

Оригинальные статьи / Original articles

<https://doi.org/10.21869/2223-1560-2019-23-3-113-134>

Применение глубоких нейронных сетей в задаче получения карты глубины из двумерного изображения

Д.И. Михальченко¹ ✉, А.Г. Ивин¹, О.Ю. Сивченко¹, Е.А. Аксаментов¹

¹ Санкт-Петербургский институт информатики и автоматизации РАН, 14 линия, 39, г. Санкт-Петербург, 199178, Российская Федерация

✉ e-mail: tekatodsham@gmail.com

Резюме

Цель исследования. исследование подходов к генерации карт глубины для проверки и обучения глубоких нейронных сетей. Рассматривается проблема получения информации о расстоянии от камеры до объекта сцены по 2D-изображению при помощи глубоких нейронных сетей без использования стереокамеры.

Методы. Генерация 3D-сцен для обучения и оценки нейронной сети осуществлялась при помощи приложения 3D-компьютерной графики Blender. Для оценки точности обучения было использовано среднеквадратическое отклонение (СКО). Машинное обучение было реализовано при помощи библиотеки Keras, а оптимизация – с использованием подхода AdaGrad.

Результаты. Представлена архитектура глубокой нейронной сети, которая на вход получает три последовательности 2D-изображений из видеопотока 3D-сцены и выдает на выходе предсказанную карту глубины для рассматриваемой 3D-сцены. Описан способ создания обучающих наборов данных, содержащих информацию о глубине карты с использованием программного обеспечения Blender. Рассматривается проблема переобучения, заключающаяся в следующем: созданные модели работают на специально сгенерированных наборах данных, но все еще не могут предсказать правильную карту глубины для случайных изображений. Представлены результаты тестирования актуальных способов создания карт глубины с использованием глубоких нейронных сетей.

Заключение. Основной проблемой предложенного метода является переобучение, которое может быть выражено в прогнозировании некоего среднего значения для разных изображений или предсказании одного и того же выхода для разных входов. Для решения данной проблемы могут быть использованы уже обученные сети или обучающие и вариационные выборки, содержащие 2D-изображения различных сцен.

Ключевые слова: компьютерное зрение; карты глубины; глубокое обучение; глубокие нейронные сети, цифровая обработка изображений; распознавание образов; нейронные сети; трехмерное очувствление.

Конфликт интересов: Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Для цитирования: Применение глубоких нейронных сетей в задаче получения карты глубины из двумерного изображения / Д.И. Михальченко, А.Г. Ивин, О.Ю. Сивченко, Е.А. Аксаментов // Известия Юго-Западного государственного университета. 2019; 23(3): 113-134. <https://doi.org/10.21869/2223-1560-2019-23-3-113-134>.

Статья поступила в редакцию 30.04.2019

Статья подписана в печать 20.05.2019

© Михальченко Д. И., Ивин А.Г., Сивченко О.Ю., Аксаментов Е.А., 2019

Application of Deep Neural Networks in the Problem of Obtaining Depth Maps from Two-Dimensional Images

Daniil I. Mikhhalchenko ¹ ✉, Arsenii G. Ivin ¹, Oleg Yu. Sivchenko ¹,
Egor A. Aksamentov ¹

¹ St Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, 39, 14 Line, St Petersburg, 199178, Russian Federation

✉ e-mail: tekatodsham@gmail.com

Abstract

Purpose of research is to study approaches to the depth map generation for deep neural networks testing and learning. The problem of obtaining information about the distance from the camera to the scenery object using a 2D image by means of deep neural networks without applying a stereocamera is considered.

Methods. Generation of 3D scenery for training and assessment of the neural network was carried out using the 3D-computer graphics application Blender. The standard deviation (RMS) was used to estimate the accuracy of learning. Machine learning was implemented using the Keras library and optimization was implemented using the AdaGrad approach.

Results. The architecture of a deep neural network which receives three sequences of 2D images from the 3D scenery video stream in the input and outputs the predicted depth map for the considered 3D scenery, is provided. The method for creating training data sets containing information about the depth of the map using Blender software is described. The problem of overtraining involving the fact that the created models work using specially generated data sets but still can not predict the correct depth map for random images is studied. The results of the testing actual methods for depth maps creation using deep neural networks are presented.

Conclusion. The main problem of the proposed method is overtraining which can be expressed in predicting a certain average value for different images or predicting the same output for different inputs. To solve this problem, we can use already trained networks or training and variation samples containing 2D images of different sceneries.

Keywords: computer vision; depth map; deep learning; deep neural networks; digital image processing; image recognition; neural networks; 3-D sensing.

Conflict of interest: The Authors declare the absence of obvious and potential conflicts of interest related to the publication of this article.

For citation: Mikhhalchenko D.I., Ivin A.G., Sivchenko O.Yu., Aksamentov E.A. Application of Deep Neural Networks in the Problem of Obtaining Depth Maps from Two-Dimensional Images. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta* = *Proceedings of the Southwest State University*. 2019, 23(3): 113-134 (In Russ.). <https://doi.org/10.21869/2223-1560-2019-23-3-113-134>.

Received 30.04.2019

Accepted 20.05.2019

Введение

Карты глубины содержат информацию о местоположении окружающих объектов и могут использоваться для получения информации о форме этих

объектов. Они находят широкое применение в робототехнике и киберфизических системах [1–3], к примеру, для виртуального измерения расстояний [4],

создания различных 3D-представлений [5], оценки поз на изображениях [6], семантической сегментации при аэрофотосъемке [7] и т.д. Карты глубины часто используются для решения задач навигации робота [8 - 10] и определения цветовых характеристик объекта [11] и распознавания ракурсов на изображении [12].

Традиционным способом получения карт глубины является использование стереокамеры. В настоящее время прикладные методы построения карт глубины на основе стереоизображений активно разрабатываются на основе технологий, ранее опробованных на платформе Microsoft Kinect [13].

Кроме того, подобные задачи могут решаться, например, с применением метода зеркального разделения изображений [11]. Несмотря на преимущества данного подхода, для его реализации требуется специфическое аппаратное и программное обеспечение, что существенно сужает область его применения.

Кроме того, задачи стереозрения успешно решаются при помощи некоторых лазерных сенсоров, которые, однако, довольно дороги. В работе [14] исследовано построение стереоизображений при эксплуатации мобильного робота с применением фреймворка ROS и реализацией на языке Python.

Анализ работ, связанных с данной темой, показывает, что существует несколько основных подходов к извлечению карты глубины из одного 2D-изображения. Ряд инновационных решений такого рода связан с использова-

нием глубоких нейронных сетей (ГНС). Некоторые подобные исследования перечислены и кратко рассмотрены ниже. Общее состояние работ по оценке глубины одного изображения с применением нейронных сетей по состоянию на 2018 год представлено в работе [15]. Авторы использовали в своих предложениях параметры для сравнения качества работы имеющихся методов по оценке глубины 2-D изображения, основанных на разных нейронных сетях. Метод Eigen и Fergus [16] работает на основе сетей AlexNET и сети, разработанной в проекте «Группа визуальной геометрии» (VGG), Liu и др. [18] на глубоких сверточных нейронных сетях (DCNN), Laina и др. [17] на CoRR, Li и др. [19] на VGG и ResNET. Перечисленные нейронные сети представляют собой новейшие архитектуры ГНС, применяемые в анализе и оценке карт глубины.

В настоящее время ГНС используются в различных областях компьютерного зрения: детектирование объектов на изображениях различного масштаба [20], распознавание зашумленных изображений в резервуарных вычислительных сетях (RCN) [21], классификация изображений [22], сжатие [23], реконструкция [24, 25] и др.

