УНИВЕРСИТЕТ ИТМО
НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
май–июнь 2020 Том 20 № 3 ISSN 2226-1494 http://ntv.itmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
May–June 2020 Vol. 20 No 3 ISSN 2226-1494 http://ntv.itmo.ru/en/
НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК УНИВЕРСИТЕТ ИТМО ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

УДК 004.272 004.032.26 doi: 10.17586/2226-1494-2020-20-3-402-409

ДИСТИЛЛЯЦИЯ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ ДЛЯ ДЕТЕКТИРОВАНИЯ И ОПИСАНИЯ КЛЮЧЕВЫХ ТОЧЕК ИЗОБРАЖЕНИЙ

А.В. Ященко[a,b], А.В. Беликов[a], М.В. Петерсон[a], А.С. Потапов[a]

[a] СингуляритиЛаб, Санкт-Петербург, 198152, Российская Федерация
[b] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация
Адрес для переписки: yashenkoxciv@gmail.com

Информация о статье
Поступила в редакцию 15.04.20, принята к печати 10.05.20
Язык статьи — русский

Ссылка для цитирования: Ященко А.В., Беликов А.В., Петерсон М.В., Потапов А.С. Дистилляция нейросетевых моделей для детектирования и описания ключевых точек изображений // Научно-технический вестник информационных технологий, механики и оптики. 2020. Т. 20. № 3. С. 402–409. doi: 10.17586/2226-1494-2020-20-3-402-409

Аннотация

Предмет исследования. Методы сопоставления и классификации изображений, а также синхронного определения местоположения и составления карты местности широко применяются на встраиваемых и мобильных устройствах. Наиболее ресурсоемкой частью их реализации является выделение и описание ключевых точек изображений. Классические методы выделения и описания ключевых точек могут исполняться в масштабе реального времени на мобильных устройствах. Вместе с тем для современных нейросетевых методов, обладающих лучшим качеством, такой подход затруднен из-за снижения быстродействия. Таким образом, задача повышения быстродействия нейросетевых моделей для детектирования и описания ключевых точек является актуальной. С этой целью выполнено исследование дистилляции — одного из способов редукции нейросетевых моделей, что позволяет получить более компактную модель детектирования и описания ключевых точек, а также процедуры получения модели. **Метод.** Предложен способ сопряжения исходной и более компактной новой модели для последующего ее обучения по выходным значениям исходной модели. С этой целью новая модель обучается реконструировать выходные данные исходной модели без использования разметки изображений. На вход обеих сетей поступают одинаковые изображения. **Основные результаты.** Протестирован способ дистилляции нейронных сетей для задачи детектирования и описания ключевых точек. Предложены целевая функция и параметры обучения, обеспечивающие наилучшие результаты в рамках выполненного исследования. Введены новый набор данных для тестирования методов выделения ключевых точек и новый показатель качества выделяемых ключевых точек и соответствующих им локальных признаков. Применение обучения новой модели предложенным способом с тем же количеством параметров позволило получить большую точность сопоставления ключевых точек по сравнению с исходной моделью. Новая модель со значительно меньшим количеством параметров обеспечивает точность сопоставления точек, близкую к исходной модели. **Практическая значимость.** Предложенным способом получена более компактная модель для детектирования и описания ключевых точек изображений. Это дает возможность применять модель на встраиваемых и мобильных устройствах для синхронного определения местоположения и составления карт местности. Применение предложенной модели может повысить эффективность работы сервиса по поиску изображений на серверной стороне.

Ключевые слова
глубокое обучение, детектирование ключевых точек, локальные признаки

doi: 10.17586/2226-1494-2020-20-3-402-409

DISTILLATION OF NEURAL NETWORK MODELS FOR DETECTION AND DESCRIPTION OF IMAGE KEY POINTS

A.V. Yashchenko[a,b], A.V. Belikov[a], M.V. Peterson[a], A.S. Potapov[a]

[a] SingularityLab, Saint Petersburg, 198152, Russian Federation
[b] ITMO University, Saint Petersburg, 197101, Russian Federation
Corresponding author: yashenkoxciv@gmail.com

Article info
Received 15.04.20, accepted 10.05.20
Article in Russian

For citation: Yashchenko A.V., Belikov A.V., Peterson M.V., Potapov A.S. Distillation of neural network models for detection and description of image key points. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2020, vol. 20, no. 3, pp. 402–409 (in Russian). doi: 10.17586/2226-1494-2020-20-3-402-409

