Matching

1. Введение. *Мэтчинг изображений* – часть многих приложений компьютерного зрения, таких как регистрация изображений, калибровка камеры и распознавание объектов, является также задачей установления соответствий между двумя изображениями одной и той сцены / объекта.

Общий подход состоит в обнаружении множества ключевых точек, каждой из которых сопоставляется дескриптор. После того как свойства и дескрипторы были извлечены из изображений, устанавливаются соответствия. Качество мэтчинга напрямую зависит от качеств соответствующих дескрипторов и детекторов.

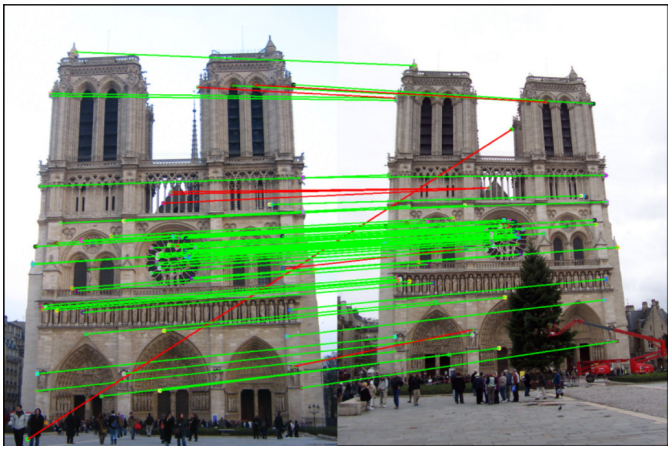

Рис. 1. Пример мэтчинга.

Как и в случае с выделением признаков, до возрастающей популярности нейросетевых подходов долгое время стандартными алгоритмами мэтчинга были:

1. Brute-Force Matcher – простой жадный перебор.
2. FLANN(Fast Library for Approximate Nearest Neighbors) – оптимизированный алгоритм ближайших соседей.

2. Исследованные работы. С развитием глубокого обучения стало появляться всё больше статей о новых подходах в мэтчинге.

2.1. SuperGlue. Learning Feature Matching with Graph Neural Networks. Подход устанавливает соответствия для уже готовых векторов признаков, изъятых из изображения. SuperGlue использует графовую нейронную сеть и механизм внимания. Помимо этого, достоинством SuperGlue является обнаружение точек, не имеющих соответствия, для чего добавляется отдельная метка.

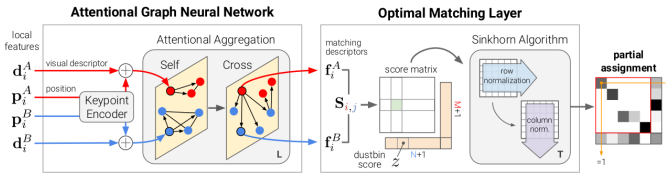

Рис. 2. Архитектура SuperGlue.

Рассматриваемая модель состоит из двух основных компонентов – нейронная графовая сеть с механизмом внимания и слой нахождения оптимального соответствия. Первый компонент кодируют информацию о позиции ключевой точки и её дескрипторе в один вектор, после чего использует механизм само-внимания и кросс-внимания для получения более комплексного представления. Второй компонент создаёт матрицу размера MxN, где M и N - количества ключевых точек на рассматриваемых изображениях. Далее ищутся оптимальные соответствия алгоритмом Sinkhorn (для 100 итераций).

Модель содержит 12M параметров и в режиме реального времени способна обрабатывать пару изображений 640x480 за 69 ms (15 fps) на NVIDIA GTX 1080 GPU. Авторы утверждают, что модель способна находить соответствия даже для достаточно разных углов ракурса. Ссылка на статью и репозиторий в списке литературы.

2.2. LoFTR. Detector-Free Local Feature Matching with Transformers. Авторы были вдохновлены успехом SuperGlue и тоже решили приспособить архитектуру Transformer для целей мэтчинга. Также предложено использовать объединённую архитектуру для обнаружения ключевых точек, сопоставления векторов признаков и мэтчинга в отличие от типичных подходов.

LoFTR имеет 4 основных компонента. В первом компоненте для двух изображений A и B свёрточная нейронная сеть (ResNet-18) конструирует по две карты признаков разных размеров. Меньшие карты признаков преобразуются в одномерные векторы и снабжаются пози-

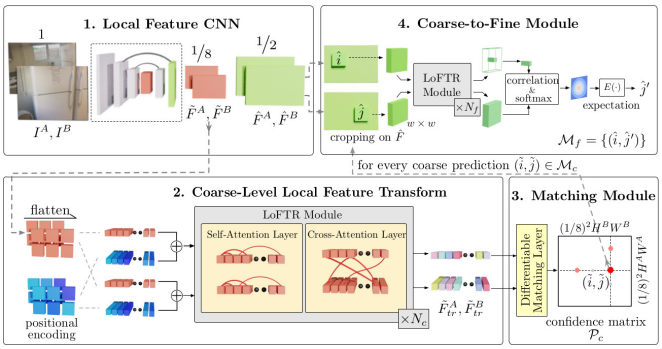

Рис. 3. Архитектура LoFTR.

ционной информацией, после чего подаются на вход модулю LoFTR с механизмами самовнимания и кроссвнимания. Из полученных представлений далее получается доверительная матрица, на основе которой извлекаются примерные сопоставления. Для каждого такого сопоставления рассматривается локальное окно на карте признаков и на основе этого окна подбираются окончательные соответствия.

Модель достигает SOTA результатов в задачах Visual Localization и Relative Pose Estimation. Для пары изображений разрешением 640x480 модель выдаёт результат за 116 мс на RTX 2080Ti.

Литература

1. Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, Andrew Rabinovich. SuperGlue: Learning Feature Matching with Graph Neural Networks // CVPR. 2020.

   Article: https://arxiv.org/abs/1911.11763

   GitHub: https://github.com/magicleap/SuperGluePretrainedNetwork

2. Jiaming Sun, Zehong Shen, Yuang Wang, Hujun Bao, Xiaowei Zhou. LoFTR: Detector-Free Local Feature Matching with Transformers. // CVPR. 2021.

   Article: https://arxiv.org/abs/2104.00680

   GitHub: https://github.com/zju3dv/LoFTR