

Кластеризация изображений при использовании предобученных нейронных сетей

А.С. Кузнецов, Е.Ю. Семенов, Л.Д. Матросова

Аннотация — В статье рассматриваются этапы решения задач кластеризации изображений с использованием предобученных нейронных сетей. Представлены некоторые составные решения задачи кластеризации, где непосредственно методы кластеризации используются на последнем этапе, а большая часть работы занимает извлечение признаков и их предобработка. Проведен анализ современных подходов к извлечению признаков из изображений, включающий в себя классические методы теории распознавания образов и компьютерного зрения и методы извлечения признаков с помощью сверточных нейросетей. В работе даются рекомендации по выбору наиболее эффективной для использования архитектуры сверточных нейронных сетей в зависимости от характера решаемых задач.

Приводится классификация методов понижения размерности изображений по типам: сохраняющим расстояние между точками при отображении из пространства высокой размерности в низкоразмерное; сохраняющим глобальную структуру данных; поиска ближайших векторов в пространствах большой размерности. Рассматривается специальный метод кластеризации изображений DeepCluster, который итеративно группирует признаки с помощью стандартного алгоритма кластеризации. Полученные результаты могут послужить в дальнейших исследованиях в данной сфере, а также при решении задач по подготовке предтренированных моделей сверточных нейронных сетей.

Ключевые слова — кластеризация изображений, нейронные сети, сверточные сети, обучение без учителя, глубокое обучение.

I. ВВЕДЕНИЕ

В современной науке о компьютерном зрении основное внимание отводится изучению методов обучения с учителем, в то время как методы кластеризации изображения недостаточно освещены. Вместе с тем, повышение качества работы современных моделей, основанных на нейронных сетях и работающих с изображениями, требует значительного количества размеченных вручную наборов данных (датасетов). Обучение без учителя в данном случае может помочь в

создании автоматизированной разметки для обучающих выборок.

Большое количество графического содержимого (медиаконтента), генерируемого пользователями в глобальной сети также требует разработки эффективных подходов к его автоматизированной сортировке для использования данных знаний в задачах обучения с учителем.

II. ЗАДАЧА КЛАСТЕРИЗАЦИИ ИЗОБРАЖЕНИЙ

Кластеризация является классической задачей методов обучения без учителя, то есть в том случае, когда структура данных представленных для анализа заранее неизвестна и исследователь не имеет возможности представить для них какую-либо разметку. Кластеризация данных в задачах, где объекты хорошо формализуются с помощью категориальных или вещественных признаков описаний достаточно хорошо изучена. Для работы с такими данными предложены целые семейства эффективных алгоритмов, реализованных в популярных пакетах прикладного анализа.

Кластеризация, зачастую, не является конечной целью, а применяется как вспомогательный этап для решения другой задачи, например, разведочного анализа данных, либо для генерации дополнительных признаков при решении задач обучения с учителем.

Кроме того, сама по себе кластеризация изображений является сложной задачей ввиду того, что в исходном виде они непригодны для применения алгоритмов и требуют проведения некоторой процедуры извлечения признаков, то есть получения некоторого вектора признакового описания.

III. ИЗВЛЕЧЕНИЕ ПРИЗНАКОВ

Решение задачи формирования признакового пространства для графических изображений является одной из самых сложных проблем при построении систем кластеризации. В общем виде группы методов можно разделить на классические и нейросетевые.

A. Классические методы извлечения признаков из изображений

Классические методы теории распознавания образов и компьютерного зрения предлагают широкий спектр по выбору метода для решения данного класса задач. Существующие мета анализы [1,2] позволяют выделить следующие способы решения задач выделения признаков из графических изображений:

Статья получена 28 марта 2019.

А.С. Кузнецов - Орловский юридический институт МВД России им. В.В. Лукьянова, (e-mail: kuznetsov_as@bk.ru).

Е.Ю. Семенов - Орловский юридический институт МВД России им. В.В. Лукьянова (e-mail: john_neg@mail.ru).

Л.Д. Матросова - Орловский юридический институт МВД России им. В.В. Лукьянова (e-mail: matrosovald@mail.ru).

1. Математическая морфология.
2. Преобразование Хафа.
3. Кратномасштабный анализ.
4. Совмещение шаблонов.
5. Динамическое программирование.

Развитие данных методов решения такого рода задач нашли отражение как в работах отечественных [3,4,5], так и зарубежных ученых [6,7].