Abstract

Subject of Research. Image matching and classification methods, as well as synchronous location and mapping, are widely used on embedded and mobile devices. Their most resource-intensive part is the detection and description of the image key points. In case of classical methods for detection and description of key points they can be executed in real time on mobile devices but for modern neural network methods with better quality, such approach is difficult due to trading off performance. Thus, the issue of speeding for neural network models applied for the detection and description of key points is currently topical. The subject of research is distillation as one of the methods for reducing neural network models. The aim of the study is to obtain more compact model for detection and description of key points and a description of the procedure for this model design. **Method.** We proposed a method for pairing the original and more compact new model for its subsequent training on the output values of the original model. In this regard, the new model is learned to reconstruct the output of the original model without using image labels. Both networks accept identical images as input. **Main Results.** Neural network distillation method for detection and description of key points is tested. The objective function and training parameters that provide the best results in the framework of the study are proposed. A new data set is created for testing key point detection methods, and a new quality indicator of the allocated key points and their corresponding local features is added. New model training in the way suggested with the same number of parameters, shows greater accuracy in key points compared to the original model. A new model with a significantly smaller number of parameters shows the accuracy of point matching close to the accuracy of the original model. **Practical Relevance.** More compact model for detection and description of image key points is created applying the proposed method. The model is applicable on embedded and mobile devices for synchronous location and mapping. Such model application can also increase the service efficiency of the image search on the server side.

Keywords

deep learning, keypoint detection, local image descriptors

Введение

Ключевые точки — это точки на изображении, характеризующие их локальные особенности и необходимые для последующего сопоставления. Из участков изображения, соответствующих ключевым точкам, выделяются признаковые описания, или дескрипторы, используемые далее для сопоставления изображений, классификации и других задач.

Основными критериями выбора ключевых точек являются:

— воспроизводимость, т. е. возможность их повторного детектирования из других положений камеры, при изменении освещенности, масштаба и других искажениях;

— качество выделяемых признаков, которые должны позволять правильно сопоставлять точки, выделенные с разных ракурсов.

Разработано множество методов детектирования ключевых точек и выделения их признаковых описаний [1–4]. Алгоритмы на основе глубокого обучения, разработанные за последние несколько лет, превосходят традиционные алгоритмы по точности детектирования и сопоставления ключевых точек [3–5].

Применение термина «ключевая точка» справедливо также при детектировании ключевых точек лица и тела человека [6, 7]. В этом случае точки связаны с определенными органами и участками тела и могут обозначать края губ, глаз и т. п. К сожалению, такой подход, а именно — описание ключевых точек заранее в виде меток на произвольных изображениях, практически неприменим, из-за сложности с определением семантических свойств таких точек.

Современные алгоритмы на основе глубокого обучения для обработки изображений чаще всего основаны на сверточных нейросетевых моделях, которые способны строить представления, не уступающие, а часто и превосходящие по качеству, представления, разработанные вручную. Передовые результаты, связанные с классификацией [8, 9], генерацией изображений [10], а также детектированием [11] и повторной идентификацией [12] пешеходов, получены с применением глубоких сверточных нейросетевых моделей. Данные модели могут содержать десятки слоев и сложно организованный процесс прямого распространения сигнала, однако для задач, связанных с одновременной локализацией и построением карт, справедливо ограничение для времени работы и используемой памяти таких моделей.

Применение детекторов ключевых точек для задачи одновременной локализации и построения карт подразумевает использование моделей на компактных устройствах, часто сильно ограниченных относительно устройств, на которых обучаются нейросетевые модели. Несмотря на то, что вывод в глубоких моделях гораздо менее требователен к вычислительным ресурсам, чем обучение, для использования нейросетевых моделей на компактных устройствах, как правило, требуется их «облегчение».

Существует несколько подходов для «облегчения», упрощения или прореживания (pruning) моделей глубокого обучения [13, 14]. Большинство из них основано на предположении, что нейросетевые модели избыточно параметризированы, и что удаление искусственных нейронов, не вносящих большой вклад в результат целевой метрики, можно осуществить без значительной потери качества, а иногда даже получить улучшение целевой метрики, например, точности распознавания. Этот эффект можно объяснить тем, что прореживание равносильно регуляризации сети. Как правило, алгоритмы такого рода прореживания включают операции «физического» удаления отдельных элементов сети — нейронов в полносвязной сети или фильтров в сверточных архитектурах.

Другие подходы могут быть основаны на сжатии фильтров сверточной сети в частотной области [15] или обучении новой, более компактной модели, используя подход на основе дистилляции исходной модели, которую необходимо упростить [16]. Последний подход применяется в данном исследовании. Кроме того, анализ моделей глубокого обучения остается особым ремеслом и, как показано в [17], может осуществляться с помощью других глубоких моделей.

Цель работы состоит в разработке и исследовании способа сопряжения исходной и новой моделей для дальнейшего обучения новой модели, а также получения целевой функции и набора данных для обучения и оценки качества работы модели.

Описание исходной модели

Представлен алгоритм, использующий базовую модель (учитель) для обучения новой модели с меньшим числом параметров (ученик), а также результаты обучения новой модели. В качестве учителя выбрана модель SuperPoint [3] — полносверточная сеть, принимающая полноразмерное изображение и порождающая два типа выходных данных:

1) тензор (многомерный массив) для последующего детектирования ключевых точек;
2) тензор, содержащий дескрипторы для каждой области 8 × 8 пикселей.

