

14. *Chastikova V.A., Mitugov A.I.* The method for detecting network attacks based on the neuroimmune approach, *Journal of Physics: Conference Series. Krasnoyarsk Science and Technology City Hall of the Russian Union of Scientific and Engineering Associations. Krasnoyarsk, Russia*, 2021, pp. 32035.
15. KDD Cup 1999 Data. Available at: <https://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html> (accessed 10 April 2025).
16. NSL-KDD. Available at: <https://www.kaggle.com/datasets/hassan06/nslkdd> (accessed 10 April 2025).
17. CSE-CIC-IDS2018. Available at: https://fkie-cad.github.io/COMIDDS/content/datasets/cse_cic_ids2018/ (accessed 10 April 2025).
18. UNSW-NB15 Network Intrusion Detection Dataset. // Fraunhofer FKIE. Available at: https://fkie-cad.github.io/COMIDDS/content/datasets/unsw_nb15/ (accessed 10 April 2025).
19. *Kingma D.P., Ba J. Adam.* A Method for Stochastic Optimization, *The 3rd International Conference for Learning Representations*. San Diego, 2015, pp. 1-15.
20. *Fawcett T.* An Introduction to ROC Analysis, *Pattern Recognition Letters*, 2006, Vol. 27, No. 8, pp. 861-874. DOI: 10.1016/j.patrec.2005.10.010.

Частикова Вера Аркадьевна – Кубанский государственный технологический университет; e-mail: chastikova_va@mail.ru; г. Краснодар, Россия; тел.: +79184635536; к.т.н.; доцент.

Козачёк Константин Валериевич – Кубанский государственный технологический университет; e-mail: Koza4ek.Konstantin@yandex.ru; г. Краснодар, Россия; тел.: +79182345367; аспирант.

Коробская Екатерина Сергеевна – Кубанский государственный технологический университет; e-mail: kate9.korobskaya@mail.ru; г. Краснодар, Россия; тел.: +79286059807; студент.

Кравцов Владислав Павлович – Кубанский государственный технологический университет; e-mail: vlad.kravtsov.1980@mail.ru; г. Краснодар, Россия; тел.: +79121585302; студент.

Chastikova Vera Arkadyevna – Kuban State Technological University; e-mail: chastikova_va@mail.ru; Krasnodar, Russia; phone: +79184635536; cand. of eng. sc.; associate professor.

Kozachek Konstantin Valerievich – Kuban State Technological University; e-mail: Koza4ek.Konstantin@yandex.ru; Krasnodar, Russia; phone: +79182345367; postgraduate student.

Korobskaya Ekaterina Sergeevna – Kuban State Technological University; e-mail: kate9.korobskaya@mail.ru; Krasnodar, Russia; phone: +79286059807; student.

Kravtsov Vladislav Pavlovich – Kuban State Technological University; e-mail: vlad.kravtsov.1980@mail.ru; Krasnodar, Russia; phone: +79121585302; student.

УДК 004.032.26

DOI 10.18522/2311-3103-2025-5-243-254

А.С. Коваленко, Я.М. Демяненко

МЕТОД ГЕНЕРАЦИИ ШУМА ПО НАБОРУ ЗАШУМЛЕННЫХ ИЗОБРАЖЕНИЙ БЕЗ ЧИСТЫХ ПРИМЕРОВ

Предлагается новый метод генерации шума по зашумленным изображениям без необходимости использования выровненных пар чистых и зашумленных данных. В отличие от традиционных подходов, требующих наличия согласованных наборов изображений или априорных моделей шума, разрабатываемый метод позволяет моделировать сложные характеристики шума, присущие конкретным КМОП-сенсорам, основываясь исключительно на наблюдаемых зашумленных данных. Для синтеза шума используется генеративно-сопоставительная архитектура U-Net-подобного типа, построенная на базе StyleGANv2 с модифицированным дискриминатором, учитывающим параметры камеры и исходных изображений. Основное внимание уделяется сохранению пространственно-цветовой структуры изображения при генерации шума, что достигается введением специализированной функции потерь, сохраняющей характеристики цветопередачи и текстурных деталей. Предлагаемый подход позволяет обучать генератор шума в условиях полного отсутствия пар чистых и зашумленных изображений, что особенно актуально при работе с реальными данными, полученными с различных камер и в различных условиях освещения. В экспе-

риментальной части проведен подробный сравнительный анализ качества синтезированных изображений по метрикам PSNR и SSIM, а также оценка распределения шума на основе статистических характеристик интенсивности и спектрального состава. Демонстрируется, что синтезированный набор изображений, созданный предложенным методом, может эффективно использоваться как самостоятельный тренировочный корпус для нейросетей подавления шума, а также в комбинации с реальным набором SIDD для повышения точности моделей подавления шума. Результаты показали, что комбинированное обучение на объединенном множестве сгенерированных и реальных примеров обеспечивает рост среднего PSNR на 1.5 дБ по сравнению с существующими методами, основанными на выровненных данных. При этом отсутствует зависимость от специфических оптических характеристик конкретного сенсора камеры, что существенно расширяет область применения разработанного метода. Полученные результаты подтверждают применимость предложенного подхода в задачах синтеза и подавления реалистичного шума в условиях отсутствия чистых эталонных изображений, а также открывают перспективы для дальнейших исследований в направлении адаптивной генерации шумовых моделей.

Нейронные сети; генерация данных; фильтрация шума; синтез шума; состязательное обучение; обработка цифровых изображений.