Наиболее популярными на сегодняшний день реализациями классических методов извлечения признаков из изображений являются алгоритмы SURF, ORB, SIFT и BRIEF. Большинство из этих алгоритмов построены на использовании градиента изображения.

1. блок сверточных слоев;
2. блок слоев объединения;
3. блок полносвязных слоев.

Наиболее часто выходом нейронной сети в задачах классификации является вероятности принадлежности поданного на вход изображения, полученного с помощью функции мягкого максимума (softmax), являющейся обобщением логистической функции для многомерного случая.

Предтренированные на датасете ImageNet сети имеют параметры блока полносвязных слоев оптимальные для решения задачи разделения конкретных классов данного датасета, что не дает ценной информации для задач

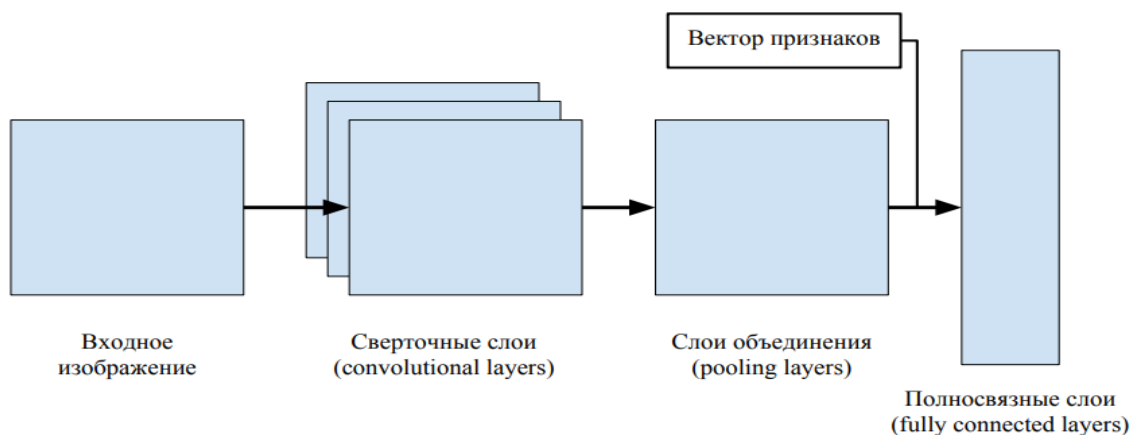


Рис. 1. Схема извлечения признаков изображения из сверточной нейронной сети

В. Извлечения признаков из изображений с помощью сверточных нейросетей

Применение глубоких нейронных сетей в задачах обучения с учителем успешно заняли свои ниши в компьютерном зрении, речи, а также большого набора задач обработки естественного языка.

Аппарат глубоких нейронных сетей показал своё превосходство в ключевых задачах, например, в задачах классификации изображений.

В частности, развитие архитектур сверточных нейронных сетей последовательно улучшали показатели ошибки на классическом для задач классификации изображений датасете ImageNet, содержащим более 14 миллионов изображений, размеченных на 1000 классов. Так AlexNet [8] в 2012 году достигла сокращения доли ошибок почти на 50%, затем последовательно улучшали этот показатель архитектуры VGG [9] в 2014 году, ResNet[10] и DenseNet[11].

Таким образом, современная практика решения задач компьютерного зрения опирается на использование глубоких нейронных сетей, которые представляются более перспективным подходом для решения задачи извлечения признаков.

Устройство приведенных выше архитектур сверточных нейронных сетей достаточно сильно различаются, однако, в общем виде их можно разделить на несколько функциональных блоков:

кластеризации в общем виде, где изображение может в принципе не относиться ни одному из этих классов.

В данных условиях лучшим решением является использование вектора признаков со входа блока полносвязных слоев, который, по сути, будет являться совокупностью ответов признаковых фильтров сверточных слоев.

Практический опыт соревновательного анализа данных также показывает эффективность данного подхода, например, лучшие решения соревнования Kaggle Avito Demand Prediction Challenge использовали сходный подход к извлечению признаков в комбинации с набором эвристик, определяющих качество изображения.

К сожалению, эффективность нейронных сетей в данных задачах достаточно часто требует значительных вычислительных ресурсов в реализации готовых решений. Реальные задачи кластеризации изображений всегда связаны с большим объемом исходного набора данных, который невозможно, либо нецелесообразно размечать вручную, что накладывает дополнительные требования к скорости извлечения признаков, особенно для систем, работающих в режиме реального времени.

