

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ

Федеральное государственное автономное образовательное  
учреждение высшего образования  
ЮЖНЫЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ

Институт математики, механики и компьютерных наук  
имени И. И. Воровича

Направление подготовки  
Фундаментальная информатика и информационные технологии

А. С. Коваленко

ОБУЧЕНИЕ ШУМОПОДАВЛЯЮЩЕЙ НЕЙРОННОЙ СЕТИ БЕЗ  
ИСПОЛЬЗОВАНИЯ ЧИСТЫХ ДАННЫХ

Магистерская диссертация

Научный руководитель:  
учёная степень, учёное звание / должность И. О. Фамилия  
Рецензент:  
ученая степень, ученое звание, должность И. О. Фамилия

Ростов-на-Дону  
2020

# ОГЛАВЛЕНИЕ

Введение	3
Глава 1. Задача шумоподавления	7
1.1. Постановка задачи . . . . .	7
1.2. Анализируемые данные . . . . .	9
1.2.1. Сбор данных . . . . .	9
1.2.2. Анализ собранных данных . . . . .	11
Список литературы	19

# ВВЕДЕНИЕ

Шумоподавление является часто встречаемой задачей в области компьютерного зрения. Так как любое изображение, полученное из картины реального мира является дискретным представлением непрерывного аналогового сигнала, то в нём будет присутствовать шум. Изначально шум появляется у полезного сигнала из-за погрешностей приёма оптического излучения фотоидами, данное явление изложено в книге *Fundamentals of linear electronics* [1]. Затем к данномуискажённому сигналу добавляются потери при процессе дискретизации. Действие шума на изображения можно легко увидеть на примере изображения 0.1, на правой части изображения приводится пример фотографии, снятой при более благоприятных условиях, при которых фотосенсор цифровой камеры порождает меньше шума, чем на фотографии слева.

Также существуют изображения, полученные не только с фотосенсоров. Примером можно привести рентгенографию, широко используемую во многих областях, таких как медицина, процессы производства и эксплуатации, криминалистика, реставрация и экспертиза художественных ценностей. Существует два варианта получения цифрового изображения рентгенографии, это оцифровка уже существующего рентгеновского снимка, и использование технологии цифровой рентгенографии, при которой сразу идет цифровая обработка получаемого рентгеновского изображения. Оба данных метода подвержены наличию шума на получаемом цифровом изображении, в большей мере шум преобладает на изображениях, получаемых первым способом, так как идёт дополнительное наложение шума сканиру-



Рисунок 0.1 — Пример зашумленного изображения.

ющим устройством. Пример подобного шума проиллюстрирован на изображении 0.2. Аналогичная ситуация наблюдается и с снимками в области радиографического контроля сварных соединений, пример проиллюстрирован на изображении 0.3.

Помимо рентгенографии в обработке медицинских изображений потребность в избавлении полезного сигнала от шума встречаются и в других методах диагностики и визуализации, таких как цифровая реконструкция компьютерной томографии, магнитно-резонансной томографии и других. Потребность в шумоподавлении на медицинских цифровых изображениях изложена в работе [2].

На данный момент с развитием технологий глубокого обучения и построения глубоких архитектур сверточных нейронных сетей, такие архитектуры применяются к решению широкого ряда задач в области обработки и анализа изображений. Существует три подхода к обучению нейронных сетей: обучение с учителем, обучение без учителя и обучение с подкреплением. Если рассматривать близкие задачи к шумоподавлению с точки зрения построения шаблона обуче-

