ОБРАБОТКА И УЛУЧШЕНИЕ КАЧЕСТВА ВИДЕО С ПОМОЩЬЮ ГЕНЕРАТИВНЫХ МОДЕЛЕЙ (GAN)

Тастанова Саида Алдияровна
Техника фанлари бўйича фалсафа доктори, PhD, Доцент, Телевизион ва медиа технологиялар кафедраси
tasmaat@mail.ru
Нарзуллоев Ойбек Мирзаевич
Техника фанлари бўйича фалсафа доктори, PhD, Доцент
Исматова Замира Аъзам қизи
Студентка магистратуры в TUIT
zamiraismatova7@gmail.com
Азларова Р.Ш
Студентка магистратуры в TUIT
studentx656@tuit.uz

ANNOTATSIYA

Maqolada videoma'lumotlarga işlov beriş bosqiçlaridan biri sifatida vizual axborot sifatini tiklaş uçun generativ qarama-qarşi tarmoqlarni (GAN) qöllaş körib çiqiladi. Video ruxsatnomasi tasvir aniqligi pasayganda, alohida kadrlar strukturasi va mayda detallari yöqoladi, şu sababli fazoviy axborotni tiklaş uçun öta yuqori ruxsatnoma super-resolution usullaridan foydalaniş mumkun. Işda eksperimental material sifatida kökrak qafasining sakkizta rentgenografik tasviridan foydalanildi. Bunday tanlov yorqin ifodalangan strukturaga va har xil darajadagi detallaşuvga ega tasvirlarda alohida kadrlarni tiklaş sifatini tahlil qiliş imkonini beradi. Eksperimentda dastlabki tasvirlarning ruxsatnomasi kamaytirildi, söngra ular ikki usulda: bikubik interpolyatsiya hamda Real-ESRGAN generativ modeli yordamida tiklandi. Miqdoriy baholaş uçun dastlabki yuqori ruxsatli tasvirlarga nisbatan hisoblangan PSNR va SSIM metrikalari qöllandi. Bicubic uçun örtacha PSNR qiymati 30,2830 dB ni, Real-ESRGAN uçun esa 29,2774 dB ni taşkil etdi. Örtacha SSIM qiymati mos ravişda 0,7815 va 0,7660 ni taşkil qildi. Şunday qilib, ötkazilgan eksperimentda an'anaviy Bicubic usuli har ikkala miqdoriy metrika böyiça yuqoriroq natijalarni körsatdi. Şuning bilan birga, vizual taqqoslaş Real-ESRGAN ança yorqin yuqori çastotali detallarni şakllantirişini va tasvirni tiklaş xarakteriga köra interpolyatsiya natijasidan farq qilişini körsatdi. Olingan natijalar GAN-modellarni faqat PSNR va SSIM böyiça baholaş tiklangan tasvirning vizual sifatini har doim ham töliq tavsiflay olmasligini tasdiqlaydi. Videoma'lumotlar uçun ketma-

ket kadrlarning vaqtinchalik mosligi qöşimça muammo hisoblanadi va uşbu işda bu keyingi tadqiqotlar yönalişi sifatida körib çiqiladi.

Kalit sözlar: *generativqarama-qarşi tarmoqlar, GAN, Real-ESRGAN, super-resolution, videoga işlov beriş, tasvirlarni tiklaş, PSNR, SSIM, rentgenografik tasvirlar.*

