УДК 004.93; 004.94; 528.92

ТЕХНОЛОГИЯ АВТОМАТИЧЕСКОГО СИНТЕЗА ОБУЧАЮЩИХ МНОЖЕСТВ В ЗАДАЧАХ СЕМАНТИЧЕСКОГО КАРТОГРАФИРОВАНИЯ

С.А.К. Диане
Институт проблем управления им. В.А. Трапезникова РАН
Россия, 117997, Москва, Профсоюзная ул., 65
E-mail: diane1990@yandex.ru

Е.А. Лесив
Институт проблем управления им. В.А. Трапезникова РАН
Россия, 117997, Москва, Профсоюзная ул., 65
E-mail: smailsbobs@gmail.com

А.В. Чехов
Институт проблем управления им. В.А. Трапезникова РАН
Россия, 117997, Москва, Профсоюзная ул., 65
E-mail: achekhov@gmail.com

Ключевые слова: анализ визуальной информации, классификация образов, синтез обучающих множеств, нейронные сети, компьютерное моделирование, семантическое картографирование.

Аннотация: Описана технология автоматизации синтеза обучающих множеств для настройки нейросетевых анализаторов изображений на базе существующих средств трехмерной графики. Предложена методика двухэтапного синтеза виртуальных сред. Рассмотрена архитектура сверточной нейронной сети, способной выполнять классификацию визуальных образов по результатам обучения на синтезируемых множествах. Обоснованы преимущества развиваемого подхода. Проведена оценка применимости разработанных средств синтеза обучающих множеств для целого ряда задач, связанных с исследованием и картографированием внешней среды автономных роботов, а также для задач поисково-спасательного характера.

1. Введение

За последние несколько десятилетий огромное количество проводимых исследований посвящено вопросам картографирования среды функционирования автономных роботов. Одновременно с этим активное внимание научного сообщества привлечено к вопросам обучения нейронных сетей для решения задач визуальной классификации и анализа геометрических параметров объектов.

С одной стороны, очевидно, что совмещение двух этих направлений обеспечит повышение информативности формируемых карт, за счет добавления в них семантической информации об объектах, расположенных в исследуемой зоне.

С другой стороны, задача визуальной классификации является комплексной и на сегодняшний день решена не полностью. С развитием вычислительной техники, стало

очевидно, что наилучшие результаты при распознавании визуальных образов дают сверточные нейронные сети (СНС). Однако привлечение искусственных нейронных сетей (ИНС) любого типа для решения конкретных прикладных задач требует не только правильного выбора архитектуры нейронной сети, но и подготовки множества данных для ее обучения.

Большинство работ в данном отношении полагаются на использование обучающих множеств, сформированных вручную [1, 2], что чревато неполнотой, а зачастую и неточностью в составлении входных и выходных образов ИНС, равно как и невозможностью оперативной подстройки сети под решение новых задач.

Альтернативный подход, предлагаемый в настоящем исследовании, связан с автоматической генерацией обучающих выборок для решения задач визуального анализа изображений.

2. Задачи визуального анализа изображений

В рамках научной проблематики визуального анализа изображений можно выделить несколько основных задачи, решение каждой из которых допускает применение технологии нейронных сетей, обучаемых на базах аннотированных примеров:

– визуальная классификация одиночных объектов;
– локализация и оценка геометрических параметров объектов;
– визуальная навигация и оценка состояния внешней среды;
– визуальная сегментация объектов на изображении;
– оценка глубины изображения и 3D-реконструкция;
– анализ топологии и лингвистическая интерпретация сцен.

Визуальная классификация необходима для определения принадлежности объекта к одной из априорно заданных категорий в условиях неопределенности внешней среды [3]. Данные неопределенности вызваны тем, что классифицируемый объект может располагаться на изображении под различными ракурсами; в различном увеличении, смещении, повороте; быть загорожен или по-разному освещен.

