

# Сегментация освещенных участков сцены с использованием полносверточных нейронных сетей и алгоритмов компьютерного зрения для систем дополненной реальности

М.И. Сорокин<sup>1</sup>, Д.Д. Жданов<sup>1</sup>, И.С. Потемин<sup>1</sup>, Б.Х. Барладян<sup>2</sup>, Н.Н. Богданов<sup>1</sup>, А.Д. Жданов<sup>1</sup>  
vergotten@gmail.com|ddzhdanov@mail.ru|ipotemin@yandex.ru|bbarladian@gmail.com  
|nnbogdanov@itmo.ru|adzhdanov@itmo.ru

<sup>1</sup>Университет ИТМО, Санкт-Петербург, Россия;

<sup>2</sup>ИПМ им. М.В.Келдыша РАН, Москва, Россия.

*Актуальность данной темы обусловлена бурным развитием систем виртуальной и дополненной реальности. Проблема заключается в формировании естественных условий освещения объектов виртуального мира в реальном пространстве. Для решения задачи определения источников света и восстановления его оптических параметров была предложена полносверточная нейронная сеть, позволяющая «схватить» особенности поведения света. На выходе нейронной сети получаем сегментированное изображение с классами и уровнями освещенности. Полносверточная нейронная сеть отлично подходит для сегментации изображения, поэтому в качестве «энкодера» была взята архитектура VGG-16 со слоями, которые сворачивают некую область входного изображения до 1x1 пикселя и классифицируют его к одному из классов.*

*Обучение полносверточной нейронной сети осуществлялось на 221 тренировочных и 39 проверочных изображениях со скоростью обучения 1E-2 на 200 эпохах. После обучения по критерию оценки «IoU» (intersection over union) точность составила порядка 70 процентов, а точность классификации к первому классу составила практически 90 процентов.*

**Ключевые слова:** классификация, освещение, свёрточные нейронные сети, сегментация.

## Segmentation of illuminated areas of scene using fully-convolutional neural networks and computer vision algorithms for augmented reality systems

Maxim I. Sorokin<sup>1</sup>, Dmitry D. Zhdanov<sup>1</sup>, Igor S. Potemin<sup>1</sup>, Boris Barladian<sup>2</sup>, Nikolai N. Bogdanov<sup>1</sup>, Andrey D. Zhdanov<sup>1</sup>  
vergotten@gmail.com|ddzhdanov@mail.ru|ipotemin@yandex.ru|bbarladian@gmail.com  
|nnbogdanov@itmo.ru|adzhdanov@itmo.ru

<sup>1</sup> ITMO University, St. Petersburg, Russia;

<sup>2</sup>Keldysh Institute of Applied Mathematics RAS, Moscow, Russia.

*The relevance of this topic is due to the rapid development of virtual and augmented reality systems. The problem lies in the formation of natural conditions for lighting objects of the virtual world in real space. To solve a light sources determination problem and recovering its optical parameters were proposed the fully-convolutional neural network, which allows catching the 'behavior of light' features. The output of FCNN is a segmented image with light levels and its strength. Naturally, the fully-convolutional neural network is well suited for image segmentation, so as an encoder was taken the architecture of VGG-16 with layers that pools and convolves an input image to 1x1 pixel and wisely classifies it to one of a class which characterizes its strength.*

*Neural network training was conducted on 221 train images and 39 validation images with learning rate 1E-2 and 200 epochs, after training the loss was 0,2. As a test was used an 'intersection over union' method, that compares the ground truth area of an input image and output image, comparing its pixels and giving the result of accuracy. The mean IoU is 0.7, almost rightly classifying the first class with a value of 90 percents of accordance and the last class with a probability of 30 percents.*

**Keywords:** classification, illumination, convolutional neural networks, segmentation.

### 1. Введение

Дополненная реальность (augmented reality, AR) — это та среда, которая создается в результате наложения информации или объектов на воспринимаемый мир в реальном времени. Для этого могут использоваться как компьютерные устройства: смартфоны и планшеты, так и различные гаджеты: очки и шлемы дополненной реальности. В отличие от виртуальной реальности человек не воспринимает какую-то другую реальность, он все так же видит реальный мир, но немного в дополненном состоянии.