A.S. Kovalenko, Ya.M. Demyanenko

NOISE GENERATION METHOD BASED ON A SET OF NOISY IMAGES WITHOUT CLEAN EXAMPLES

In this work, a novel method is proposed for noise generation from noisy images that does not require aligned pairs of clean and noisy data. Unlike traditional approaches demanding matched image sets or a priori noise models, the developed technique models complex noise characteristics intrinsic to specific CMOS sensors solely from observed noisy data. Noise synthesis is achieved via a U-Net-like generative adversarial architecture based on StyleGANv2, featuring a modified discriminator conditioned on camera parameters and input image metadata. Special emphasis is placed on preserving the spatial-color structure and textural details of each image, enforced through a dedicated loss function that ensures fidelity to the original color rendering and fine-grained patterns. Training of the noise generator is performed without any paired clean and noisy images, which proves particularly valuable when handling real-world datasets acquired from multiple camera models under varied lighting conditions. The experimental section presents a detailed comparative analysis of the synthesized images using PSNR and SSIM metrics, along with an evaluation of the noise distribution based on intensity statistics and spectral characteristics. It is demonstrated that the generated dataset functions effectively as a standalone training corpus for denoising neural networks and, when combined with a real dataset (e.g., SIDD), yields further enhancements in denoising performance. Results indicate that combined training on the union of generated and real examples produces an average PSNR improvement of 1.5 dB compared to existing methods reliant on aligned data. Independence from the specific optical characteristics of any given sensor significantly broadens the method's applicability. These findings confirm the utility of the proposed approach for realistic noise synthesis and removal in scenarios lacking clean reference images, and they open avenues for future research into adaptive noise-model generation.

Neural networks; data generation; noise filtering; noise synthesis; adversarial training; digital image processing.

Введение. Задача подавления шума на цифровых изображениях является одной из самых распространенных в области обработки изображений, поскольку, как большинство методов анализа изображений, так и глубокие нейронные сети имеют чувствительность к наличию шума в обрабатываемых данных [1]. Современные подходы устранения шума на изображениях основаны на применении глубоких сверточных нейронных сетей. В отличие от традиционных подходов, глубокие сети требуют этапа обучения для настройки параметров слоев [2]. Данный этап сильно зависит от используемых наборов данных, состоящих, как правило, из пар изображений – чистое и с шумом. Сбор таких наборов данных для обучения нейронных сетей шумоподавления представляет собой значительную сложность. Процесс получения эталонных «чистых» изображений требует использования высококачественного оборудования и тщательных условий съемки, что ограничивает доступность подобных данных. Генерация зашумленных изображений, реалистично отражающих разнообразие шумов, встречающихся в реальных условиях, требует моде-

лирования различных источников и типов шумов, таких как фотонный шум матрицы и шум квантования при дискретизации аналогового сигнала [3]. При этом синтетически созданные шумы не всегда точно соответствуют характеристикам реальных искажений, что может снижать эффективность обученных моделей. А также масштабирование таких наборов данных для охвата большого числа сценариев и типов изображений требует значительных временных и вычислительных ресурсов, делая процесс их формирования дорогостоящим и трудоемким.

В связи с перечисленными сложностями получения новых обучающих данных перспективным подходом является генерация зашумленных изображений, с распределением шума из заданной выборки. Данная задача может решаться с помощью генеративных состязательных сетей [4]. Они позволяют изучать характеристики шума в наборе данных и синтезировать шум для входных чистых изображений максимально приближенный к оригинальным шумным кадрам.

Мотивация. Методы генерации зашумленных изображений на основе генеративных подходов требуют выровненных наборов данных для обучения, содержащих пары: чистое – зашумленное. Использование простых моделей шума, таких как добавочный гауссов шум или пуассоновский шум, не учитывает всех физических особенностей КМОП-сенсора камеры [5]. В связи с этим, актуальной является задача разработки метода, позволяющего на основе только зашумленных данных строить генеративную модель, позволяющую генерировать изображения со схожим шумом.

Существующие подходы. Подходы, основанные на состязательном обучении используют, как правило, две модели: генератор, обучающийся генерировать добавочный шум или зашумленное изображение, и дискриминатор, обучающийся определять, что изображение сгенерировано генератором. Данный подход может улучшаться добавлением дополнительных моделей в схему обучения. Так в работе [6] добавляется дополнительная предобученная модель для подавления шума на изображении. В данном подходе генератор учится синтезировать такой шум, чтобы при наложении он был не отличим от реального шума ни дискриминатором, ни методом подавления шума.

Подходы на диффузионных моделях предоставляют высокое качество генерации шума без использования дополнительных моделей в схеме обучения, а также при использовании условной генерации позволяют генерировать шум с заданными параметрами сенсора. Такой подход предлагают авторы работы *Realistic Noise Synthesis with Diffusion Models* [7].

Наряду с методами генерации зашумленных примеров существуют подходы, позволяющие переносить шум из изученного распределения на изображение. Это позволяет производить расширение наборов данных для обучения шумоподавляющих сетей. Такой метод предложен авторами работы *NoiseTransfer: Image Noise Generation with Contrastive Embeddings* [8], однако для обучения модели переноса шума все еще необходимы выровненные обучающие данные.

В подходе [9] генерация шума вводит дополнительный стохастический элемент, позволяя модели лучше захватывать распределение реального шума. Данный подход обучается подавлять шум на изображении без использования чистых данных, но генератор шума в подходе не учится генерировать распределения шума из используемого набора данных. Он может использоваться для получения «псевдочистых» изображений на основе зашумленных, что позволит использовать методы генерации шума, требующие наличия выровненных данных.

Для генерации и подавления шума может использоваться подход циклического состязательного обучения (*CycleGAN*) [10]. Так авторы работы [11] применяют данный подход для решения задачи восстановления снимков компьютерной томографии. При этом во время обучения на реальные чистые изображения накладывается синтетический шум. Данный подход имеет существенные ограничения: он требует наличия реальных чистых изображений, а также может привести к переобучению модели на искусственно смоделированном шуме, что снизит качество восстановления реальных зашумленных данных.

