



УДК 004.896

С. В. Сай, А. А. Кудяшов

ОЦЕНКА КАЧЕСТВА РЕСТАВРАЦИИ ЗАШУМЛЕННЫХ ИЗОБРАЖЕНИЙ НЕЙРОСЕТЕВЫМ АВТОЭНКОДЕРОМ

Сай С. В. – д-р техн. наук, проф., зав. кафедрой «Вычислительная техника», (ТОГУ), e-mail: 000493@pnu.edu.ru; *Кудяшов А. А.* – асп. кафедры «Вычислительная техника», (ТОГУ), e-mail: 2015101665@pnu.edu.ru.

В статье рассматривается применение нейронной сети автоэнкодера в задаче подавления шумов в полутоновых изображениях. Приводится описание архитектуры модели и обучения автоэнкодера на основе сверточной сети. Модель обучена на примере тренировочной выборки из 130 изображений природного ландшафта с разрешением 512×512 пикселей. В результате тестирования автоэнкодера получены усредненные коэффициенты традиционных метрик PSNR и SSIM, а также по метрике MFSD учитывающей пороги зрительного восприятия изменения контраста мелких структур и фоновых участков изображения.

Ключевые слова: нейронные сети, машинное обучение, автоэнкодер, шумоподавление, анализ качества изображений.

Введение

Автоэнкодер – это тип нейронных сетей с обучением без учителя, которые принимают входной набор данных, сжимают (кодируют) эти данные до представления скрытого пространства (закодированного состояния) и реконструируют (декодируют) эти данные на выходе. Данные всегда реконструируются с потерями при этом сохраняя важные признаки изображения [1].

Кодирующая часть автоэнкодера изучает, как шум добавляется к исходному изображению. Шум генерируется по заранее заданной функции генерации $f(X) = Y$, где X – исходное изображение, а Y – зашумленное изображение. Декодирующая часть удаляет шум из изображений и реконструирует аппроксимированное входное изображение X .

В процессе обучения на вход сети подаются зашумлённые изображения и на выходе сравниваются с исходными изображениями на основе среднеквадратичной ошибки (MSE). Обучение происходит до тех пор, пока значение MSE не достигнет минимального значения. После обучения автоэнкодер может использоваться в задачах реставрации зашумленных изображений. Упрощенная модель автоэнкодера показана на рис. 1.

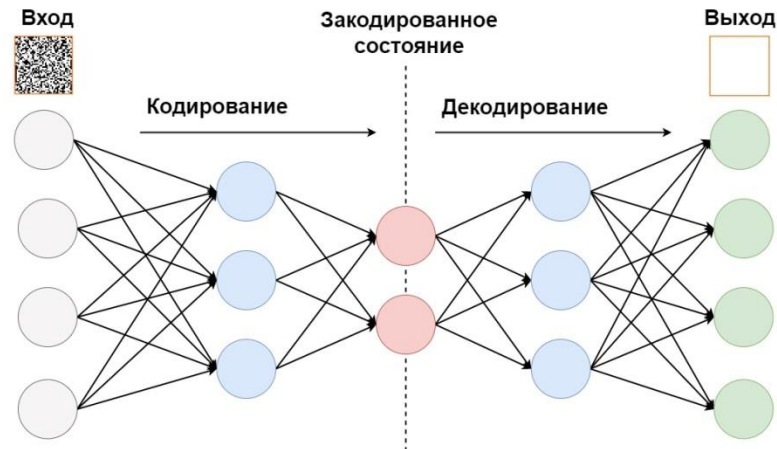


Рис. 1. Упрощенный вид модели автоэнкодера

Добавление шума

Перед началом обучения элементы исходных изображений из обучающего набора суммируются со случайно сгенерированными значениями шума по нормальному распределению вероятностей:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

где $\mu = 0$ и $\sigma = 1$.

Каждый элемент зашумлённого изображения с пространственными координатами (i, j) вычисляется по формуле:

$$\tilde{I}_{i,j} = C\left(\frac{I_{i,j}}{255} + k_m M_{i,j}\right)$$

где $\tilde{I}_{i,j}$ – элемент зашумлённого изображения; $I_{i,j}$ – элемент исходного изображения; k_m – коэффициент шума; $M_{i,j}$ – случайно сгенерированная матрица шума по нормальному закону распределения; $C(x_{i,j})$ – функция ограничения значений $\tilde{I}_{i,j}$, описываемая выражением:

$$C(x_{i,j}) = \begin{cases} 0, & x < 0 \\ 1, & x > 1 \end{cases}$$

Архитектура модели

Как и в случае с другими свёрточными нейронными сетями обучение автоэнкодера сводится к минимизации функции ошибки путем корректировки весовых коэффициентов синаптических связей между нейронами. Таким образом, чем меньше возвращаемое значение функцией потерь, тем точнее предсказания модели. Архитектура модели представлена в табл. 1.