Создание ГНС для получения карты глубины из 2D-изображения без использования методов стереозрения является актуальной задачей, поскольку в условиях постоянного нарастания объемов анализируемых данных требуются

более эффективные методы распознавания образов, в частности, их классификации и кластеризации. В настоящее время такие задачи все чаще решаются на практике при помощи ГНС и специализированных библиотек – в частности, OpenCV и Keras. Подобная тенденция четко прослеживается в актуальных работах [17 - 19], но другие решения в данной предметной области также сохраняются. Накопление все новых множеств данных, используемых для практической реализации алгоритмов компьютерного зрения, способствует распространению методов глубокого обучения при решении таких задач, как построение карт глубины. Целью данной работы является разработка способа создания моделей для генерации 3D-сцен, который позволил бы решить проблемы переобучения и отсутствия универсального набора данных для обучения ГНС. В следующем разделе представлен обзор современных методов построения карт глубины по одному двумерному изображению. В разделе 2 рассматриваются возможные пути решения сопутствующей задачи – самостоятельной подготовки подходящей выборки для обучения ГНС. Актуальность рассматриваемой задачи обусловлена дефицитом и несовершенством подобных датасетов, имеющихся сегодня в наличии в свободном доступе. В разделе 3 обсуждаются возможные пути решения проблемы переобучения ГНС.

Материалы и методы решения задачи

Существует несколько работ, описывающих оценку карты глубины по

одному изображению. Например, метод, представленный в [16] и связанный с интеграцией локальной и глобальной информации со стереоизображений, хорошо работает на материале баз данных Нью-Йоркского университета [26]. Авторы предлагают использовать многоуровневую глубокую сеть для получения карты глубины из одного изображения. В этой работе сеть имеет два входа для одного и того же изображения. Один из входов используется для так называемого процесса переработки, который позволяет получить карту глубины с четкими границами объектов. Эта работа также предполагает использование масштабно-инвариантной ошибки для обучения. В описанной сети используются только стандартные сверточные и полносвязные слои.

В [27] для распознавания стереоизображений предлагается использовать сверточную нейронную сеть (CNN), уделяя особое внимание слиянию признаков, полученных при разных масштабах, для чего спроектирована улучшенная архитектура, состоящая из четырех модулей: энкодера, декодера, модуля слияния многомасштабных признаков и модуля переработки.

В [28] была предложена модель для оценки глубины, состоящая из двух частей: одна часть представляет собой сеть, предварительно подготовленную Оксфордской группой визуальной геометрии (VGG) [29], причем, эта сеть первоначально была обучена классифицировать объекты. Обучив модель, ав-

торы добавили два полносвязных слоя собственной разработки, чтобы проверить оценку глубины в режиме реального времени. Заявленное значение СКО составило 0,833.

При этом, в работе [30] авторы уделяют особое внимание неустранимым неоднозначностям, возникающим при восстановлении 3D-изображений из 2D-изображения. Такие неоднозначности иллюстрируются при сравнении методов обучения с учетом фокусного расстояния и без него, подтверждается критическое влияние фокусного расстояния на оценку глубины изображения. Предложен метод генерации набора данных с переменным фокусным расстоянием из набора данных с фиксированным фокусным расстоянием. Таким образом глубокая нейронная сеть фактически объединяет информацию из двух этих наборов данных, заполняя лакуны в сгенерированных изображениях и превосходит по определению глубины лучшие результаты, достигнутые на материале предобученной VGG.

Авторы [15] также делают акцент на использовании сверточных нейронных сетей и предлагают самостоятельно разработанную систему критериев оценки признаков, разработанную на материале множества данных RGB-D. В частности, авторы уделяют внимание сохранению граней, плоских поверхностей, согласованности глубин (depth consistency) и определению абсолютной точности расстояний.

В [31] представлен метод оценки глубины одного изображения. Проблема оценки глубины формулируется как задача дискретно-непрерывной оптимизации, где непрерывные переменные кодируют глубину суперпикселей во входном изображении, а дискретные переменные представляют собой отношения между соседними суперпикселями. Эта проблема определена как условное случайное поле. Авторы используют многочастичный алгоритм распространения доверия для вывода графической модели. В статье представлены результаты значения СКО 1,06-1,08. В более позднем исследовании достигнутое значение СКО этого метода составило 0,824 [17].

В [32] авторы также описывают оценку глубины с использованием методов глубокого обучения. Однако авторы оценивают глубину с помощью стереовхода, т.е. для создания карты глубины они используют стереоизображение. Кроме того, в статье показан хороший пример использования ГНС для создания алгоритма быстрой оценки глубины.

В [33] многомасштабная глубокая сверточная сеть используется для выполнения трех задач: прогнозирование глубины, оценка поверхности и семантическая маркировка. Предложенная модель изначально основана на архитектуре, представленной в [34], но с несколькими улучшениями: больше сверточных слоев, более высокое разрешение третьего слоя, прохождение многоканальных карт признаков вместо пере-

дачи выходных прогнозов из первого во второй слой. СКО предсказания глубины этой модели достигает 0,641.

В более старых работах, например в [35], авторы используют Марковскую сеть и обучение с учителем для создания карт глубины по одному изображению. Представленная модель не делает явных предположений о структуре изображения и способна захватывать гораздо более подробную трехмерную структуру. В результате авторы создают качественно правильные 3D-модели для 64,9% из 588 изображений.

Результаты и их обсуждение

Для создания валидной базы данных карты глубины требуются дорогие и не всегда доступные датчики, например лазерный дальномер. Кроме того, в открытом доступе можно найти только несколько существующих серий изображений с картами глубины. Для решения этой проблемы предлагается создание искусственных изображений с соответствующими картами глубины.

Созданный набор данных для обучения и проверки ГНС включает три набора изображений в градациях серого размером 160×120 . Входной слой принимает простые изображения в формате PNG, а выходной слой должен предоставлять карту глубины в формате OpenEXR [36]. Этот формат был выбран для изображений карты глубины, поскольку он позволяет хранить ненормализованные данные. Первые два набора данных для обучения были сгенериро-

ваны с применением 3D-компьютерной графики приложения Blender [37, 38]. После визуализации кадра сцены изображения и данные из канала глубины автоматически сохраняются в отдельных файлах.

Первый набор данных содержит только изображения лестницы в квадратной комнате с двумя источниками света. Второй набор был расширен другой сценой, включающей пять объектов разного масштаба, имеющих форму бутылочной крышки, и тремя объектами, имеющими сложную цилиндрическую форму с «ржавой» текстурой. На рисунке 1 показана вторая сцена, а также траектория камеры, фиксирующей сцену с разных позиций. Эти два обучающих набора содержат 500 изображений, сделанных с разных точек обзора. Валидационный набор данных был сформирован на тех же сценах, но с другой траекторией камеры.

Третий набор данных для обучения содержит 500 изображений реального мира. Существует два набора данных изображений и соответствующих им карт глубины в Make3D [35,39,40]. Эти наборы данных были сделаны с помощью лазерного сканера и камеры. Для хранения и обработки наборов данных в файлах HDF5 был написан инструмент на Python (рис. 2 и 3). В предлагаемых моделях ГНС используются сверточные и полносвязные слои. Инициализация Glorot [41] используется в сверточных слоях. Для инициализации в полностью подключенных слоях ис-

пользуется блок линейной ректификации (ReLU, rectified linear unit), как это было предложено в [42], а также Инициализация, предложенная в [43].