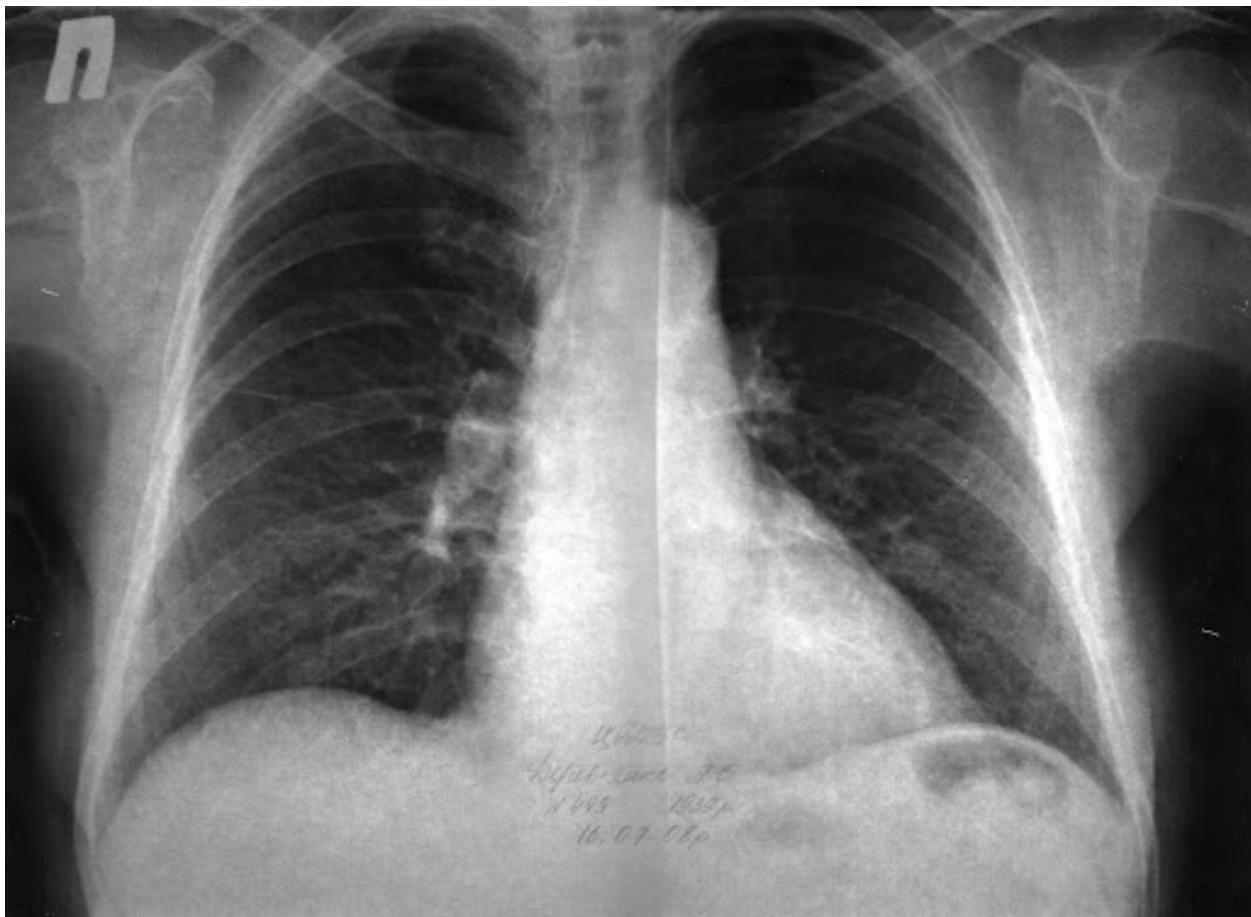


Рисунок 0.2 — Сканированный рентгеновский снимок грудной клетки человека.

ния сети, это задача увеличения разрешения изображения с помощью нейронных сетей [3] и задача восстановления изображения [4]. Для решения данных задач нейросетевые архитектуры обучаются с помощью подхода обучения с учителем [5]. При процессе обучения таких архитектур в качестве входных данных выступают сжатые варианты изображений, подаваемых, как ожидаемые при предсказании сети. Если рассматривать в данном контексте обучение для задачи классификации, то на вход сети необходимо подавать изображение с присущим на нём шумом, и обучать сеть предсказывать уже само чистое изображение без шума.

Все перечисленные способы в этом разделе получения цифрового изображения объединяет один недостаток, это невозможность

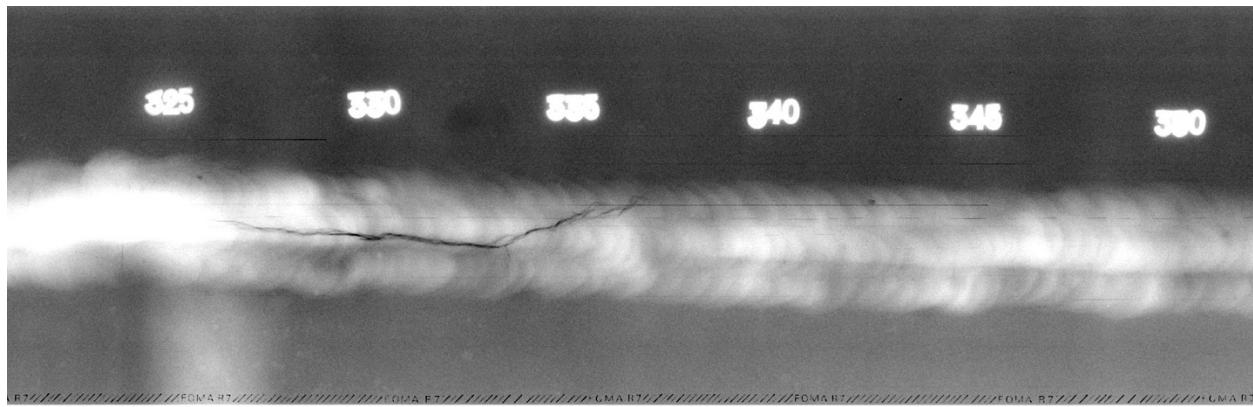


Рисунок 0.3 – Цифровой рентгеновский снимок сварного соединения.

получения чистого изображения без шума для обучения шумоподавляющей нейронной сети. В данной работе рассматривается подход к построению архитектуры шумоподавляющей нейронной сети и построения обучающего процесса, основанного на методе обучения без учителя (неконтролируемое обучение).

