**Анотация**

Одной из востребованных областей поиска с Интернете является поиск по изображениям. Современные методы индексации данных показывают отличный результат поиска в многомиллиардных коллекциях и повсеместно применяются для данной задачи. В этой этой работе предлагается изучить существующие подходы индексации больших коллекций изображений и посмотреть на поведение этих алгоритмов в задаче распознавания лиц. Будем считать, что современные нейросетевые алгоритмы достаточно хорошо выделяют признаки лица на изображении и лишь немного затронем этот вопрос. Основной задачей ставим сравнение скорости поиска индексных методов на наборах данных лиц. Начнем рассмотрение с простейших алгоритмов поиска, таких как поиск по точному расстоянию и простая индексная структура. Затем на более продвинутых алгоритмах будем увеличивать скорость поиска и исследовать ухудшение точности. Эксперименты показывают, что в больших коллекция изображений лиц можно добиться приемлемой скорости поиска с хороших показателем точности. Сильное же ускорение приводит к большим потерям.

**Введение**

В последнее десятилетие визуальный поиск стал широко распространенной функцией многих поисковых систем. В связи с этим эффективный поиск ближайших соседей стал серьезной исследовательской проблемой [1-6]. Потребность быстрого поиска похожих изображений занимает большую нишу в современных приложениях компьютерного зрения, в том числе в задаче распознавания лиц [8]. Социальные сети, правоохранительные органы имеют огромные коллекции изображений лиц, среди которых надо уметь быстро извлекать нужную информацию. Для решения данной проблемы требуются эффективные и масштабируемые алгоритмы поиска с низкими временными затратами. Ожидается, что ответ на запросы к базам данных из миллиардов элементов будет занимать несколько миллисекунд.

Решение данной задачи можно рассматривать с двух сторон. Во-первых, даже самые современные алгоритмы поиска и обработки лиц на изображении не идеальны. Это открывает просторы для исследований. Во-вторых, подходящая структура данных для поиска может давать многократное увеличение скорости. Среди всех алгоритмов распознавания сверточные нейронные сети (CNN) показывают лучшие результаты поиска лиц на изображении и используются в большинстве исследованиях этой области [8,10]. Современные CNN достигают точности 97-99%. В связи с этим основной задачей исследования будем считать проблему выбора поисковой структуры данных, а дескрипторы лиц для экспериментов будем строить по одной из общедоступных CNN.

Все существующие крупномасштабные поисковые системы избегают исчерпывающего поиска путем ограничения конечного набора кандидатов, который рассматривается для запроса. Данный подход называют приближенным поиском ближайших соседей (ANN). Современные алгоритмы ANN имеют три основных реализации: инвертированная индексация [1-6], хеширование [9], многомерная инвертированная индексация основанная на квантовании произведения (PQ) [4,5]. В этой работе основное внимание будет уделено инвертированному индексу и его оптимизации с помощью PQ.

Структуры индексации разбивают пространство поиска на большое количество непересекающихся областей, и в процессе поиска используются только малая часть коллекции, наиболее близкая к конкретному запросу. Отобранная часть данных образует короткий список кандидатов, и поисковая система исчерпывающе рассчитывает расстояния между запросом и всеми кандидатами. На этом этапе важно, чтобы список кандидатов был коротким, так как вычисление расстояния имеет линейную сложность по данной длине. Метод PQ для ANN используется в двух видах: для построения многомерного инвертированного индекса для приближенного поиска или для кодирования векторов в компактные коды для точного поиска. Идея этих подходов состоит в том, чтобы разложить пространство векторов на большое количество подпространств и обучить запросы получать доступ к ближайшим подпространствам.

Первая структура индексации, способная работать с миллиардным набором данных, была представлена в [1]. Она была основана на структуре инвертированного индекса, которая разбивает пространство признаков на диаграмму Вороного. Каждая область задается свои центроидом, который предварительно обучили алгоритмом K-средних. Было показано, что эта система достигает разумных скоростей поиска порядка нескольких десятков миллисекунд. Позже обобщение структуры инвертированного индекса было предложено в [3]. В этой работе был представлен инвертированный многомерный индекс или мультииндекс (IMI), который разбивает пространство признаков на несколько ортогональных подпространств и разбивает каждое подпространство на диаграммы Вороного независимо. Декартово произведение такого разбиения образует неявное разбиение всего пространства поиска. Обе эти структуры обладают своими недостатками, которые можно устранить с помощью различных оптимизаций PQ [6,7].