В качестве метрики для оценки точности обучения используется показатель СКО. Библиотека Keras [44] применяется для реализации машинного обучения, поскольку она проста в использовании, гибка и позволяет использовать различные фреймворки, к примеру, TensorFlow или Theano [45–47]. В нашей работе использовались машинные интерфейсы TensorFlow. Для оптимизации используется подход AdaGrad [48].

На рисунке 4 показан один из типов обученной модели. Этот рисунок демонстрирует принцип работы, но не конкретное количество слоев, поскольку

ку в эксперименте было разное количество сверточных и полносвязных слоев. На рисунке показано, что для изображений есть три входа. Эти изображения представляют собой короткий поток кадров, захваченных во время движения камеры.

Пока камера движется, объекты перемещаются с разной скоростью в зависимости от расстояния от камеры, что сравнимо с эффектом параллакса. Мы исходили из того, что сеть сможет научиться использовать этот эффект для создания карт глубины. После этого каждый из сверточных и плотных слоев обрабатывает каждое изображение, после чего они связываются и обрабатываются еще раз несколькими сверточными и полносвязными слоями.

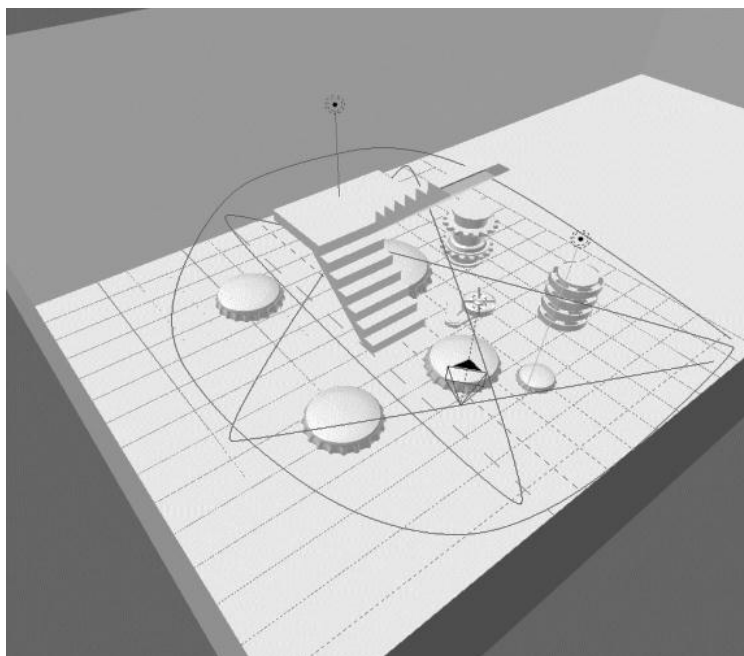


Рис. 1. Траектория камеры на второй моделируемой сцене

Fig. 1. Camera tracking in the second simulated scenery

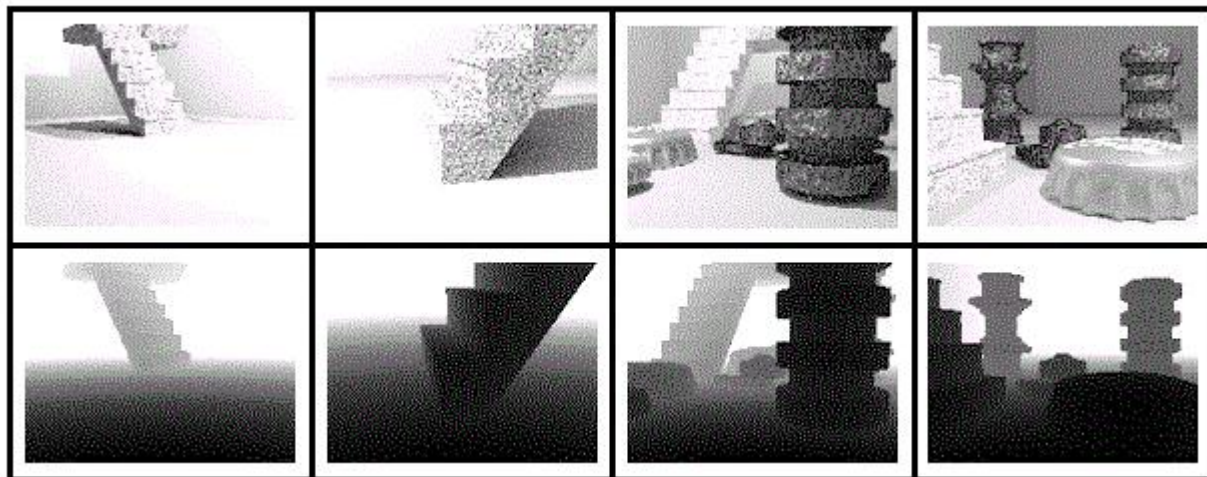


Рис. 2. Сгенерированный набор данных для обучения

Fig. 2. Generated training data set

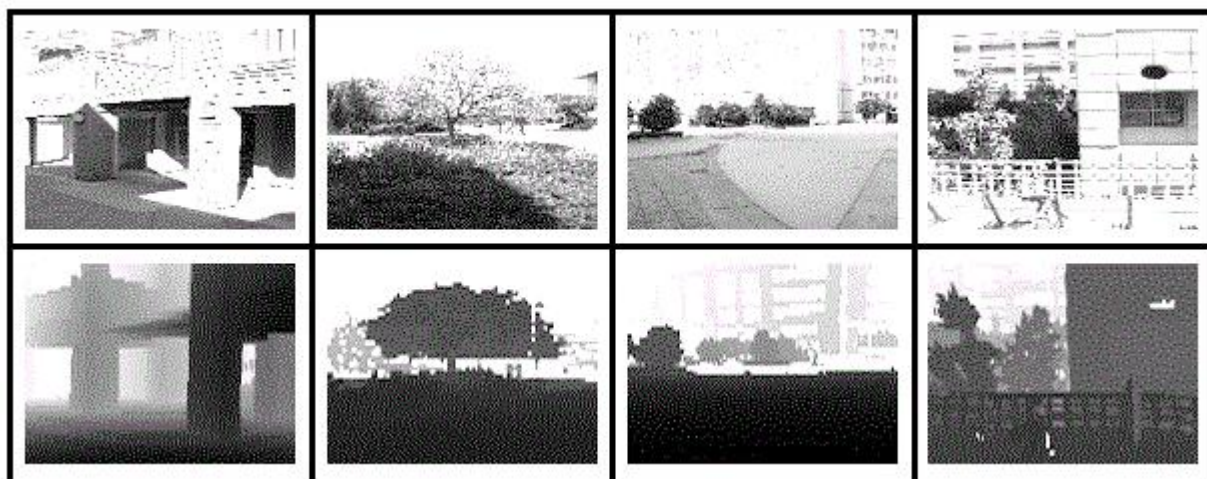


Рис. 3. Набор данных для обучения в формате Make3D

Fig. 3. Training data set in Make3D

В каждом коротком потоке кадров вывод должен давать карту глубины, соответствующую последнему изображению. Для уменьшения вероятности переобучения использовались прореживающие слои (dropout layers). Это известный метод решения проблемы переобучения [49]. Пример входных потоков кадров показан на рисунке 5, а. Соответствующие им карты глубины на

выходе представлены на рис 5, б. Как показано на рисунке 4, существует множество других слоев для изменения формы и сглаживания. Эти слои используются только для правильного подключения выходов и входов разных слоев. Графики моделей созданы с использованием встроенных инструментов Keras.