Постановка задачи. Модели генерации шума обучаются методами обучения с учителем [7], с частичным привлечением учителя [11] и без учителя [9]. Все перечисленные подходы к обучению требуют наличия выровненного с зашумленными или невыровненного множества чистых кадров. В работе рассматривается задача, когда чистые примеры взяты из наборов кадров, снятых на одной и той же камере, а чистые кадры берутся из независимых наборов, полученных при помощи других камер. Такой подход исключает утечку информации о специфических оптических характеристиках устройства, на котором сняты шумные кадры, и вынуждает модель учиться генерировать и удалять шум, не полагаясь на прямые соответствия между чистыми и зашумленными изображениями.

Добавляемый шум к изображениям должен принадлежать распределению $\Omega(H)$, где H это множество матриц шума из набора зашумленных изображений X :

$$x = y + \eta, x \in X, \eta \in H, y \in Y,$$

причем y – неизвестные матрицы соответствующих чистых изображений для примеров из X .

Дополнительно имеются изображения $b \in B$, которые не содержат шума, и не являются чистыми парами для изображений из X , то есть не принадлежат множеству Y .

Для наложения шума из распределения $\Omega(H)$ на изображение b необходимо построить генеративную модель G , такую, что:

$$\eta = \hat{b} - b, \eta \sim \Omega(H): \hat{b} = G(b).$$

В решение добавляется дополнительная модель для подавления шума на изображении F . Таким образом, решение задачи можно свести к использованию подхода циклического состязательного обучения [10].

Предлагаемый подход. Для обучения моделей в данной работе используется модифицированная функция ошибки из подхода CycleGAN, имеющая вид:

$$\mathcal{L}_{CycleGAN} = \mathcal{L}_{GAN}(G, D_B, B, X) + \mathcal{L}_{GAN}(F, D_X, X, B) + \lambda_{cycle} \mathcal{L}_{cycle}(G, F) + \lambda_{color} \mathcal{L}_{color}(G, F). \quad (1)$$

В классической функции ошибки циклического состязательного обучения добавляется ошибка λ_{color} для корректировки цвета при переходе из доменов. Чистые изображения и зашумленные получены из разных камер, имеющие разные настройки цветопередачи, при обучении с помощью дискриминаторов может происходить смещение цветового домена от одного распределения к другому. Цветовая функция ошибки вычисляется между результатами работы моделей G , F и входными в них данными:

$$\mathcal{L}_{color}(G, F) = \mathcal{L}_{kde}^{b \sim B}(G(b), b) + \mathcal{L}_{kde}^{x \sim X}(F(x), x),$$

Цвета сравниваются на основе оценки плотности ядра (Kernel Density Estimation, KDE). Эта функция потерь позволяет измерять расхождение в цветовых распределениях предсказанного и целевого изображений, игнорируя пространственные особенности. Также такая оценка изображений является устойчивой к шуму [12]. \mathcal{L}_{kde} определяется как:

$$\mathcal{L}_{kde} = \frac{1}{BS \cdot C \cdot K} \sum_{b=1}^{BS} \sum_{c=1}^C \sum_{k=1}^K (\hat{f}_{I^1}(k, c, b) - \hat{f}_{I^2}(k, c, b))^2, \quad (2)$$

где

$$\hat{f}_{I^1}(k, c, b) = \frac{1}{H \cdot W} \sum_{i=1}^H \sum_{j=1}^W \frac{1}{\sqrt{2\pi}h} \exp\left(-\frac{(I_{b,c,i,j}^1 - \xi_k)^2}{2h^2}\right),$$

$$\hat{f}_{I^2}(k, c, b) = \frac{1}{H \cdot W} \sum_{i=1}^H \sum_{j=1}^W \frac{1}{\sqrt{2\pi}h} \exp\left(-\frac{(I_{b,c,i,j}^2 - \xi_k)^2}{2h^2}\right),$$

I^1 – предсказанное изображение, $I^1 \in \mathbb{R}^{BS \times C \times H \times W}$, I^2 – целевое изображение, $I^2 \in \mathbb{R}^{BS \times C \times H \times W}$, BS – размер пакета обучения, C – число каналов в изображении, H, W – высота и ширина изображения, K – число интервалов KDE, ξ_k – центральная точка k -го интервала, равномерно распределенная в $[0,1]$, h – ширина ядра KDE.

В циклической потере необходимо учитывать характеристики чистой компоненты предсказываемого зашумленного изображения x . Чтобы признаки изображения x учитывались к циклической функции ошибки добавляется дополнительная перцептивная функция потерь [13], обозначаемая P . Предлагаемый вариант функции циклической потери будет иметь вид:

$$\mathcal{L}_{cycle}(G, F) = \mathbb{E}_{b \sim B} [\|F(G(b)) - b\|_{L_1}] + \mathbb{E}_{x \sim X} [\|P(G(F(x))) - P(x)\|_{L_1}].$$

Функция ошибки \mathcal{L}_{GAN} , вычисляемая с помощью дискриминаторов, без изменений взята из оригинального подхода [10].

Общие схемы обучения моделей с помощью циклической соревновательной функции ошибки представлены на рис. 1.