Задача локализации объектов на изображении, полученном с бортовой камеры автономного робота, подразумевает оценку их положения на растре и одновременно угловых координат: азимутального угла φ и нормального угла θ между направлением на центр объекта и оптической осью камеры. Более того, определяя данные углы в последовательные моменты времени, возможно рассчитать положение объекта в трехмерном пространстве и нанести информацию о типе объекта и о его положении на семантическую карту [4].

В рамках оценки геометрических параметров визуальных образов одной из важнейших задач является определение размера и пространственной ориентации объекта. Помимо этих задач, относящихся сугубо к области обработки информации, перед исследователями встают и задачи, приближенные к управлению – определение ориентации и возможности захвата целевого объекта при помощи манипуляционного устройства [6, 7].

Кроме того, для мобильных роботов особую важность играет задача визуальной локальной навигации и оценки состояния внешней среды на предмет проходимости поверхности (или, в случае БПЛА, безопасности пролета через некоторый объем пространства) и формирование соответствующих управляющих команд [8].

Еще одно направление визуального анализа изображений – сегментация объектов пересекается с задачей классификации, но дает более полные с геометрической точки зрения результаты по различению точек, принадлежащих тем или иным объектам

внешней среды [9]. Сегментация так же, как и локализация, может использоваться для решения задач управления мобильными и манипуляционными роботами, таких как выбор маршрута перемещения на основе анализа поверхности передвижения или для захвата целевого предмета из неструктурированного набора объектов.

Особый интерес представляет задача сегментации объектов, тип которых неизвестен. Такая постановка задачи дает возможность разработки механизма формирования обучающего множества для решения задач классификации и последующей реализации самообучения на борту автономных роботов.

Тесно переплетается с вопросами локализации и сегментации задача оценки глубины сцены. По аналогии с биологическими прототипами ИНС способны с некоторой достоверностью вычислять дальность от объектива камеры до предметов в сцене. Наилучшие результаты дает анализ стереоизображений, однако известны решения, основанные на монокулярной оценке глубины [10]. Карту глубину сцены не составляет труда трансформировать в трехмерное облако вершин. Объединение нескольких таких множеств точек дает возможность реконструкции 3D модели наблюдаемого объекта [11].

Перспективным направлением является анализ топологии визуальных сцен. Так, например, в работе [12] подобная функциональность достигнута за счет совмещения технологии сегментации и нейросетевой реализации метода максимального правдоподобия. Для каждой пары смежных суперпикселей изображения рассчитывается оценка вероятности их принадлежности к одному родительскому сегменту, и далее вверх по иерархии – оценивается принадлежность сегментов к объектам и объектов к группам объектов. В результате формируется дерево отношений между объектами, которое может быть интерпретировано в виде лингвистического описания. Альтернативный подход к формированию текстовых описаний изображений базируется на сочетании сверточных и рекуррентных нейронных сетей [13].

3. Программно-алгоритмическое обеспечение для автоматического формирования обучающих выборок

Наличие качественного обучающего множества во многом определяет работу алгоритмов машинного обучения. Следует отметить, что при составлении обучающей выборки следует уделять внимание не только объему данных, но и таким моментам, как сбалансированность классов и порядок их следования. Данные должны содержать сопоставимый объем экземпляров для каждого класса и должны быть перемешаны. Целесообразно включение в обучающую выборку данных, которые описывают близкие условия к условиям дальнейшего использования ИНС.

Широкие возможности для автоматизации формирования обучающих множеств в задачах анализа изображений дают библиотеки компьютерной графики. Широкими возможностями обладают находящиеся в открытом доступе библиотеки OpenGL [14], включающие необходимые функции для рисования сложных трёхмерных сцен из простых примитивов.

С точки зрения настройки информационно-управляющих систем, библиотеки OpenGL пригодны для имитации основных неопределенностей, которые возникают во внешнем виде объектов, находящихся в поле зрения робота.

В настоящем исследовании предлагается технология синтеза обучающих множеств, базирующаяся на применении разработанного комплекса программно-алгоритмических средств, структурная схема которого представлена на рис. 1.