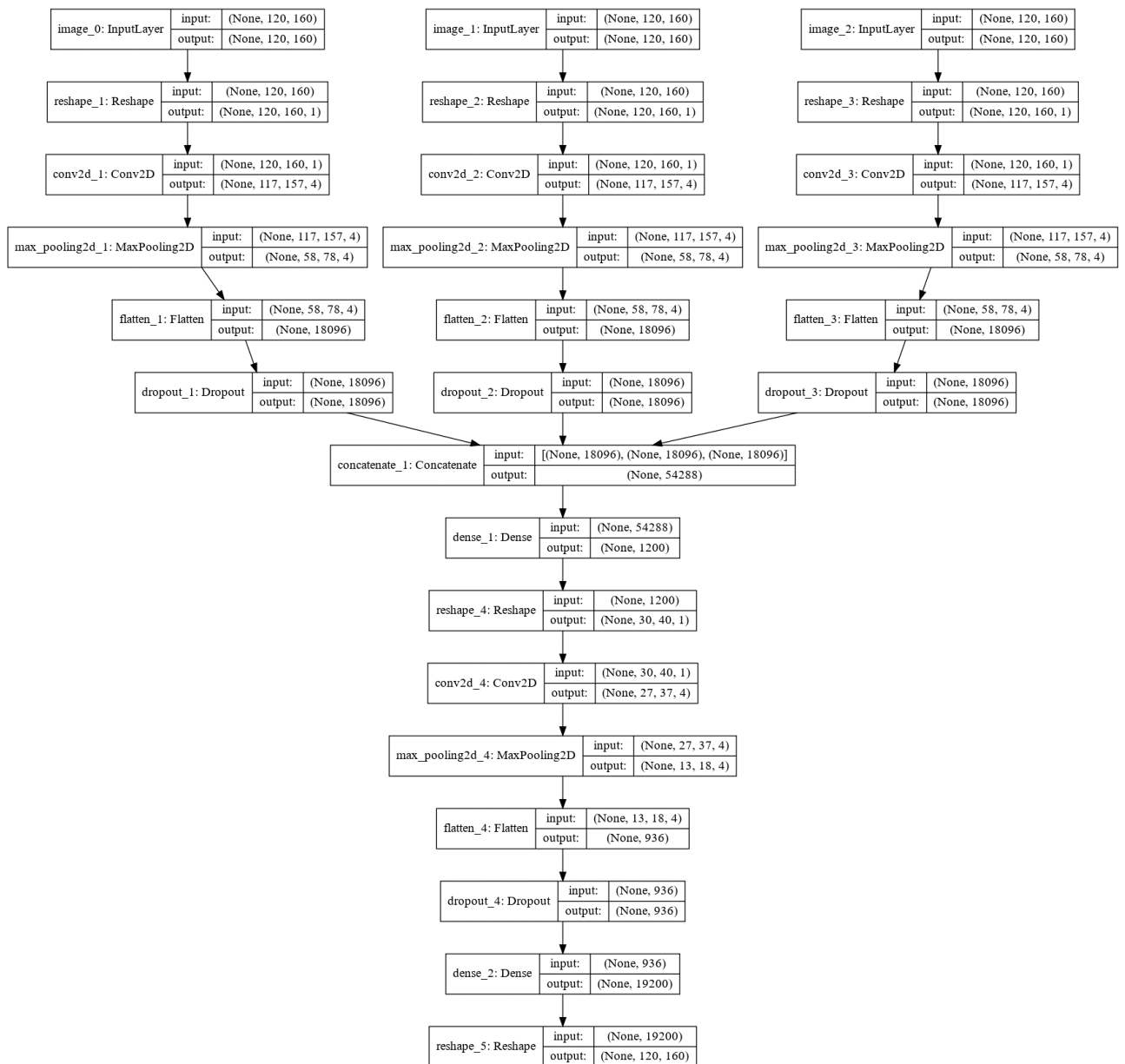


Рис. 4. Тип модели, принимающей три входа

Fig. 4. Three-input model type

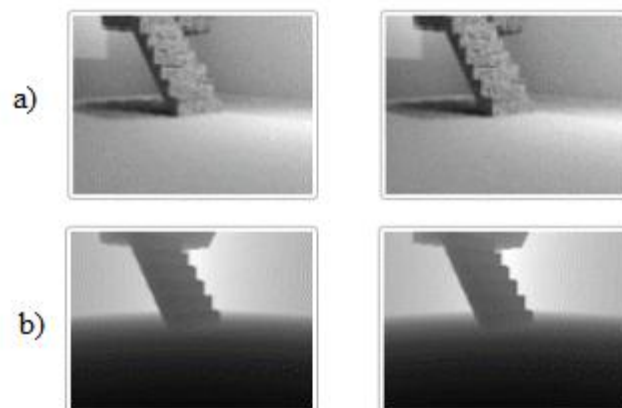


Рис. 5. Пример двух мини-поток в наборе данных

Fig. 5. Examples of two mini streams in a data set

Основная проблема предложенного подхода – переобучение. Типичная ситуация переобучения моделей представлена на рисунке 6. Пунктирной линией на графике обозначены потери значений, рассчитанные для валидационного набора данных. Из графика видно, что модель быстро начинает прогнозировать некое среднее значение и таким образом сохраняет его. Варьируя различные параметры модели, можно

добиться более низких значений потерь, однако это не решает проблему переобучения, поскольку ГНС все еще учится прогнозировать усредненное значение для разных изображений. Кроме того, существует еще одна проблема переобучения: модель всегда предсказывает один и тот же выход для каждого входа - обычно это первое изображение в наборе данных (рис. 7–8, а).

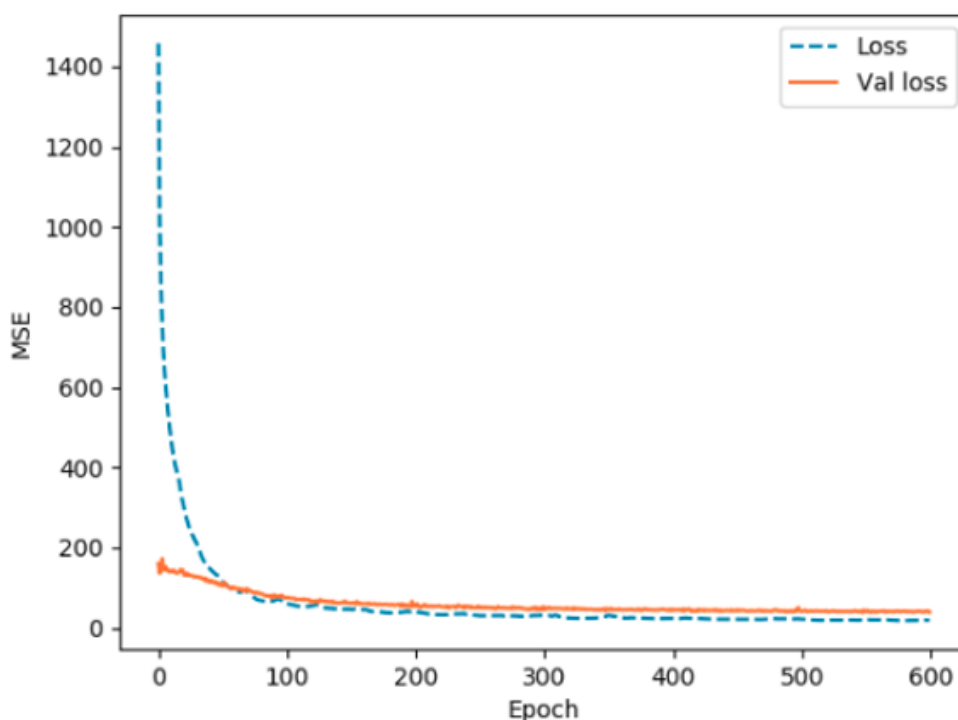


Рис. 6. Пример переобучения модели

Fig. 6. Example of model overtraining



Рис. 7. Результат эксперимента для модели с одним входом: (а) выход модели, (b)-(d) входные мини-потoki кадров, (е) ожидаемый результат