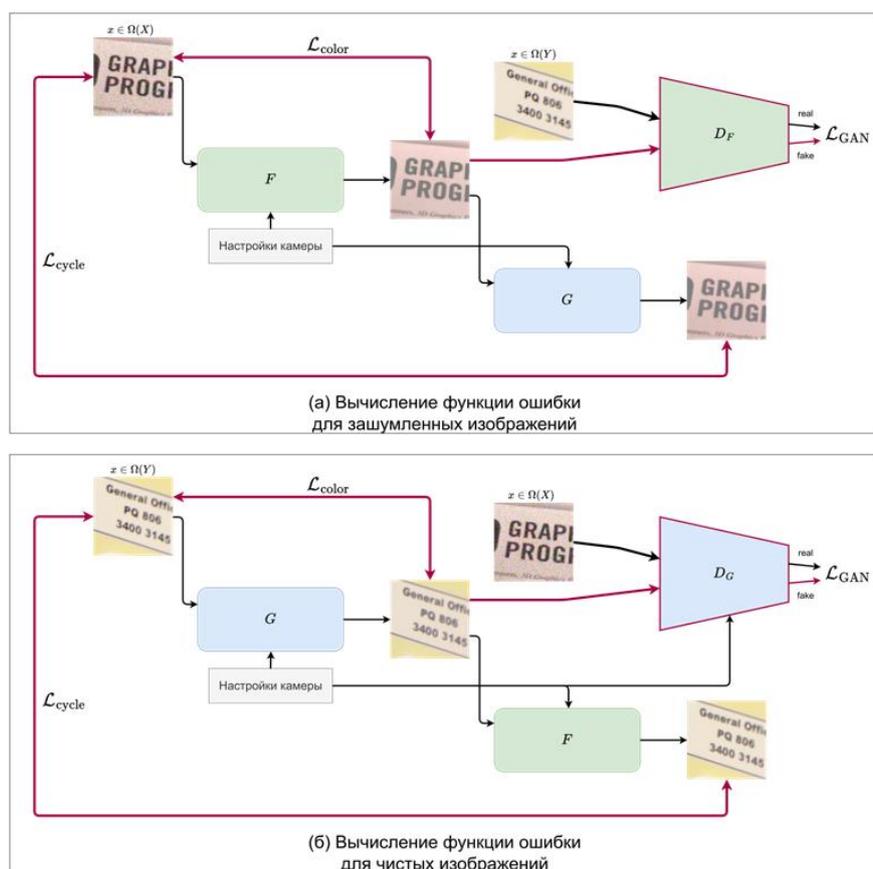


Рис. 1. Вычисление предлагаемой функции ошибки: (а) – для зашумленных изображений, (б) – для чистых изображений

Построение U-Net с механизмом адаптивной нормализации. Использование сверточных слоев с механизмом адаптивной нормализации (AdaIN) и многоуровневой генерации изображений в разных масштабах позволяет архитектуре генератора из подхода StyleGANv2 интегрировать стилевые параметры в каждый слой модели. Применение AdaIN обеспечивает генерацию изображений с разными масштабами и характеристиками, сохраняя при этом структурную согласованность.

В контексте генерации шума с помощью U-Net-подобной архитектуры AdaIN может позволить управлять, как высокоуровневыми характеристиками шума (общая текстура), так и локальными деталями (шумовые паттерны в конкретных областях).

В оригинальном подходе StyleGANv2 вектора стилей генерируются из случайного вектора z , имеющего многомерное нормальное распределение. В предлагаемой архитектуре стилиевые параметры предсказываются на основе параметров, полученных кодировщиком входного изображения. Кодировщик изображения строится на сверточных слоях с алгоритмом понижения размера признаков из дискриминатора StyleGANv2. Ключевым шагом алгоритма является применение фильтра низких частот к входным признакам с последующим применением сверточного слоя для устранения искажений после сжатия. Данный шаг позволит эффективно извлекать сильные признаки из входных зашумленных изображений.

Дополнительно ко всем масштабам признаков в декодирующей части предлагаемой архитектуры добавляются сквозные признаки кодировщика, как в оригинальных U-Net-подобных моделях. В предлагаемом подходе обучения, по аналогии с работой [7] используется кодирование параметров камеры для их передачи в диффузионную модель. С помощью полносвязных слоев кодируются следующие параметры, предоставляемые в наборе данных [14]: уровень чувствительности (ISO), скорость затвора, название камеры, цветовая температура, уровень яркости светового прибора. Кодировка данных параметров также передается и в модель дискриминатора во время обучения. Это позволит определять дискриминатору характеристику шума, которую генерирует модель $SG\mathcal{G}$. Передача дополнительной информации о генерируемом изображении повышает стабильность обучения [15]. Поэтому в предлагаемую модель добавлен блок предсказаний характеристик шума на основе кодировки настроек камеры. Данный блок добавлен в механизм генерации векторов стиля.

Общая схема разработанной архитектуры приведена на рис. 2.

Наборы данных. Информация о настройках камер во время съемки и соответствующие зашумленные изображения брались из открытого набора данных Smartphone Image Denoising Dataset (SIDDD) [14]. Набор SIDDD предоставляет реальные зашумленные изображения и соответствующие им чистые изображения. Обучающая часть набора содержит 320 изображений высокого разрешения, а проверочная часть содержит 1280 пар изображений, имеющих размер 256 на 256 точек. Съемка проводилась авторами на 5 мобильных устройств с КМОП-сенсорами.



Рис. 2. Схема построенной U-Net-подобной архитектуры для зашумления и восстановления изображения

Также для обучения использовались изображения из набора DIV2K [16]. Данный набор содержит только чистые изображения и, как правило, используются для обучения моделей, повышающих разрешение изображения. Так как изображения в наборе не содержат шума [17], в данной работе они используются в качестве множества чистых изображений для обучения генеративной модели, накладывающей шум.

Параметры обучения. Код обучения реализован на фреймворке глубокого обучения PyTorch [18].