Качество получаемого вектора признака можно опосредованно оценить по характеристикам архитектур для исходной задачи классификации, а скорость вычисления соответствует только процедуре прямого прохождения (forward propagation) без необходимости вычисления градиентов, необходимых для процедуры вычисления обратного распространения ошибки

(backpropagation) ввиду отсутствия задачи обучения нейронной сети.

Таблица 1. Параметры VGG, ResNet, среднее время

Архитектура сети	Топ-5 ошибка	Скорость (мс)	Число признаков
VGG-16	8.80	128.62	512
VGG-19	9.00	147.32	512
ResNet-18	10.76	31.54	512
ResNet-34	8.74	51.59	512
ResNet-50	7.02	103.58	2048
ResNet-101	6.21	156.44	2048
ResNet-152	6.16	217.91	2048

прямого прохода [12]

Из представленной таблицы можно сделать вывод о том, что наиболее оптимальным выбором являются нейронные сети архитектуры ResNet с числом слоев 18,

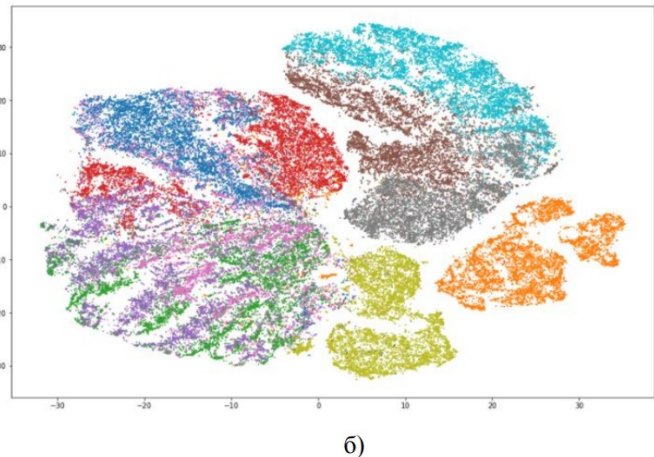
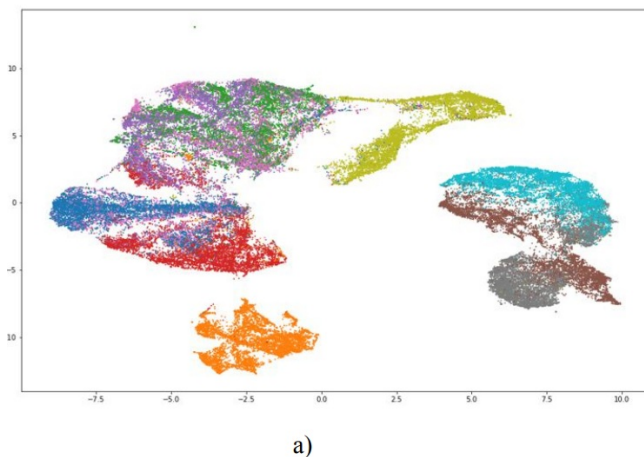


Рис. 2. Результаты работы методов (а) UMAP и (б) t-SNE на датасете Fashion MNIST [14].

IV МЕТОДЫ ПОНИЖЕНИЯ РАЗМЕРНОСТИ И ПОИСК БЛИЖАЙШИХ ВЕКТОРОВ В ПРОСТРАНСТВАХ БОЛЬШОЙ РАЗМЕРНОСТИ

Методы понижения размерности можно классифицировать по порядку типу работы с данными на две большие группы.

А. Методы, сохраняющие расстояние между точками при отображении из пространства высокой размерности в низкоразмерное.

Типичными представителями данной группы являются: метод главных компонент (PCA) и многомерное шкалирование (MDS - Multidimensional Scaling).

34 и 50. Однако, в зависимости от конечной задачи следует попробовать использовать VGG, например, для формирования взвешенного предсказания использованием признаков пространств нескольких нейросетей.

Полученное признаковое описание зависит от выбранной архитектуры нейронной сети, но в любом случае оно обладает достаточно большой размерностью, что затрудняет использование классических методов кластеризации в силу проявления эффекта “проклятия размерности”.

Существующие группы методов кластеризации высокоразмерных данных построенных на принципах подпространств (Subspace clustering) или проекций (Projected clustering) обладают рядом недостатков, затрудняющих их применение на реальных данных, в частности высокой вычислительной сложностью.

Для решения исходной задачи кластеризации необходимо решить задачу понижения размерности с сохранением максимального количества информации в новом признаковом описании.