Fig. 7. The outcome of the experiment for the one input model: (a) model output, (b)-(d) input frame mini streams, (e) the expected result

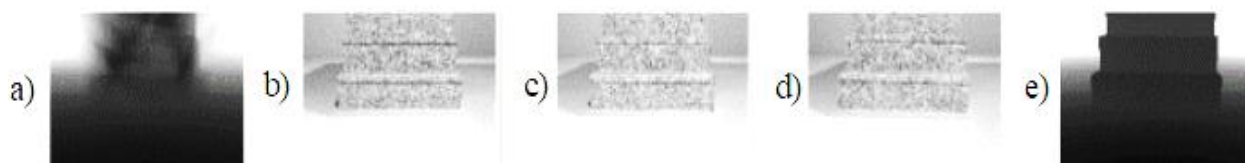


Рис. 8. Результат эксперимента для модели с тремя входами: (a) выход модели, (b)-(d) входные мини-потoki кадров, (e) ожидаемый результат

Fig. 8. The outcome of the experiment for three input model: (a) model output, (b)-(d) input frame mini streams, (e) the expected result

При размере пакета, равного 1, было достигнуто значение потерь 0,5 (рис. 9) и 33,77 на валидационной выборке (рис. 10). Пример работы модели на тестовой выборке показан на рисунке 3.

Несмотря на то, что визуализация такой карты легко распознаваема человеческим глазом, она не дает правильной оценки глубины (см. рис. 10).


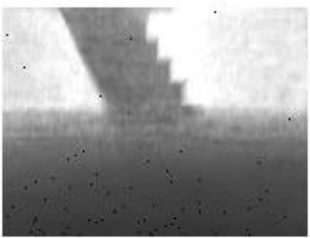
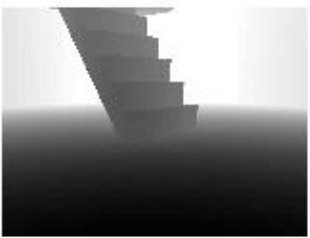






Изображение	Предсказание модели	Контрольные данные
		
		
		

Рис. 9. Результат работы модели на тестовой выборке

Fig. 9. Model work output for test sample


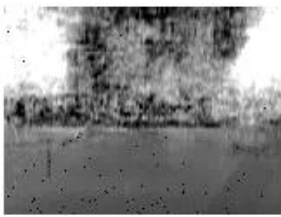

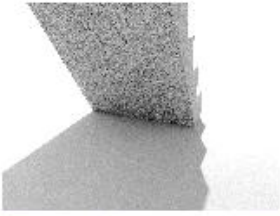
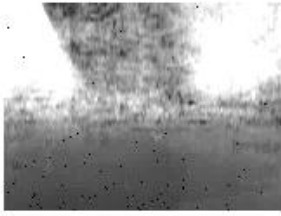
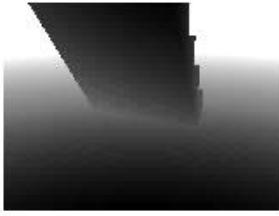

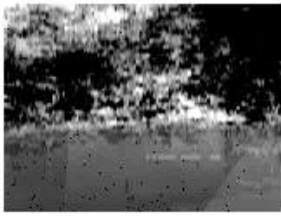

Изображение	Предсказание модели	Контрольные данные
		
		
		

Рис. 10. Результат работы модели на валидационной выборке

Fig. 10. Model work output for validation sample

Выводы

Основные проблемы построения карты глубины из одного изображения связаны с отсутствием универсального набора данных для обучения ГНС и с переобучением моделей. В данной работе был предложен способ создания моделей на основе специального программного обеспечения для генерации 3D-сцен, например, при помощи Blender: данный способ решает первую проблему, однако не устраняет проблему переобучения. В частности, из предсказанных изображений на рисунке 10 видно, что модель учится прогнозировать местоположение горизонта в одном и том же месте для каждого входа,

что, вероятно, связано с особенностями генерируемых наборов данных. При использовании модели с тремя входами переполнение может происходить из-за небольшого смещения объектов во входных мини-потоках, что дает практически одинаковые изображения. С другой стороны, использование обучающих и валидационных выборок, содержащих изображения различных сцен с различными объектами и видами, позволит создать наборы данных с более выраженным смещением между одними и теми же объектами на изображениях во входных мини-потоках для обучения модели с несколькими входами. Другое возможное решение – использовать уже обученную сеть, как это описано в [17].

Список литературы

1. Levonevskiy D., Vatamaniuk I., Saveliev A. Integration of corporate electronic services into a smart space using temporal logic of actions // International Conference on Interactive Collaborative Robotics. Springer, Cham. 2017. 10459. P. 134-143. DOI: 10.1007/978-3-319-66471-2_15.
2. Conceptual model of cyberphysical environment based on collaborative work of distributed means and mobile robots / A. Ronzhin, A. Saveliev, O. Basov, S. Solyonyj // International Conference on Interactive Collaborative Robotics. Springer, Cham. 2016. Vol. 9812. P. 32-39. DOI: 10.1007/978-3-319-43955-6_5.
3. Scenarios of multimodal information navigation services for users in cyberphysical environment / I. Vatamaniuk, D. Levonevskiy, A. Saveliev, A. Denisov // International Conference on Speech and Computer. Springer, Cham. 2016. Vol. 9811. P. 588-595. DOI: 10.1007/978-3-319-43958-7_71.
4. Richards H.W. Method and apparatus for user interaction for virtual measurement using a depth camera system // U.S. Patent Application No 20170302908. 2017. T. 15/132. C. 822.
5. An efficient hole-filling method based on depth map in 3D view generation / H. Liang, X. Su, Y. Liu, H. Xu, Y. Wang, X. Chen // 2017 International Conference on Optical Instruments and Technology: Optoelectronic Imaging/Spectroscopy and Signal Processing Technology. 2018. Vol. 10620, DOI: 10.1117/12.2293301.
6. Girshick R. Efficient regression of general-activity human poses from depth images //2011 International Conference on Computer Vision. IEEE, 2011. P. 415-422.
7. Sun Y. Developing a multi-filter convolutional neural network for semantic segmentation using high-resolution aerial imagery and LiDAR data //ISPRS journal of photogrammetry and remote sensing. 2018. Vol. 143. P. 3-14.
8. Watts K.W. , Konolige K. Ground plane detection to verify depth sensor status for robot navigation // US Patent No. 9886035. 2018.
9. Tee Kit Tsun M., Lau B.T., Jo H. Siswoyo An improved indoor robot human-following navigation model using depth camera // Active IR marker and proximity sensors fusion, Robotics. 2018. Vol. 7. № 1. P. 4. DOI: 0.3390/robotics7010004.
10. Горобцов А.С. Особенности решения уравнений метода обратной задачи для синтеза устойчивого управляемого движения шагающих роботов // Труды СПИИРАН. 2019. Т. 18, №1. С. 85-122. DOI: 10.15622/sp.18.1.85-122.

