

МЕТОДОЛОГИЯ АВТОМАТИЧЕСКОГО СИНТЕЗА ОБУЧАЮЩИХ МНОЖЕСТВ В ЗАДАЧАХ НЕЙРОСЕТЕВОГО УПРАВЛЕНИЯ АВТОНОМНЫМИ РОБОТАМИ

Диане С.А.К., Лесив Е.А., Зинченко И.В.

Институт проблем управления им. В.А. Трапезникова РАН,

Россия, г. Москва, ул. Профсоюзная д.65

diane1990@yandex.ru, smailsbobs@gmail.com, ziv97@mail.ru

Аннотация: Описана технология автоматического синтеза обучающих множеств для настройки нейросетевых анализаторов изображений в виртуальной среде с использованием средств трехмерной графики. Предложены принципы формирования обучающих выборок для задач классификации объектов. Исследована возможность применения нейросетевых классификаторов, обученных на виртуальных множествах, при решении реальных прикладных задач.

Ключевые слова: классификация визуальных образов, синтез обучающих множеств, трехмерная графика, нейронные сети

Введение

За последние несколько десятилетий проведено огромное количество исследований, посвященных 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.

1 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: 1) визуальная классификация одиночных объектов; 2) локализация и оценка геометрических параметров объектов; 3) визуальная навигация и оценка состояния внешней среды; 4) визуальная сегментация объектов на изображении; 5) оценка глубины изображения и 3D-реконструкция; 6) анализ топологии и лингвистическая интерпретация сцен.

2 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.

В настоящем исследовании предлагается технология синтеза обучающих множеств, базирующаяся на применении средств трехмерной графики (библиотека OpenGL) и разработанного на их основе программно-алгоритмического комплекса (рис. 1).

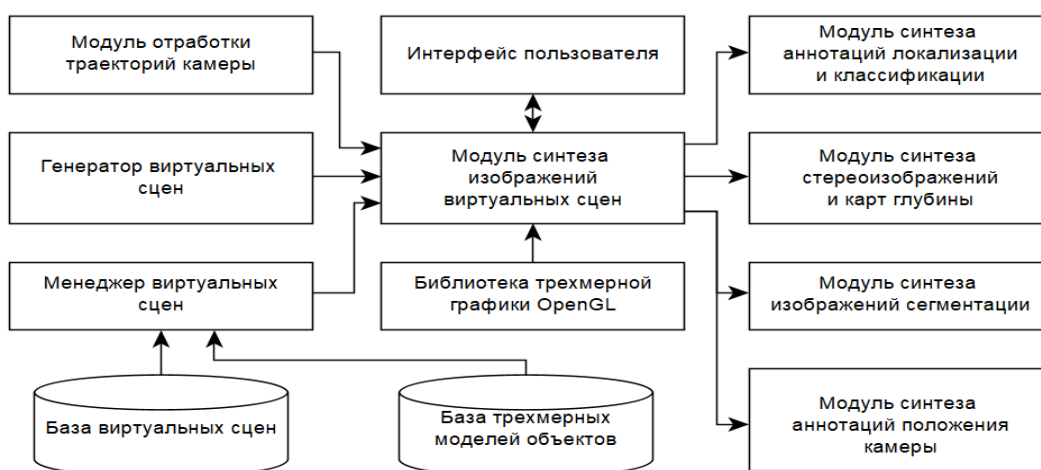


Рис. 1. Структура программного комплекса для генерации обучающих множеств

В основе программного комплекса для синтеза обучающих множеств (КСОМ) 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.

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 классификаторов, когда взаимное положение различных объектов непринципиально 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 $P = \{p_1, \dots, p_N\}$. 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 классификаторов 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.

На первом этапе решения конкретной прикладной задачи или группы прикладных задач используется 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.

Аннотации, содержащие желаемые результаты анализа типа и положения объектов, сохраняются совместно с изображениями в каталог на диске для дальнейшей настройки нейронных сетей.

3 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 [1] на грубо проработанных виртуальных сценах с лесными массивами настроить нейросетевой классификатор на распознавание человека в реальных условиях леса (рис. 2).

В робототехнических приложениях, когда наряду с визуальной информацией доступны навигационные данные, результаты классификации могут быть нанесены на семантическую карту [2-4] или же переданы напрямую оператору, контролирующему процесс картографирования.

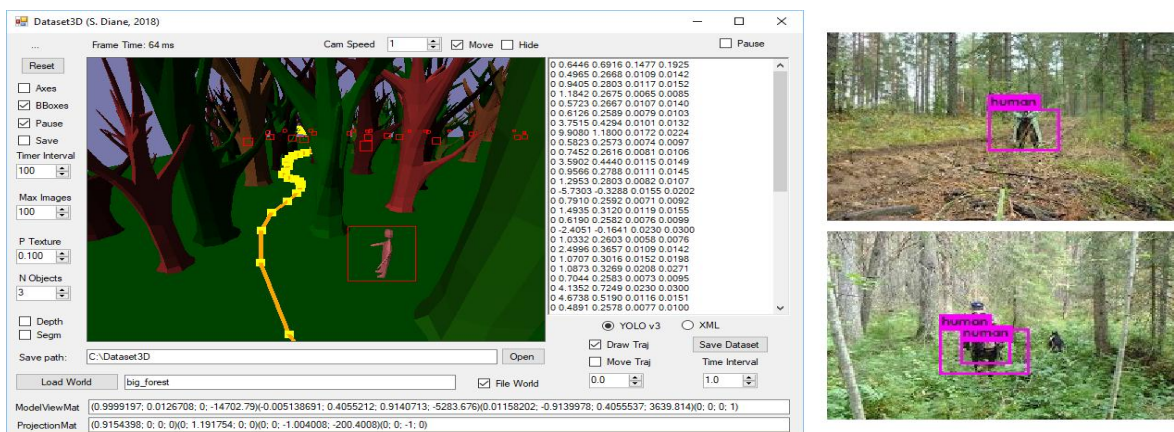


Рис. 2. Результаты моделирования виртуальных сцен (слева) и распознавания схожих объектов в реальной среде (справа)

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.

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

Литература

1. *J. Redmon, S. Divvala, R. Girshick, and A. Farhadi*, "You Only Look Once: Unified, Real-Time Object Detection", in CVPR 2016
2. *S. Diane, E. Lesiv, I. Pesheva, A. Neschetnaya*, Multi-Aspect Environment Mapping with a Group of Mobile Robots. 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EConRus), pp. 474-478
3. *C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka*, Multi-hierarchical semantic maps for mobile robotics, in 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems
4. *Legovich Yu.S., Diane S.A.K., Rusakov K.D.* Integration of modern technologies for solving territory patrolling problems with the use of heterogeneous autonomous robotic systems / Proceedings of the 11th International Conference "Management of Large-Scale System Development" (MLSD). Moscow: IEEE, 2018.