Уже сегодня дополненная и виртуальная реальность сильно повлияли на образование и медицину. С помощью дополненной реальности хирурги повышают уровень своего

профессионализма, а методы лечения и диагностики заболеваний меняются в лучшую сторону. Более того, благодаря виртуальной реальности возможности врачей сильно увеличиваются, а с помощью дистанционных браслетов врачи осуществляют полный контроль состояния пациентов на любом расстоянии. Хорошим примером являются микрофоны и микрокамеры, встроенные в очки дополненной реальности, которые помогают при реабилитации глухих или немых людей, способствуя скорейшей социальной адаптации.

В сфере образования очки дополненной реальности позволяют разнообразить учебный процесс, удерживая внимание школьников и студентов, а в науке очки дополненной реальности можно использовать для виртуальных экспериментов. Виртуальная реальность также

используется в компьютерных играх: в онлайн-играх есть целые виртуальные галактики, которые можно посетить, не выходя из дома.

Кроме того, уже появились виртуальные помощники для водителей. Голосовые команды, подключенные к смартфону, и отображение информации в очках дополненной реальности помогают в управлении автомобилем и решении неотложных вопросов.

Распространение в повседневной жизни технологий AR и VR приведет к массовому использованию интерактивного цифрового телевидения. Уже возможно посетить виртуальные экскурсии по городам и музеям, используя экранные гаджеты, со смартфонами возможно видеть виртуальный мир, а технологии дополненной реальности позволяют совершать очень даже реальные покупки в виртуальных магазинах.

Однако, какими бы оптимистичными ни были все эти утверждения, проблема реалистичного рендеринга еще не полностью решена. Дело в том, что все объекты дополненной реальности должны соответствовать условиям окружающей среды. И если мы говорим об освещении, то все объекты виртуального мира должны соответствовать освещению реального мира, должны быть правильно освещены и давать тени в противоположном направлении от источника света. Следовательно, в этой статье решается проблема восстановления оптических источников света.

Также в этой статье рассматриваются методы фильтрации. Приведены примеры рассмотренных методов решения задач улучшения визуального представления и выделения контуров объектов на изображениях.

## 2. Анализ существующих работ

Поскольку данный подход фокусируется на сегментации освещенных областей, существует несколько решений, которые только частично решают поставленную задачу. В настоящее время сверточные нейронные сети активно используются для решения различных задач, связанных непосредственно с анализом и обработкой изображений, будь то классификация или распознавание каких-либо отдельных участков. Сверточные нейронные сети уже доказали свою силу в различных классификациях изображений, и освещение не является исключением.

Например, в работе [1] представлен анализ освещения на открытом воздухе с использованием полносверточной сети. Следующая работа [2] в качестве входных данных анализирует панорамные изображения окружающей среды под открытым небом и встраивает изображение под эти условия среды. Как и ранее рассмотренные работы, работа [3] также анализирует окружение и строит тени объектов так, как они должны быть. В работе [4] также используется сверточная сеть, но для определения того, где находится объект: на открытом воздухе или в помещении. Следующая статья [5] представляет свою собственную архитектуру и решает три различные задачи: прогнозирование глубины, оценка нормалей поверхности и семантическая маркировка.

Многие работы [6,7,8,9] были направлены на обнаружение объектов с использованием сверточных нейронных сетей. Отличие нейронной сети, описанной в данной работе, заключается в том, что набор данных генерируется с использованием мощного рендерера - «Lumicert» [10], который использовался для обучения нейронной сети, восстанавливая сегментированные участки света, похожие на эталонные изображения (ground truth), с помощью объектной функции «categorical crossentropy».