11. Алтухов В.Г. Вычисление расстояния до объекта на основе карты глубин полученной методом зеркального разделения изображений // Автоматика и программная инженерия. 2017. Т. 1. С. 65-69.
12. Ерофеев М.В. Многослойное решение проблемы полупрозрачных границ при построении стереоскопических изображений // International Journal of Open Information Technologies. 2016. Т. 4. № 8.
13. Lin K.Y., Hang H.M. Depth Map Enhancement on RGB-D Video Captured by Kinect V2 // 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). IEEE, 2018. С. 1530-1535.
14. Ульянов С.В. Разработка системы стереозрения для мобильного робота // Программные продукты и системы. 2017. Т. 30. № 3.
15. Koch T. Evaluation of CNN-based single-image depth estimation methods // Proceedings of the European Conference on Computer Vision (ECCV). 2018. С. 0-0.
16. Eigen D., Puhrsch C., Fergus R. Depth map prediction from a single image using a multiscale deep network // Advances in Neural Information Processing Systems. 2014. С. 2366-2374. arXiv: 1406.2283v1.
17. Liu F., Shen C., Lin G. Deep convolutional neural fields for depth estimation from a single image // The IEEE Conference on Computer Vision and Pattern Recognition. 2015. С. 5162-5170. DOI: 10.1109/ CVPR.2015.7299152.
18. Laina I., Rupprecht C., Belagiannis V., Tombari F., Navab N. Deeper depth prediction with fully convolutional residual networks. CoRR, abs/1606.00373. 2016. DOI: 10.1109/3dv.2016.32. arXiv: 1606.00373.
19. Li J., Klein R., Yao A. A two-streamed network for estimating fine-scaled depth maps from single rgb images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017. p. 3372–3380.
20. Применение нечётких нейронных сетей для определения типа кристаллических решёток, наблюдаемых на наномасштабных изображениях / О.П. Солдатова, И.А. Лёзин, И.В. Лёзина, А.В. Куприянов, Д.В. Кирш // Компьютерная оптика. 2015. Т. 39, № 5. С. 787-795. DOI: 10.18287/0134-2452-2015-39-5-787-794.
21. On the application of reservoir computing networks for noisy image recognition / A. Jalalvand, K. Demuynck, W.D. Neve, J.-P. Martensa // Neurocomputing. 2018. Т. 277. С. 237-248. DOI: 10.1016/j.neucom.2016.11.100.
22. Dutta S. Classification of diabetic retinopathy images by using deep learning models / S. Dutta, B. CS Manideep, S.M. Basha, R.D. Caytiles, N.Ch.S.N. Iyengar // International

Journal of Grid and Distributed Computing. 2018. Т. 11. № 1. С. 89-106. DOI: 10.14257/ijgdc.2018.11.1.09.

23. Сирота А.А., Дрюченко М.А. Обобщённые алгоритмы сжатия изображений на фрагментах произвольной формы и их реализация с использованием искусственных нейронных сетей // Компьютерная оптика. 2015. Т. 39, № 5. С. 751-761. DOI: 10.18287/0134-2452-2015-39-5-751-761.

24. Реконструкция изображений в дифракционно-оптических системах на основе сверточных нейронных сетей и обратной свертки / А.В. Никоноров, М.В. Петров, С.А. Бибииков, В.В. Кутикова, А.А. Морозов, Н.Л. Казанский // Компьютерная оптика. 2017. Т. 41, № 6. С. 875-887. DOI: 10.18287/2412-6179-2017-41-6-875-887.

25. Олейник А.Л. Алгоритмы взаимной реконструкции изображений лиц на основе методов проекции в собственные подпространства // Труды СПИИРАН. 2018. Т. 2, №57. С. 45-74. DOI: 10.15622/sp.57.3

26. Silberman N., Fergus R. Indoor scene segmentation using a structured light sensor // ICCV Workshop on 3D Representation and Recognition. 2011. P. 601-608.

27. Hu J. et al. Revisiting single image depth estimation: toward higher resolution maps with accurate object boundaries //2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2019. P. 1043-1051.

28. Zhu J. Ma R. (2016), Real-time depth estimation from 2D images. Available at: http://cs231n.stanford.edu/reports/2016/pdfs/407_Report.pdf (accessed May 1, 2018).

29. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition // 3rd International Conference on Learning Representations (ICLR). 2015. P. 1-14. arXiv: 1409.1556.

30. He L., Wang G., Hu Z. Learning depth from single images with deep neural network embedding focal length // IEEE Transactions on Image Processing. 2018. Vol. 27. № 9. P. 4676-4689.

31. Liu M., Salzmann M., He X. Discrete-continuous depth estimation from a single image // The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2014. С. 716-723. DOI: 10.1109 / CVPR.2014.97.

32. Luo W., Schwing A.G., Urtasun R. Efficient deep learning for stereo matching // The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 5695-5703. DOI: 10.1109/CVPR.2016.614.

33. Eigen D., Fergus R. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture // The IEEE International Conference on Computer Vision (ICCV). 2015. P. 2650-2658. DOI: 10.1109 / ICCV.2015.304.

34. Vision meets robotics: the KITTI dataset / A. Geiger, A. Lenz, C. Stiller, R. Urtasun // The International Journal of Robotics Research. 2013. Vol. 32. № 11. P. 1231-1237. DOI: 10.1177/0278364913491297.
35. Saxena A., Sun M., Ng A.Y. Make3d: learning 3D scene structure from a single still image // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2009. Vol. 31. № 5. P. 824-840. DOI: 10.1109 / TPAMI.2008.132.
36. Bogart R., Kainz F., Hess D. OpenEXR image file format / // ACM SIGGRAPH. 2003. Sketches & Applications.
37. Kent B.R. 3D Scientific Visualization with Blender®, Morgan & Claypool, San Rafael, 2015.
38. Valenza E. Blender 2.6 Cycles: Materials and Textures Cookbook – Third Edition. Packt Publishing Ltd. Birmingham. Mumbai. 2013. 280 c.
39. Saxena A. Learning depth from single monocular images / A. Saxena, S.H. Chung, Y. Ng. Andrew // Neural Information Processing Systems (NIPS). 2005. P. 1161-1168. DOI: 10.1109/TPAMI.2015.2505283.
40. Saxena A., Chung S.H., Andrew Y. Ng. 3D depth reconstruction from a single still image // International Journal of Computer Vision. 2008. Vol. 76. № 1. P. 53-69. DOI: 10.1109 / TPAMI.2008.132.
41. Glorot X., Bengio Y. Understanding the difficulty of training deep feedforward neural networks // Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. 2010. Vol. 9. P. 249-256.
42. Heaton J. Artificial Intelligence For Humans: Deep Learning and Neural Networks / J. Heaton Heaton Research. Inc., St Louis, MO, 2015. Vol. 3. 375 p.
43. Delving deep into rectifiers: surpassing human-level performance on imagenet classification / K. He, X. Zhang, S. Ren, J. Sun // Proceedings of the IEEE international conference on computer vision. 2015. P. 1026-1034. DOI: 10.1109 / ICCV.2015.123
44. Keras: The Python Deep Learning library. Available at: <https://keras.io/> (accessed 31.08.2018)
45. Backends – TensorFlow or Theano. Available at: <https://www.tensorflow.org/> (accessed 31.08.2018)
46. TensorFlow: a system for large-scale machine learning / M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D.G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, G. Brain // the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16). Nov. 2-4. 2016. P. 265-283.