Таким образом, для задачи детектирования и описания ключевых точек большая часть сети объединена, однако для порождения ключевых точек и дескрипторов, сеть-учитель разделяется на два «рукава», содержащие данные для извлечения ключевых точек и дескрипторов. Дескрипторы для точек получаются с помощью билинейной интерполяции тензора дескрипторов в координатах полученных ключевых точек.

Процесс вывода в модели SuperPoint (SP) проиллюстрирован на рис. 1. Этот процесс включает несколько этапов:

— входное изображение подается кодировщику (рис. 1, *а*);
— используя представление (блок *z*), порожденное кодировщиком, вычисляются выходные карты признаков, используемые для детектирования ключевых точек (блок КТ) и выделения дескрипторов (блок ДК) (рис. 1, *б*);
— выходные карты (рис. 1, *в*) признаков КТ и ДК обрабатываются дополнительно.

Кодировщик, производящий общее представление, позволяет последовательно снижать размерность входного изображения с помощью блоков, включающих сверточные слои, операции субдискретизации и применение нелинейной активационной функции. Кодировщик разработан таким образом, что каждый элемент карты признаков содержит «ячейку» размером 8 × 8 пикселей, а детектирование ключевых точек проводится с помощью нормированной экспоненциальной функции (softmax) к каждой «ячейке». Далее после удаления канала, указывающего на отсутствие ключевой точки, карта признаков преобразуется к форме входного изображения.

Обучение модели предполагает оптимизацию сложной функции ошибки, которая состоит из двух компонент:

1) ошибки детектирования ключевых точек *Lp*;
2) ошибки соответствия дескрипторов *Ld*.

Для каждого входного изображения формируется пара — второе изображение, полученное после применения гомографического искажения к первому изображению, параметризованного случайно. Таким образом, появляется пара изображений и соответствие ключевых точек для этих изображений. Формируется часть функции ошибки $Lp = Lp(X,Y) + Lp(X',Y')$, где *Lp* — кроссэнтропия, минимизация которой означает, что сеть справляется с детектированием точек на исходном и искаженном изображении, а X, X', Y, Y' обозначают: первое входное изображение, искаженное первое входное изображение, второе входное изображение и искаженное второе входное изображение соответственно.

Процесс создания функции ошибки для дескрипторов состоит из трех этапов:

1) получение ошибки для ожидаемых сопоставлений $d_{\text{прав}}(desc1, desc2) = 1 - \cos(desc1[ids], desc2)$, где *ids* — ожидаемые индексы сопоставлений дескрипторов;
2) получение ошибки для неправильно сопоставленных дескрипторов $d_{\text{ошиб}}(desc1, desc2) = \cos(desc1[ids], desc2)$, где *ids* — индексы сопоставлений дескрипторов, не совпадающие с ожидаемыми;
3) получение ошибки для случайных сопоставлений $d_{\text{rand}}(desc1[i], desc2[j]) = \cos(desc1[i], desc2[j])$, если $\cos(desc1[i], desc2[j]) > 0{,}2$, иначе 0, где i, j — случайные индексы; cos — косинус угла между дескрипторами.

Итоговая функция ошибки $Ld = d_{\text{прав}} + d_{\text{ошиб}} + d_{\text{rand}}$.

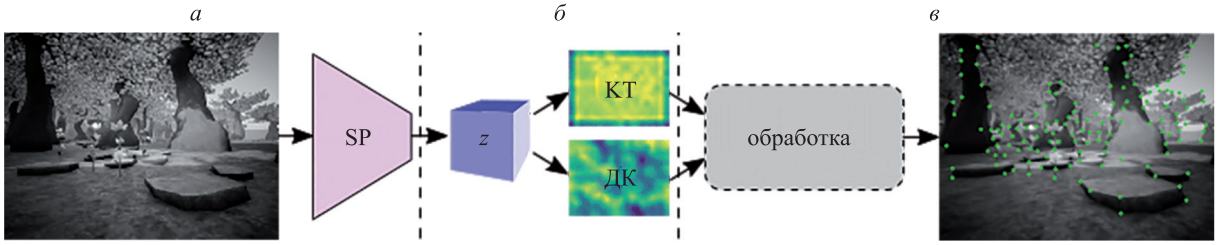

Рис. 1. Процесс прямого распространения (сигнала) в модели SuperPoint: получение общего представления *z* (*а*); получение данных для выделения ключевых точек и локальных признаков; *в* — обработка выходов SuperPoint (*б*)

Ожидаемые сопоставления вычисляются путем проецирования точек из плоскости изображения X в плоскость изображения X' и вычисления ближайших соседей.

