

СРАВНЕНИЕ МЕТОДОВ ПОСТРОЕНИЯ КАРТ ГЛУБИНЫ ИЗ ИЗОБРАЖЕНИЙ

Аннотация. В статье представлен сравнительный анализ методов построения карты глубины из изображения и определен, относительно рассмотренных, лучший для использования в системах реального времени. Рассмотрены методы: сверточные нейронные сети, капсульные нейронные сети, построение по стереопаре.

Ключевые слова: сравнительный анализ, нейронные сети, карта глубины, система реального времени, сверточные нейронные сети, CNN, капсульные нейронные сети, CapsNet.

Введение. В последнее время стало достаточно популярно использовать для повышения качества машинного зрения карты глубин, построенные нейронными сетями по одному изображению [1-3]. Карта глубины — матрица, содержащая данные, о расстоянии до поверхностей объектов сцены от точки обзора. Данный подход выделяется заметно за счет отсутствия необходимости в стереопаре, следовательно, дает возможность использовать старые наборы данных и сокращать общий объем данных для обработки. Но имеются и недостатки: невозможность вычислить масштаб и отсутствие разницы для алгоритмов между фотографией места и реальным пространством. В случаях, не требующих вышеперечисленных аспектов нейронные сети справляются с достаточной точностью. Для задач, требующих построения пространства на основе изображений, более качественным считается вычисление по стереопаре [4-9]. Различают несколько способов построения карт глубины с использованием нейронных сетей, их классификация производится относительно архитектуры сетей, так как именно она влияет на способ обработки поступающих сигналов.

Сверточные нейронных сети (CNN) — при обучение данного типа сетей первично производится построение карты несоответствия. Она представляет собой матрицу, содержащую разницу пик-

селей или движений между парой стереоизображений [1]. На ее основе создается карта глубины, которая через заданную функцию потерь позволяет определить, насколько близким было текущее предсказание относительно реальных данных. От стандартного обучения сверточных нейронных сетей, данный метод отличается только дополнительной предобработкой данных: анализ стереопары для построения истинной карты глубин и построение карты смещений перед при обучении. Недостатком данных сетей является возможность регистрировать только наличие элемента во входных данных, без учета его положения в пространстве. Используются CNN рассматривают: Ибрагим Альхашим и Питер Вонка в статье «High Quality Monocular Depth Estimation via Transfer Learning» [1]; Клеман Годар, Ойсин Мак Аодха, Габриэль Дж. Бростоу в статье «Unsupervised Monocular Depth Estimation with Left-Right Consistency» [2].

Капсульные нейронные сети (CapsNet) — в отличие от сверточных нейронных сетей, капсульные не просто сравнивают входные данные с одним фильтром или шаблоном, а продолжают обработку сигналов, передавая от слоев свертки на специальные группы, они же капсулы [3].

Каждая капсула характеризуется некоторым положением и вектором, не зависимо от других капсул того же уровня. Это позволяет увеличить точность в обмен на более длительный процесс обработки. Но все так же сохраняется проблема интерпретации в случае наличия изображения пространства на рассматриваемом снимке.

Проблема исследования. С ростом числа роботизированных средств способных к свободному перемещению растет и потребность в быстрых и точных методах оценки пространства. Построение по стереопаре просто в реализации, не требует подготовки, но сильно зависит от техники и требует применять дополнительно алгоритмы для подавления шумов. В тоже время нейронные сети позволяют получать достаточно точную карту глубины, но требуют настройки и обучения перед использованием. Цель данной работы — сравнить применимость представленных методов при использовании в системах реального времени для построения модели пространства.

Материалы и методы. Одной из отличительных особенностей систем реального времени является скорость предоставления данных, вне зависимости от объема требуемых для этого аппаратных ресурсов. Таким образом в качестве метрик будут использованы следующие характеристики:

- скорость получения результатов,
- точность предоставляемого ответа,
- объем ресурсов необходимый для вычислений (оперативная память).

В связи с тем, что показатели основаны не только на качестве методов, но и на характеристиках вычислительного аппарата, они будут рассматриваться относительно эталонных показателей, в качестве которого будет выступать метод вычисления по стереопаре, все измерений будут проводиться на одном аппаратном обеспечении. Кроме того, все используемые далее нейронные сети обучены на одном наборе данных NYUv2 [10], с целью сравнить методы в равных условиях. Замеры будут проводиться на коллекции наборов данных Middlebury [5-9]. Это позволит определить лучший метод при одинаковых обучающем наборе данных, тестовом наборе данных и технических характеристиках аппаратного обеспечения.