АННОТАЦИЯ

В статье рассматривается применение генеративных состязательных сетей (GAN) для восстановления качества визуальной информации как одного из этапов обработки видеоданных. При снижении разрешения видео происходит потеря мелких структур и деталей отдельных кадров, поэтому методы сверхразрешения могут использоваться для восстановления пространственной информации. В качестве экспериментального материала в работе использованы восемь рентгенографических изображений грудной клетки. Такой выбор позволяет исследовать качество восстановления отдельных кадров на изображениях с выраженной структурой и различными уровнями детализации. В эксперименте исходные изображения подвергались уменьшению разрешения, после чего восстанавливались двумя способами: с использованием бикубической интерполяции и генеративной модели Real-ESRGAN. Для количественной оценки применялись метрики PSNR и SSIM, рассчитанные относительно исходных изображений высокого разрешения. Среднее значение PSNR для Bicubic составило 30,2830 дБ, а для Real-ESRGAN — 29,2774 дБ. Среднее значение SSIM составило 0,7815 и 0,7660 соответственно. Таким образом, в проведённом эксперименте традиционный метод Bicubic показал более высокие значения обеих количественных метрик. Вместе с тем визуальное сравнение показало, что Real-ESRGAN формирует более выраженные высокочастотные детали и отличается от результата интерполяции характером восстановления изображения. Полученные результаты подтверждают, что оценка GAN-моделей только по PSNR и SSIM не всегда позволяет полностью описать визуальное качество восстановленного изображения. Для видеоданных дополнительной проблемой является временная согласованность последовательных кадров, которая в данной работе рассматривается как направление дальнейшего исследования.

Ключевые слова: *генеративные состязательные сети, GAN, Real-ESRGAN, суперразрешение, обработка видео, восстановление изображений, PSNR, SSIM, рентгенографические изображения.*

ANNOTATION

The article examines the use of generative adversarial networks (GANs) to restore the quality of visual information as one of the stages of video data processing. When video resolution is reduced, fine structures and details of individual frames are lost, so super-resolution methods can be used to restore spatial information. Eight radiographic images of the chest were used as experimental material. This choice allows us to study the quality of restoration of individual frames in images with a pronounced structure and varying levels of detail. In the experiment, the original images were subjected to resolution reduction and then restored using two methods: with bicubic interpolation and the generative model Real-ESRGAN. For quantitative assessment, the PSNR and SSIM metrics were used, calculated relative to the original high-resolution images. The average SSIM value was 0.7815 and 0.7660, respectively. Thus, in the experiment conducted, the traditional Bicubic method showed higher values for both quantitative metrics. At the same time, visual comparison showed that Real-ESRGAN generates more pronounced high-frequency details and differs from the interpolation result in the nature of image restoration. The obtained results confirm that evaluating GAN models based solely on PSNR and SSIM does not always allow for a complete description of the visual quality of the restored image. For video data, an additional challenge is the temporal consistency of consecutive frames, which is considered in this work as a direction for further research.

Keywords: *generative adversarial networks, GAN, Real-ESRGAN, super-resolution, video processing, image restoration, PSNR, SSIM, radiographic images.*

ВВЕДЕНИЕ

В последние годы методы глубокого обучения получили широкое распространение в задачах обработки изображений и видео. Одним из направлений, в котором активно применяются нейронные сети, является восстановление изображений с низким разрешением, или super-resolution. Основная задача super-resolution заключается в получении изображения более высокого разрешения на основе изображения, в котором часть пространственной информации была потеряна. Проблема повышения качества особенно актуальна для видеоданных [9]. Видеопоследовательность состоит из большого количества кадров, и снижение разрешения каждого отдельного кадра приводит к потере мелких деталей всей последовательности. Кроме того, при обработке видео необходимо учитывать не только пространственное качество изображения, но и его согласованность во времени. Поэтому методы, разработанные первоначально для обработки отдельных изображений, могут рассматриваться как один из базовых этапов более сложной задачи video super-