Из подхода к состязательному обучению StyleGANv2 [19] используются методы регуляризации параметров генераторов и дискриминаторов для стабилизации их обучения.

Модели обучались на случайных срезах из изображений размером 256×256 пикселей.

Для обучения параметров моделей генератора и дискриминатора применялся метод стохастической оптимизации, основанный на адаптивной оценке моментов первого и второго порядка Adam [20] с параметром скорости обучения равным 0.0002.

Параметр ширины ядра h в формуле (2) задан значением 0.01, а количество интервалов KDE задано 256. Коэффициент λ_{cycle} в основной функции ошибки (1) задан как в оригинальной работе CycleGAN [10], значением 10, а коэффициент λ_{color} дополнительной функции потерь имеет значение 10, как и у циклической ошибки.

Эксперименты проводились на вычислительной машине с графическим ускорителем NVidia RTX 4090, процессором Intel i9-10920X и объемом оперативной памяти 128 Гб. При размере входных изображений 256×256 в процессе обучения использовались пакеты размером 2 (batch size).

Метрики для оценки. Для оценки точности работы обученной модели для генерации шума использовались метрики среднего значения расстояния Кульбака-Лейблера и модуль разности значений пикового отношения сигнала к шуму между предсказанным зашумленным изображением и чистым и между реальным зашумленным и чистым.

При оценке эффективности работы обучаемых моделей для подавления шума дополнительно использовалась метрика структурного сходства изображений.

Расстояние Кульбака-Лейблера (Kullback-Leibler Divergence, KLD) является мерой различия между двумя вероятностными распределениями. В контексте задачи генерации шума AKLD используется для сравнения распределения интенсивностей пикселей в предсказанном зашумленном изображении I_{pred} и реальном зашумленном изображении I_{real} . Формально, KLD определяется как:

$$KDL(P||Q) = \sum_x P(x) \cdot \log \frac{P(x)}{Q(x)},$$

где $P(x)$ и $Q(x)$ — вероятностные распределения интенсивностей пикселей в реальном и предсказанном изображениях соответственно.

Использование среднего значения KLD (Average KLD, AKLD) позволяет учесть глобальные различия между распределениями шума на всем изображении. Эта метрика особенно полезна для оценки точности генерации шума, так как она чувствительна к отклонениям в распределении интенсивностей, что важно для создания реалистичного шума [21].

Пиковое отношение сигнал/шум (Peak Signal-to-Noise Ratio, PSNR) является одной из наиболее распространенных метрик для оценки качества изображений. Оно измеряет соотношение между максимальной возможной мощностью сигнала и мощностью шума. Для двух изображений I_1 и I_2 размера $M \times N$ PSNR вычисляется по формуле:

$$PSNR(I_1, I_2) = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE(I_1, I_2)} \right),$$

где MAX_I — максимальное значение интенсивности пикселей (например, 255 для 8-битных изображений), а $MSE(I_1, I_2)$ — среднеквадратичная ошибка между изображениями:

$$MSE(I_1, I_2) = \frac{1}{M \cdot N} \sum_{i=1}^M \sum_{j=1}^N (I_1(i, j) - I_2(i, j))^2.$$

В данной работе используется модуль разности значений PSNR между предсказанным зашумленным изображением и чистым ($PSNR_{pred}$), и между реальным зашумленным и чистым ($PSNR_{real}$):

$$PGap = |PSNR_{pred} - PSNR_{real}|.$$

Метрика PGap позволяет оценить, насколько точно модель воспроизводит уровень шума, характерный для реальных данных. Чем меньше значение PGap, тем ближе предсказанный шум к реальному [22].

Метрика структурного сходства изображений (Structural Similarity Index Measure, SSIM) оценивает визуальное сходство между двумя изображениями, учитывая яркость, контраст и структуру. Для двух изображений x и y SSIM определяется как:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)},$$

где μ_x и μ_y – средние значения интенсивностей пикселей, σ_x^2 и σ_y^2 – дисперсии, σ_{xy} – ковариация, а C_1 и C_2 – константы для стабилизации деления.

SSIM особенно полезна при оценке точности восстановления зашумленных изображений, так как она фокусируется на сохранении структурных особенностей изображений, которые могут быть потеряны при агрессивном удалении шума [21].

Результаты обучения. Сравнение эффективности наложения шума производилось на валидационном множестве набора данных SIDD [14]. При генерации шума не использовалась информация о настройках камеры, при которых был произведен снимок.

Для сравнения точности моделирования шума с помощью обученных моделей были взяты современные решения, основанные на разных методах генерации шума: на основе состязательно-генеративном подходе DANet [21], на основе нормализации потоков sRGBFlow [23], корреляции чистого и зашумленного сигналов NeCA [24], и диффузионных моделях RNSD [7]. Результаты приведены в табл. 1.

Таблица 2

Сравнение метрик точности генерации шума

Метод	Тип обучения	AKLD ↓	PGap ↓
sRGB2Flow [23]	С учителем	0,237	6,3
DANet [21]	С учителем	0,212	0,206
NeCA [24]	С учителем	0,156	0,97
RNSD [7]	С учителем	0,117	0,54
Предлагаемый подход	Без учителя	0,421	3,48
Предлагаемый подход без \mathcal{L}_{kde}	Без учителя	1,465	7,76
Предлагаемый подход с ошибкой [25]	Без учителя	1,046	5,97

По сходству распределения генерируемого шума с реальным, предлагаемый подход не превосходит модели, обучаемые с учителем (использующие чистые данные при обучении), но при этом не требует наличия выровненного набора данных для обучения. Метрике PGap обученная модель показывает лучшее моделирование шума в сравнении с подходом sRGBFlow [23].