Результаты. CNN. Ибрагим Альхашим и Питер Вонка представили свой дизайн CNN в статье «High Quality Monocular Depth Estimation via Transfer Learning» [1]. Отличительной чертой представленного дизайна является применение трансферного обучения (рис. 1). Используя кодировщик изображений на базе DenseNet-169 [11], предназначенный для задач классификации, создается вектор признаков. Этот вектор передается в последовательную серию слоев с повышающей дискретизацией [12], чтобы построить окончательную карту глубины с половинным входным разрешением.

CapsNet. Сунил Пракаш и Гаэлан Гу в работе «Simultaneous Localization And Mapping with depth Prediction using Capsule Networks for UAVs» [3] рассматривают задачу создания карты глубины, как задачу реконструкции изображения. Особенности дизайна предложенной ими CapsNet (рис. 2) являются: использование

метода «Left-Right Consistency» [2] для создания карты несоответствия, что позволяет обучать сеть без использования истинных карт глубины, на основе только стереопар; применение фильтра Калмана [13] для постобработки результатов CapsNet, сглаживающего шумы в пределах определенных поверхностей.

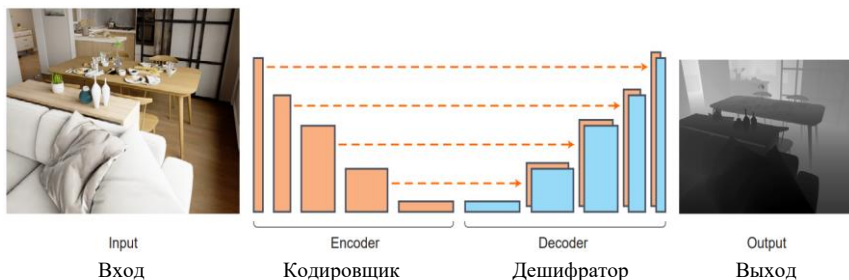


Рис. 1. Схема CNN Ибрагима Альхашим и Питера Вонка

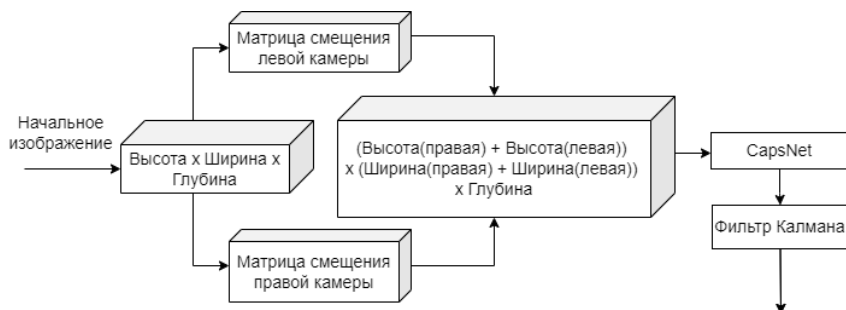


Рис. 2. Схема CapsNet Сунил Пракаш и Гаэлан Гу

Вычисление по стереопаре — классический метод, основанный стереометрии. При его реализации в первую очередь необходимо вычислить совпадающие зоны на изображениях, для этого используют метрику суммы абсолютных разностей (SAD) [14]. В результате будет получена матрица с глубиной для каждого участка изображения. Следует отметить, что несколько важных аспектов. Во-первых, при условии, что изображения выравнены на один уровень по вертикали, поиск схожих зон можно производить только в рамках

одной строки матрицы. Во-вторых, важным аспектом является подбор размера зоны. Так при слишком большом размере может быть получена значительная погрешность в дальнейших расчетах, а при малых размерах начинают проявляться шумы, представляющие собой ложный вывод пар зон. В-третьих, важными аспектами для всех вычислений параметры стереокамеры: фокусное расстояние линз и диспозиция между оптическими центрами камер. Помимо прямого влияния на точность вычислений, не зная точно диспозиции, нельзя определить в матрицах зоны, выходящие за общую видимую зону, а, следовательно, будут ложные срабатывания в этих зонах, что приведет к росту шума.