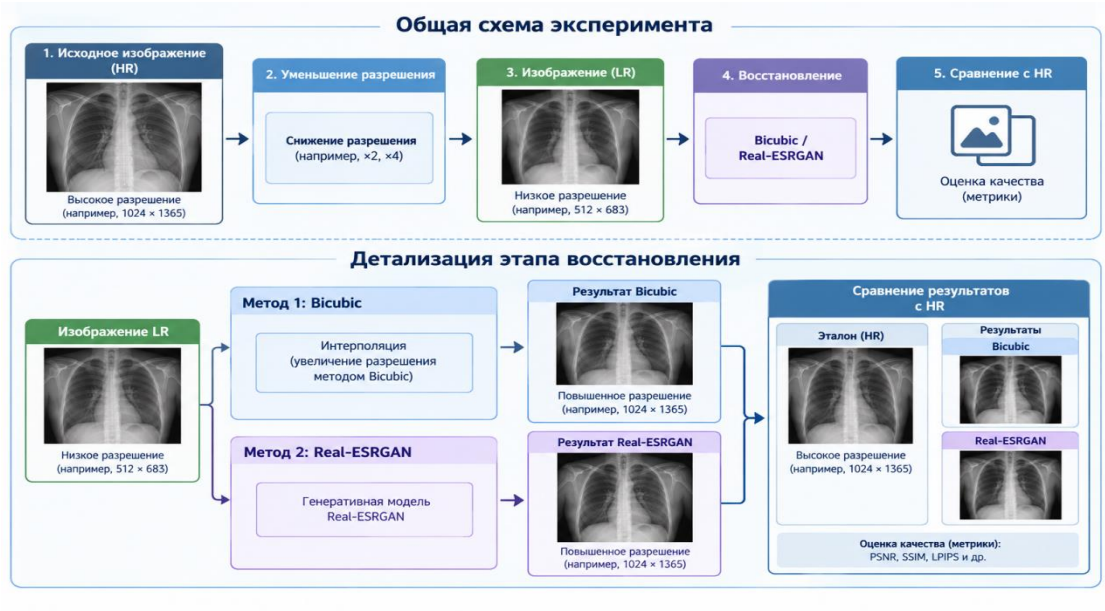
resolution. Одним из первых значимых подходов к использованию генеративных состязательных сетей для сверхразрешения стала модель SRGAN, предложенная C. Ledig и соавторами. В данной работе авторы показали, что использование adversarial loss и perceptual loss позволяет получать более детализированные изображения при увеличении разрешения в четыре раза. При этом авторы отмечали различие между высокой пиксельной точностью и визуальной реалистичностью восстановленного изображения [1]. Следующим важным этапом развития данного направления стала модель ESRGAN. В ней были изменены архитектура генератора, структура дискриминатора и perceptual loss. В частности, авторами были предложены Residual-in-Residual Dense Blocks без batch normalization. Эти изменения были направлены на повышение визуального качества и более реалистичное восстановление текстур. В публикации ESRGAN также подчёркивается проблема появления искусственных деталей при генеративном восстановлении [2]. Дальнейшим развитием данного направления стала модель Real-ESRGAN, предназначенная для восстановления изображений с более сложными и неизвестными деградациями. Авторы предложили использовать многоэтапное моделирование деградаций при создании синтетических обучающих данных. В модели также применяется U-Net-дискриминатор со spectral normalization. Такой подход был разработан для более практического восстановления изображений в условиях, когда тип исходной деградации заранее неизвестен [3,7]. Однако применение генеративных моделей к видео имеет дополнительную проблему. Если каждый кадр обрабатывать независимо, модель может восстанавливать детали немного по-разному на соседних кадрах. В результате возможно появление визуального мерцания или изменения текстур во времени. Для решения данной проблемы были предложены специальные методы, учитывающие временную информацию. Например, TecoGAN использует пространственно-временный дискриминатор, рекуррентную структуру и специальную Ping-Pong loss для повышения временной согласованности [6,9]. Таким образом, при исследовании GAN для обработки видео целесообразно разделять две связанные задачи. Первая заключается в восстановлении пространственной информации отдельного кадра, а вторая — в сохранении согласованности восстановленных кадров во времени. В настоящей работе исследуется первая задача. Для экспериментальной проверки были выбраны рентгенографические изображения грудной клетки. Использование таких изображений позволяет исследовать восстановление структур с различной контрастностью и детализацией. При этом медицинская область в данной работе

рассматривается только как экспериментальный сценарий. Исследование не ставит целью диагностику заболеваний, интерпретацию рентгенограмм или доказательство клинической эффективности рассматриваемого метода. Важной особенностью эксперимента является использование не только визуального сравнения, но и количественных метрик. В работе применяются PSNR и SSIM. PSNR характеризует степень численного различия между восстановленным и эталонным изображением, тогда как SSIM оценивает сходство их структурных характеристик. Исследование различий между этими метриками было представлено, в частности, в работе Horé и Ziou [3].