Поскольку реалистичные цифровые изображения синтезируются методами стохастической трассировки лучей, они могут содержать шумовую составляющую. Для

устранения шума в системах компьютерного зрения используются различные алгоритмы фильтрации, например, детектор краев «Канни», оператор Собеля или медианный фильтр. Однако эти фильтры не всегда подходят для решения проблемы устранения шума, возникающего в результате процесса стохастической трассировки лучей. Для устранения такого шума наиболее подходящими будут фильтры, интегрированные в процесс трассировки лучей [11, 12].

Работы [13,14,15,16] так же, как и данная работа направлены на восстановление освещения для систем дополненной реальности, но для различных целей и задач. Среди данных статей рассматриваются методы анализа прямого освещения, не принимая во внимание вторичного для объектов виртуальной реальности. Также, в рассматриваемых работах рассматривается проблема, связанная с нахождением непосредственно самого наблюдателя в системе дополненной реальности и освещения относительно его самого. Был предложен метод восстановления источников освещения по теням объектов, но это лишь может дать направление луча, а не оптические свойства.

Главное отличие данной работы от вышеперечисленных заключается в том, что в существующих работах, используя огромные наборы данных, нейронную сеть учат не только распознавать освещенные участки изображения, но и в ряде случаев встраивать изображения под конкретные условия окружающей среды. Главная задача настоящей работы — это определения и классификация реальной освещенности помещения в световых единицах. Данная работа сосредоточена на определении освещенности и яркости видимых участков сцены. Для этого была использована вручную синтезированная выборка изображений с реалистичными оптическими параметрами сцены. Несмотря на то, что выборка состоит всего из 260 изображений (221 использовалось для обучения, а 39 для проверки), на выходе нейронная сеть с хорошей точностью классифицирует реальные оптические параметры освещения среды, которые было принято разделить по значению освещенности на 5 классов, где первый - 0 люкс, что означает не освещено вовсе, а 5 класс – освещенность, создаваемая от источника света в виде обычной комнатной лампы. Главное достоинство нейронных сетей, помимо гибкой настройки алгоритма и хороших результатов на выходе — это простота их использования. В отличие от сложных алгоритмов, требующих значительных ресурсов, нейронную сеть после обучения можно использовать даже на мобильных устройствах и других платформах.

## 3. Реализация

Преимущества сверточных нейронных сетей уже не раз доказаны на многих соревнованиях по компьютерному зрению и хакатонам. Они заключаются в способности сетей изучать и строить сложные карты признаков, основанные на данных предыдущих сверточных слоев, которые на практике позволяют изучать, распознавать и строить сложные иерархии признаков.

В качестве архитектуры сверточной сети было решено брать архитектуру VGG16 Net, которая с успехом использовалась в следующих работах [17, 18]. Она состоит из 5 блоков со слоями свертки, пулинга и функцией активации ReLU между слоями. В отличие от данной архитектуры было решено не использовать метод регуляризации «dropout», а в качестве метода оптимизации был выбран «Nesterov», это связано с небольшим набором данных для обучения.

Архитектура данной сети и ее детали представлены на рисунке 1.

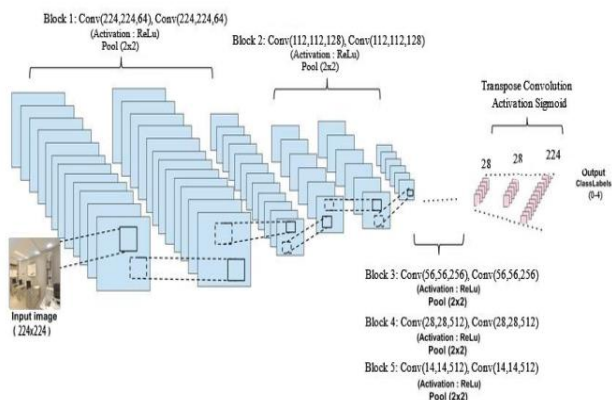


Рис. 1. Архитектура и описание полносверточной нейронной сети

Поскольку использовался метод обучения «с учителем», то изображения и их соответствующие маски в представлении псевдоцветов подавались на вход нейронной сети, где более «холодные» цвета отвечают за менее интенсивное освещение, в то время как «более теплые» цвета указывают на более освещенные участки.