Таблица 1

Архитектура модели		
№ слоя	Слой	Параметры
Кодирование		
1	Input	



	[512, 512, 1] – [Ширина, высота, размерность]	
2	Conv2D (Операция свёртки) [512, 512, 128]	Фильтры: 128 Ядро: 3×3 Активация: ReLU
3	Conv2D (Операция свёртки) [512, 512, 64]	Фильтры: 64 Ядро: 3×3 Активация: ReLU
4	Conv2D (Операция свёртки) [512, 512, 32]	Фильтры: 32 Ядро: 3×3 Активация: ReLU
Декодирование		
5	Conv2DTranspose (Операция обратная свёртке) [512, 512, 32]	Фильтры: 32 Ядро: 3×3 Активация: ReLU
6	Conv2DTranspose (Операция обратная свёртке) [512, 512, 64]	Фильтры: 64 Ядро: 3×3 Активация: ReLU
7	Conv2DTranspose (Операция обратная свёртке) [512, 512, 128]	Фильтры: 128 Ядро: 3×3 Активация: ReLU
8	Output [512, 512, 1]	Фильтры: 1 Ядро: 3×3 Активация: linear

На всех слоях сети, кроме выходного слоя, используется линейная функция активации ReLU (Линейный выпрямитель). Модель обучалась с использованием оптимизатора Adam и адаптивным уменьшением скорости обучения.

Кодирующая часть использует слои свёртки (конволюционные слои) для определения признаков с ядром 3×3, а декодирующая слои с обратной свёрткой (деконволюционные слои) для построения признаков с ядром 3×3.

Три обратных слоя (Conv2DTranspose), выступают в качестве декодера в сети. Они учатся преобразовывать скрытое пространство, являющееся выходом последнего слоя свёртки, в выходное изображение без шума.

На рис. 2 показан пример выполнения прямой (Conv2D) и обратной (Conv2DTranspose) свёртки фрагмента изображения 3×3 с ядром 2×2.

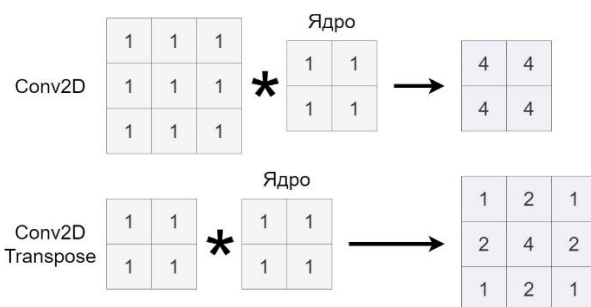


Рис. 2. Выполнение Conv2D и Conv2DTranspose с ядром 2×2

Наборы изображений и параметры сети

В модели используется обучающий и проверочный набор:
– обучающий набор состоит из 130 изображений;

– проверочный из 10 новых изображений, не входящих в обучающий.
Изображения для этих наборов сгенерированы в приложении Nvidia Canvas [2]. На рис. 3. показаны примеры изображений обучающей выборки.



Рис. 3. Примеры тестовых изображений

Зададим начальные параметры сети:

- количество эпох $E = 1000$;
- размер выборки $B = 1$;
- коэффициент шума $k_m = 0,05$.

Модель, построенная с этими параметрами, является основной для получения результатов.

Параметр E указывает на число итераций прямого и обратного прохода по всем предоставленным данным в сети. В случае если данных слишком много для хранения в памяти, их можно разделить на подгруппы – выборки. Размер выборки B определяет часть от общего количества изображений, которая будет использоваться в эпохе в течении одного шага итерации.

На практике было замечено, что использование большой группы данных приводит к ухудшению качества предсказанного моделью изображения, измеряемое ее способностью к обобщению [3] и сходимостью.