МЕТОДЫ

Для эксперимента использовались восемь рентгенографических изображений грудной клетки. Изображения имели различное пространственное разрешение. После обработки итоговые размеры изображений составляли от 312 до 512 пикселей по одной из координат. В эксперименте каждое исходное изображение рассматривалось как изображение высокого разрешения, обозначаемое далее как HR (High Resolution). На его основе формировалась версия с более низким разрешением LR (Low Resolution), после чего выполнялось восстановление. Такой подход позволяет использовать исходное изображение в качестве эталона и непосредственно сравнивать результаты различных методов восстановления. Следует отметить, что экспериментальные изображения использовались исключительно для оценки алгоритмов обработки изображений. Медицинская интерпретация полученных изображений в исследовании не проводилась.

Схема эксперимента



Таким образом, для одного исходного изображения можно было сравнить четыре варианта: исходное изображение HR; изображение LR; результат

Bicubic; результат Real-ESRGAN. На рисунке 1 представлено визуальное сравнение результатов обработки.

Формирование изображения низкого разрешения

Для моделирования потери пространственной информации исходные изображения уменьшались по разрешению. Полученные изображения LR использовались как входные данные для последующего восстановления. Уменьшение разрешения необходимо для имитации ситуации, при которой часть информации исходного изображения была потеряна. В реальных видеосистемах подобные изменения могут возникать, например, вследствие уменьшения разрешения, сжатия или ограничений устройства записи. В данном эксперименте исходные изображения после уменьшения восстанавливались до исходного размера. Для Real-ESRGAN использовался коэффициент масштабирования ×4.

Бикубическая интерполяция

Бикубическая интерполяция была выбрана в качестве базового метода сравнения. Данный метод является традиционным способом изменения разрешения изображения и использует значения соседних пикселей для вычисления новых значений. Преимущество такого baseline заключается в том, что он позволяет определить, действительно ли использование сложной генеративной модели приводит к улучшению результата относительно стандартного метода масштабирования. При этом Bicubic не создаёт новые семантические структуры изображения. Его задача заключается в интерполяции существующей информации при изменении размера изображения.

Real-ESRGAN модель

В качестве генеративного метода использовалась модель Real-ESRGAN. Она является развитием архитектуры ESRGAN и предназначена для восстановления изображений при наличии сложных деградаций. В отличие от простой интерполяции, GAN-модель использует обученное представление визуальных структур для генерации изображения сверхразрешения [3]. Обобщённо процесс можно представить следующим образом (1):

$$I_{LR} \xrightarrow{\text{предобработка}} I_{LR}^{*} \xrightarrow{G_\theta(\cdot)} I_{SR} \xrightarrow{\text{постобработка}} I_{SR}^{*} \quad (1)$$

где (I_{LR}) — изображение низкого разрешения, G— генератор модели Real-ESRGAN, а (I_{SR})— восстановленное изображение сверхразрешения. В отличие от методов, оптимизируемых только по пиксельной ошибке, GAN-подход позволяет ориентироваться также на визуальные характеристики изображения. Это является одной из причин, по которой GAN-модели могут формировать более резкие и текстурированные результаты. Однако такая

особенность одновременно создаёт риск появления деталей, которые не совпадают с исходным изображением. Подобная проблема обсуждается в работах SRGAN и ESRGAN [1,8].

Оценка качества

Для количественной оценки использовались две метрики: **Peak Signal-to-Noise Ratio (PSNR)** и **Structural Similarity Index Measure (SSIM)**.PSNR основан на среднеквадратичной ошибке между эталонным и восстановленным изображениями. Для изображений с одинаковым диапазоном значений более высокое значение PSNR означает меньшее численное отличие от эталона. SSIM был предложен Wang и соавторами как метрика, ориентированная на структурное сходство изображений. В отличие от простой пиксельной ошибки, SSIM учитывает характеристики яркости, контраста и локальной структуры [4]. В общем виде PSNR может быть записан как (2):

$$PSNR=10log10(\frac{MAXI2}{MSE}) \qquad (2)$$

где$MAXI$— максимальное значение пикселя, аMSE — среднеквадратичная ошибка между эталонным и восстановленным изображениями. SSIM оценивает сходство по нескольким компонентам, связанным с яркостью, контрастом и структурой изображения. Поэтому использование обеих метрик позволяет получить более полное представление о результате [5].