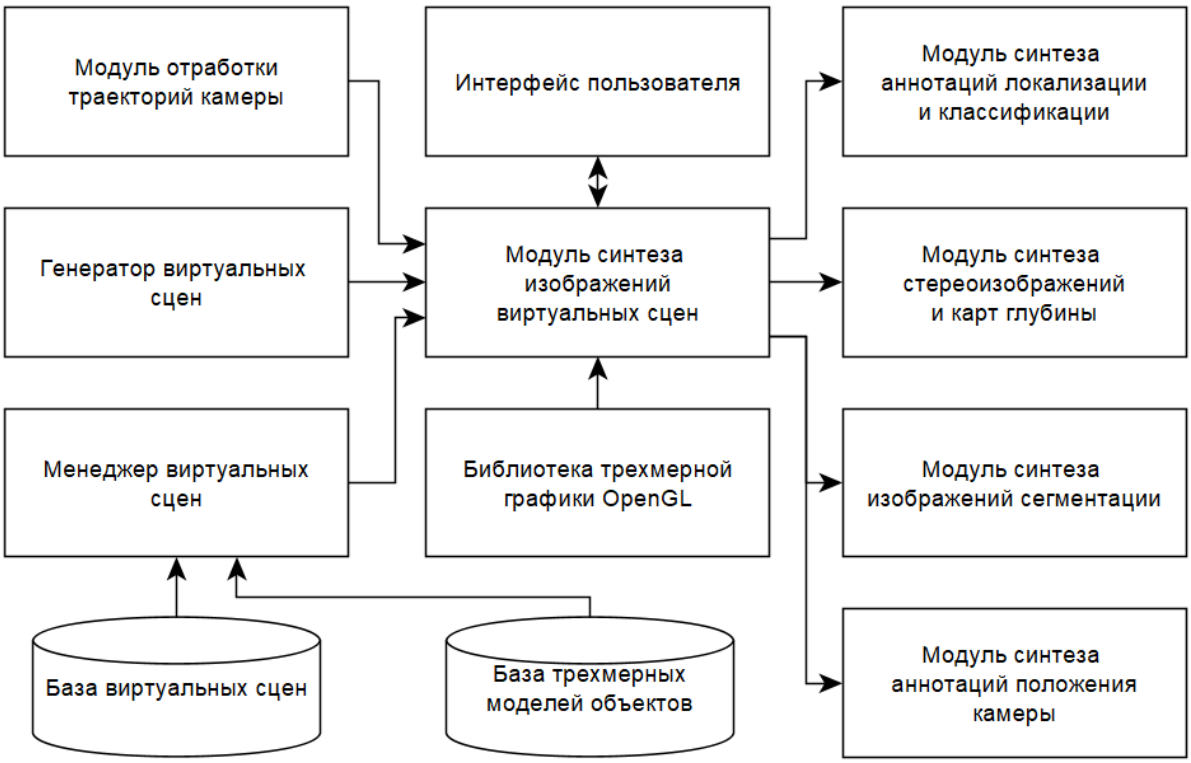

Рис. 1. Структура комплекса программно-алгоритмических средств для генерации обучающих множеств.

В основе программного комплекса для синтеза обучающих множеств (КСОМ) наряду с библиотекой OpenGL лежит модуль синтеза изображений виртуальных сцен. Виртуальная сцена представляет из себя совокупность трехмерных объектов различных категорий, снабженных описанием положения в пространстве, ориентации и цветовых характеристик.

Модули, изображенные в правой части структурной схемы КСОМ (рис. 1), отражают возможности комплекса по синтезу различных типов обучающих множеств. В соответствии с вышеперечисленными задачами визуального анализа изображений КСОМ позволяет генерировать обучающие выборки для решения задач визуальной классификации, локализации, сегментации, оценки глубины изображений. Кроме того, виртуальная среда предоставляет доступ к точному положению камеры в последовательные моменты времени, что дает возможность синтеза обучающих выборок и для решения задачи визуальной одометрии.