При размере выборки равной $B = 1$ обучение нейросети нестабильно, а на графике истории обучения отображаются резкие подъёмы значений функции потерь (Loss). По мере работы сети скорость обучения уменьшается и значения функции потерь сглаживаются, таким образом обучение стабилизируется, при этом сходимость в сети постепенно начинает улучшаться (рис. 4).

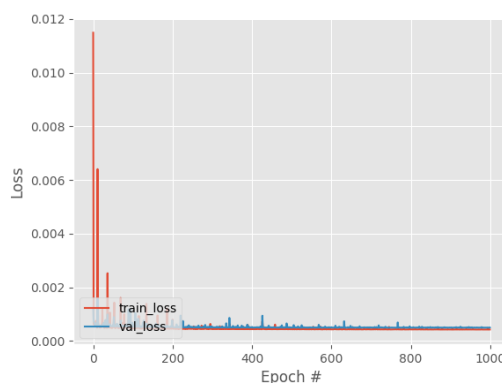


Рис. 4. График значений функций потерь на обучающем и проверочном наборе



Минимальное значение за историю потерь H по проверочному набору $\min(H_v) = 0,000497$, по обучающему $\min(H_t) = 0,000435$. Использование основных параметров и уменьшение размера обучающего набора до 32 приводит к следующим минимальным значениям потерь $\min(H_v) = 0,000366$ и $\min(H_t) = 0,000516$. Для подтверждения вывода о независимости сходимости от размера выборки и обобщения возьмём $B = 13$. В итоге получим $\min(H_v) = 0,000506$ и $\min(H_t) = 0,000445$, что превышает полученные значения для $B = 1$. В графике истории обучения значения потерь относительно стабильно снижаются за исключением эпох с 316 по 322 (рис. 5).

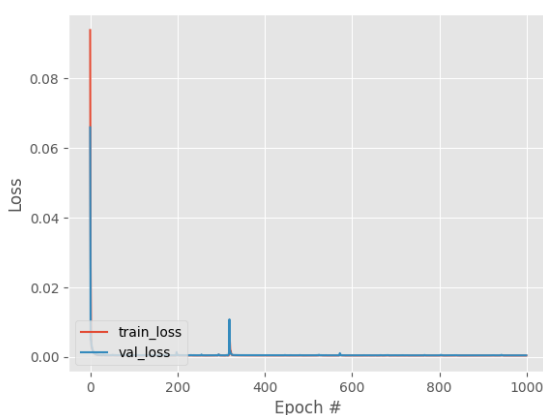


Рис. 5. График функции потерь при $B = 13$

Результаты анализа

Для анализа качества шумоподавления были использованы метрики $PSNR$, $SSIM$ и $MFSD$.

Метрика $PSNR$ позволяет оценивать искажения изображения в децибелах по следующей формуле [4]:

$$PSNR = 20 \log_{10} \left(\frac{255}{\sqrt{MSE}} \right). \quad (1)$$

Средний квадрат ошибки (MSE) изображения с размерами $W \times H$ находится как

$$MSE = \frac{1}{WH} \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} (Y_{i,j} - \tilde{Y}_{i,j})^2,$$

где $Y_{i,j}$ – элементы яркости оригинального изображения и $\tilde{Y}_{i,j}$ – элементы изображения после шумоподавления. На практике значение $PSNR$ сопоставляется с субъективными оценками и, при этом считается, что хорошее качество изображения обеспечивается при $PSNR > 40$ дБ.

Метрика $SSIM$ сводится к определению степени сходства соответствующих участков (окон) сравниваемых изображений. Разница между двумя окнами x и y с размером $N \times N$ определяется как [5]

$$SSIM_{x,y} = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (2)$$

где x, y – окна оригинального и искаженного изображений; μ_x, μ_y – средние значения цветовых координат пикселей в окнах x и y ; σ_x, σ_y – среднеквадратичные отклонения; σ_{xy} – значение ковариации; c_1, c_2 – постоянные коэффициенты.

Как правило, метрика (2) рассчитана на окно с размером 8×8 пикселей. Для анализа всего изображения окно смещается на заданное число пикселей последовательно по горизонтали и по вертикали. Интегральная оценка определяется как

$$SSIM = \frac{1}{L} \sum_{l=1}^L SSIM_l, \quad (3)$$

где l – номер окна; L – количество окон. Чем ближе значение $SSIM$ к 1, тем ближе качество изображения к исходному.

