

УДК 004.932.4

АППАРАТНАЯ НЕЗАВИСИМОСТЬ И ТОЧНОСТЬ НЕЙРОСЕТЕВОГО ШУМОПОДАВЛЕНИЯ НА ИЗОБРАЖЕНИЯХ КАК ФУНКЦИЯ ОБЪЕМА ОБУЧАЮЩИХ ДАННЫХ

© 2021 г. Н. И. Попов^{1,2,*}, А. С. Григорьев²¹ Московский физико-технический институт (государственный университет),
141701 Московская обл., г. Долгопрудный, Институтский переулок, д. 9, Россия² Институт проблем передачи информации имени А.А. Харкевича РАН,
127051 Москва, Большой Каретный переулок, д. 19, Россия

*E-mail: popov.n@phystech.edu

Поступила в редакцию 06.10.2020 г.

После доработки 15.10.2020 г.

Принята к публикации 02.11.2020 г.

Данная работа представляет собой исследование нейросетевого подхода к улучшению изображений (повышению экспозиции и устранению шума), предложенного в работе (Chen et al., 2018). Анализируется применимость нейросети, обученной на датасете с одной фотокамеры, к данным с другой камеры. Исследуется возможность снижения размера датасета для обучения нейросети для новой камеры. Был собран датасет из 27 пар фотографий в сыром формате, сделанных с экспозициями в 0.01 и 1 с и соответствующих друг другу попиксельно. Путем тестирования на наборах фотографий с разных камер, использующих одинаковый тип мозаики цветочных фильтров, сравниваются две модели, обученные на наборах, размеры которых отличаются в 8 раз. Из зависимости метрик качества PSNR и SSIM от размера обучающего набора выяснено, что для размера в 25–30 сцен метрики превосходят 90% от значений, заявленных авторами вышеупомянутой статьи для модели, обученной на 160 сценах. Также эти метрики сравниваются для модели, обученной на фотографиях с одной камеры, и модели, обученной на фотографиях с новой камеры, после их тестирования на наборе из фотографий со второй камеры. Демонстрируется такая переносимость результата, что качество деталей на изображениях для модели, обученной на относительно большом датасете из фотографий с другой камеры, заметно лучше.

Ключевые слова: обработка изображений, нейросетевое шумоподавление, вычислительная фотография

DOI: 10.31857/S0235009221010078

ВВЕДЕНИЕ

Фотосъемка темных сцен, как и фотосъемка с короткой экспозицией, обладают неудовлетворительным качеством из-за наличия шума, неверного отображения цветов и недостаточной яркости изображения. В частности, шумоподавление изображений было в центре внимания многих исследований в области компьютерного зрения и обработки изображений. Классические методы, такие как анизотропная диффузия (Perona et al., 1990) и использование вейвлетов (Simoncelli et al., 1996), преобразуют шумный вход в качественное изображение и работают в предположении, что сигнал и шум подчиняются некоторой статистической закономерности. Несмотря на то что эти параметрические модели просты в использовании, они ограничены в своих возможностях. Это привело к повышенному интересу к методам, ис-

пользующим похожие участки изображения, таким как BM3D (Dabov et al., 2007). Переход от простых аналитических методов к подходам, основанным на извлечении закономерности из данных, продолжался путем использования словарей для обучения и алгоритмов поиска базисов, таких как KSVD (Aharon et al., 2006). Примерно с 2009 г. большинство алгоритмов шумоподавления одиночных изображений полностью основаны на данных и состоят из глубоких нейронных сетей, обученных удалять шум с изображений (Gharbi et al., 2016; Guo et al., 2019; Zhang et al., 2017).

Тогда как в классических работах шум на изображениях представляется пуассоновским, гауссовым (Foi et al., 2008) или по модели Яне (Bernd, 2005), в современных подходах, основанных на данных, шум создается генеративно-состязательной сетью (Kim et al., 2019). Также существуют решения, в которых используются пары изображе-

ний с реальным шумом и без него. Такой датасет может быть получен путем сбора данных, состоящих из пар фотографий, одно из которых снято с короткой экспозицией и поэтому шумное, а другое – с долгой экспозицией и поэтому в значительной степени без шума (Anaya, Barbu, 2014; Plotz, Roth, 2017). В качестве зашумленных изображений также могут быть использованы фотографии, сделанные при слабом освещении (Chen et al., 2018; Hsinoff et al., 2016; Zhang et al., 2019).