Формирование виртуальных сцен на этапе, предшествующем отрисовке, может выполняться двумя способами.

Для задач грубой настройки нейросетевых анализаторов, когда взаимное положение различных объектов не принципиально и, напротив, требуется как можно большее разнообразие перемещений объектов по сцене, применяется подход, суть которого в следующем. Задается или случайным образом выбирается число N объектов, одновременно наблюдаемых в сцене. Формируется вектор случайных положений для данных объектов $P = \{p_1, ..., p_N\}$. Производится устранение ситуаций взаимопроникновения объектов на основе метода потенциальных полей:

$$p_i' = p_i + \min(d_{\max}, \sum_{j=1, j\neq i}^{N} \eta / (p_j - p_i)^2),$$

где $p_i' = \{x', y', z'\}$ – обновленное положение объекта; $d_{\max}$ – максимальное смещение объектов; η – коэффициент силы отталкивания.

Для задач более точной настройки нейросетевых анализаторов на решение конкретных прикладных задач применяется подход, основанный на загрузке заблаговременно подготовленных виртуальных сцен. Разнообразие обучающих примеров при этом достигается уже не вариацией положения предметов в сцене, а движением камеры по указанной траектории.

При использовании последнего подхода, следует учитывать, что разработка трехмерных сцен может быть столь же утомительна для человека-эксперта, как и ручное формирования обучающих примеров. Предлагаемая методика двухэтапного синтеза виртуальных сцен направлена на устранение данного недостатка.

На первом этапе решения конкретной прикладной задачи или группы прикладных задач формируется программное обеспечение, позволяющее формировать описания сцен произвольного размера и плотности расположения объектов. Обеспечивается совместимость форматов хранения описаний сцен с программно-алгоритмическим комплексом синтеза обучающих множеств.

На втором этапе автоматически сформированные сцены загружаются в КСОМ: производится интерпретация текстовых описаний сцен и формирование соответствующих программных представлений для объектов, перечисленных в файле сцены. Так, например, на рис. 2 представлен внешний вид КСОМ с загруженной сценой и траекторией движения камеры.

