

МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра математического обеспечения и применения ЭВМ

ОТЧЕТ
по научно-исследовательской работе
Тема: Использование методов ML в задачах визуальной одометрии

Студент гр. 3303

Гэ Као

Руководитель

Кринкин К.В.

Санкт-Петербург

2018

1. Наименование научной работы

Использование методов ML в задачах визуальной одометрии

2. Исследовательская часть

2.1. Постановка задачи исследования

Алгоритм глубокого обучения (такой как LSTM, сеть CNN) используется для одновременной оценки позы 6-DoF монокулярной камеры и глубины сцены.

2.2. Актуальность работы

Вся сетевая структура включает в себя сеть оценки позы и сеть оценки глубины. Сеть оценки поз представляет собой глубоко повторяющуюся сверточную нейронную сеть (RCNN), которая реализует оценку монокулярной позы от конца до конца, состоящую из выделения признаков на основе сверточных нейронных сетей и моделирования временных рядов на основе повторяющихся нейронных сетей (RNN). Сеть оценки глубины генерирует плотные карты глубины, главным образом, на основе архитектуры кодера-декодера.

2.3. Проведённые исследования

Эта статья оценивает сквозную оценку позы и оценку глубины сцены монокулярной камеры путем создания сети глубокого обучения. Нейронная сеть состоит из сети оценки позы и сети оценки глубины. Как показано, каждая из двух сетей глубины принимает непрерывное монокулярное изображение в качестве входных данных и создает позу 6-DoF и глубину в качестве выходных сигналов соответственно.

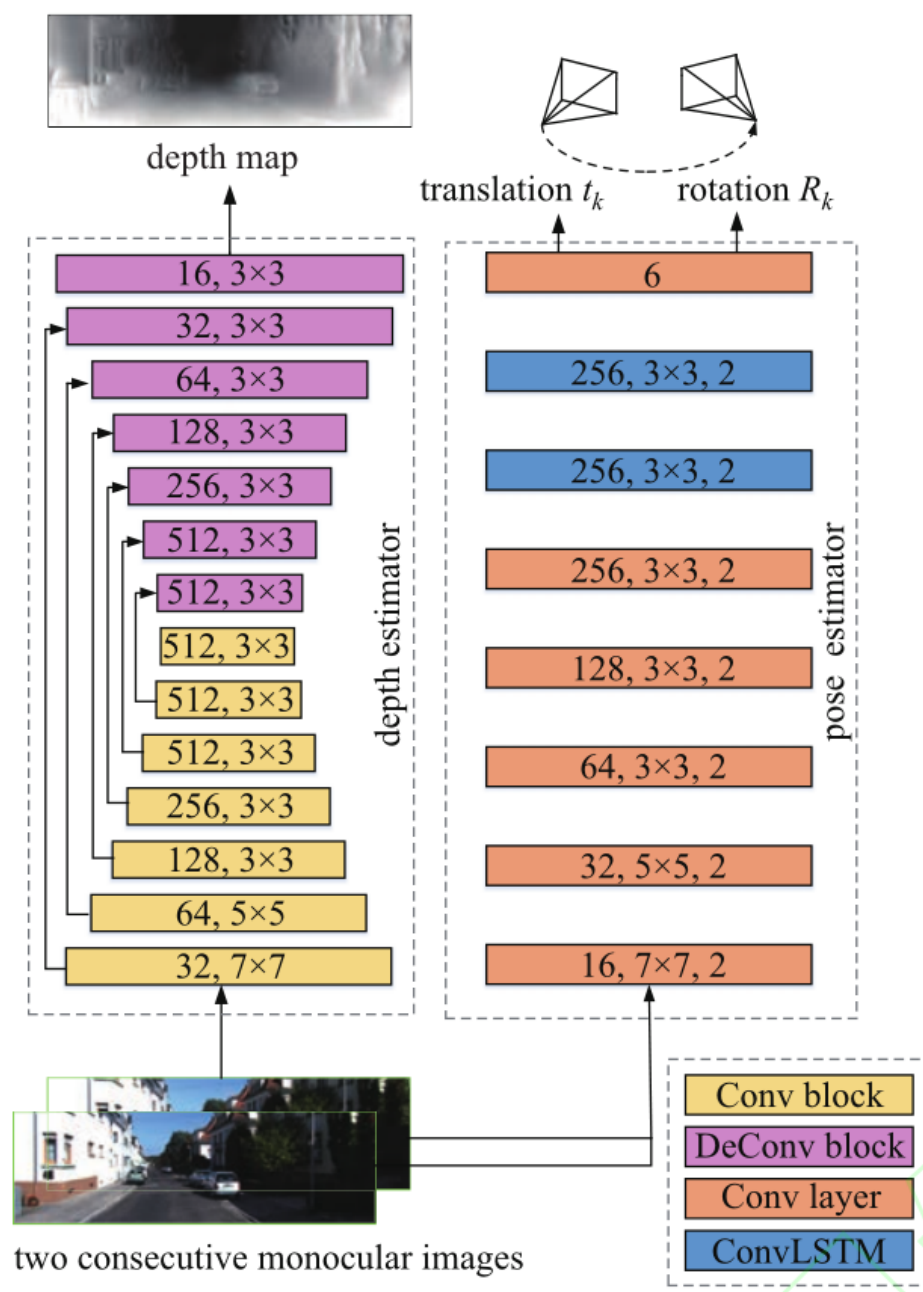


Рисунок 1. Архитектура оценки позы и оценки глубины

Сеть оценки поз представляет собой архитектуру CNN на основе VGG[1]. Он принимает 2 последовательных монокулярных изображения в качестве входных данных и предсказывает преобразование 6-DoF между ними. Кроме того, сеть оценки позы опирается на сеть оценки глубины на основе кодера и декодера для формирования карты плотной глубины, изображения параллакса (обратной глубины). Функция потерь строится и минимизируется с помощью потери синхронизации, потери глубины сглаживания и потери согласованности

спереди и сзади последовательности монокулярного изображения. Веса и пороги получены для оценки позы и карты глубины 6-DoF сквозным способом.

Эта статья будет использовать тензорный поток для проектирования RCNN, чтобы достичь оценки позы и карты глубины.

ЗАКЛЮЧЕНИЕ

Мы можем оценить предложенный метод сквозного монокулярного VO на хорошо известном тесте KITTI VO / SLAM [2].

Так как большинство существующих алгоритмов монокулярного VO не оценивают абсолютную шкалу, результаты их локализации должны быть вручную выровнены с основополагающей правдой, библиотекой VO с открытым исходным кодом LIBVISO2 [3], которая использует фиксированную высоту камеры для восстановления масштаба для Монокуляр VO принят для сравнения, также используется его стерео версия, которая может напрямую получать абсолютные позы.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- [1] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C/OL]. (2015-04-10) [2017-01-01]. <https://arxiv.org/pdf/1409.1556.pdf>.
- [2] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the KITTI vision benchmark suite,” in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012.
- [3] A. Geiger, J. Ziegler, and C. Stiller, “Stereoscan: Dense 3D reconstruction in real-time,” in Intelligent Vehicles Symposium (IV), 2011.