В недавних работах (Lehtinen et al., 2018) предлагается использовать несколько зашумленных изображений одной и той же сцены в качестве обучающих данных вместо парных изображений с шумом и без него, но это не значительно облегчает трудоемкость сбора датасета. Иным новаторским подходом является использование нейросетевой архитектуры GAN, которая может быть обучена без пар изображений с низким и высоким уровнями шума, но которая хорошо обобщается на различных реальных тестовых изображениях (Schwartz et al., 2019; Ren et al., 2019).

Современные методы обработки сырых изображений с сенсора фотокамеры состоят из нескольких этапов, преобразующих интенсивность изображения, тем самым осуществляя изменение входного шума и получение конечного изображения из измерений сенсора. Работа (Nam et al., 2016) определяет ограничения традиционной модели шума и в ней рассмотрены модели для удаления шума с sRGB изображений. Такой же подход используется в работе (Wang et al., 2019), где нейросетевая модель вычисляет коэффициенты, умножение которых на значения пикселей исходного шумного изображения по трем каналам приводит к шумоподавлению. Мы же сосредотачиваемся на шумоподавлении сырых (raw) изображений. В данной работе изучаются точность нейросетевого шумоподавления на изображениях в зависимости от объема обучающих данных и аппаратная независимость (применение модели для одной камеры после обучения на другой) по методу работы (Chen et al., 2018).

МЕТОД НЕЙРОСЕТЕВОГО ШУМОПОДАВЛЕНИЯ

В статье (Chen et al., 2018) решалась проблема улучшения фото при фотографировании темных сцен с короткой экспозицией при помощи end-to-end нейросетевой модели с архитектурой U-net. Модель обучается на парах фото в raw формате с короткой и долгой экспозициями. На каждой итерации обучения из произвольно выбранного участка изображения составляется тензор, состоящий из четырех слоев, каждый из которых соответствует одному из компонентов байеровской мозаики. После вычитания уровня черного (темнового тока) и умножения на отношение экспозиций

двух фотографий одной сцены, находящейся в пределах от 100 до 300, полученный тензор подается на вход сети. Результатом работы модели является sRGB изображение, улучшенное относительно входного, с качеством, оцениваемым по средней L1-норме попиксельной разности между ним и фотографией с долгой экспозицией. Путем обучения этой нейросети на собранном датасете из 424 пар фотографий авторы получили модель, которая удаляет шум, повышает яркость фотографии и демонстрирует качество (метрики качества PSNR/SSIM = 28.88/0.787), не сильно проигрывающее последующим работам (Wang et al., 2019; Schwartz et al., 2019; Guo et al., 2019) в данном направлении. Данный подход превосходит классические методы, использующие серии фотографий (Hsinoff et al., 2016; Liu et al., 2014), которые могут не сработать при крайне низком освещении, как отмечают авторы работы (Chen et al., 2018). Набор данных содержит изображения как внутри помещений, так и снаружи, причем последние фото, как правило, снимались ночью, при лунном свете или при уличном освещении. Несмотря на то что фото с долгой экспозицией могут содержать некоторый шум, их воспринимаемое качество достаточно высоко, чтобы эти изображения служили эталоном качества. В статье рассматривались две камеры с соответственно различными типами мозаики цветных фильтров (Байер и X-Trans), и проведены предварительные эксперименты на фотографиях с iPhone 6S, показывающие потенциальную переносимость данной модели между различными камерами.

ЭКСПЕРИМЕНТЫ

В этом контексте анализируется применимость нейронной сети, уже обученной на сравнительно большом датасете с одной фотокамеры, к данным с другой камеры и возможностью снижения требований к размеру датасета для обучения модели для новой камеры. При использовании датасетов особое внимание уделяется попиксельному соответствию фотографий, соответствующих одной сцене, что критически важно для корректного обучения модели. Архитектура сети в трех ниже описанных экспериментах не меняется и соответствует архитектуре, выбранной в работе в качестве оптимальной (Chen et al., 2018).