Описанный подход используется для обучения исходной модели – учителя. При этом исходный код недоступен, и воспроизвести результаты, используя исходную кодовую базу, продемонстрированные в [3] невозможно. Более того, обучение модели с меньшим количеством параметров «с нуля» может привести к менее стабильному процессу обучения и потребует корректировки значений гиперпараметров. Для решения указанных проблем предлагается способ на основе дистилляции нейронных сетей.

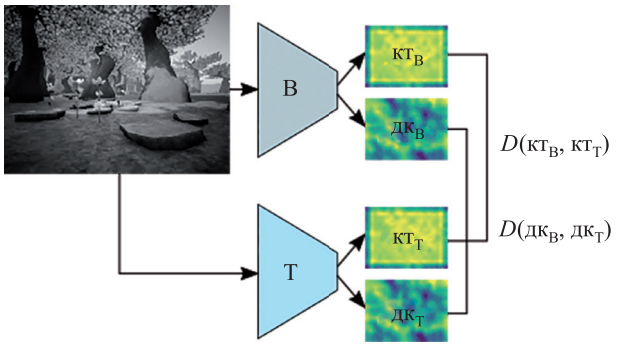

Рис. 2. Сопряжение базовой и новой моделей. D — среднеквадратичное отклонение многомерных массивов $кт_B$ и $кт_T$, $дк_B$ и $дк_T$, где $кт_B$ и $кт_T$ — выходные данные для детектирования ключевых точек; $дк_B$ и $дк_T$ — выходные данные для локальных признаков

Описание предлагаемого способа и эксперименты

Основная цель экспериментов заключается в получении более компактной модели, позволяющей эффективно детектировать и сопоставлять ключевые точки. В качестве базовой модели, подлежащей «сжатию», используется SuperPoint, описанный ранее. Для получения более компактной модели используется алгоритм дистилляции или разновидность обучения с учителем (rich-supervision), суть которого заключается в попытке обучить новую, обычно более компактную, модель по выходам базовой сети. Для входного изображения базовая модель порождает некоторое представление, используемое в функции ошибки для обучения новой модели.

Функция ошибки для обучения или дистилляции новой модели, как правило, включает некоторую меру расстояний, минимизируя которую, обучается новая модель. Один из экспериментов подразумевает воспроизведение базовой модели с новым алгоритмом обучения. Однако основная цель заключается в получении меньшей модели со сравнимым критерием качества относительно базовой модели.

Для воспроизведения модели предлагается использовать такую же архитектуру сети как для базовой модели. Обозначим базовую модель как B, а ее выходы для ключевых точек и дескрипторов — $кт_B$ и $дк_B$ соответственно. Новая модель T воспроизводит работу B, но параметры T представлены случайными числами, близкими к нулю. Подход к сопряжению моделей T и B проиллюстрировано на рис. 2.

Для обучения модели T предлагается минимизировать среднеквадратичное отклонение выходных тензоров $кт_B$ и $кт_T$, как и $дк_B$ и $дк_T$.

Таким образом, функция ошибки для обучения модели T имеет вид $L = D(кт_B, кт_T) + D(дк_B, дк_T)$. Функция ошибки является дифференцируемой и может использоваться вместе с алгоритмом градиентного спуска для оптимизации параметров модели T.

Для обучения новой модели использовались синтетические изображения, полученные с помощью средств моделирования AirSim [18].

Точность сопоставления для базовой модели составляет 0,7456, а для новой модели — 0,7515. Примечательно, что модель с идентичной архитектурой и описанным алгоритмом обучения позволяет получить несколько лучший результат метрики полноты. Однако основная цель экспериментов заключается в получении более компактной модели T. Для этого была обучена модель с вдвое меньшим количеством фильтров, чем базовая модель, метрика полноты для этой модели составила 0,7410. В результате уменьшения количества фильтров осталось только 23 % от общего количества параметров исходной модели.

Так как базовая модель является полносверточной и производит карты признаков, содержащие «ячейки» исходного изображения, то минимизация среднеквадратичного отклонения по каждому элементу таких карт признаков может приводить к «размытию» карт признаков, производимых моделью T. Предлагается минимизировать среднеквадратичное отклонение пространственных градиентов (применяется фильтр Собеля) для $кт_B$ и $кт_T$. В дополнительном слагаемом целевой функции взятие пространственных градиентов обозначено как G.

Тогда новая функция ошибки:

$$L = D(кт_B, кт_T) + D(дк_B, дк_T) + D(G(кт_B), G(кт_T)).$$

При дополнении функции ошибки модель T, являясь идентичной B, дает наилучший результат метрики полноты — 0,7584.

