженной текстурой эхокардиограммы, во-вторых, увеличения размера генерируемых изображений до 512×512 , путем изменения архитектуры нейронной сети.

Для подтверждения высказанной гипотезы в архитектуре генератора реализованы соединения быстрого доступа (skip-connections) между блоками энкодера и декодера, а транспонированные сверточные слои (transposed

convolutions) заменены на блоки из сверточных слоев с размером ядра 4×4 , с увеличением количества карт признаков в 4 раза, слоя sub-pixel convolution layer раза.

Результаты

Рассмотрим результаты сравнительного анализа синтезированных УЗИ-изображений с помощью генератора в [14] и модифицированного в данной работе генератора. Здесь генераторы и дискриминаторы обучались с помощью оптимизатора Adam со скоростью обучения 0.00013 и 0.00015 соответственно. Размер батча был равен 8. Размер патча для дискриминатора был равен 16×16 пикселей. Обучение происходило в течение 200 эпох. Типичные УЗИ-изображения ЛЖ синтезированные модифицированного G sGAN, представлены на рисунке 4.

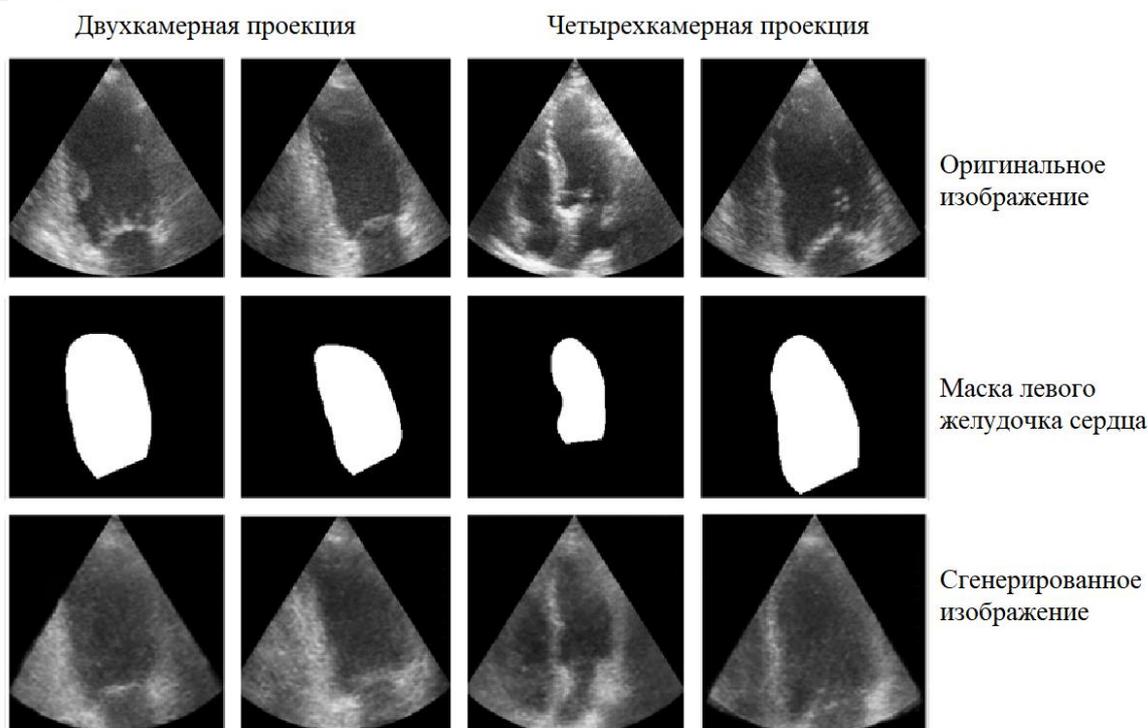


Рисунок 4 – Типичные результаты, возвращаемые сетью улучшенной архитектуры

Из рисунка 4 видно, что предложенная модификация генератора sGAN, действительно, позволила уменьшить эффект “шахматной доски”. Для количественного сравнения синтезированных УЗИ-изображений использовались пиковое отношение сигнала к шуму (PSNR) и индекс структурного сходства (SSIM). Результаты представлены в таблице 1, метрики сглажены с помощью экспоненциального сглаживания ввиду наличия осцилляций. Результаты приведены отдельно для двухкамерной (2ch) и четырехкамерной (4ch) проекций.