Задача снижения размерности признакового пространства может достаточно успешно решаться с помощью линейных методов, вроде метода главных компонент (PCA) и его модификаций:

1. Incremental PCA используется для больших наборов данных, когда невозможно вместить весь датасет. Метод позволяет вычислять компоненты частями и незначительно отличается от классического PCA по качеству.

2. Kernel PCA применяется для получения нелинейного снижения пространства с помощью функций ядра. Вид применяемых ядер подбирается индивидуально, однако, наиболее часто применяют ядро радиальных гауссовских функций (RBF).

В. Методы, сохраняющие глобальную структуру данных.

Популярность методов данной группы определяется нуждами задач визуализации данных в ходе разведочного анализа данных. Наиболее часто

используемыми являются метод стохастического вложения соседей с t -распределением (t-SNE) и метод аппроксимаций и проекций равномерного пространства (UMAP) [13].

В общем случае применение данных методов дает схожие результаты, однако, UMAP имеет несколько преимуществ, наиболее значимым из которых является то, что он более вычислительно эффективен, чем t-SNE.

На сегодняшний день активно развивается проект t-SNE-CUDA [15], который отчасти решает проблему относительно низкой производительности t-SNE благодаря использованию графических процессоров (GPU) в расчетах, однако, на момент написания статьи его работа в некоторых из протестированных авторами задач была нестабильной.

С. Методы поиска ближайших векторов в пространствах большой размерности

Требования к скорости работы систем, работающих в реальном времени, а также большие объемы требуют разработки специальных методов способных решать задачу поиска ближайших векторов в многомерных пространствах.

В качестве примера таких систем можно привести системы видеонаблюдения с распознаванием лиц. В базе данных таких систем хранится большое число обработанных лиц в формате векторов больших размерностей. Основной задачей является быстрый поиск, оценка сходства сличенного в режиме реального

времени лица с имеющимися в базе и принятие решения по значению некоторого установленного порога метрики сходства (в изображениях часто используют евклидово расстояние).

Наиболее популярными алгоритмами решения данного типа задач является аппроксимированный метод ближайших соседей и его реализация Annoy, а также библиотека Faiss, использующая набор алгоритмов [16] на основе модифицированного метода K-средних и позволяющая проводить расчеты на GPU.

V СПЕЦИАЛЬНЫЕ МЕТОДЫ КЛАСТЕРИЗАЦИИ ИЗОБРАЖЕНИЙ НА ОСНОВЕ НЕЙРОННЫХ СЕТЕЙ

Рассмотренные выше подходы представляют собой некоторые составные решения для решения задачи кластеризации изображений, где непосредственно методы кластеризации используются на последнем этапе, а большая часть работы занимает извлечение признаков и их предобработка.

DeepCluster – это специальный метод кластеризации изображений, который итеративно группирует признаки с помощью стандартного алгоритма кластеризации (метода K-средних) и последовательно использует метки псевдо-классов, полученных при кластеризации для обучения параметров сверточной нейронной сети [17].

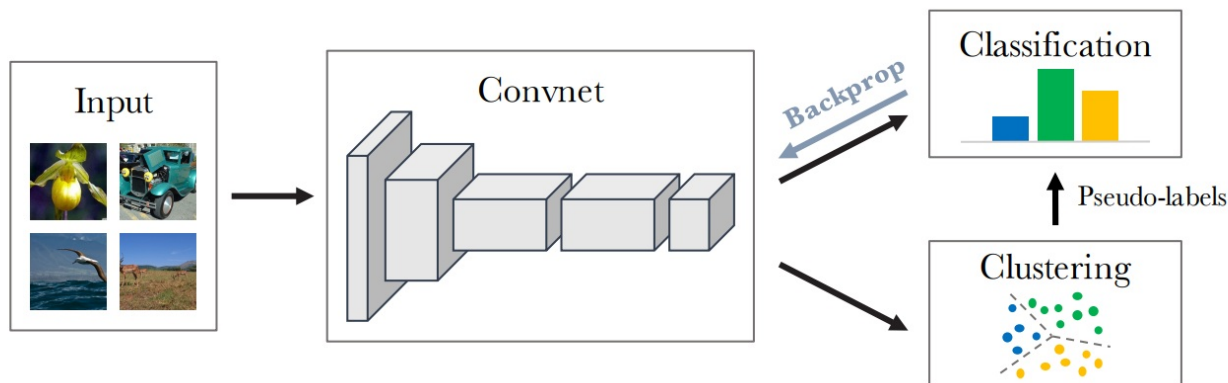


Рис. 3. Архитектура DeepCluster

Архитектура модели DeepCluster основана на классической AlexNet архитектуре с пятью сверточными и тремя полносвязными слоями, однако метод в целом робастен к конкретной архитектуре, в частности, в работе указан пример использования VGG, что дало значительный прирост к качеству генерируемых признаков слоев.