В данной работе будет описано несколько современных архитектур индексирования и путем экспериментов исследована применимость данного подходя для задачи распознавания лиц.

**Постановка задачи**

Главной задачей данной работы является проверка современных индексных структур быстрого поиска в задаче распознавания лиц. Считается, что выделением лица на изображении и построением его признаков занимаются алгоритмы общедоступных сверточных нейронных сетей. Работа будет производится в рамках заранее подготовленных 128-мерных векторов, по одному для каждого изображения. Время построения каждого вектора учитывать не будем, так как во всех алгоритмах будет использоваться одна и та же нейронная сеть. Также не будем учитывать, но обратим внимание на время обучения индексных структур, так как для разных алгоритмов оно может отличаться на порядки.

Проверка заключается в измерении среднего времени поиска похожих лиц в большой коллекции изображений. Время, затраченное на поиск k (k = 1, 5, 10, 30, 50, 100) ближайших соседей будет основным критерием скорости алгоритма. Помимо скорости надо учитывать точность поиска. Во всех приведенных статьях точность измерялась как процент истинных ближайших соседей среди k найденных, где истинные ближайшие соседи определялись точным евклидовым расстоянием. В связи со спецификой нашей задачи, наиболее правильным вариантом измерения точности будет получение процента лиц, отмеченных так же, как и запрос, среди k найденных ближайших соседей.

Для применимости данного подхода к реальным задачам требуемая скорость поиска не должна превышать нескольких миллисекунд. А точность должна быть в пределе допустимой для выбранного алгоритма, то есть не сильно отличаться от приводимой в статьях. В качестве вывода следует оценить эти показатели и сформулировать возможные пути улучшения.

**Обзор существующих решений**

**Исследование и построение решений задачи**

Из-за огромного количества кластеров в IMI некоторые области могут содержать содержать сравнительно малое количество кандидатов или не содержать совсем. Следовательно, IMI в процессе поиска тратит много времени на посещение пустых областей. Фактически, причина этого недостатка состоит в том, что IMI при обучении алгоритмом К-средних не учитывает зависимость подпространств при разделении, которые зачастую зависимы на практике. В частности, существуют значительные корреляции между различными подпространствами дескрипторов, построенных с помощью сверточных нейронных сетей,которые наиболее актуальны в наши дни. Для решения проблемы адаптации алгоритмов к коррелированным данным используют различные оптимизации PQ [4,5].

**Описание практической части**

**Заключение**

Литература (По номерам статей из файла TextNotes):

[1] Jegou, H., Tavenard, R., Douze, M., Amsaleg, L.: Searching in one billion vectors: Re-rank with source coding. In: ICASSP. (2011)

[2]H. Jegou, M. Douze, and C. Schmid, “Product quantization for ´ nearest neighbor search,” TPAMI, vol. 33, no. 1, 2011.

[3]A. Babenko and V. Lempitsky, “The inverted multi-index,” 2014 IEEE

[4]Ge, T., He, K., Ke, Q., Sun, J.: Optimized product quantization. Technical report (2013)

[5]Kalantidis, Y., Avrithis, Y.: Locally optimized product quantization for approximate nearest neighbor search. In: in Proceedings of International Conference on Computer Vision and Pattern Recognition (CVPR 2014), IEEE (2014)

[6]Babenko, A., Lempitsky, V.S.: Efficient indexing of billion-scale datasets of deep descriptors. In: CVPR. (2016)

[7]Revisiting the Inverted Indices for Billion-Scale Approximate Nearest Neighbors Dmitry Baranchuk1,2 , Artem Babenko1,3 , Yury Malkov4

[8]Wang, D., Otto, C., Jain, A.K.: Face search at scale. TPAMI (2017)

[9]] W. Dong, Z. Wang, W. Josephson, M. Charikar, and K. Li, “Modeling LSH for performance tuning,” CIKM, 2008. 2, 4

[10]D. Yi, S. Liao, and S. Z. Li, “Learning face representation from scratch,” arXiv:1411.7923v1, 2014

Двухуровневая индексация

[11] D. Nister and H. Stew ´ enius, “Scalable recognition with a vocab- ´ ulary tree,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2161–2168, 2006