Был использован модифицированный вариант метода, включающий предел отклонения зон соответствия стереопары, что позволяет уменьшить необходимое число зон для обхода в пределах одной строки, особенно при обработке изображений высокого разрешения. Сглаживание шумов не проводилось. Фокусное расстояние использовалось фиксированное, так как невозможен замер измененного фокусного расстояния во время съемки.

Используя код, приложенный к статьям «Simultaneous Localization And Mapping with depth Prediction using Capsule Networks for UAVs» и «High Quality Monocular Depth Estimation via Transfer Learning», были воспроизведены и протестированы дизайны CNN и CapsNet, собраны замеры целевых метрик (табл. 1).

Таблица 1

Результаты замеров метрик

<i>Метрики</i>	<i>Стереопара</i>	<i>CNN</i>	<i>CapsNet</i>
Accuracy	62%	96.8%	98.7%
Затрачиваемая оперативная память, МБ	373.	700.1	717.8
Время обработки, сек.	0.9909	3.014	1.9386

Заключение. В статье представлен сравнительный анализ методов построения карты глубины: построение по стереопаре, CNN, CapsNet, — проведены замеры ключевых метрик. Из результатов

следует, что лучшим методом для использования в системах реального времени является CapsNet дизайна Сунил Пракаш и Гаэлан Гу, так как является оптимальным в соотношении время/качество в сравнении с другими методами. Построение по стереопаре, имея преимущество в скорости, не может быть использовано из-за низкой точности, пока не будет решена проблема с шумами.

СПИСОК ЛИТЕРАТУРЫ

1. Ibraheem Alhashim High Quality Monocular Depth Estimation via Transfer Learning / Ibraheem Alhashim, Peter Wonka. — URL: <https://arxiv.org/abs/1812.11941> (date of applification 05.05.2023). — Text: electronic
2. Clement Godard Unsupervised Monocular Depth Estimation with Left-Right Consistency / Clement Godard, Oisín Mac Aodha, Gabriel J. Brostow. — URL: <https://arxiv.org/abs/1609.03677> (date of applification 05.05.2023). — Text: electronic
3. Sunil Prakash Simultaneous, Localization And Mapping with depth Prediction using Capsule Networks for UAVs / Sunil Prakash, Gaelan Gu.— URL: <https://arxiv.org/abs/1808.05336> (date of applification 05.05.2023). — Text: electronic
4. Fei Wu Depth Estimation maps of lidar and stereo images / Fei Wu, Luoyu Chen. — URL: <https://arxiv.org/abs/1609.03677> (date of applification 05.05.2023). — Text: electronic
5. D. Scharstein A taxonomy and evaluation of dense two-frame stereo correspondence algorithms / D. Scharstein, R. Szeliski. — Текст: непосредственный // International Journal of Computer Vision — 2002 — № 47, — P. 7-42.
6. D. Scharstein High-accuracy stereo depth maps using structured light / D. Scharstein, R. Szeliski. — Direct text // IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, 2003. — P. 195-202.
7. D. Scharstein Learning conditional random fields for stereo / D. Scharstein, C. Pal. — Direct text // IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2007.
8. H. Hirschmüller Evaluation of cost functions for stereo matching / H. Hirschmüller, D. Scharstein. — Direct text // IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2007.
9. D. Scharstein High-resolution stereo datasets with subpixel-accurate ground truth / D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nesić,

- X. Wang, P. Westling. — Direct text // German Conference on Pattern Recognition, 2014.
10. Nathan Silberman Indoor Segmentation and Support Inference from RGBD Images Courant Institute / Nathan Silberman, Derek Hoiem, Pushmeet Kohli, Rob Fergu. — Direct text. // European conference on computer vision, Springer, 2012. — P. 746-760.
11. G. Huang Densely connected convolutional networks, / G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger. — Direct text // IEEE Conference on Computer Vision and Pattern Recognition, 2017 — P. 2261-2269.
12. J. Lehtinen Noise2Noise: Learning image restoration without clean data / J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, T. Aila — Текст: непосредственный // Proceedings of the 35th International Conference on Machine Learning, Vol. 80, 2018, — P. 2965-2974.
13. Kalman R. E. A new approach to linear filtering and prediction problems, / Kalman R. E. — Direct text // Journal of Basic Engineering. — 1960. — № 82. — Pp. 35-45.
14. E. G. Richardson H.264 and MPEG-4 Video Compression: Video Coding for Next-generation Multimedia / E. G. Richardson. 2003. — Direct text.