В данной работе было сделано сравнение качества работы двух моделей, первая из которых обучена авторами статьи (Chen et al., 2018) на наборе из 320 пар фото (датасет “See-in-the-Dark”), сделанных на камеру Sony A7SII, а вторая – на случайной выборке из этого набора размером 40 пар фото. При этом для каждой сцены в обучающем датасете присутствуют два фото с короткой экспозицией и одно фото с долгой экспозицией,

Таблица 1. Метрики качества работы нейросетевой модели для разных обучающих и тестовых наборов

(PSNR/SSIM)	Обучение на Sony A7SII (160 сцен)	Обучение на Sony A7SII (20 сцен)	Обучение на Sony A5100 (20 сцен)
Тестирование на Sony A7SII (50 сцен)	28.62/0.78	24.50/0.67	—
Тестирование на Sony A5100 (12 сцен)	28.17/0.85	29.69/0.85	—
Тестирование на Canon 6D (65 сцен)	25.42/0.80	25.79/0.78	—
Тестирование на Sony A5100 (7 сцен)	27.40/0.82	—	30.06/0.83

образующих две пары. Эти модели после обучения тестируются на наборах фотографий с разных камер, но с сенсорами одного типа Bayer: той же камеры Sony A7SII, Sony Alpha A5100 и Canon 6D.

Из количественных результатов, отображенных в первых трех строках табл. 1, можно сделать вывод, что при применении двух рассматриваемых моделей на фото с новых камер метрики качества либо ухудшаются незначительно в сравнении с применением для той же камеры, на которой проходило обучение, либо даже улучшаются. Можно заметить, что размер обучающего набора влияет на качество при тестировании моделей на той же камере, чего нельзя сказать о тестировании на новых двух камерах, для которых качество мало отличается, что можно заметить также на рис. 1. Таким образом, при использовании чужого датасета для обучения модели с целью использования на своей фотокамере нет необходимости в большом его объеме (в частности, показано, что при использовании обучающих выборок с размерами, отличающимися в 8 раз, качество для них отличается не более, чем на 6%).

Для получения зависимости метрик PSNR и SSIM от размера обучающего набора был проведен следующий эксперимент. Упомянутая выше нейросеть была обучена на наборах разного размера из пар фото, случайно выбранных в датасете

для камеры Sony A7SII. Для каждого из размеров выборки, относительно небольших по сравнению с размером авторского обучающего набора в работе (Chen et al., 2018), модель обучалась на данной выборке, после чего она тестировалась на авторском тестовом наборе для той же камеры. Затем были вычислены две метрики качества между результатами работы сети и соответствующими фото с долгой экспозицией путем усреднения этих показателей по всем таким парам.

Для выполнения следующего эксперимента был собран датасет из 27 пар фотографий в сыром формате, сделанных на камеру Sony A5100. Каждой сцене в нем соответствует фотография с короткой экспозицией в 0.01 с и фотография с длинной экспозицией в 1 с, которые соответствуют друг другу попиксельно (камера оставалась неподвижной). Просмотр датасета доступен по ссылке <https://rb.gy/kmt6ar>.

В рамках данной работы сравниваются модель, обученная авторами статьи (Chen et al., 2018) на фотографиях с камеры Sony A7SII, и модель, обученная на фотографиях с новой камеры Sony A5100. Критерием сравнения является качество работы на тестовом наборе из фотографий, сделанных на Sony A5100.

Результаты этого сравнения, приведенные в нижней строке табл. 1 и на рис. 2, демонстрируют



Рис. 1. Изображение, сделанное на Canon 6D. *a* – фото с экспозицией 2.5 с; *б* и *в* – результаты работы моделей, обученных на 160 и 20 сценах, сфотографированных на Sony A7SII, получившей на вход фото с экспозицией 0.03 с.



Рис. 2. Изображение из собранного на Sony A5100 датасета. *а* – фото с экспозицией 1 с; *б* – результат работы модели, обученной на 160 сценах, сфотографированных на Sony A7SII, получившей на вход фото с экспозицией 0.01 с; *в* – результат работы модели, обученной на 20 сценах, сфотографированных на Sony A5100, получившей на вход то же фото.