РЕЗУЛЬТАТЫ

В результате проведённого эксперимента были получены значения PSNR и SSIM для восьми изображений. Результаты представлены в таблице 1.

	Снимок	Размер	PSNR Bicubic	PSNR Real-ESRGAN	SSIM Bicubic	SSIM Real-ESRGAN	ΔPSNR (ESRGAN − Bicubic)	ΔSSIM (ESRGAN − Bicubic)
0	xray_01	512x448	29.9064	29.1705	0.7531	0.7400	-0.7359	-0.0131
1	xray_02	472x512	33.0473	32.1950	0.8335	0.8227	-0.8523	-0.0108
2	xray_03	512x312	29.4959	27.7506	0.7731	0.7473	-1.7454	-0.0258
3	xray_04	512x348	30.6489	29.3502	0.7736	0.7615	-1.2987	-0.0120
4	xray_05	512x396	29.7757	28.7506	0.7849	0.7545	-1.0251	-0.0304
5	xray_06	508x460	29.6441	28.5139	0.7578	0.7382	-1.1302	-0.0197
6	xray_07	512x412	29.4930	28.0715	0.7890	0.7753	-1.4215	-0.0137
7	xray_08	512x448	30.2530	30.4167	0.7873	0.7885	0.1637	0.0012
8	СРЕДНЕЕ	-	30.2830	29.2774	0.7815	0.7660	-1.0057	-0.0155

Таблица 1 — Сравнение результатов Bicubic и Real-ESRGAN

Из таблицы видно, что для семи из восьми изображений значение PSNR после обработки Real-ESRGAN оказалось ниже, чем после Bicubic interpolation. Аналогичная ситуация наблюдалась для SSIM.Среднее значение PSNR для Bicubic составило:

$$PSNRBicubic=30{,}2830dB.$$

Для Real-ESRGAN среднее значение составило:

$$PSNRReal-ESRGAN=29{,}2774dB.$$

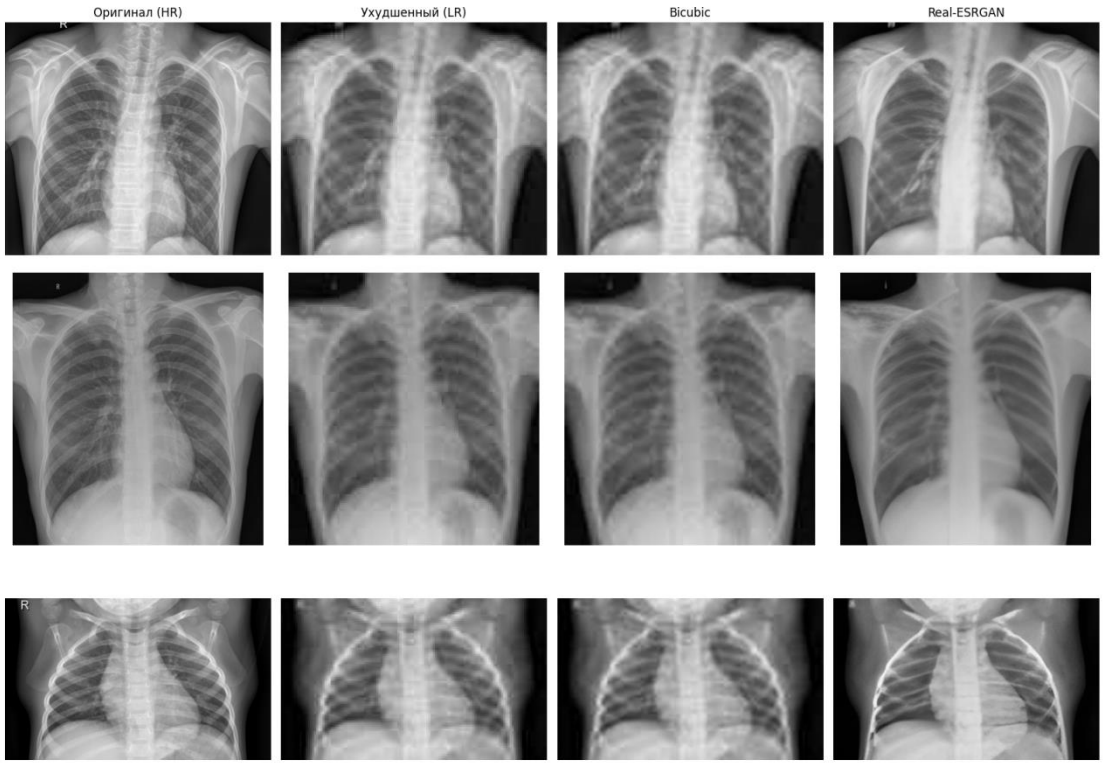


Рисунок 1 — Сравнение исходного изображения, изображения низкого разрешения, Bicubic и Real-ESRGAN.

Разница между методами составила:

$$\Delta PSNR=-1{,}0057dB.$$

Таким образом, по средней величине PSNR результат Real-ESRGAN был примерно на 1,01 дБ ниже результата Bicubic. Для SSIM были получены следующие средние значения:

$$SSIMBicubic=0{,}7815,$$

$$SSIMReal-ESRGAN=0{,}7660.$$

Разность составила:

$$\Delta SSIM=-0{,}0155.$$

Следовательно, и по структурному сходству средний результат Real-ESRGAN оказался ниже результата базового метода. Однако результаты отдельных изображений не были одинаковыми. Наиболее заметным

исключением стало изображение Xray_08. Для него Real-ESRGAN показал более высокие значения обеих метрик: PSNR — 30,4167 дБ против 30,2530 дБ; SSIM — 0,7885 против 0,7873. Это показывает, что результат генеративной модели зависит от характеристик конкретного входного изображения. Наибольшая разница в PSNR наблюдалась для Xray_03. Для него значение Bicubic составило 29,4959 дБ, а Real-ESRGAN — 27,7506 дБ. Разница составила −1,7454 дБ. Для Xray_05 наблюдалось наиболее выраженное различие SSIM: 0,7849 для Bicubic против 0,7545 для Real-ESRGAN. Таким образом, эксперимент не показал общего преимущества Real-ESRGAN по используемым количественным метрикам. Однако это не означает, что генеративная модель не выполняет восстановление изображения. Полученные значения необходимо рассматривать совместно с визуальным результатом. Визуальный анализ показал различия между результатами двух методов. После уменьшения разрешения исходные изображения становились заметно более размытыми, особенно в областях с большим количеством мелких структур. Bicubic позволял восстановить исходный размер изображения, однако результат сохранял часть сглаженности, характерной для интерполяции. Real-ESRGAN формировал более выраженные границы и локальные детали. Визуально некоторые области выглядели более резкими по сравнению с результатом Bicubic. Однако визуальное увеличение резкости не обязательно означает более точное восстановление исходного изображения. Если генеративная модель создаёт структуру, которая отличается от исходной, PSNR и SSIM могут снижаться, несмотря на субъективное ощущение большей детализации. Именно поэтому результаты данного эксперимента представляют интерес. Получена ситуация, при которой более сложная генеративная модель не показала более высоких значений традиционных метрик относительно простого основание [1,2,5].

ОБСУЖДЕНИЕ

Полученные результаты показали, что в проведённом эксперименте Real-ESRGAN не превзошёл Bicubic по средним значениям PSNR и SSIM. Средний PSNR составил 30,2830 дБ для Bicubic и 29,2774 дБ для Real-ESRGAN, а SSIM — 0,7815 и 0,7660 соответственно. Одной из возможных причин является различие между пиксельной точностью и визуальной детализацией. GAN-модели ориентированы не только на минимизацию ошибки, но и на восстановление визуально правдоподобных деталей. Поэтому изображение после Real-ESRGAN может выглядеть более резким, хотя его численное сходство с исходным изображением оказывается ниже. Подобная особенность отмечалась и в исследованиях SRGAN и ESRGAN . Следует