Данная модель предполагает применение метода K-средних из-за простоты его реализации и вычисления, однако, на его месте может быть использован другой алгоритм кластеризации.

Авторами были получены значения точности линейной классификации на обученных активациях сверточных слоев превосходящие существующие

результаты на датасете YFCC100M, а также задачах детекции, сегментации и классификации на датасете Pascal VOC 2007.

Данные результаты могут послужить в дальнейших исследованиях в задачах подготовки предтренированных моделей сверточных нейронных сетей.

VI ЗАКЛЮЧЕНИЕ И ВЫВОДЫ

Таким образом, рассмотренные подходы к кластеризации изображений позволяют эффективно решать данную задачу и могут выбираться в зависимости от конкретных условий применения.

В случае наличия конечной задачи визуализации структуры данных представляется целесообразным использование последовательного извлечения признаков из изображения посредством предобученной нейронной

сети и использования методов UMAP и t-SNE для понижения размерности.

В случае, когда кластеризация изображения является процедурой предобработки изображения для обучения поверх этих данных, например, классификатора в первую очередь следуют использовать методы PCA, особенно в случаях большого объема данных ввиду его относительной скорости и возможности подобрать число компонент под необходимую долю объясненной дисперсии при максимальном использовании доступных ресурсов памяти вычислительной системы.

Архитектуру нейронной сети необходимо подбирать в зависимости от объема данных и наличия вычислительных ресурсов. Наиболее приемлемым по показателю отношения скорости вычисления к качеству формируемых признаков, по мнению авторов, является использование архитектуры ResNet-50, понижение размерности методов UMAP и кластеризация полученных данных любым устойчивым к выбросам алгоритмам, например, HDBSCAN.

БИБЛИОГРАФИЯ

- [1] Quackenbush, Lindi J. "A review of techniques for extracting linear features from imagery." *Photogrammetric Engineering & Remote Sensing* 70.12 (2004): 1383-1392.
- [2] Kumar, Gaurav, and Pradeep Kumar Bhatia. "A detailed review of feature extraction in image processing systems." 2014 Fourth international conference on advanced computing & communication technologies. IEEE, 2014.
- [3] Васильев, В.Н., И. П. Гуров, and А. С. Потапов. "Математические методы и алгоритмическое обеспечение анализа и распознавания изображений в информационно-телекоммуникационных системах." *Всероссийский конкурсный отбор обзорноаналитических статей по приоритетному направлению "Информационно-телекоммуникационные системы"* 46 (2008).
- [4] Гладилин, Сергей Александрович, et al. "Построение устойчивых признаков детекции и классификации объектов, не обладающих характерными яркостными контрастами." *Информационные технологии и вычислительные системы I* (2014): 53-60.
- [5] Кучуганов, Александр Валерьевич. "Биоинспирированные алгоритмы выделения информативных признаков изображений." *Известия Томского политехнического университета. Инжиниринг георесурсов* 321.5 (2012).
- [6] De, Sourav, et al. "Dimension Reduction Using Image Transform for Content-Based Feature Extraction." *Feature Dimension Reduction for Content-Based Image Identification*. IGI Global, 2018. 26-40.
- [7] Zhao, Wenzhi, and Shihong Du. "Spectral-spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach." *IEEE Transactions on Geoscience and Remote Sensing* 54.8 (2016): 4544-4554.
- [8] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012.
- [9] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).
- [10] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- [11] Huang, Gao, et al. "Densely connected convolutional networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
- [12] Benchmarks for popular CNN models. URL: <https://github.com/jcjohnson/cnn-benchmarks>
- [13] McInnes, Leland, John Healy, and James Melville. "Umap: Uniform manifold approximation and projection for dimension reduction." *arXiv preprint arXiv:1802.03426* (2018).
- [14] Обзор нового алгоритма уменьшения размерности UMAP. Действительно ли он лучше и быстрее, чем t-SNE? <https://habr.com/ru/company/newprolab/blog/350584/>
- [15] Chan, David M., et al. "t-SNE-CUDA: GPU-Accelerated t-SNE and its Applications to Modern Data." *arXiv preprint arXiv:1807.11824*(2018).
- [16] Johnson, Jeff, Matthijs Douze, and Hervé Jégou. "Billion-scale similarity search with GPUs." *arXiv preprint arXiv:1702.08734* (2017).
- [17] Caron, Mathilde, et al. "Deep clustering for unsupervised learning of visual features." *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.