# ГЛАВА 1

## ЗАДАЧА ШУМОПОДАВЛЕНИЯ

### 1.1. Постановка задачи

Пусть имеется изображение в формате непрерывного сигнала, полученного с АЦП об сигналах матрицы фотосенсора фиксированного цифрового устройства в момент времени  $t$  (DNG формат [6]):

$$S(t) \quad (1.1)$$

Сигнал  $S(t)$  (1.1) состоит из полезного сигнала  $G(t)$  и шума  $r(t)$ , порождаемым фотосенсором:

$$S(t) = G(t) + r(t) \quad (1.2)$$

При применении преобразования сырого сигнала  $S(t)$  к трехмерному дискретному изображению в цветовой схеме RGB получаем матрицу изображения  $\tilde{I}$ , с влиянием шума квантования при округлении сигнала при его дискретизации. Подход преобразования  $p$  описан в работе Processing RAW Images in MATLAB [7].

$$\tilde{I}_{i,j} = p(S(t))_{i,j} + q, \quad q \sim \mathcal{N}\left(-\frac{1}{2}, \frac{1}{12}\right), \quad (1.3)$$

где  $q$ , это шум квантования, семплируемый из нормального распределения с параметрами  $\mathcal{N}\left(-\frac{Q}{2}, \frac{Q}{12}\right)$  при шаге квантования  $Q = 1$ , более

подробно можно ознакомиться с этим в главе Оценки ошибок (шумов) квантования выходного сигнала в цифровом фильтре из книги Цифровая обработка сигналов [8].

Для упрощения постановки задачи будем полагать, что матрица изображения  $\tilde{I}$  состоит из суммы изображения, полученного из преобразования чистого (полезного) сигнала  $G(t)$  (1.2) в матрицу  $I$  и шума  $\alpha$ , полученного из случайного распределения  $\mathbb{P}$ , так как появление шума  $r(t)$  (1.3) можно считать абсолютно случайным процессом.

$$\tilde{I} = I + \alpha, \quad \alpha \sim \mathbb{P} \quad (1.4)$$

Таким образом, имея серию из  $N$  изображений одной и той же сцены, снятых в разный момент времени  $t$  имеем следующую выборку:

$$\{\tilde{I}_k\}_{k=1}^N, \text{ где } \tilde{I}_k = I + \alpha_k, \quad \alpha_k \sim \mathbb{P} \quad (1.5)$$

Причём для каждого элемента из набора  $\tilde{I}_k$  компонента  $I$  является одним и тем же значением, так как при съёмке одной сцены полезный сигнал  $G(t)$  (1.2) остаётся неизменным.

Целью данной работы является построение с помощью нейронной сети приближения отображения  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , обладающим следующим свойством:

$$\forall \tilde{I} \in \{\tilde{I}_k\}_{k=1}^N \implies \phi(\tilde{I}) = I \quad (1.6)$$

Для построения приближения отображения  $\phi$ , нейронная сеть  $f$  будет обучаться решать следующую задачу оптимизации:

$$\min_w \|f(\tilde{I}, w) - I\|_{L_2}, \quad (1.7)$$

где  $w$  - параметры сети  $f$ ,  $L_2$  - евклидова норма [9].

## 1.2. Анализируемые данные

### 1.2.1. Сбор данных

Так как в данной работе поставлена задача подавления шума у сигнала с заданного сенсора, то изображения для проведения исследования были получены с помощью фиксированного устройства, *Apple iPhone X*, имеющего камеру, состоящую из двух сенсоров.

Таблица 1.1. — Характеристики первого сенсора

Характеристика	Значение	Единица измерения
Разрешение	12	<i>MP</i>
Апертура	f/1.8	
Фокусное расстояние	28	<i>mm</i>
Размер сенсора	1/3"	
Размер пикселя	1.22	$\mu m$
Стабилизация изображения	OIS	

Таблица 1.2. — Характеристики второго сенсора

Характеристика	Значение	Единица измерения
Разрешение	12	<i>MP</i>
Апертура	f/2.4	
Фокусное расстояние	52	<i>mm</i>
Размер сенсора	1/3.4"	
Размер пикселя	1.0	$\mu m$
Стабилизация изображения	OIS	

С данного устройства были сделаны серии изображений семи сцен с различным освещением и цветовым наполнением. Данные серии получены стандартными средствами для съёмки на устройстве. Под сценой подразумевается съёмка неизменяемой картины реального мира, получая матрицу (1.4). Для этого устройство фиксировалось на штативе в неподвижном состоянии и запуск процесса съёмки

кадра производился с беспроводного устройства, тем самым получая набор изображений  $\{\tilde{I}_k^q\}_{k=1}^N$  (1.5), где  $q$  - порядковый номер снимаемой сцены. Для каждой сцены при съемке были зафиксированы значения ISO, фокуса и цветовой температуры.