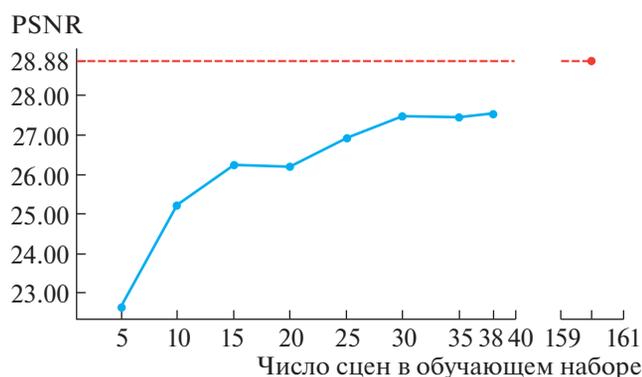


Рис. 3. Зависимость метрики качества PSNR от размера обучающей выборки при обучении и тестировании на фото с камеры Sony A7SII.

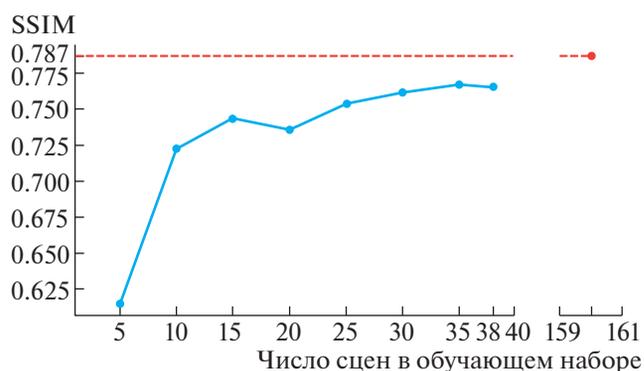


Рис. 4. Зависимость метрики качества SSIM от размера обучающей выборки при обучении и тестировании на фото с камеры Sony A7SII.

такую переносимость модели, что, несмотря на более высокие метрики для модели, обученной на фотографиях с той же камеры, на которой проведено тестирование, качество деталей на изображении, созданном моделью авторов статьи (Chen et al., 2018), заметно лучше. Метрики качества во втором эксперименте при размере обучающего набора, равном 25–30 пар фотографий, превосходят 90% от значений метрик, заявленных авторами статьи (Chen et al., 2018) для своей модели, обученной на 160 парах фото (точки красного цвета на рис. 3 и 4, содержащих графики зависимостей метрик от размера обучающего набора). Исходя из этого, можно заключить, что размер обучающего набора в 25–30 пар является достаточным для обучения модели с целью ее применения на той же камере.

ЗАКЛЮЧЕНИЕ

Получение хороших изображений при фотосъемке в условиях недостаточного освещения яв-

ляется трудной задачей из-за низкого отношения сигнала к шуму на изображениях. Способы такие, как фотографирование с долгой экспозицией или использование вспышки, имеют существенные недостатки, например, необходимость в неподвижности камеры для избежания размытий и появление бликов. Съемка в темноте с высокой скоростью в условиях отсутствия света, считается нецелесообразной при использовании традиционных методов обработки сигналов. Предложенный метод демонстрирует успешное подавление шума, а проведенные эксперименты подтверждают переносимость результата статьи (Chen et al., 2018) на новые камеры, заявленную ее авторами, а также определяют примерный диапазон в 25–30 пар фото для уменьшения размеров обучающего набора при незначительной потере в качестве, что выражается в значительном ускорении обучения.

Hardware independence and accuracy of neural network denoising of images depending on training set size

N. I. Popov and A. S. Grigoryev

^a *Moscow Institute of Physics and Technology, 171401 Dolgoprudny, Institutskiy Pereulok, 9, Russia*

^b *Institute for Information Transmission Problems, 127051 Moscow, Bolshoy Karetny per., 19, Russia*

[#] *E-mail: popov.n@phystech.edu*

This work is an investigation of image enhancement (increasing exposure and noise reduction) using neural networks proposed in Chen et al. Learning to See in the Dark (2018). The applicability of a neural network trained on the dataset from a camera to images from another camera and possibility to reduce the size of dataset to train neural network for a new camera are analyzed. A dataset of 27 aligned pairs of raw photos was collected with Exposures of 0.01s and 1s. The model trained on dataset of 320 scenes is compared with the model trained on randomly sampled 40 scenes by testing on photos from multiple cameras with the same type of colour filter. Considering the dependence of such quality metrics as PSNR and SSIM on training set size it was concluded that for the size of 25–30 scenes metrics are greater than 90% of the values claimed by the authors of the above-mentioned article for their model trained on 160 scenes. Also these metrics are compared for the model trained on the photos from the first camera, and model trained on new camera photos after their testing on a set of photos from the second camera. Despite the numerical results (PSNR/SSIM = 27.40/0.82 for the first and 30.06 / 0.83 for the second model) visual comparison of the quality of the details on the images for the model trained on a relatively large dataset from another camera are markedly better that confirms the cross-sensor generalization of the method of the mentioned article, as claimed by its authors.