Генерация базы изображений и тестирование способа

Для обучения модели на изображениях с различными условиями освещения в работе использовался набор виртуальных сцен на базе программного обеспечения Unreal Engine, а также библиотека AirSim [18]. Данная библиотека предоставляет инструментарий для получения изображений виртуальных сцен и соответствующих карт глубины с задаваемыми внутренними и внешними параметрами камеры. Эти данные использовались авторами в качестве опорных для оценки точности сопоставления локальных признаков, полученных на выходе исследуемых моделей. Пример изображений и карт глубины, полученных из базы AirSim представлен на рис. 3.

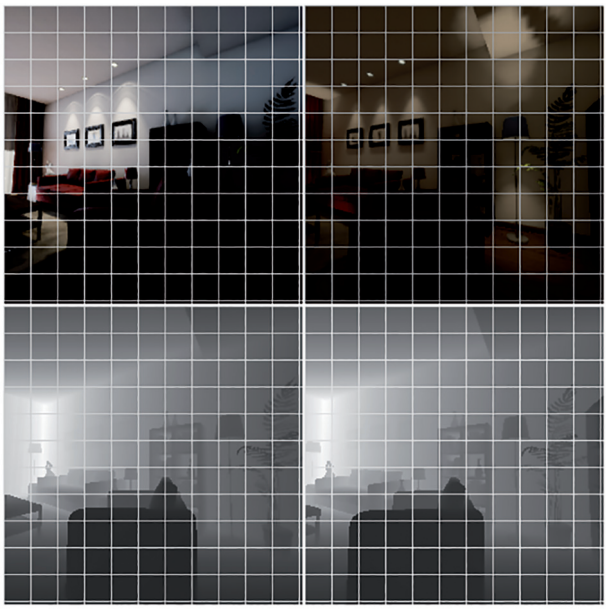

Рис. 3. Визуализированная в AirSim сцена с картами глубины.
Размер масштабной сетки 23 × 23 пикселей

Для каждого изображения I_i в базе хранится соответствующая ему карта глубины D_i, внутренние параметры камеры K_i, внешние параметры: матрица поворота $\mathbf{R}_i$, задающая ориентацию камеры в мировых координатах, а также вектор смещения $\mathbf{t}_i$. Траектории камеры для последовательностей при разных условиях освещения совпадают. Матрица внутренних параметров камеры [19] фиксирована и определена в AirSim как: $\mathbf{K} = [[W/2, 0, W/2], [0, W/2, H/2], [0, 0, 1]]$, где W, H — ширина и высота изображения.

Стоит отметить, что используемая версия AirSim v1.2.0 по умолчанию вместо нормальных расстояний от точек сцены по оси Z в системе отсчета камеры возвращает расстояния до проекционного центра камеры, поэтому для получения Z координат наблюдаемых точек сцены необходима дополнительная конвертация. Тогда, исходя из известных параметров камеры, можно: восстановить трехмерные координаты ключевых точек изображения I_i в системе отсчета i-ой камеры; перенести их в систему отсчета $(i + 1)$-й камеры; спроецировать эти точки на изображение I_{i+1}; оценить расстояния до соответствующих сопоставленных точек; в итоге рассчитать точность сопоставления для данной пары изображений.

Для обучения модели, совмещающей выделение ключевых точек и вычисление дескрипторов, желательно иметь единственный критерий качества модели, так как при обучении модели существует баланс между компонентами целевой функции. Предложено применить гармоническое среднее с использованием точности дескрипторов и полноты детектора. Полнота детектора соответствует воспроизводимости ключевых точек.

Гармоническое среднее дает большие значения для множества незначительно отличающихся величин, чем для множества с большой разницей, в отличие от арифметического среднего. Значения гармонического среднего для ключевых показателей представлены в табл. 1.

В работе [5] приведены точность и полнота для оценки качества дескрипторов, и воспроизводимость для оценки качества детектора ключевых точек. Полнота определяется как:

$$\text{recall} = \frac{\text{количество правильных сопоставлений}}{\text{количество соответствий}}.$$

Эта величина обладает следующими особенностями. При попиксельном вычислении числа соответствий в знаменателе всегда будут большие числа, пропорциональные площади изображения. Использование всех участков изображения для сравнения дескрипторов не отражает качество работы системы, так как в практических приложениях, как правило, дескрипторы выделяются только с релевантных точек.

Метрика для оценки качества сопоставления – это отношение количества правильно сопоставленных пар локальных признаков к общему количеству сопоставленных пар. Данная величина аналогична точности (precision), применяемой при оценке результата информационного поиска. Здесь необходимо использовать термин «точность» для применяемой метрики. В таком определении точность будет зависеть от выбора исходного изображения, так, например, если на изображении I_1 выделена одна точка (и она правильно сопоставлена), а на I_2 — сто, то точность будет 1,0 для сопоставления из $I_1 \rightarrow I_2$ и 0,01 для $I_2 \rightarrow I_1$.

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}),$$

где TP — число правильно сопоставленных по дескрипторам точек; FP — неправильно сопоставленных (по дескрипторам точек). Эту зависимость можно устранить, взяв среднее значение точности от прямого и обратного сопоставлений.