Задача полносверточной сети - классифицировать каждый пиксель изображения к одному конкретному классу. То есть, проходя через все сверточные слои, сеть относит определенную область изображения к одному классу в соответствии с интенсивностью освещения. В данной реализации присутствуют 5 классов освещения. После того, как сеть привязала анализируемую область к определенному классу освещения, в действие вступает функция «upsample». Задача этой функции – использовать метод «ближайшего соседа» для восстановления исходной размерности изображения. Более подробно о слое «upsample». «Upsample» это своего рода функция, которая конвертирует изображение с низкой размерностью в изображение с более высокой размерностью путем дублирования каждого пикселя, что называется методом ближайшего соседа (nearest neighbor approach).

Обучение полносверточной нейронной сети проводилось на 260 изображениях (221 для обучения, 39 для тестирования) с 200 эпохами. Для проверки точности определения был использован метод «intersection over union», который сравнивает оригинальное изображение и предсказанное. При этом значение средней точности составило порядка 70 процентов, что является хорошим результатом для небольшого набора данных.

Для устранения шума синтезированного изображения использовался специализированный фильтр подавления шума вторичного освещения. Кроме того, чтобы не загромождать изображение ненужными деталями, было решено использовать алгоритмы для уменьшения шума «sobel» и «thresholding», которые подчеркивают градиенты яркости на краях изображений, оставляя тем самым только очертания интерьера. Это помогает значительно уменьшить вычислительную нагрузку, а также выделить внутренние границы.



Рис. 2. Алгоритмы фильтрации для уменьшения шума

На рисунке 2 представлены алгоритмы фильтрации, используемые совместно с нейронными сетями, а на рисунке 3 представлен пример тестового набора данных в виде «heatmap» маски оригинального изображения, который содержит различные виды освещения (темные и освещенные) для полноты обучения.

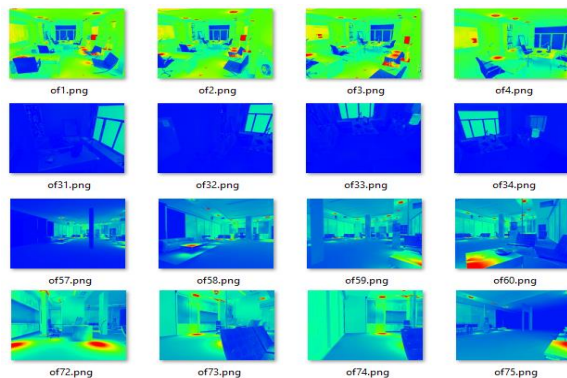


Рис. 3. Тестовый набор данных для обучения

Тестовые изображения и их маски создавались с помощью программного средства «Lumiscert» методом стохастической трассировки лучей. Каждое изображение получалось путем поворота и наклона камеры наблюдателя, тем самым охватывая и источники освещения, и отраженные участки для более полной выборки. Были выставлены границы от 0 до 550 люкс, что характеризует собой минимальные и самые освещенные участки помещения. Помимо этого, всего выделены 5 основных диапазонов освещения, так называемые «классы освещения», которые и были указаны как параметры обучения и нахождения на тестовых данных. На рисунке 4 представлен схематический рисунок предложенного метода.