Для демонстрации эффективности использования предлагаемой функции ошибки \mathcal{L}_{kde} (2) были обучены варианты модели без ее использования и с использованием функции ошибки на основе UV-гистограммы изображения [25]. Использование предлагаемой функции ошибки для сохранения характеристик цветопередачи демонстрирует прирост точности генерации шума по обоим метрикам оценки.

Для оценки эффективности подавления шума с помощью обученной модели были сгенерированы наборы данных на основе чистых примеров из обучающей части набора данных SIDD [14] и 1000 изображений из набора LSDIR [26], используемого для сравнений в работе [7]. Набор, обозначаемый M^c , сгенерирован из чистых изображений набора SIDD и при синтезе зашумленных кадров использовалась информация о настройках камеры. Во втором варианте набора использовались изображения из LSDIR, аналогично подходу [7], настройки камеры брались случайным образом, этот вариант обозначается как M^u . На сгенерированных наборах были обучены следующие архитектуры нейронных сетей: DnCNN [27], U-Net [28], NAFNET [29].

Сравнение точности восстановления зашумленных изображений обученными моделями на валидационной части набора SIDD приведено в табл. 2. Если авторы не предоставили результаты по определенным моделям, в соответствующей ячейке таблицы ставится прочерк.

Таблица 3

Сравнение точности восстановления зашумленных изображений из набора SIDD при использовании различных способов получения обучающей выборки

Обучающий набора данных	DnCNN		U-Net		NAFNET	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
SIDD	37,73	0,941	37,92	0,944	39,96	0,960
sRGB2Flow	34,74	-	-	-	-	-
SIDD + sRGB2Flow	37,79	0,950	-	-	39,39	0,957
NeCA	37,65	-	-	-	-	-
RNSD	38,11	-	-	-	-	-
SIDD + RNSD	38,27	0,952	-	-	39,56	0,958
SIDD + DANet	38,05	0,951	-	-	39,05	0,956
M^c	34,54	0,845	33,87	0,811	33,21	0,849
SIDD + M^u	38,81	0,897	38,71	0,900	39,62	0,918

Предлагаемый подход генерации зашумленных изображений позволяет создавать обучающие выборки для обучения шумоподавляющих нейронных сетей без наличия чистых кадров в изначальном наборе. При использовании только набора M^c , состоящего из сгенерированных изображений, точность восстановления зашумленных изображений ниже, чем у подходов, использующих выровненные данные. Обученная модель генератора шумных примеров позволяет добавлять в обучающий набор больше информации о распределении шума, за счет новых примеров. За счет этого результаты комбинации сгенерированного набора с набором SIDD, превосходят рассматриваемые в сравнении подходы.

Для качественного анализа результатов генерации и оценки визуальной достоверности синтезированных изображений были проведены дополнительные эксперименты по сравнению предсказанных шумных кадров с реальными примерами изображений. На рис. 3 представлены примеры зашумленных изображений, сгенерированных моделью на основе чистых входных данных, и их сопоставление с реальными шумными кадрами из набора SIDD [14]. Визуально можно отметить, что предложенная модель успешно воспроизводит пространственно-частотные свойства шума.

Для иллюстрации разнообразия синтезируемого шума и демонстрации способности модели формировать реалистичные шумовые структуры при различных параметрах съёмки дополнительно построен набор синтетических примеров, представленных на рис. 4. В данных примерах для каждого исходного чистого изображения показаны результаты зашумления, полученные при варьировании параметров чувствительности ISO, выдержки и условий освещения, используемых в качестве входных условных признаков модели. Можно отметить, что генератор способен корректно адаптировать интенсивность и структуру шума в зависимости от изменения параметров сенсора, сохраняя при этом визуальную согласованность и отсутствие артефактов в текстурных областях.

Таким образом, представленные визуальные результаты подтверждают, что разработанная архитектура корректно моделирует характеристики шума сенсоров без необходимости привлечения выровненных данных и обеспечивает визуально реалистичное наложение шумовых компонентов, согласованных с физическими параметрами съёмки.

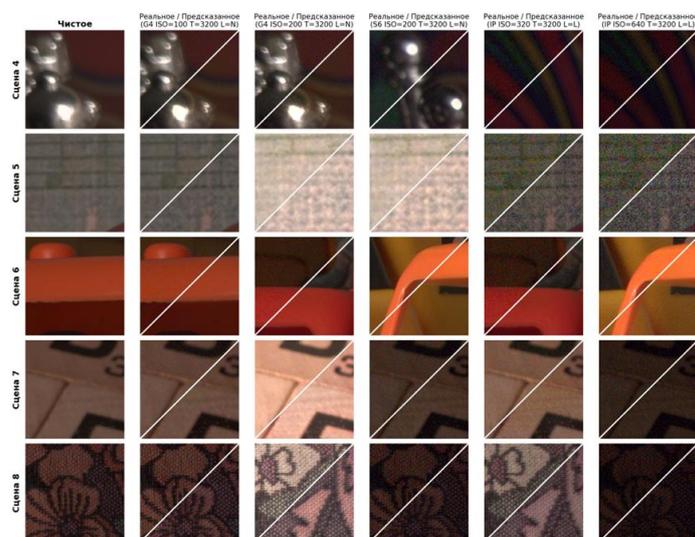


Рис. 3. Сравнение реальных и синтезированных зашумлённых изображений на примере нескольких сцен из набора SIDD. Визуализация представлена по диагонали: в верхней левой части каждого фрагмента показан реальный шум, в нижней правой – синтезированный моделью шум при тех же параметрах съёмки. Первая колонка содержит чистое изображение, остальные – пары реальный / предсказанный для различных настроек камеры. Над каждым столбцом указаны параметры: обозначение устройства, ISO – светочувствительность сенсора, T – цветовая температура источника света, L – уровень освещения (L – низкий, N – нормальный, H – высокий)

