

Переносимость нейросети U-net для улучшения фотографий с короткой экспозицией

Н. И. Попов^{1,2}, А. С. Григорьев²

¹ Московский физико-технический институт (государственный университет)

² Институт проблем передачи информации имени А. А. Харкевича РАН

popov.n@phystech.edu

me@ansgri.com

Абстракт. Для изучения переносимости статьи Chen Chen, Qifeng Chen “Learning to See in the Dark” на новые фотокамеры был собран датасет из 27 пар фотографий в сыром формате. 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, путем их тестирования на наборе из фотографий со второй камеры. Несмотря на численные результаты этого сравнения (PSNR / SSIM = 27.40 / 0.82 для первой и 30.06 / 0.83 для второй модели) демонстрируется такая переносимость результата, что качество деталей на изображениях для модели, обученной на относительно большом датасете из фотографий с другой камеры, заметно лучше. Таким образом, подтверждается обобщаемость результата упомянутой статьи, заявленной ее авторами.

Keywords: U-net, raw format, denoising.

Фотографирование темных сцен, как и фотографирование с короткой экспозицией, обладают неудовлетворительным качеством из-за наличия шума, неверного отображения цветов и недостаточной яркости изображения. Данная работа представляет собой продолжение подхода, полученных авторами статьи [1], путем изучения переносимости их результатов на новые камеры и упрощения процесса получения качественных фото при фотографировании в описанных условиях.

С этой целью анализируется применимость нейронной сети, уже обученной на сравнительно большом датасете с одной фотокамеры, к данным с другой камеры и возможность снижения требований к размеру датасета для обучения модели для новой камеры. Используемые датасеты состоят из пар сырых (raw) фотографий, соответствующих одной сцене, одно из которых снято с долгой, а другое - с короткой экспозицией. При этом особое внимание уделяется пиксельному соответствию этих двух фотографий, что критически важно для корректного обучения модели.

Архитектура сети соответствует выбранной авторами U-net архитектуре, входом для которой являются сырые фото, снятые с короткой экспозицией, а результатом работы - изображения, качество которых оценивается по близости к соответствующим фото с долгой экспозицией. Для оценки последнего используются метрики Peak Signal-to-Noise Ratio (PSNR) и Structural Similarity Index Measure (SSIM), а также сравнение фрагментов изображений.

В данной работе было сделано сравнение качества работы двух моделей, первая из которых обучена авторами статьи [1] на наборе из 320 пар фото, 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, образующих две пары. Данные 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 Таблице 1,

Таблица 1.

PSNR / SSIM	Модель, обученная на Sony A7SII (160 сцен)	Модель, обученная на Sony A7SII (20 сцен)
Sony A7SII (50 сцен)	28.62 / 0.78	24.50 / 0.67
Sony A5100 (12 сцен)	28.17 / 0.85	29.69 / 0.85
Canon 6D (65 сцен)	25.42 / 0.80	25.79 / 0.78

можно сделать вывод, что при применении данных моделей на фото с новых камер метрики качества либо ухудшаются незначительно в сравнении с камерой Sony A7SII, на которой проходило обучение, либо даже улучшаются. Также важно заметить, что размер обучающего набора влияет на качество при тестировании моделей на той же камере Sony A7SII, чего нельзя сказать о тестировании на новых двух камерах, для которых качество мало отличается, что можно заметить на Рисунке 1.



(a)

(b) 160 сцен

(c) 20 сцен

Рис. 1. Изображение из датасета авторов статьи [2], собранного на Canon 6D. (a) - фото с экспозицией 2.5 с. (b) и (c) - результаты работы сети, обученной на Sony A7SII, получившей на вход фото с экспозицией 0.03 с.

Для получения зависимости метрик PSNR и SSIM от размера обучающего набора модели были обучены на наборах разного размера из пар фото, случайно выбранных в датасете для камеры Sony A7SII, упомянутом выше. Для каждого из размеров выборки, относительно небольших по сравнению с размером авторской обучающей выборки в [1], после обучения модели на данной выборке и ее тестирования на авторском тестовом наборе для той же камеры были вычислены две метрики качества между результатами работы сети и соответствующими фото с долгой экспозицией путем усреднения этих показателей по всем таким парам. Основываясь на графиках, демонстрирующих эту зависимость на Рисунках 2 и 4