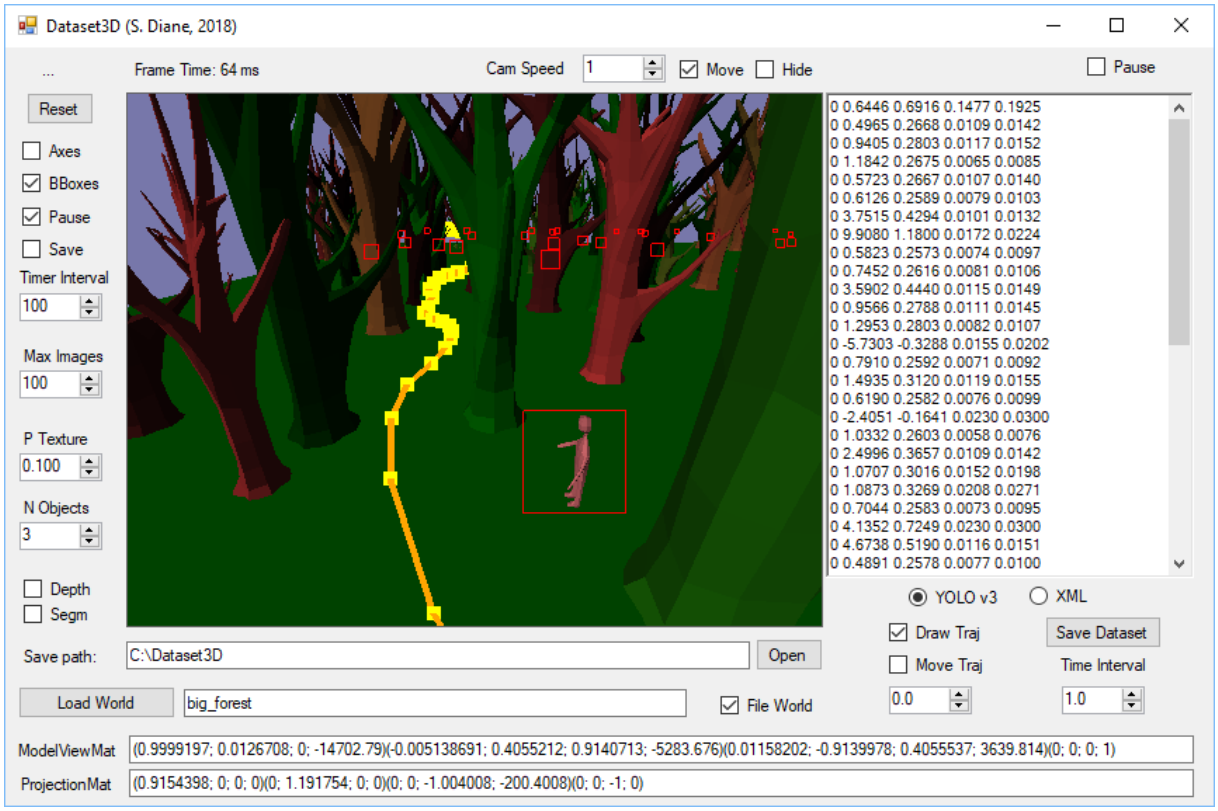

Рис. 2. Структура комплекса программно-алгоритмических средств для генерации обучающих множеств.

Как видно из рисунка, сцена содержит объекты различного типа. Причем лишь некоторые из загруженных объектов подлежат визуальному анализу. Так, модель человека на переднем плане обведена рамкой локализации, в то время как деревья – нет.

Определение положений и размера рамок основывается на расчете обращенной видовой матрицы OpenGL. В режиме сегментации отключается отрисовка теней и текстур объектов. Аннотации, содержащие желаемые результаты анализа типа и положения объектов, сохраняются совместно с изображениями в каталог на диске для дальнейшей настройки нейронных сетей.

4. Экспериментальная оценка применимости технологии

Важнейшим вопросом в оценке применимости развиваемой технологии является проверка способности нейронных сетей, обученных на виртуальных примерах, осуществлять визуальный анализ изображений, полученных в реальной среде функционирования автономных роботов.

При поиске ответа на данный вопрос в качестве примера была рассмотрена задача поиска человека в лесу. По результатам обучения нейронной сети YOLO v2 [3] на грубо проработанных виртуальных сценах с лесными массивами, подобных той, что изображена на рис. 1, удалось настроить нейросетевой классификатор на распознавание человека в реальных условиях леса (рис. 3). Архитектура использованной сверточной нейронной сети приведена в таблице 1.

Таблица 1. Архитектура сверточной ИНС YOLO v2.

№ слоя	Тип слоя	Число ядер	Размер / сдвиг ядра	Число выходов
1	Convolutional	32	3 × 3	224 × 224
2	Maxpool		2 × 2/2	112 × 112
3	Convolutional	64	3 × 3	112 × 112
4	Maxpool		2 × 2/2	56 × 56
5	Convolutional	128	3 × 3	56 × 56
6	Convolutional	64	1 × 1	56 × 56
7	Convolutional	128	3 × 3	56 × 56
8	Maxpool		2 × 2/2	28 × 28
9	Convolutional	256	3 × 3	28 × 28
10	Convolutional	128	1 × 1	28 × 28
11	Convolutional	256	3 × 3	28 × 28
12	Maxpool		2 × 2/2	14 × 14
13	Convolutional	512	3 × 3	14 × 14
14	Convolutional	256	1 × 1	14 × 14
15	Convolutional	512	3 × 3	14 × 14
16	Convolutional	256	1 × 1	14 × 14
17	Convolutional	512	3 × 3	14 × 14
18	Maxpool		2 × 2/2	7 × 7
19	Convolutional	1024	3 × 3	7 × 7
20	Convolutional	512	1 × 1	7 × 7
21	Convolutional	1024	3 × 3	7 × 7
22	Convolutional	512	1 × 1	7 × 7
23	Convolutional	1024	3 × 3	7 × 7
24	Convolutional	1000	1 × 1	7 × 7
25	Avgpool		Global	1000
26	Softmax			1000

Рис. 3. Результаты распознавания и локализации человека сверточной ИНС в условиях леса. Корректное распознавания (слева). Ошибочное двойное распознавание (справа).