Каждая сцена содержит в среднем по 14 фотографий, суммарное количество кадров из всех наборов составляет 95 изображений.

Так как стандартные средства для съемки не позволяют получать RAW (изображение с сырого сигнала АЦП (1.1)) изображение, то были произведены съемки дополнительных сцен с использованием стороннего программного обеспечения, позволяющего получать «сырые» кадры в формате TIFF [10]: Simple Raw camera [11]. С помощью данного приложения были сняты серии трёх дополнительных сцен. Данные серии содержат суммарно 30 изображений, ровно по 10 кадров в каждой серии.

Максимальное количество изображений в одной серии ограничено 20-ю кадрами, так как при продолжении процесса съемки фотосенсор нагревается и из-за теплового воздействия появляются дополнительные искажения сигнала, из-за которых теряется пиксельное соответствие кадров в серии между собой.

В итоге получается набор данных, состоящий из 125 изображений, снятых на один сенсор. Визуальное различие шума на изображениях, полученных разными подходами можно рассмотреть на рисунке 1.1.

Также для сравнения результатов были сделаны 6 серий, снятых следующей web камерой [12] в разрешении  $1920 \times 1080$ . Снимались видеозаписи короткой длительности и разбивались на кадры с помощью библиотеки FFmpeg [13]. Суммарное количество кадров в данных сериях составило 811.



Рисунок 1.1 — Сравнение получаемого шума на изображении, снятого стандартными средствами устройства (слева) и RAW изображения (справа)

### 1.2.2. Анализ собранных данных

Во время процесса съемки при нагреве фотосенсора могут происходить незначительные искажения кадра. Или также возможны незаметные смещения устройства при съемке, а также смена внешних условий, таких как освещение или движение объектов в кадре. Для собираемого набора изображений сдвиг более чем на один пиксель между кадрами одной серии уже критичен.

Для анализа качества полученных серий изображений были построены распределения  $e^q$  (1.9) отклонений каждого изображения от усредненного по всем изображениям из серии  $\hat{I}^q$  (1.8) по евклидовой метрике.

$$\hat{I}_{i,j}^q = \sum_k \tilde{I}_k^q i,j \quad (1.8)$$

$$e_k^q = \|\hat{I}^q - \tilde{I}_k^q\|_{L_2} \quad (1.9)$$

Дополнительно производится нормировка выборки по метрике  $L_1$ :

$$e_k^q = \frac{e_k^q}{\sum_i (e_i^q)}$$

Далее по значениям выборки  $e^q$  (1.9) строятся графики плотности нормального распределения  $\mathcal{N}(\mu, \sigma)$  с параметрами:

$$\mu^q = \mu(e^q), \sigma^q = \sigma(e^q) \quad (1.10)$$

Таким образом математическое ожидание  $\mu^q$  (1.10) показывает насколько усредненное изображение  $\hat{I}^q$  (1.8) близко к изображению, теоретически, получаемому из чистого сигнала  $I^q$  (1.5):

$$\hat{I}^q \xrightarrow{\mu^q \rightarrow 0} I^q$$

А также, чем ближе к 0 значение среднеквадратичного отклонения выборки  $\sigma^q$  (1.10), тем на большем количестве изображений из серии присутствует сигнал  $G(t)$  (1.2), зафиксированный в момент времени  $t = t_0$ , то есть насколько сцена неизменна между кадрами.

Графики распределений для серий, снятых стандартными средствами устройства *Apple iPhone X* изображен на рисунке 1.2. Из графи-

ков можно увидеть, что чем больше размер серии, тем ближе усредненное изображение к чистому.