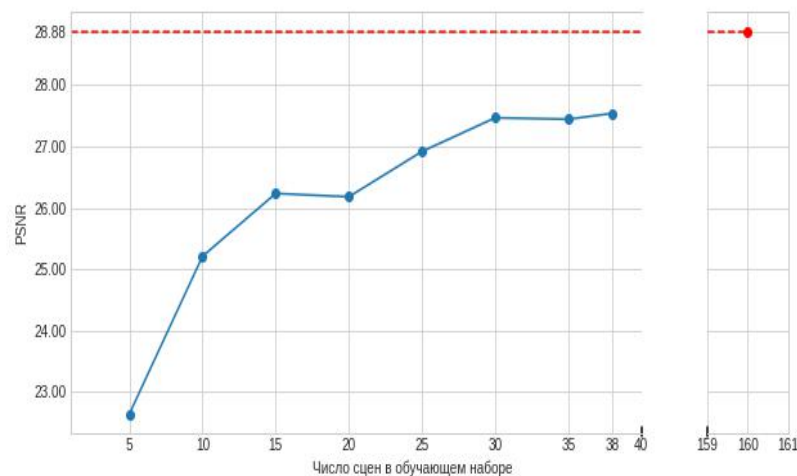


Рис. 2. Зависимость метрики качества PSNR от размера обучающей выборки при обучении и тестировании на фото с камеры Sony A7SII.

можно заключить, что размер обучающего набора в 25-30 пар фотографий является вполне достаточным для обучения модели с целью ее применения на той же камере, поскольку метрики качества в таких экспериментах превосходят 90% от значений метрик, заявленных авторами статьи [1] для своей модели, обученной на 160 парах фото (точки красного цвета на графиках).



(a)

(b) 160 сцен

(c) 20 сцен

Рис. 3. Изображение из собранного на Sony A5100 датасета. (a) - фото с экспозицией 1 с. (b) - результат работы сети, обученной на Sony A7SII, получившей на вход фото с экспозицией 0.01 с. (c) - результат работы сети, обученной на Sony A5100, получившей на вход то же фото.

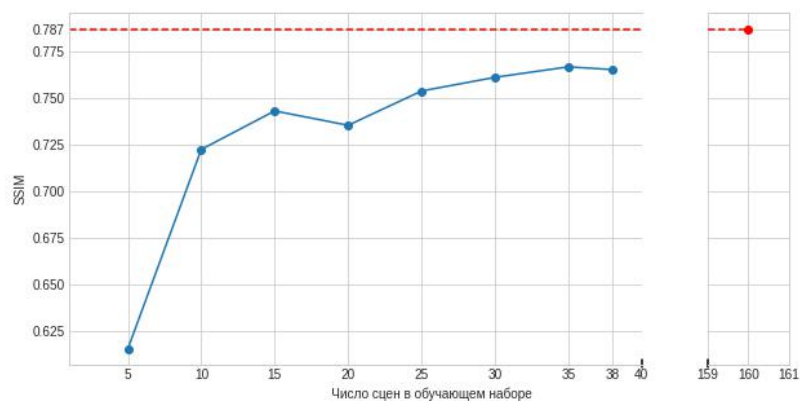


Рис. 4. Зависимость метрики качества SSIM от размера обучающей выборки при обучении и тестировании на фото с камеры Sony A7SII.

В рамках данной работы также сравниваются модель, заимствованная у авторов статьи [1] и обученная 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 следующей таблице, и Рисунок 3 демонстрируют такую переносимость результата, что, несмотря на более высокие метрики для модели, обученной на фотографиях с той же камеры, на которой проведено тестирование, качество деталей на изображении для модели авторов статьи [1] заметно лучше.

Таблица 2.

PSNR / SSIM	Модель, обученная на Sony A7SII (160 сцен)	Модель, обученная на Sony A5100 (20 сцен)
Sony A5100 (7 сцен)	27.40 / 0.82	30.06 / 0.83

Для выполнения последнего 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 секунду и фотография с длинной экспозицией в 1 секунду, которые соответствуют друг другу попиксельно (камера оставалась неподвижной).

Таким образом, данная работа подтверждает переносимость результата статьи [1], заявленную ее авторами.

Литература

1. Chen Chen, Qifeng Chen, Jia Xu, Vladlen Koltun, Learning to See in the Dark // Conference on Computer Vision and Pattern Recognition 2018.
2. Report at www.github.com/jconenna/Canon-6D-Datasets-For-Learning-to-See-in-the-Dark