В робототехнических приложениях информация об обнаруженных объектах может быть нанесена на семантическую карту или же передана напрямую оператору.

6. Заключение

Полученные результаты подтверждают перспективность развиваемого подхода. Интеграция технологий трехмерной графики и экспертных знаний о предметной области позволяет осуществить эффективную и вычислительно быструю генерацию обучающих множеств для решения задач визуального анализа в необходимом объеме.

Следует, однако, понимать, что обучение на виртуальных множествах не дает стопроцентной точности в настройке ИНС под конкретную задачу. Тем не менее, дальнейшее повышение качества функционирования нейросетевых анализаторов изображений возможно путем дообучения сети на небольшом множестве реальных изображений из выбранной предметной области.

Нет сомнения, что роль систем виртуального моделирования как мощного инструмента машинного обучения и сточки зрения обработки информации, и с точки зрения синтеза целесообразных управляющих процедур будет возрастать.

Работа выполнена при поддержке программы президиума РАН № 30 «Теория и технологии многоуровневого децентрализованного группового управления в условиях конфликта и кооперации».

Список литературы

1. Deng J., Dong W., Socher R., Li L.-J., Li K., Fei-Fei L. ImageNet: A Large-Scale Hierarchical Image Database // IEEE Computer Vision and Pattern Recognition (CVPR). 2009.
2. Tsung-Yi Lin et al. Microsoft COCO: Common Objects in Context. 2015. arXiv:1405.0312.
3. Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection // IEEE Computer Vision and Pattern Recognition (CVPR). 2016.
4. Diane S., Lesiv E., Pesheva I., Neschetnaya A. Multi-Aspect Environment Mapping with a Group of Mobile Robots // 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus). P. 474-478.
5. Galindo C., Saffiotti A., Coradeschi S., Buschka P. Multi-hierarchical semantic maps for mobile robotics // IEEE/RSJ International Conference on Intelligent Robots and Systems. 2005.
6. Pinto L., Gupta A. Supersizing Self-supervision: Learning to Grasp from 50K Tries and 700 Robot Hours // arXiv. 2015. 1509.06825.
7. Redmon J., Angelova A. Real-Time Grasp Detection Using Convolutional Neural Networks // arXiv. 2014. 1412.3128.

8. Loquercio A., Maqueda A., Del Blanco C.R., Scaramuzza D. DroNet: Learning to Fly by Driving. IEEE Robotics and Automation Letters. 2018. P. 1-1. 10.1109/LRA.2018.2795643.
9. Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, Jiaya Jia. Path Aggregation Network for Instance Segmentation // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.
10. Eigen D., Puhrsch C., Fergus R., Depth Map Prediction from a Single Image using a Multi-Scale Deep Network // arXiv. 2014. 1406.2283.
11. Jeong D., Y. Li, Lee H.J. et al. Efficient 3D Volume Reconstruction from a Point Cloud Using a Phase-Field Method // Mathematical Problems in Engineering. 2018. Article ID 7090186.
12. Socher R., Lin C.C.-Y., Ng A., Manning C. Parsing Natural Scenes and Natural Language with Recursive Neural Networks // Proc. of the 26th International Conference on Machine Learning (ICML), 2011.
13. Karpathy A., Li Fei-Fei. Deep Visual-Semantic Alignments for Generating Image Descriptions // IEEE Trans. Pattern Anal. Mach. Intell. 2017. Vol. 39, No. 4. P. 664-676.
14. Херн Д., Бейкер П.М.. Компьютерная графика и стандарт OpenGL = Computer Graphics with OpenGL. 3-е изд. М.: Вильямс, 2005. 1168 с.