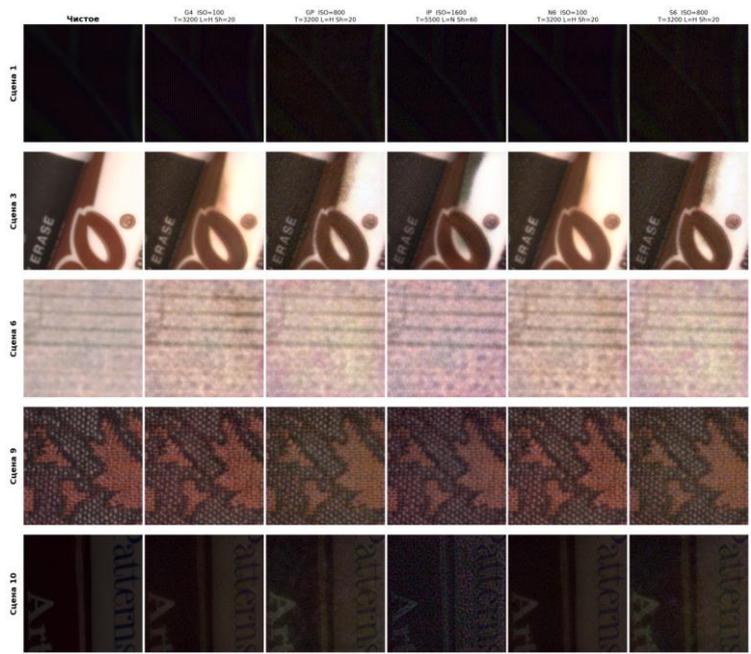


Рис. 4. Примеры синтетических изображений, полученных с помощью обученной модели при различных параметрах съёмки. Над каждым столбцом указаны параметры: обозначение устройства, ISO – светочувствительность сенсора, T – цветовая температура источника света, L – уровень освещения (L – низкий, N – нормальный, H – высокий), Sh – выдержка, выраженная в секундах как 1/Sh

Реализация предлагаемой модели и код обучения приведен в следующем репозитории: https://gitfllic.ru/project/alexeykov/unsupervised_noise_generation.

Заключение. В работе рассмотрен метод синтеза шума с помощью U-Net-подобной архитектурой, построенной с применением механизмов из подхода StyleGANv2. Данная модель обучалась генеративно-сопоставительным подходом с применением функции ошибки, сохраняющей цветопередачу при генерации. Для поддержки изображения и настроек камеры в качестве условных параметров были модифицирована архитектура дискриминатора StyleGANv2. В отличие от других подходов это позволило обучать модель генерации шума без наличия выровненного набора данных.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Al Mudhafar R.A., El Abbadi N.K. Noise in Digital Image Processing: A Review Study, 2022 *3rd Information Technology To Enhance e-learning and Other Application (IT-ELA)*, 2022, pp. 79-84. DOI: 10.1109/IT-ELA57378.2022.10107965.
2. Srujana P., et al. Comparison of Image Denoising using Convolutional Neural Network (CNN) with Traditional Method, 2021 *5th International Conference on Computing Methodologies and Communication (ICCMC)*, 2021, pp. 826-831. DOI: 10.1109/ICCMC51019.2021.9418244.
3. Brouk I., Nemirovsky A., Nemirovsky Y. Analysis of noise in CMOS image sensor, 2008 *IEEE International Conference on Microwaves, Communications, Antennas and Electronic Systems*, 2008, pp. 1-8. DOI: 10.1109/COMCAS.2008.4562800.
4. Bernardo Henz, Eduardo S.L., Gastal M.M.O. Synthesizing Camera Noise using Generative Adversarial Networks, *IEEE Trans. Vis. Comput. Graph.*, 2021, Vol. 27, No. 3, pp. 2123-2135. DOI: 10.1109/TVCG.2020.3012120.
5. Hasino S.W., Durand F., Freeman W.T. Noise-optimal capture for high dynamic range photography, 2010 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 553-560. Available at: <https://api.semanticscholar.org/CorpusID:7762067>.
6. Zhang F., et al. Towards General Low-Light Raw Noise Synthesis and Modeling, *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 10820-10830.
7. Wu Q., et al. Realistic Noise Synthesis with Diffusion Models, *arXiv preprint*, 2023. arXiv:2305.14022 [cs.CV]. Available at: <https://arxiv.org/abs/2305.14022>.
8. Lee S., Kim T. H. NoiseTransfer: Image Noise Generation with Contrastive Embeddings, *Proc. Asian Conf. Comput. Vis. (ACCV)*, Dec. 2022, pp. 3569-3585.
9. Lin X., et al. Unsupervised Image Denoising in Real-World Scenarios via Self-Collaboration Parallel Generative Adversarial Branches, 2023 *IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2023, pp. 12608-12618. DOI: 10.1109/ICCV51070.2023.01162.
10. Zhu J.-Y., et al. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks, 2017 *IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 2242-2251. DOI: 10.1109/ICCV.2017.244.
11. Kwon T., Ye J.C. Cycle-Free CycleGAN Using Invertible Generator for Unsupervised Low-Dose CT Denoising, *IEEE Trans. Comput. Imaging*, 2021, Vol. 7, pp. 1354-1368. DOI: <https://doi.org/10.1109/TCI.2021.3129369>.
12. Gevers T., Stokman H. Robust Histogram Construction from Color Invariants for Object Recognition, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2004, Vol. 26, No. 1, pp. 113-117. DOI: 10.1109/TPAMI.2004.1261083. Available at: <https://doi.org/10.1109/TPAMI.2004.1261083>.
13. Zhang R., et al. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric, *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018.
14. Abdelhamed A., Lin S., Brown M.S. A High-Quality Denoising Dataset for Smartphone Cameras, *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018.
15. Zhang Q., et al. Conditional Adversarial Domain Generalization With a Single Discriminator for Bearing Fault Diagnosis, *IEEE Trans. Instrum. Meas.*, 2021, Vol. 70, pp. 1-15. DOI: 10.1109/TIM.2021.3071350.
16. Agustsson E., Timofte R. NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study, *Proc. CVPR Workshops*, Jul. 2017.
17. Huang J.-B., Singh A., Ahuja N. Single Image Super-Resolution From Transformed Self-Exemplars, *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 5197-5206.
18. Paszke A., et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library, *Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2019, Vol. 32, pp. 8024-8035. Available at: <http://papers.neurips.cc/paper/9015>.
19. Karras T., et al. Analyzing and Improving the Image Quality of StyleGAN, *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 8107-8116. DOI: 10.1109/CVPR42600.2020.00813.