учитывать, что эксперимент проводился только на восьми изображениях. Кроме того, Real-ESRGAN является предварительно обученной моделью, а характер деградации экспериментальных изображений может отличаться от тех данных, на которых она обучалась . Поэтому полученные результаты нельзя рассматривать как общее сравнение эффективности двух методов. Для медицинских изображений также важна осторожность при интерпретации результатов. Генеративная модель может создавать визуально правдоподобные детали, которые отсутствовали в исходном изображении. Поэтому повышение резкости не следует автоматически считать повышением диагностической ценности. В данной работе оценивалось только качество восстановления изображения по техническим метрикам. С точки зрения исходной задачи обработки видео данный эксперимент представляет собой исследование восстановления отдельного кадра. В дальнейшем исследование можно расширить до последовательности кадров и дополнительно оценить временную согласованность, поскольку независимая обработка кадров может приводить к мерцанию и изменению деталей между соседними кадрами. Для этой задачи могут использоваться специализированные методы video super-resolution, например TecoGAN .

ЗАКЛЮЧЕНИЕ

В работе было исследовано применение Real-ESRGAN для восстановления изображений как одного из этапов улучшения качества видеоданных. Эксперимент проведён на восьми рентгенографических изображениях с использованием Bicubic interpolation в качестве базового метода. Средние значения PSNR составили **30,2830 дБ для Bicubic** и **29,2774 дБ для Real-ESRGAN**, а SSIM — **0,7815 и 0,7660** соответственно. Таким образом, в данной экспериментальной выборке Bicubic показал более высокие значения обеих количественных метрик. При этом Real-ESRGAN изменял характер визуального восстановления и позволял получать более выраженные локальные детали. Полученный результат показывает, что количественное и визуальное качество изображения не всегда совпадают, поэтому при исследовании GAN целесообразно использовать несколько критериев оценки. В дальнейшем работу можно расширить за счёт увеличения объёма данных, использования дополнительных метрик и перехода от отдельных изображений к видеопоследовательностям с оценкой временной согласованности кадров.

FOYDALANILGAN ADABIYOTLAR

[1] Ledig C., Theis L., Huszár F., Caballero J., Cunningham A., Acosta A., Aitken A., Tejani A., Totz J., Wang Z., Shi W. **Photo-Realistic Single Image**

Super-Resolution Using a Generative Adversarial Network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. DOI: 10.1109/CVPR.2017.19.

[2] Wang X., Yu K., Wu S., Gu J., Liu Y., Dong C., Qiao Y., Loy C. C. **ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks.** In: *Computer Vision – ECCV 2018 Workshops*. Lecture Notes in Computer Science, vol. 11133, pp. 63–79, 2019. DOI: 10.1007/978-3-030-11021-5_5.

[3] Wang X., Xie L., Dong C., Shan Y. **Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data.** *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*,pp.1905–1914,2021. DOI: 10.1109/ICCVW54120.2021.00217.

[4] Wang Z., Bovik A. C., Sheikh H. R., Simoncelli E. P. **Image Quality Assessment: From Error Visibility to Structural Similarity.** *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004. DOI: 10.1109/TIP.2003.819861.

[5] Horé A., Ziou D. **Image Quality Metrics: PSNR vs. SSIM.** *2010 20th International Conference on Pattern Recognition (ICPR)*, pp. 2366–2369, 2010. DOI: 10.1109/ICPR.2010.579.

[6] Chu M., Xie Y., Mayer J., Leal-Taixé L., Thuerey N. **Learning Temporal Coherence via Self-Supervision for GAN-based Video Generation.** *ACM Transactions on Graphics*, 2020. DOI: 10.1145/3386569.3392457.

[7] Dusanov X. T. Sun’iy intellekt algoritmlari asosida tasvirlar sifatini oshirish // Management and Economics Scientific Research Journal. — 2026. — Vol. 3, No. 2.—P.49–55. — URL: https://journals.timeedu.uz/index.php/mesr/article/view/112

[8] Qosimjonov, S. (2025). **GAN yordamida sun’iy rasmlar generatsiya qilish**. *Research and Implementation*, Spec 3(3), 183–186. **https://rai-journal.uz/index.php/rai/article/view/2222**

[9] Ismatova, Z. A., & Tastanova, S. A. (2025). **Обработка и улучшение качества видео с помощью генеративных моделей (GAN)**. *International Journal of Science and Technology*, 2(16), 10–12. DOI: https://doi.org/10.70728/tech.v02.i16.003