Также важным показателем качества является воспроизводимость, равная доле точек одного изображения, детектируемых на другом. Эта метрика идентична точности при использовании идеального генератора дескрипторов, т. е. когда все точки сопоставляются правильно. Повторяемость для точек соответствует полноте в оценке информационного поиска. Для повторяемости также имеет смысл считать среднюю величину для сопоставлений $I_1 \rightarrow I_2$ и $I_2 \rightarrow I_1$.

При информационном поиске $\text{recall} = \text{TP}/(\text{TP} + \text{FN})$, для ключевых точек обозначим $\text{recall}(\text{points}(I_1), \text{points}(I_2)) = (\text{points}(I_1) \cap \text{points}(I_2))/\text{points}(I_2)$, т. е. доля точек на I_2 воспроизведенная на I_1. Имея точность и полноту можно посчитать $F1 = 2 \times (\text{precesion}(d, p) \times \times \text{recall}(p))/(\text{precesion}(d, p) + \text{recall}(p))$.

Таблица 1. Сравнение интегральных критериев $F1$ и арифметического среднего

Точность	Полнота	Гармоническое среднее ($F1$)	Арифметическое среднее
0,61	0,55	0,58	0,58
0,65	0,51	0,57	0,58
0,16	1,00	0,28	0,58

Таблица 2. Сравнение редуцированной и оригинальной модели в условиях яркостной или ракурсной изменчивости

Модель	Воспроизводимость детектирования, отн. ед.		Точность сопоставления, отн. ед.	
	Яркостная изменчивость	Ракурсная изменчивость	Яркостная изменчивость	Ракурсная изменчивость
Оригинальная модель SuperPoint	0,6108	0,5370	0,7909	0,7139
Редуцированная модель	0,5658	0,4905	0,7953	0,6802

Таблица 3. Результаты тестирования модели на AirSim Village

Модель	Метрика		
	Точность дескрипторов	Повторяемость точек	Гармоническое среднее
Оригинальная модель SuperPoint	0,7515	0,4482	0,5615
Редуцированная модель	0,7192	0,4492	0,5530

Таблица 4. Результаты тестирования модели на AirSim Fantasy Village

Модель	Метрика		
	Точность дескрипторов	Повторяемость точек	Гармоническое среднее
Оригинальная модель SuperPoint	0,8829	0,5500	0,6778
Редуцированная модель	0,8563	0,5436	0,6650

Пара локальных признаков считается верно сопоставленной, если ошибка репроекции ключевых точек с первого изображения в паре на второе изображение не превышает заданного порога. В рассмотренном случае был установлен порог в три пикселя.

База [20] содержит два типа наборов изображений: изображения с яркостной и ракурсной изменчивостями. Результаты тестирования представлены в табл. 2, а для базы AirSim [18] в табл. 3 и 4. Таким образом, видно, что качество выделения ключевых точек изменилось незначительно.

Заключение

Рассмотрено применение способа дистилляции к моделям на основе глубокого обучения для задачи детектирования и описания ключевых точек. Предложен алгоритм обучения новой модели детектирования и описания ключевых точек на основе существующей. В алгоритме новая модель обучается воспроизводить (дистиллировать) выходные данные базовой модели. Для этого минимизируется сложная функция ошибки, включающая меру расстояния для тензоров ключевых точек и дескрипторов, а также регуляризация в виде пространственного градиента для карт признаков ключевых точек. К недостатку текущего алгоритма можно отнести зависимость от качества синтетических данных. Дальнейшим направлением исследований авторами ставится переход к обучению без учителя, что может устранить данный недостаток.

В результате экспериментов не удалось получить идентичную модель – значения функции ошибки всегда больше нуля. Однако по целевой метрике – точности сопоставления ключевых точек – новая модель превосходит базовую. Точность сопоставления более компактной новой модели отличается незначительно от точности исходной модели.

Литература

1. Bay H., Tuytelaars T., Van Gool L. Surf: Speeded up robust features // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2006. V. 3951. P. 404–417. doi: 10.1007/11744023_32
2. Rublee E., Rabaud V., Konolige K., Bradski G. ORB: An efficient alternative to SIFT or SURF // Proc. of the International Conference on Computer Vision (ICCV 2011). 2011. P. 2564–2571. doi: 10.1109/ICCV.2011.6126544
3. DeTone D., Malisiewicz T., Rabinovich A. SuperPoint: Self-supervised interest point detection and description // Proc. 31st Meeting of the IEEE/CVF IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2018. P. 337–349. doi: 10.1109/CVPRW.2018.00060