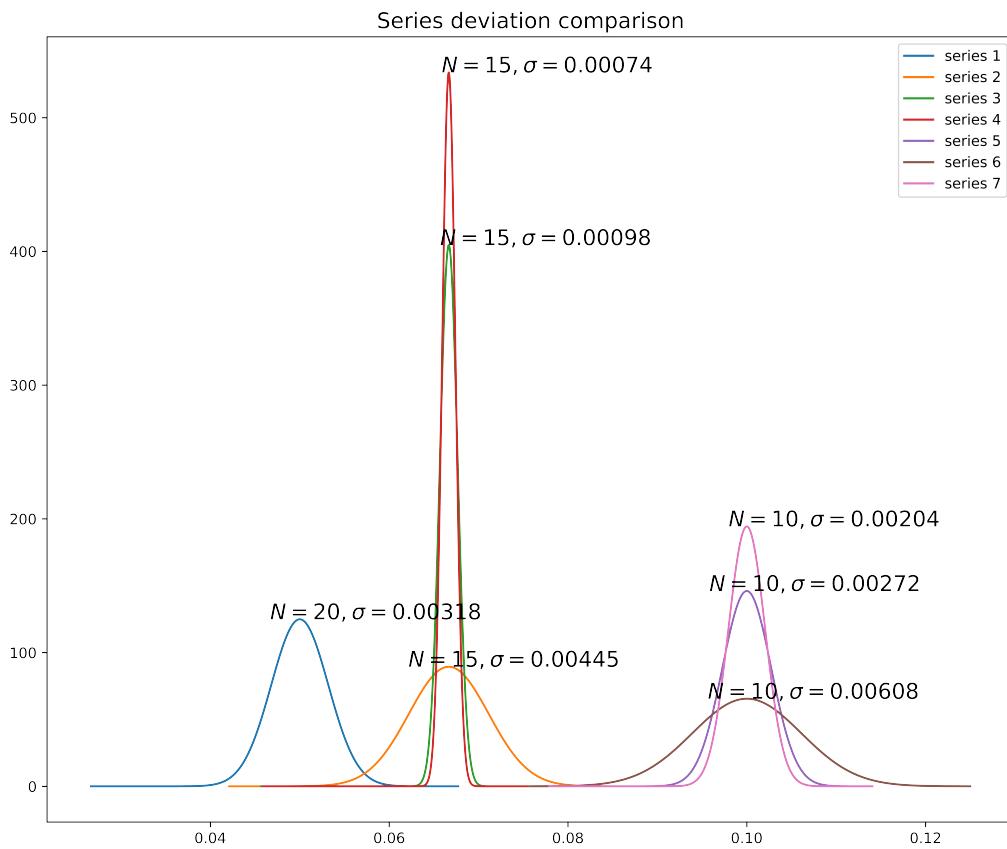


Рисунок 1.2 — Графики плотностей нормальных распределений для серий, снятых стандартными средствами устройства *Apple iPhone X*

Рассматриваемый параметр  $\sigma$  (1.10) сильно зависит не только от движения объектов на снимаемой сцене, но и от уровня освещения. Сдвиг объектив особо критичен, так как обученная нейросеть на такой выборке будет размывать результирующее изображение. Сравнение усредненного изображения, имеющего сдвиг сцены с изображе-

нием, полученным из усреднения кадров статичной сцены изображено на рисунке 1.3.

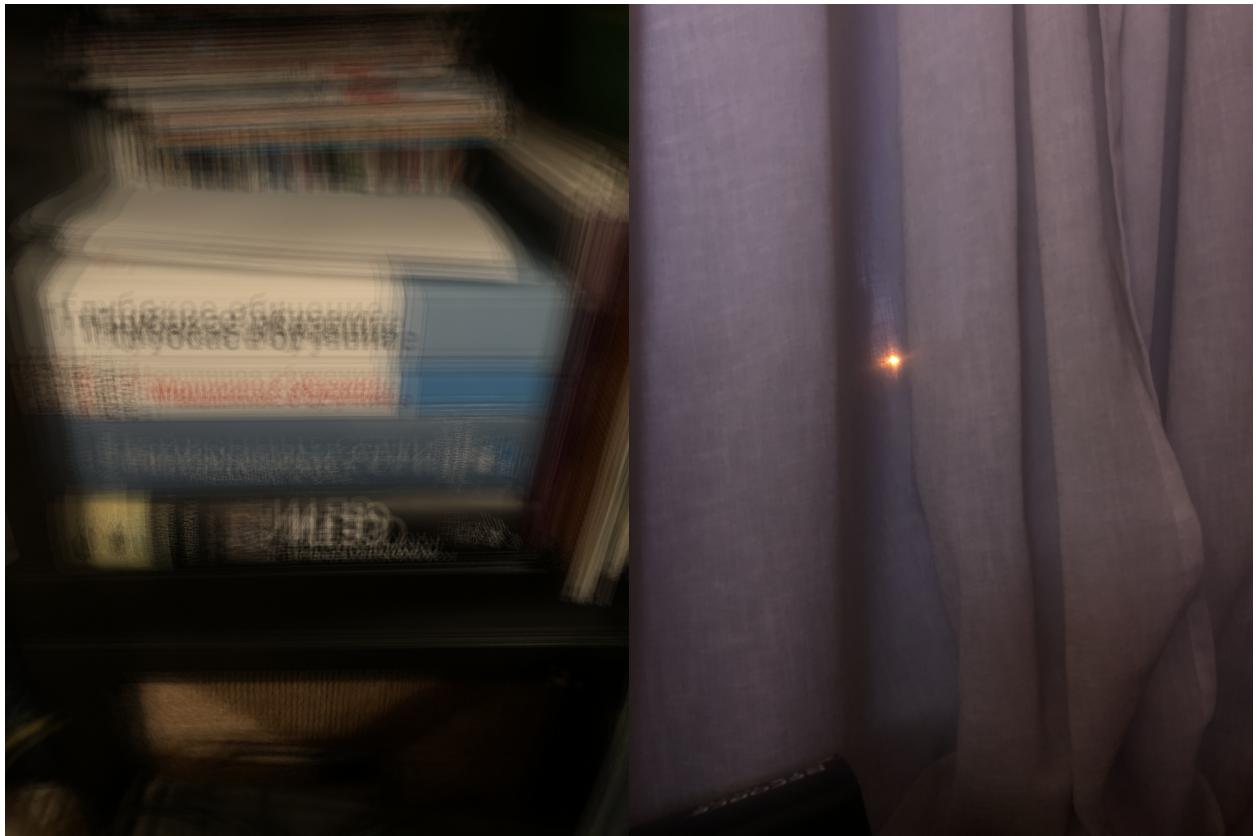


Рисунок 1.3 – Для среднего изображения серии слева  $\sigma = 0.00891$ , для среднего изображения серии справа  $\sigma = 0.00272$