20. Kingma D.P., Ba J. Adam: A Method for Stochastic Optimization, *Int. Conf. Learn. Represent. (ICLR)*, May 2015. Available at: <http://arxiv.org/abs/1412.6980>.
21. Yue Z., et al. Dual Adversarial Network: Toward Real-World Noise Removal and Noise Generation, In: Vedaldi A., et al. (Eds.) *Computer Vision – ECCV 2020*. Cham: Springer, 2020, pp. 41-58. ISBN: 978-3-030-58607-2.
22. Wang Z., et al. Image Quality Assessment: From Error Visibility to Structural Similarity, *IEEE Trans. Image Process*, 2004, Vol. 13, No. 4, pp. 600-612.
23. Kousha S., et al. Modeling sRGB Camera Noise with Normalizing Flows, *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17442-17450. DOI: 10.1109/CVPR52688.2022.01694.
24. Fu Z., Guo L., Wen B. sRGB Real Noise Synthesizing with Neighboring Correlation-Aware Noise Model, *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2023, pp. 1683-1691. DOI: 10.1109/CVPR52729.2023.00168.
25. Aff M., Brubaker M. A., Brown M.S. HistoGAN: Controlling Colors of GAN-Generated and Real Images via Color Histograms, *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 7937-7946. Available at: <https://api.semanticscholar.org/CorpusID:227151819>.
26. Li Y., et al. LSDIR: A Large Scale Dataset for Image Restoration, *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2023, pp. 1775-1787. DOI: 10.1109/CVPRW59228.2023.00178.
27. Zhang K., et al. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising, *IEEE Trans. Image Process*, 2017, Vol. 26, No. 7, pp. 3142-3155. DOI: 10.1109/TIP.2017.2662206.
28. Komatsu R., Gonsalves T. Comparing U-Net Based Models for Denoising Color Images, *AI*, 2020, Vol. 1, No. 4, pp. 465-486. ISSN: 2673-2688. DOI: 10.3390/ai1040029. Available at: <https://www.mdpi.com/2673-2688/1/4/29>.
29. Chu X., Chen L., Yu W. NAFSSR: Stereo Image Super-Resolution Using NAFNet, *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 1239-1248.

Коваленко Алексей Сергеевич – Южный федеральный университет; e-mail: akov@sfedu.ru; г. Ростов-на-Дону, Россия; тел.: +79281217747; кафедра прикладной математики и программирования; ассистент.

Демяненко Яна Михайловна – Южный федеральный университет; e-mail: demyana@sfedu.ru; г. Ростов-на-Дону, Россия; тел.: +78632975111; кафедра прикладной математики и программирования; к.т.н.; доцент.

Kovalenko Aleksei Sergeevich – Southern Federal University; e-mail: akov@sfedu.ru; Rostov-on-Don, Russia; phone: +79281217747; the Department of Applied Mathematics and Programming; assistant lecturer.

Demyanenko Yana Mikhailovna – Southern Federal University; e-mail: demyana@sfedu.ru; Rostov-on-Don, Russia; phone: +78632975111; the Department of Applied Mathematics and Programming; cand. of eng. sc.; associate professor.

УДК 004.896

DOI 10.18522/2311-3103-2025-5-254-276

О.Б. Лебедев, Р.И. Черкасов

ПРИМЕНЕНИЕ ТЕХНОЛОГИЙ КОМПЬЮТЕРНОГО ЗРЕНИЯ В СИСТЕМАХ ОБРАБОТКИ ВИЗУАЛЬНОЙ ИНФОРМАЦИИ

Рассмотрено применение технологий искусственного интеллекта, в частности компьютерного зрения в системах обработки визуальной информации. Проведен комплексный анализ нейросетевых подходов к решению задач компьютерного зрения, включая систематизацию ключевых типов задач: классификацию изображений, детектирование объектов и семантическую сегментацию. Детально исследованы архитектурные принципы сверточных нейронных сетей с акцентом на механизмы извлечения пространственных признаков через сверточные слои, оптимизацию представления данных посредством операций пулинга и преобразование признаков в полностью связанные слои. Особое внимание уделено эволюции методов обнаружения объектов, где задача выбора модели рассмотрена как расширение классификации за счет интеграции регрессии пространственных координат, а также проведена оценка эффективности детекторов на основе метрик IoU, Precision, Recall и F1-score, демонстрирующих фундаментальный компромисс между точностью локализации и скоростью обработки. В качестве оптимального решения для систем реального времени представлен алгоритм YOLOv7, архитектура которого основана на разбиении входного