В настоящей работе дополнительно к традиционным метрикам использована альтернативная метрика *MFSD* (Metric of Fine Structures Distortion), описание которой приведено в работе [6]. Способ оценки искажений можно разделить на два этапа. На первом этапе выполняется алгоритм идентификации мелких структур оригинального изображения, который основан на результатах анализа структуры микроблоков с размером 3×3 пикселя, с помощью вычисления контраста между соседними пикселями в нормированной системе N-CIELAB:

$$K_{i,j} = \sqrt{(\Delta \bar{L}_{i,j}^*)^2 + (\Delta \bar{a}_{i,j}^*)^2 + (\Delta \bar{b}_{i,j}^*)^2}, \quad (4)$$

где $\Delta \bar{L}_{i,j}^*$, $\Delta \bar{a}_{i,j}^*$ и $\Delta \bar{b}_{i,j}^*$ разностные, нормированные к порогам контрастной чувствительности зрения, значения цветовых координат по яркости и по цветности. Идентификация блока выполняется по следующему критерию – если количество переходов с цветовым контрастом $K_{i,j} > 1$ превышает единицу, то в микроблоке присутствует мелкая структура различимая глазом. Если данное условие не выполняется, то мелкая структура не различается глазом и такой блок относится к фоновым участкам изображения.

На втором этапе оцениваются искажения реставрированного изображения блока m на основе вычисления максимального отклонения контраста между цветовыми переходами p оригинального $K_{m,p}$ и искаженного $\hat{K}_{m,p}$ изображений:

$$dE_m = \max_p (|K_{m,p} - \hat{K}_{m,p}|).$$

Если выполняется условие $dE_m > 0,5$, принимается решение о том, что отклонения цветового контраста в микроблоке m заметны глазом. Средняя оценка искажений по всем N_m микроблокам изображения с мелкими структурами вычисляется как:

$$dE_N = \frac{1}{N_m} \sum_{m=1}^{N_m} dE_m. \quad (5)$$

Для фоновых участков использована дополнительная оценка:

$$dE_F = \frac{1}{N_F} \sum_{k=1}^{N_F} \left(\max_i \left(\sqrt{(\Delta \bar{L}_{k,i}^*)^2 + (\Delta \bar{a}_{k,i}^*)^2 + (\Delta \bar{b}_{k,i}^*)^2} \right) \right) \quad (6)$$

где отклонение цветовых координат в блоке k вычисляется в обычной системе CIELAB. Экспериментально доказано [6], что высокое качество реставрированных изображений без заметных для глаза искажений обеспечивается при выполнении критерия:



$$dE_N < 0,5 \text{ и } dE_F < 2,5. \quad (7)$$

В табл. 2 приведены результаты анализа качества зашумленных $\tilde{I}_{i,j}$ и реставрированных $\tilde{\tilde{I}}_{i,j}$ изображений из проверочного набора по метрикам $PSNR$, $SSIM$ и $MFSD$ при заданном коэффициенте шума $k_m = 0,05$. Также в таблице приведены средние значения метрик по всем изображениям.

На рис. 6 показан пример предсказания модели по первому зашумленному изображению из проверочного набора.

Анализ результатов показывает, что качество реставрированных изображений невысокое. В частности, средние значения $dE_N \approx 1,18$ и $dE_F \approx 3,86$ практически в два раза превышают пороговые значения согласно критерию высокого качества (7). При этом, искажения заметны глазом. Тем не менее, по отношению к зашумленному изображению качество реставрированных изображений улучшается примерно в два раза.

Таблица 2

Результаты анализа								
№	$PSNR$		$SSIM$		dE_N		dE_F	
	$\tilde{I}_{i,j}$	$\tilde{\tilde{I}}_{i,j}$	$\tilde{I}_{i,j}$	$\tilde{\tilde{I}}_{i,j}$	$\tilde{I}_{i,j}$	$\tilde{\tilde{I}}_{i,j}$	$\tilde{I}_{i,j}$	$\tilde{\tilde{I}}_{i,j}$
1	26,06	33,02	0,56	0,94	1,91	1,16	9,12	4,51
2	26,11	33,29	0,51	0,93	1,94	1,28	11,2	3,03
3	26,06	33,37	0,50	0,92	1,91	1,22	10,17	3,81
4	26,04	32,78	0,52	0,94	1,88	1,20	12,13	3,59
5	26,05	31,83	0,57	0,94	1,89	1,27	12,68	4,61
6	26,05	33,96	0,48	0,91	2,00	1,08	11,28	3,38
7	26,08	31,88	0,58	0,91	1,96	1,27	11,68	4,79
8	26,03	33,36	0,52	0,90	1,92	1,05	11,10	4,19
9	26,03	33,48	0,48	0,93	1,90	1,22	11,09	3,14
10	26,02	33,81	0,50	0,92	1,92	1,11	10,52	3,56
Mean	26,05	33,07	0,52	0,92	1,92	1,18	11,09	3,86