Также большое значение отклонения  $\sigma$  (1.10), полученное из-за сильного изменения освещения при съемке также неблагоприятно скажется на качестве обучения нейронной сети, так как нейронная сеть, обученная на таких сериях, не сможет корректно предсказывать цветовую гамму результирующего изображения. Пример зависимости параметра отклонения от уровня освещения снимаемой сцены при естественном освещении изображен на рисунке 1.4. Из данного примера можно заметить, что незначительные отклонения в освещении, различимые по небольшим отличиям в цветовых гистограммах изображений из серии сильно влияют на параметр  $\sigma$  (1.10), и для дан-

ной статичной сцены данный параметр близок к значению, получаемому у сцены с сильным сдвигом 1.3.

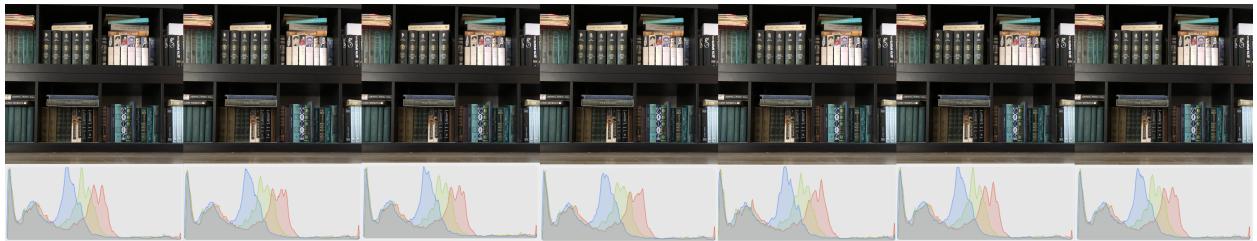


Рисунок 1.4 — Серия изображений с цветовыми гистограммами, параметр  $\sigma$  данной серии равен 0.00608

Для остальных условий съемки аналогичным образом были построены графики с распределениями, изображенные на рисунках 1.5, 1.6 и 1.7. Серии, полученные с web камеры содержат относительно большое количество кадров каждой сцены (в среднем по 135) и не подвергаются искажениям из-за нагрева сенсора за счет большего шага дискретизации изображения реального мира (меньшего разрешения).

При выборе данных для обучения нейронных сетей подавляющих шум на изображении задается условие, что параметр  $\sigma$  (1.10) не должен превышать значения 0.007