Image clustering using pretrained neural networks

A.S. Kuznetsov, E.Y. Semenov, L.D. Matrosova

Abstract — The article deals with the stages of solving image clustering problems using pre-trained neural networks. Some composite solutions of the clustering problem are presented, where clustering methods are used at the last stage, and most of the work is the extraction of features and their preprocessing. The analysis of modern approaches to feature extraction from images, including classical methods of pattern recognition theory and computer vision and feature extraction methods using convolutional neural networks. The paper provides recommendations for choosing the most effective architecture for convolutional neural networks, depending on the nature of the tasks. A classification of methods for reducing the dimension of images by types: preserving the distance between points when displaying from high-dimensional to low-dimensional space; preserving the global structure of data; search for the nearest vectors in large-dimensional spaces. We consider a special method of clustering images DeepCluster, which iteratively groups the features using a standard clustering algorithm. The obtained results can serve in further research in this area, as well as in solving problems in the preparation of pre-trained models of convolutional neural networks.

Keywords — *images clustering, convolutional neural networks, unsupervised learning, deep learning.*

REFERENCES

- [1] Quackenbush, Lindi J. "A review of techniques for extracting linear features from imagery." *Photogrammetric Engineering & Remote Sensing* 70.12 (2004): 1383-1392.
- [2] Kumar, Gaurav, and Pradeep Kumar Bhatia. "A detailed review of feature extraction in image processing systems." 2014 Fourth international conference on advanced computing & communication technologies. IEEE, 2014.
- [3] Vasil'ev, V.N., I. P. Gurov, and A. S. Potapov. "Matematicheskie metody i algoritmicheskoe obespechenie analiza i raspoznavanija izobrazhenij v informacionno-telekommunikacionnyh sistemah." *Vserossijskij konkursnyj otbor obzornoanaliticheskikh statej po prioritetnomu napravleniju* Informacionno-telekommunikacionnye sistemy 46 (2008).
- [4] Gladilin, Sergej Aleksandrovich, et al. "Postroenie ustojchivyh priznakov detekcii i klassifikacii ob"ektov, ne obladajushhij harakternymi jarkostnymi kontrastami." *Informacionnye tehnologii i vychislitel'nye sistemy* 1 (2014): 53-60.
- [5] Kuchuganov, Aleksandr Valer'evich. "Bioinspirirovannye algoritmy vydelenija informativnyh priznakov izobrazhenij." *Izvestija Tomskogo politehnicheskogo universiteta. Inzhiniring georesursov* 321.5 (2012).
- [6] De, Sourav, et al. "Dimension Reduction Using Image Transform for Content-Based Feature Extraction." *Feature Dimension Reduction for Content-Based Image Identification*. IGI Global, 2018. 26-40.
- [7] Zhao, Wenzhi, and Shihong Du. "Spectral-spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach." *IEEE Transactions on Geoscience and Remote Sensing* 54.8 (2016): 4544-4554.
- [8] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012.
- [9] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).
- [10] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- [11] Huang, Gao, et al. "Densely connected convolutional networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
- [12] Benchmarks for popular CNN models. URL: <https://github.com/jcjohnson/cnn-benchmarks>
- [13] McInnes, Leland, John Healy, and James Melville. "Umap: Uniform manifold approximation and projection for dimension reduction." *arXiv preprint arXiv:1802.03426* (2018).
- [14] Obzor novogo algoritma umen'shenija razmernosti UMAP. Dejstvitel'no li on luchshe i bystree, chem t-SNE? <https://habr.com/ru/company/newprolab/blog/350584/>
- [15] Chan, David M., et al. "t-SNE-CUDA: GPU-Accelerated t-SNE and its Applications to Modern Data." *arXiv preprint arXiv:1807.11824*(2018).
- [16] Johnson, Jeff, Matthijs Douze, and Hervé Jégou. "Billion-scale similarity search with GPUs." *arXiv preprint arXiv:1702.08734* (2017).
- [17] Caron, Mathilde, et al. "Deep clustering for unsupervised learning of visual features." *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.