Keywords: image processing, denoising with neural network, computational photography

REFERENCES

- Aharon M., Elad M., Bruckstein A. KSVD: An algorithm for designing overcomplete dictionaries for sparse representation. *Trans. Sig. Proc.* 2006. V. 54 (11). P. 4311–4322.
- Anaya J., Barbu A. Renoir - a dataset for real low-light noise image reduction. arXiv preprint arXiv:1409.8230. 2014. P. 6.
- Bernd J. Digital Image Processing. 6th revised and expanded edition. ISBN 3-540-24035-7 Springer, Berlin, Heidelberg. New York. 2005.
- Chen C., Chen Q., Xu J., Koltun V. Learning to see in the dark. *CVPR*. 2018. P. 3291–3300.
- Dabov K., Foi A., Katkovnik V., Egiazarian K. Image denoising by sparse 3-d transformdomain collaborative filtering. *TIP*. 2007. V. 16 (8). P. 2080–2095.
- Foi A., Trimeche M., Katkovnik V., Egiazarian K. Practical Poissonian-Gaussian Noise Modeling and Fitting for Single-Image Raw-Data. *TIP*. 2008. V. 17 (10). P. 1737–1754.
- Gharbi M., Chaurasia G., Paris S., Durand F. Deep joint demosaicking and denoising. *SIGGRAPH*. 2016. V. 35 (6). P. 1–12.
- Guo S., Yan Z., Zhang K., Zuo W., Zhang L. Toward convolutional blind denoising of real photographs. *CVPR*. 2019. P. 1712–1722.
- Hasinoff S.W., Sharlet D., Geiss R., Adams A., Barron J.T., Kainz F., Chen J., Levoy M. Burst photography for high dynamic range and low-light imaging on mobile cameras. *SIGGRAPH Asia*. 2016. V. 35 (6). P. 1–12.
- Kim D., Chung J.R., Jung S. Grdn: Grouped residual dense network for real image denoising and gan-based real-world noise modeling. *CVPR*. 2019.
- Lehtinen J., Munkberg J., Hasselgren J., Laine S., Karras T., Aittala M., Aila T. Noise2Noise: Learning image restoration without clean data. arXiv preprint arXiv:1803.04189. 2018.
- Liu Z., Yuan L., Tang X., Uyttendaele M., Sun J. Fast burst images denoising. *ACM Transactions on Graphics*. 2014. V. 33 (6). P. 1–9.
- Nam S., Hwang Y., Matsushita Y., Kim S.J. A holistic approach to cross-channel image noise modeling and its application to image denoising. *CVPR*. 2016. P. 1683–1691.
- Perona P., Malik J. Scale-space and edge detection using anisotropic diffusion. *TPAMI*. 1990. V. 12 (7). P. 629–639.
- Plotz T., Roth S. Benchmarking denoising algorithms with real photographs. *CVPR*. 2017. P. 1586–1595.
- Ren W., Liu S., Ma L., Xu Q., Xu X., Cao X., Du J., Yang M.H. Low-Light Image Enhancement via a Deep Hybrid Network. *TIP*. 2019. V. 28 (9). P. 4364–4375.
- Schwartz E., Giryes R., Bronstein A.M. DeepISP: Toward Learning an End-to-End Image Processing Pipeline. *TIP*. 2019. V. 28 (2). P. 912–923.
- Simoncelli E.P., Adelson E.H. Noise removal via bayesian wavelet coring. *ICIP*. 1996. V. 1. P. 379–382.
- Wang R., Zhang Q., Fu C., Shen X., Zheng W., Jia J. Underexposed Photo Enhancement Using Deep Illumination Estimation. *CVPR*. 2019. P. 6849–6857.
- Zhang Y., Zhang J., Guo X. Kindling the Darkness: A Practical Low-light Image Enhancer. *CVPR*. 2019. P. 1632–1640.
- Zhang K., Zuo W., Chen Y., Meng D., Zhang L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *TIP*. 2017. V. 26 (7). P. 3142–3155.