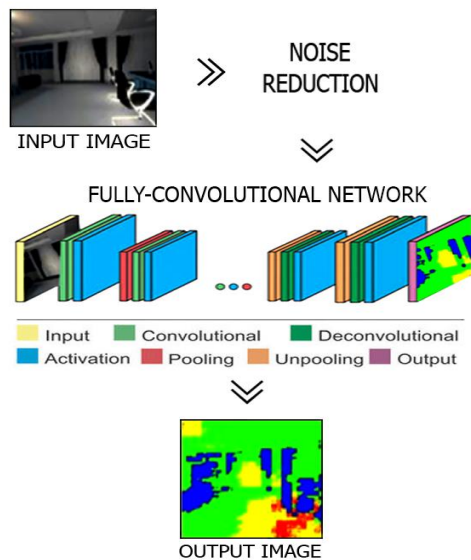


Рис.4. Схематический рисунок метода определения освещенных участков помещения

На рисунке 5 показан пример работы обученной полносверточной нейронной сети. Где на левом изображении исходное оригинальное изображение, правое изображение – эталонное (ground truth), а среднее - предсказанное нейросетью.

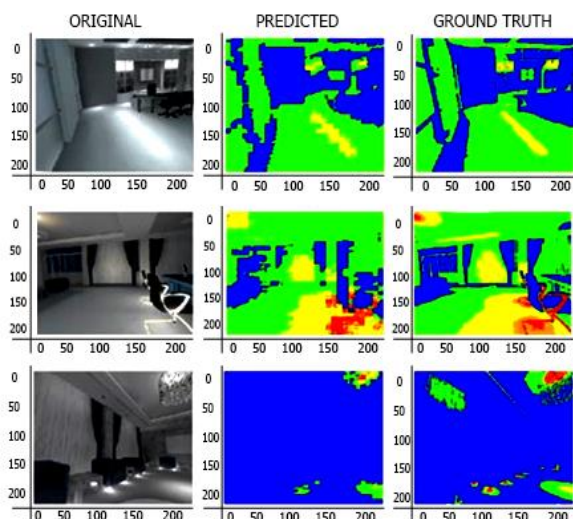


Рис. 5. Пример работы обученной полносверточной нейронной сети

Как можно видеть, результаты довольно близки к эталонным изображениям, но сеть по-прежнему не очень хорошо схватывает самые яркие участки. Это связано с обучающей выборкой, мы считаем, что увеличение обучающей выборки должно решить эту проблему. На данном этапе нейронная сеть работает достаточно хорошо, но есть ряд моментов, которые можно улучшить. Например, добавить алгоритм определения источников света [18] и разработать алгоритм определения диффузных поверхностей.

Полученные изображения будут использоваться непосредственно в задачах восстановления источников освещения и реальных оптических свойств для задач дополненной реальности. Данные изображения необходимы для разработки качественных методов и алгоритмов определения и классификации источников освещения и их отражений, с дальнейшим восстановлением реалистичного освещения для объектов виртуальной и дополненной реальности.

## Заключение

В рамках этой работы была обучена полносверточная нейронная сеть, состоящая из 5 блоков для классификации освещенных областей комнаты. В дополнение к обучению в сети также использовались алгоритмы компьютерного зрения, позволяющие фильтровать изображения, выделяя основные зоны помещения, что предполагает плавный переход к семантической составляющей комнаты. В дальнейшем планируется усовершенствовать семантическую составляющую нейронной сети в сочетании с алгоритмами фильтрации для определения освещенных областей и различий между ними.

## Благодарности

Работа выполнена при частичной государственной финансовой поддержке грантами РФФИ № 18-08-01484 и № 18-01-00569.

## Литература

[1] Hold-Geoffroy, Y., Sunkavalli, K., Hadap, S., Gambaretto, E. and Lalonde, J.-F., "Deep outdoor illumination estimation," In Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR) (2017).

[2] Lalonde, J.-F., Efros, A. A. and Narasimhan, S. G., "Estimating the natural illumination conditions from a single outdoor image," International Journal of Computer Vision, 98(2), 123–145 (2012).

[3] Gardner, M.-A., Sunkavalli, K., Yumer, E., Shen, X., Gambaretto, E., Gagné, C. and Lalonde, J.-F., "Learning to predict indoor illumination from a single image," ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia), preprints (2017).