47. Tensorflow: large-scale machine learning on heterogeneous distributed systems / M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng // (2016b). Available at: <https://arxiv.org/pdf/1603.04467.pdf> (accessed May 1, 2018).
48. Duchi J., Hazan E., Singer Y. Adaptive subgradient methods for online learning and stochastic optimization // *Journal of Machine Learning Research*. July. 2011. Vol. 12. P. 2121-2159.
49. Reduction of overfitting in diabetes prediction using deep learning neural network / A. Ashiquzzaman, A.K. Tushar, Md. R. Islam, D. Shon, K. Im, J.-H. Park, D.-S. Lim, J. Kim // *2017 IT Convergence and Security*. Springer, Singapore. 2018. Vol. 449. P. 35-43. DOI: 10.1007/978-981-10-6451-7_5.

Reference

1. Levonevskiy D., Vatamaniuk I., Saveliev A. Integration of corporate electronic services into a smart space using temporal logic of actions. *International Conference on Interactive Collaborative Robotics*, Springer, Cham 2017; 10459, pp. 134-143. DOI: 10.1007/978-3-319-66471-2_15.
2. Ronzhin A., Saveliev A., Basov O., Solyonyj S. Conceptual model of cyberphysical environment based on collaborative work of distributed means and mobile robots. *International Conference on Interactive Collaborative Robotics*. Springer, Cham 2016; 9812, pp. 32-39. DOI: 10.1007/978-3-319-43955-6_5.
3. Vatamaniuk I., Levonevskiy D., Saveliev A., Denisov A. Scenarios of multimodal information navigation services for users in cyberphysical environment. *International Conference on Speech and Computer*. Springer, Cham 2016; 9811, pp. 588-595. DOI: 10.1007/978-3-319-43958-7_71.
4. Richards HW. Method and apparatus for user interaction for virtual measurement using a depth camera system. U.S. Patent Application no. 20170302908, 2017, no. 15(132), p. 822.
5. Liang H., Su X., Liu Y., Xu H., Wang Y., Chen X. An efficient hole-filling method based on depth map in 3D view generation. *2017 International Conference on Optical Instruments and Technology: Optoelectronic Imaging/Spectroscopy and Signal Processing Technology*, 2018; 10620. DOI: 10.1117/12.2293301.

6. Girshick R., Shotton J., Kohli P., Criminisi A., Fitzgibbon A. Efficient regression of general-activity human poses from depth images. 2011 International Conference on Computer Vision, 2011, pp. 415-422.
7. Sun Y., Zhang X., Xin Q., Huang, J. Developing a multi-filter convolutional neural network for semantic segmentation using high-resolution aerial imagery and LiDAR data. *ISPRS journal of photogrammetry and remote sensing*, 2018, vol. 143, pp. 3-14.
8. Watts K.W., Konolige K., Konolige K. Ground plane detection to verify depth sensor status for robot. US Patent no. 9886035, 2018.
9. Tee Kit Tsun M., Lau B., Siswoyo Jo H. An improved indoor robot human-following navigation model using depth camera, active IR marker and proximity sensors fusion. *Robotics*, 2018, no. 7(1), 4 p. DOI: 0.3390/robotics7010004.
10. Gorobtsov A.S., Andreev A.E., Markov A.E., Skorikov A.V., Tarasov P.S. Osobennosti resheniya uravnenii metoda obratnoi zadachi dlya sinteza ustoichivogo upravlyaemogo dvizheniya shagayushchikh robotov [Features of solving the inverse dynamic method equations for the synthesis of stable walking robots controlled motion]. *Trudy SPIIRAN = SPIIRAS Proceedings*, 2019, no. 18, pp. 85-122. DOI: 10.15622/sp.18.1.85-122 (In Russ.).
11. Altukhov V.G., Kolker A.B. Vychislenie rasstoyaniya do ob"ekta na osnove karty glubin poluchennoi metodom zerkal'nogo razdeleniya izobrazhenii [Calculation of the distance to the object based on the depth map obtained by the method of mirror image separation]. *Avtomatika i programmnyaya inzheneriya = Automation and Software Engineering*, 2017, no. 1, pp. 65-69 (In Russ.).
12. Erofeev M., Vatolin D. Mnogosloinoe reshenie problemy poluprozrachnykh granits pri postroenii stereoskopicheskikh izobrazhenii [A multi-layered solution to the problem of translucent boundaries when building stereoscopic images]. *International Journal of Open Information Technologies*, 2016, no.4(8) (In Russ.).
13. Lin K.Y., Hang H.M. Depth Map Enhancement on RGB-D Video Captured by Kinect V2. 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2018, pp. 1530-1535.
14. Ulyanov S.V., Reshetnikov A.G., Koshelev K.V. Razrabotka sistemy stereozreniya dlya mobil'nogo robota [Development of a stereo vision system for a mobile robot]. *Programmnye produkty i sistemy = Software products and systems*. 2017; no. 30(3) (In Russ.).
15. Koch T. Evaluation of CNN-based single-image depth estimation methods. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.
16. Eigen D., Puhrsch C., Fergus R. Depth map prediction from a single image using a multiscale deep network. *Advances in Neural Information Processing Systems*, 2014, pp. 2366-2374. arXiv: 1406.2283v1.

17. Liu F., Shen C., Lin G., Deep convolutional neural fields for depth estimation from a single. The IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 5162-5170. DOI: 10.1109/ CVPR.2015.7299152.
18. Laina I., Rupprecht C., Belagiannis V., Tombari F., Navab N. Deeper depth prediction with fully convolutional residual networks. CoRR, abs/1606.00373. 2016. DOI: 10.1109/3dv.2016.32. arXiv: 1606.00373.
19. Li J., Klein R., Yao A. A two-streamed network for estimating fine-scaled depth maps from single rgb images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3372–3380.
20. Soldatova O.P., Lyozin I.A., Lyozina I.V., Kupriyanov A.V., Kirsh D.V. Primenenie nechetkikh neironnykh setei dlya opredeleniya tipa kristalliche-skikh reshetok, nablyudayemykh na nanomasshtabnykh izobrazheniyakh [Application of fuzzy neural networks for defining crystal lattice types in nanoscale images]. *Komp'yuternaya optika = Computer Optics*. 2015; no. 39(5): 787-795. (In Russ.). DOI: 10.18287/0134-2452-2015-39-5-787-794.
21. Jalalvand A., Demuynck K., Neve W.D., Martensa J.-P. On the application of reservoir computing networks for noisy image recognition. *Neurocomputing*. 2018; 277: 237-248. DOI: 10.1016/j.neucom.2016.11.100.
22. Dutta S., Manideep B.CS., Basha S.M., Caytiles R.D., Iyengar N.Ch.S.N. Classification of diabetic retinopathy images by using deep learning models. *International Journal of Grid and Distributed Computing*. 2018; no. 11(1), pp. 89-106. DOI: 10.14257/ijgdc.2018.11.1.09.
23. Sirota A.A., Dryuchenko M.A. Obobshchennye algoritmy szhatiya izobrazhenii na fragmentakh proizvol'noi formy i ikh realizatsiya s ispol'zovaniem iskusstvennykh neironnykh setei [Generalized image compression algorithms for arbitrarily-shaped fragments and their implementation using artificial neural networks]. *Komp'yuternaya optika = Computer Optics*. 2015; 39(5): 751-761. (In Russ.). DOI: 10.18287/0134-2452-2015-39-5-751-761.
24. Nikonorov A.V., Petrov M.V., Bibikov S.A., Kutikova V.V., Morozov A.A., Kazanskij N.L. Rekonstruktsiya izobrazhenii v difraktsionno-opticheskikh sistemakh na osnove svertochnykh neironnykh setei i obratnoi svertki [Image restoration in diffractive optical systems using deep learning and deconvolution]. *Komp'yuternaya optika = Computer Optics*. 2017; no. 41(6), pp. 875-887. (In Russ.). DOI: 10.18287/2412-6179-2017-41-6-875-887.
25. Oleinik A.L., Kukharev, G.A. Algoritmy vzaimnoi rekonstruktsii izobrazhenii lits na osnove metodov proektsii v sobstvennye podprostranst [Algorithms for Face Image Mutual Reconstruction by Means of Two-Dimensional Projection Methods]. *Trudy SPIIRAN = SPIIRAS Proceedings*. 2018; 2(57): 45-74. (In Russ.). DOI: 10.15622/sp.57.3.