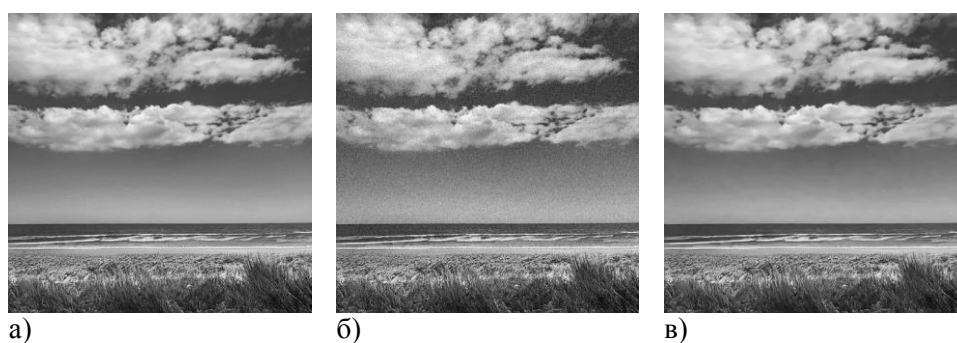


Рис. 6. Примеры изображений:
(а) оригинал; (б) зашумленное; в) реставрированное



Заключение

В результате исследования модели автоэнкодера получено, что большой объём набора изображений, и большая размерность сверточных слоев не дает значительных улучшений значений функции потерь.

Реализованная модель автоэнкодера позволяет удалять шумы из полутонных изображений, однако качество реставрированных изображений невысокое: $PSNR < 40\text{дБ}$; $SSIM < 0,98$; $dE_N > 0,5$ и $dE_F > 2,5$.

К дальнейшим направлениям исследований относится: поиск и разработка новых моделей автоэнкодера для решения задач реставрации искаженных или масштабированных изображений с повышенным разрешением и с высоким качеством.

Благодарности

Исследование выполнено при финансовой поддержке Российского научного фонда в рамках научного проекта № 22-21-00394 «Развитие нейросетевых методов повышения качества передачи цифровых изображений в интеллектуальных видеосистемах».

Библиографические ссылки

1. Building Autoencoders in Keras. URL: <https://blog.keras.io/building-autoencoders-in-keras.html> (дата обращения : 20.12.2022).
2. NVIDIA Canvas. URL: <https://www.nvidia.com/en-us/studio/canvas/> (дата обращения : 21.12.2022).
3. On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima / Keskar N.S, Mudigere D, Nocedal J, Smelyanskiy M., Ping Tang P. // Published as a conference paper at ICLR 2017. P. 1-16.
4. Pratt W.K. Digital Image Processing // John Wiley & Sons, 2001.
5. Image Quality Assessment: From Error Visibility to Structural Similarity / Wang Zhou, Bovik A.C., Sheikh H.R., Simoncelli E.P. // IEEE Transactions on Image Processing. 2004. Vol. 13. P. 600-612.
6. Сай С.В. Метрика искажений мелких структур компрессированных изображений // Компьютерная оптика. 2018. № 5. С. 829-837.

Title: Quality Evaluation of Noisy Images Restoration with a Neural Network Autoencoder

Authors' affiliation:

Sai S. V. – Pacific National University, Khabarovsk, Russian Federation

Kudyashov A. A. – Pacific National University, Khabarovsk, Russian Federation

Abstract: In the article, the authors discuss the use of an autoencoder neural network in the problem of noise suppression in grayscale images. The description of the architecture of the model and training of the autoencoder based on the convolutional network is given. The model has been trained by the example of a training set of 130 natural landscape images with resolution of 512×512 pixels. As a result of testing the autoencoder, the averaged coefficients of the traditional PSNR and SSIM metrics have been obtained, as well as the MFSD metric, which takes into account the thresholds of visual perception of changes in the contrast of small structures and background areas of the image.

Keywords: neural networks, machine learning, autoencoder, noise reduction, image quality analysis.