[4] Lombardi, S. and Nishino, K., "Reflectance and Illumination Recovery in the Wild," IEEE Transactions on Pattern Analysis and Machine Intelligence 38, 129– 141 (2016).

[5] Eigen, D. and Fergus, R., "Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-Scale Convolutional Architecture," International Conference on Computer Vision (2015).

[6] Girshick, R. B., Donahue, J., Darrell, T. and Malik, J., "Rich feature hierarchies for accurate object detection and semantic segmentation," CVPR (2014).

[7] Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R. and LeCun, Y., "Overfeat: Integrated recognition, localization and detection using convolutional networks," ICLR (2013).

[8] Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition," CoRR, abs/1409.1556 (2014).

[9] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., "Going deeper with convolutions," CoRR, abs/1409.4842 (2014).

[10] "Lumicept | Integra Inc.," Integra Inc., 2019, <<https://integra.jp/en/products/lumicept>> (April 12, 2019).

[11] Д.Д. Жданов, С.В. Ершов, А.Г. Волобой. Метод подавления стохастического шума изображения, сгенерированного Монте-Карло трассировкой лучей, сохраняющий мелкие детали // Препринты ИПМ им. М.В. Келдыша. 2018. № 194. 15 с.

[12] С.В. Ершов, Д.Д. Жданов, А.Г. Волобой. Модификация стохастической трассировки лучей для снижения шума на диффузных поверхностях // Препринты ИПМ им. М.В. Келдыша. 2018. № 204. 17 с.

[13] Heymann, S., Smolic, A., Müller, K., Froehlich, B., "Illumination reconstruction from real-time video for interactive augmented reality," International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), (2005)

[14] Bruno Augusto Dorta, M., Rafael Rego, D., Cristina Nader, Vasconcelos., Esteban, C., "Deep light source estimation for mixed reality," VISIGRAPP 2018 - Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, (303-311), (2018)

[15] Salma, J., Philippe, R., Eric, M., "Illumination Estimation Using Cast Shadows for Realistic Augmented Reality Applications," Adjunct Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct (2017)

[16] Frahm, Jan-Michael., Koeser, K., Grest, D., Koch, R., "Markerless Augmented Reality with Light Source Estimation for Direct Illumination," European Conference on Visual Media Production, (211-220), (2005)

[17] Long, J., Shelhamer, E. and Darrell T., "Fully Convolutional Networks for Semantic Segmentation," The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431-3440 (2015).

[18] Xu-yang Wang; Dmitry D. Zhdanov; Igor S. Potemin; Ying Wang and Han Cheng. The efficient model to define a single light source position by use of high dynamic range image of 3D scene // Proc. SPIE 10020, Optoelectronic Imaging and Multimedia Technology IV, 100200I (October 31, 2016)

## Об авторах

Сорокин Максим Игоревич, аспирант кафедры визуализации (базовой) факультета программной инженерии и компьютерной техники Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики. E-mail [vergotten@gmail.com](mailto:vergotten@gmail.com).

Жданов Дмитрий Дмитриевич, к.ф.-м.н., заведующий кафедрой технологий визуализации Университета ИТМО. E-mail [ddzhdanov@mail.ru](mailto:ddzhdanov@mail.ru).

Жданов Андрей Дмитриевич, аспирант кафедры технологий визуализации Университета ИТМО. E-mail [adzhdanov@corp.ifmo.ru](mailto:adzhdanov@corp.ifmo.ru).

Потемин Игорь Станиславович, к.т.н., доцент кафедры технологий визуализации Университета ИТМО. E-mail [ipotemin@yandex.ru](mailto:ipotemin@yandex.ru).

Богданов Николай Николаевич, аспирант кафедры технологий визуализации Университета ИТМО. E-mail [pnbogdanov@corp.ifmo.ru](mailto:pnbogdanov@corp.ifmo.ru).

Барладян Борис Хаимович, к.т.н., старший научный сотрудник ИПМ им. М.В.Келдыша РАН. E-mail [bbarladian@gmail.com](mailto:bbarladian@gmail.com).