References

1. Bay H., Tuytelaars T., Van Gool L. Surf: Speeded up robust features. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2006, vol. 3951, pp. 404–417. doi: 10.1007/11744023_32
2. Rublee E., Rabaud V., Konolige K., Bradski G. ORB: An efficient alternative to SIFT or SURF. *Proc. of the International Conference on Computer Vision (ICCV 2011)*, 2011, pp. 2564–2571. doi: 10.1109/ICCV.2011.6126544
3. DeTone D., Malisiewicz T., Rabinovich A. SuperPoint: Self-supervised interest point detection and description. *Proc. 31st Meeting of the IEEE/CVF IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018, pp. 337–349. doi: 10.1109/CVPRW.2018.00060

4. Ono Y., Fua P., Trulls E., Yi K. LF-Net: learning local features from images // Advances in Neural Information Processing Systems. 2018. P. 6234–6244.
5. Mikolajczyk K., Schmid C. A performance evaluation of local descriptors // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2005. V. 27. N 10. P. 1615–1630. doi: 10.1109/TPAMI.2005.188
6. Cao Z., Hidalgo G., Simon T., Wei S., Sheikh Y. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2019. Early access. doi: 10.1109/TPAMI.2019.2929257
7. Baltrušaitis T., Robinson P., Morency L.-P. Openface: an open source facial behavior analysis toolkit // Proc. IEEE Winter Conference on Applications of Computer Vision (WACV). 2016. P. 7477553. doi: 10.1109/WACV.2016.7477553
8. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition // Proc. 3rd International Conference on Learning Representations (ICLR). 2015.
9. Iandola F., Moskewicz M., Karayev S., Girshick R., Darrell T., Keutzer K. Densenet: Implementing efficient convnet descriptor pyramids [Электронный ресурс]. URL: https://arxiv.org/abs/1404.1869, свободный. Яз. англ. (дата обращения: 17.01.2020).
10. Brock A., Donahue J., Simonyan K. Large scale gan training for high fidelity natural image synthesis // Proc. 7th International Conference on Learning Representations (ICLR). 2019.
11. Redmon J., Divvala S., Girshick R., Farhadi A. You only look once: Unified, real-time object detection // Proc. 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 779–788. doi: 10.1109/CVPR.2016.91
12. Zheng Z., Yang X., Yu Z., Zheng L., Yang Y., Kautz J. Joint discriminative and generative learning for person re-identification // Proc. 32nd IEEE /CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019. P. 2133–2142. doi: 10.1109/CVPR.2019.00224
13. Huang Q., Zhou K., You S., Neumann U. Learning to prune filters in convolutional neural networks // Proc. 18th IEEE Winter Conference on Applications of Computer Vision (WACV). 2018. P. 709–718. doi: 10.1109/WACV.2018.00083
14. Gomez A.N., Zhang I., Kamalakara S.R., Madaan D., Swersky K., Gal Y., Hinton G.E. Learning sparse networks using targeted dropout [Электронный ресурс]. URL: https://arxiv.org/abs/1905.13678, свободный. Яз. англ. (дата обращения: 18.03.2020).
15. Wang Y., Xu C., You S., Tao D., Xu C. CNNpack: Packing convolutional neural networks in the frequency domain // Advances in Neural Information Processing Systems. 2016. P. 253–261.
16. Hinton G., Vinyals O., Dean J. Distilling the knowledge in a neural network [Электронный ресурс]. URL: https://arxiv.org/abs/1503.02531, свободный. Яз. англ. (дата обращения: 06.02.2020).
17. Wang J., Gou L., Zhang W., Yang H., Shen H.-W. Deepvid: Deep visual interpretation and diagnosis for image classifiers via knowledge distillation // IEEE Transactions on Visualization and Computer Graphics. 2019. V. 25. N 6. P. 2168–2180. doi: 10.1109/TVCG.2019.2903943
18. Shah S., Dey D., Lovett C., Kapoor A. AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles // Field and Service Robotics. Springer, 2018. P. 621–635. doi: 10.1007/978-3-319-67361-5_40
19. Hartley R., Zisserman A. Multiple View Geometry in Computer Vision. Cambridge University Press, 2003. 178 p. doi: 10.1017/CBO9780511811685
20. Balntas V., Lenc K., Vedaldi A., Mikolajczyk K. HPatches: A benchmark and evaluation of handcrafted and learned local descriptors // Proc. 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. P. 3852–3861. doi: 10.1109/CVPR.2017.410