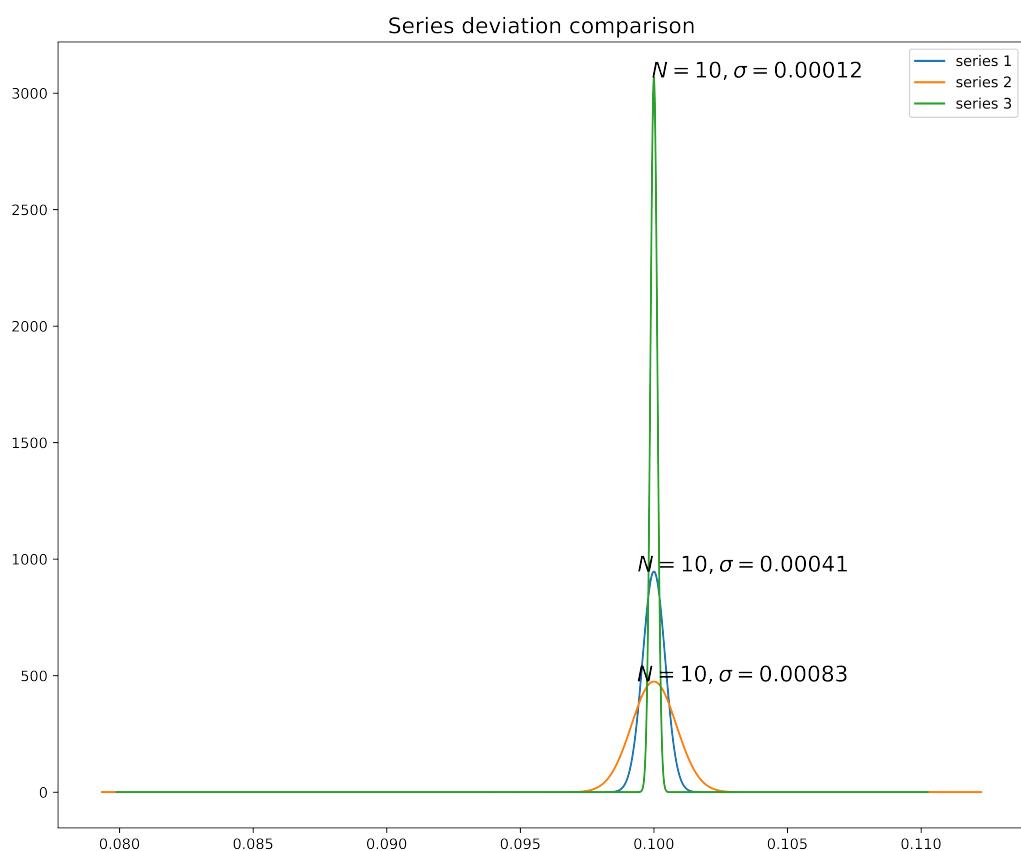


Рисунок 1.5 — Графики плотностей нормальных распределений для серий, снятых при условиях плохого искусственного освещения с помощью приложения [11]

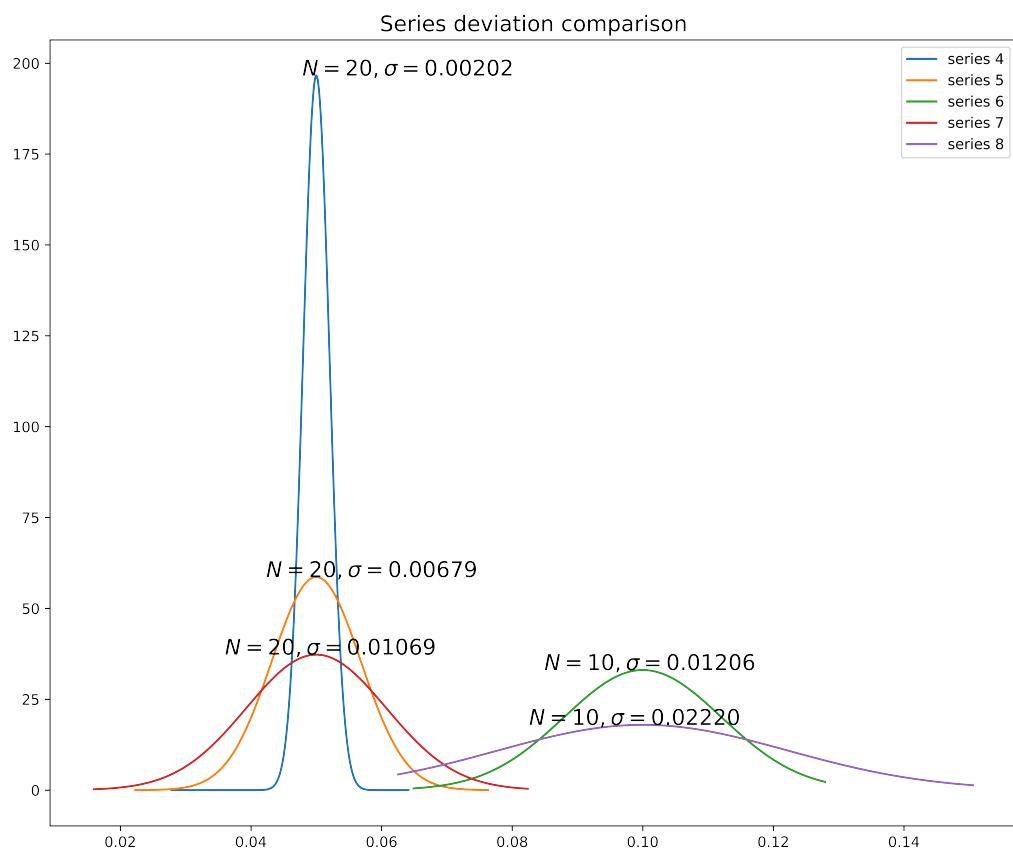


Рисунок 1.6 — Графики плотностей нормальных распределений для серий, снятых при условиях естественного освещения с помощью приложения [11]