Таблица 1 – PSNR и SSIM.

	PSNR		SSIM	
	2ch	4ch	2ch	4ch
cGAN [14] (256×256)	8.404	8.387	0.5569	0.5498
Улучшенный cGAN (256×256)	8.411	8.407	0.5694	0.5680
Улучшенный cGAN (512×512)	8.396	8.389	0.6033	0.6002

Генеративная модель выучилась отображению из сегментационных масок в соответствующие структуры сердца и может генерировать изображения размера 512x512 с сохранением структур сердца.

Заключение

Предложена модификация генератора cGAN, используемого для синтеза УЗИ-изображений на основе экспертных масок ЛЖ в двух и четырехкамерных проекциях сердца. Представлены результаты, подтверждающие, что модифицированный генератор cGAN позволяет уменьшить на синтезированных УЗИ-изображениях эффект “шахматной доски”, а также обеспечить близость их структуры к структуре реальных УЗИ-изображений.

В дальнейших исследованиях планируется разработать подход, обеспечивающий синтез УЗИ-изображений различного качества, необходимых для проверки работоспособности автоматических алгоритмов сегментации областей сердца.

Библиографический список

1. Lang R., Badano L., Mor-Avi V. Recommendations for Cardiac Chamber Quantification by Echocardiography in Adults: An Update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging // European Heart Journal - Cardiovascular Imaging. 2015. N. 16(3). С. 233–271.
2. Yu J., Lin Z., Yang J., Shen X. Generative Image Inpainting With Contextual Attention // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018. С. 5505-5514.
3. Liu G., Reda F., Shih K. Image Inpainting for Irregular Holes Using Partial Convolutions // Proceedings of the European Conference on Computer Vision (ECCV). 2018. С. 85-100.
4. Xie J., Xu L., Chen E. Image Denoising and Inpainting with Deep Neural Networks//Advances in Neural Information Processing Systems 25 (NIPS 2012). 2012.

5. Zhang K., Zuo W., Chen Y. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising // IEEE Transactions on Image Processing. 2017. N. 26(7). C. 3142-3155.
6. Zhang R., Zhu J., Isola P. Real-Time User-Guided Image Colorization with Learned Deep Priors // arxiv.org. 2017. URL: <https://arxiv.org/abs/1705.02999>
7. Nazeri K., Ng E., Ebrahimi M. Image Colorization Using Generative Adversarial Networks. 2018. Articulated Motion and Deformable Objects. AMDO 2018. Lecture Notes in Computer Science. N. 10945.
8. Semmo A., Isenberg T., Döllner J. Neural style transfer: a paradigm shift for image-based artistic rendering? // Proceedings of the Symposium on Non-Photorealistic Animation and Rendering (NPAR '17). 2017. C. 1-13.
9. Karras T., Laine S., Aila T. A Style-Based Generator Architecture for Generative Adversarial Networks // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019. C. 4401-4410.
10. Hong S., Yan X., Huang T., Lee H. Learning Hierarchical Semantic Image Manipulation through Structured Representations // Advances in Neural Information Processing Systems 31. 2018. C. 2713-2723.
11. Lee C., Liu Z., Wu L., Luo P. MaskGAN: Towards Diverse and Interactive Facial Image Manipulation // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020. C. 5549-5558.
12. Johnson J., Alahi A., Fei-Fei L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution // European Conference on Computer Vision. 2016. C. 694-711.
13. Jafari M., Girgis H., Abdi A. Semi-Supervised Learning For Cardiac Left Ventricle Segmentation Using Conditional Deep Generative Models as Prior // 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). 2019. C. 649-652.
14. Abdi A., Tsang T., Abolmaesumi P. GAN-enhanced Conditional Echocardiogram Generation // arxiv.org. 2019. URL: <https://arxiv.org/abs/1911.02121>
15. Leclerc S., Smistad E., Pedrosa J. Deep Learning for Segmentation Using an Open Large-Scale Dataset in 2D Echocardiography // IEEE Transactions on Medical Imaging. 2019. N 38(9). C. 2198-2210.
16. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation // MICCAI. 2015. C. 234-241.
17. Ledig C., Theis L., Huszar F. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. C. 4681-4690.