4. Ono Y., Fua P., Trulls E., Yi K. LF-Net: learning local features from images. *Advances in Neural Information Processing Systems*, 2018, pp. 6234–6244.
5. Mikolajczyk K., Schmid C. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, vol. 27, no. 10, pp. 1615–1630. doi: 10.1109/TPAMI.2005.188
6. Cao Z., Hidalgo G., Simon T., Wei S., Sheikh Y. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, Early access. doi: 10.1109/TPAMI.2019.2929257
7. Baltrušaitis T., Robinson P., Morency L.-P. Openface: an open source facial behavior analysis toolkit. *Proc. IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016, pp. 7477553. doi: 10.1109/WACV.2016.7477553
8. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition. *Proc. 3rd International Conference on Learning Representations (ICLR)*, 2015.
9. Iandola F., Moskewicz M., Karayev S., Girshick R., Darrell T., Keutzer K. *Densenet: Implementing efficient convnet descriptor pyramids*. Available at: https://arxiv.org/abs/1404.1869 (accessed: 17.01.2020).
10. Brock A., Donahue J., Simonyan K. Large scale gan training for high fidelity natural image synthesis. *Proc. 7th International Conference on Learning Representations (ICLR)*. 2019.
11. Redmon J., Divvala S., Girshick R., Farhadi A. You only look once: Unified, real-time object detection. *Proc. 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91
12. Zheng Z., Yang X., Yu Z., Zheng L., Yang Y., Kautz J. Joint discriminative and generative learning for person re-identification. *Proc. 32nd IEEE /CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 2133–2142. doi: 10.1109/CVPR.2019.00224
13. Huang Q., Zhou K., You S., Neumann U. Learning to prune filters in convolutional neural networks. *Proc. 18th IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2018, pp. 709–718. doi: 10.1109/WACV.2018.00083
14. Gomez A.N., Zhang I., Kamalakara S.R., Madaan D., Swersky K., Gal Y., Hinton G.E. *Learning sparse networks using targeted dropout.* Available at: https://arxiv.org/abs/1905.13678 (accessed: 18.03.2020).
15. Wang Y., Xu C., You S., Tao D., Xu C. CNNpack: Packing convolutional neural networks in the frequency domain. *Advances in Neural Information Processing Systems*, 2016, pp. 253–261.
16. Hinton G., Vinyals O., Dean J. *Distilling the knowledge in a neural network*. Available at: https://arxiv.org/abs/1503.02531 (accessed: 06.02.2020).
17. Wang J., Gou L., Zhang W., Yang H., Shen H.-W. Deepvid: Deep visual interpretation and diagnosis for image classifiers via knowledge distillation. *IEEE Transactions on Visualization and Computer Graphics*, 2019, vol. 25, no. 6, pp. 2168–2180. doi: 10.1109/TVCG.2019.2903943
18. Shah S., Dey D., Lovett C., Kapoor A. AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles. *Field and Service Robotics*. Springer, 2018, pp. 621–635. doi: 10.1007/978-3-319-67361-5_40
19. Hartley R., Zisserman A. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003, 178 p. doi: 10.1017/CBO9780511811685
20. Balntas V., Lenc K., Vedaldi A., Mikolajczyk K. HPatches: A benchmark and evaluation of handcrafted and learned local descriptors. *Proc. 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3852–3861. doi: 10.1109/CVPR.2017.410

Авторы

Ященко Артем Владимирович — инженер, СингуляритиЛаб, Санкт-Петербург, 198152, Российская Федерация; аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, ORCID ID: 0000-0001-7292-2301, yashenkoxciv@gmail.com

Беликов Анатолий Владимирович — инженер, СингуляритиЛаб, Санкт-Петербург, 198152, Российская Федерация, Scopus ID: 57210427029, ORCID ID: 0000-0002-9081-642X, awbelikov@gmail.com

Authors

Artem V. Yashchenko — Engineer, SingularityLab, Saint Petersburg, 198152, Russian Federation; Postgraduate, ITMO University, Saint Petersburg, 197101, Russian Federation, ORCID ID: 0000-0001-7292-2301, yashenkoxciv@gmail.com

Anatoly V. Belikov — Engineer, SingularityLab, Saint Petersburg, 198152, Russian Federation, Scopus ID: 57210427029, ORCID ID: 0000-0002-9081-642X, awbelikov@gmail.com

Петерсон Максим Владимирович — кандидат технических наук, инженер, СингуляритиЛаб, Санкт-Петербург, 198152, Российская Федерация, Scopus ID: 36721957600, ORCID ID: 0000-0003-2945-6856, maxim.peterson@gmail.com
Потапов Алексей Сергеевич — доктор технических наук, профессор, ведущий научный сотрудник, СингуляритиЛаб, Санкт-Петербург, 198152, Российская Федерация, Scopus ID: 7201761961, ORCID ID: 0000-0001-6013-8843, pas.aicv@gmail.com

Maxim V. Peterson — PhD, Engineer, SingularityLab, Saint Petersburg, 198152, Russian Federation, Scopus ID: 36721957600, ORCID ID: 0000-0003-2945-6856, maxim.peterson@gmail.com

Alexey S. Potapov — D.Sc., Professor, Leading Scientific Researcher, SingularityLab, Saint Petersburg, 198152, Russian Federation, Scopus ID: 7201761961, ORCID ID: 0000-0001-6013-8843, pas.aicv@gmail.com