26. Silberman N, Fergus R. Indoor scene segmentation using a structured light sensor. Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on; 2011, pp. 601-608.
27. Hu J., Ozay M., Zhang Y., Okatani T. Revisiting single image depth estimation: toward higher resolution maps with accurate object boundaries. 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), 2019, pp. 1043-1051.
28. Zhu J. Ma R. (2016), Real-time depth estimation from 2D images. Available at: http://cs231n.stanford.edu/reports/2016/pdfs/407_Report.pdf (accessed May 1, 2018).
29. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations (ICLR), 2015, pp. 1-14. arXiv: 1409.1556.
30. He L., Wang G., Hu Z. Learning depth from single images with deep neural network embedding focal length. *IEEE Transactions on Image Processing*, 2018, no. 27(9), pp. 4676-4689.
31. Liu M., Salzmann M., He X. Discrete-continuous depth estimation from a single image. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 716-723. DOI: 10.1109 / CVPR.2014.97.
32. Luo W., Schwing A.G., Urtasun R. Efficient deep learning for stereo matching. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 5695-5703. DOI: 10.1109/CVPR.2016.614.
33. Eigen D., Fergus R. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. The IEEE International Conference on Computer Vision (ICCV), 2015, pp. 2650-2658. DOI: 10.1109 / ICCV.2015.304.
34. Geiger A., Lenz A., Stiller C., Urtasun R. Vision meets robotics: the KITTI dataset. *The International Journal of Robotics Research*, 2013, vol. 32, no. 11, pp. 1231-1237. DOI: 10.1177/0278364913491297.
35. Saxena A., Sun M., Ng A. Y. Make3d: learning 3D scene structure from a single still image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, no. 31(5), pp. 824-840. DOI: 10.1109 / TPAMI.2008.132.
36. Bogart R., Kainz F., Hess D. OpenEXR image file format. ACM SIGGRAPH 2003, Sketches & Applications, 2003.
37. Kent B.R. 3D Scientific Visualization with Blender, Morgan & Claypool, San Rafael, 2015.
38. Valenza E. Blender 2.6 Cycles: Materials and Textures Cookbook – Third Edition. Packt Publishing Ltd. Birmingham, Mumbai, 2013, 280 p.

39. Saxena A., Chung S.H., Andrew Y. Ng. Learning depth from single monocular images. *Neural Information Processing Systems (NIPS)*, 2005, pp. 1161-1168. DOI: 10.1109/TPAMI.2015.2505283.

40. Saxena A., Chung S.H., Andrew Y. Ng. 3D depth reconstruction from a single still image. *International Journal of Computer Vision*, 2008, no. 76(1), pp. 53-69. DOI: 10.1109/TPAMI.2008.132.

41. Glorot X., Bengio Y. Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010, no. 9, pp. 249-256.

42. Heaton J. Artificial Intelligence For Humans: Deep Learning and Neural Networks, Heaton Research. Inc., St Louis, MO 2015; no. 3.

43. He K., Zhang X., Ren S., Sun J. Delving deep into rectifiers: surpassing human-level performance on imagenet. *Proceedings of the IEEE international conference on computer vision*, 2015; 1026-1034. DOI: 10.1109 / ICCV.2015, 123 p.

44. Keras library. Available at: <https://keras.io/> (acceessed 31.08.2018).

45. Backends—TensorFlow or Theano. Available at: www.tensorflow.org/ (acceessed 31.08.2018).

46. Abadi M., Barham P., Chen J., Chen Z., Davis A., Dean J., Devin M., Ghemawat S., Irving G., Isard M., Kudlur M., Levenberg J., Monga R., Moore S., Murray D.G., Steiner B., Tucker P., Vasudevan V., Warden P., Wicke M., Yu Y., Zheng X., Brain G. TensorFlow: a system for large-scale machine learning. The 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16), Nov. 2-4, 2016; pp. 265-283.

47. Abadi M., Agarwal A., Barham P., Brevdo E., Chen Z., Citro C., Corrado G.S., Davis A., Dean J., Devin M., Ghemawat S., Goodfellow I., Harp A., Irving G., Isard M., Jia Y., Jozefowicz R., Kaiser L., Kudlur M., Levenberg J., Mane D., Monga R., Moore S., Murray D., Olah C., Schuster M., Shlens J., Steiner B., Sutskever I., Talwar K., Tucker P., Vanhoucke V., Vasudevan V., Viegas F., Vinyals O., Warden P., Wattenberg M., Wicke M., Yu Y., Zheng X. Tensorflow: large-scale machine learning on heterogeneous distributed systems. (2016b). Available at: <https://arxiv.org/pdf/1603.04467.pdf> (accessed May 1, 2018).

48. Duchi J., Hazan E., Singer Y. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*. July 2011; 12: pp. 2121-2159.

49. Ashiquzzaman A., Tushar A.K., Islam MdR., Shon D., Im K., Park J.-H, Lim D.-S, Kim J, Reduction of overfitting in diabetes prediction using deep learning neural network. 2017 IT Convergence and Security, Springer, Singapore 2018; 449, pp. 35-43. DOI: 10.1007/978-981-10-6451-7_5.

Информация об авторах / Information about the Authors

Михальченко Даниил Игоревич, аспирант лаборатории автономных робототехнических систем, Санкт-Петербургский институт информатики и автоматизации Российской академии наук, г. Санкт-Петербург, Российская Федерация, e-mail: tekatodsham@gmail.com

Daniil I. Mihalchenko, Post-Graduate Student, Laboratory of Autonomous Robotic Systems, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg, Russian Federation e-mail: tekatodsham@gmail.com

Ивин Арсений Григорьевич, аспирант лаборатории автономных робототехнических систем, Санкт-Петербургский институт информатики и автоматизации Российской академии наук, разработчик программного обеспечения, ООО Яндекс.Технологии, г. Санкт-Петербург, Российская Федерация, e-mail: arssivka@yandex.ru

Arsenii G. Ivin, Post-Graduate Student, Laboratory of Autonomous Robotic Systems, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, Software Developer, Yandex.Technology LLC. St. Petersburg, Russian Federation e-mail: arssivka@yandex.ru

Сивченко Олег Юрьевич, программист лаборатории автономных робототехнических систем Санкт-Петербургского института информатики и автоматизации Российской академии наук, г. Санкт-Петербург, Российская Федерация, e-mail: [E-mail: adrelan@gmail.com](mailto:adrelan@gmail.com)

Oleg Y. Sivchenko, Programmer, Laboratory of Autonomous Robotic Systems, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences. St. Petersburg, Russian Federation e-mail: adrelan@gmail.com

Аксаментов Егор Алексеевич, младший научный сотрудник лаборатории технологий больших данных социкиберфизических систем Санкт-Петербургского института информатики и автоматизации Российской академии наук, г. Санкт-Петербург, Российская Федерация, e-mail: Egor.aksamentov.96@mail.ru

Egor A. Aksamentov, Junior Research Fellow, Laboratory of Autonomous Robotic Systems, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg, Russian Federation e-mail: Egor.aksamentov.96@mail.ru