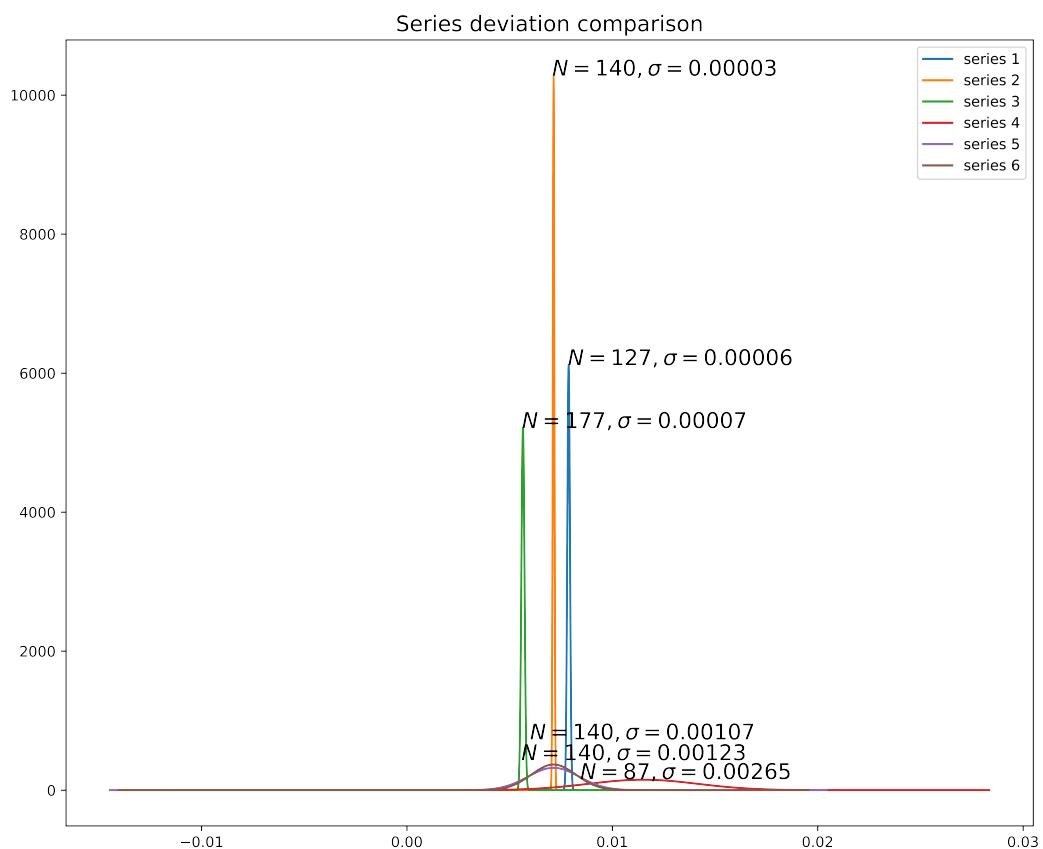


Рисунок 1.7 — Графики плотностей нормальных распределений для серий, снятых при разных условиях освещения с помощью web камеры [12]

## СПИСОК ЛИТЕРАТУРЫ

1. Cox J. Fundamentals of linear electronics: integrated and discrete. — US : DELMAR, 2001.
2. Авшаров Е. М., Абгарян М. Г., Сафарянц С. А. Обработка медицинских изображений как необходимый инструментарий медицинского диагностического процесса // Вестник рентгенологии и радиологии. — 2010. — № 3. — С. 54—61.
3. Image Super-Resolution Using Very Deep Residual Channel Attention Networks / Y. Zhang [и др.] // CoRR. — 2018. — T. abs/1807.02758. — arXiv: 1807 . 02758. — URL: <http://arxiv.org/abs/1807.02758>.
4. You S., You N., Pan M. PI-REC: Progressive Image Reconstruction Network With Edge and Color Domain // CoRR. — 2019. — T. abs/1903.10146. — arXiv: 1903 . 10146. — URL: <http://arxiv.org/abs/1903.10146>.
5. Nasteski V. An overview of the supervised machine learning methods// HORIZONS.Б. — 2017. — Дек. — Т. 4. — С. 51—62. — DOI: 10.20544/HORIZONS.Б.04.1.17.P05.
6. Adobe. Digital negative (DNG) Specification. — US, 2012. — URL: [https://www.adobe.com/content/dam/acom/en/products/photoshop/pdfs/dng\\_spec\\_1.4.0.0.pdf](https://www.adobe.com/content/dam/acom/en/products/photoshop/pdfs/dng_spec_1.4.0.0.pdf).
7. Sumner R. Processing RAW Images in MATLAB. — 2014. — URL: [https://rcsumner.net/raw\\_guide/Rawguide.pdf](https://rcsumner.net/raw_guide/Rawguide.pdf).

8. Гондельберг Л. М., Матюшкин Б. Д., Поляк И. Н. Цифровая обработка сигналов. — Москва : Радио и связь, 1985. — URL: <http://optic.cs.nstu.ru/files/Lit/Math/gold.pdf>.
9. Хайкин С. Нейронные сети : Полный курс. — Москва : Вильямс, 2016.
10. Parsons G., Rafferty J. Tag Image File Format (TIFF) - image/tiff // Adobe Systems. — 1998. — URL: <https://tools.ietf.org/rfc/rfc2302.txt>.
11. Simple Raw camera. — URL: <https://apps.apple.com/ru/app/simple-raw-camera/id1286921686> (дата обр. 09.05.2020).
12. Logitech B525 HD Webcam. — URL: <https://www.logitech.com/assets/64667/b525-datasheet.ENG.pdf> (дата обр. 09.05.2020).
13. FFmpeg : A complete, cross-platform solution to record, convert and stream audio and video. — URL: <http://ffmpeg.org/> (дата обр. 09.05.2020).