

Государственное образовательное учреждение высшего профессионального
образования

«Московский государственный технический
университет имени Н.Э. Баумана»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ ИНФОРМАТИКИ И СИСТЕМ УПРАВЛЕНИЯ
КАФЕДРА ТЕОРЕТИЧЕСКОЙ ИНФОРМАТИКИ И КОМПЬЮТЕРНЫХ
ТЕХНОЛОГИЙ

Отчет о НИРС

Студент _____ Батусов П. В.

Научный руководитель _____ Вишняков И. Э.

Москва 2016

Содержание

1	Цели обзора	2
2	Существующие подходы определения и классификации объектов на изображении	2
2.1	Классификация по признакам	2
2.1.1	Выделение признаков по трехмерной структуре	2
2.1.2	Применение цепочки классификаторов	3
2.1.3	Байесовские сети	3
2.1.4	Поиск выделенных признаков в базе данных	4
2.2	Классификация без явного определения признаков	4
2.2.1	Глубокие сети доверия	5
2.2.2	Сверточные нейронные сети	5
2.2.3	Глубокие сверточные нейронные сети	6
2.2.4	Плиточные сверточные нейронные сети	6
3	Выбранная модель	7
3.1	Обучение нейронных сетей	7
3.1.1	Эффективное обратное распространение ошибки	7
3.1.2	Использование графического процессора	8
3.1.3	Дообучение сети	8
3.2	Определение параметров сети	9
3.2.1	Размер сети	9
3.2.2	Формирование обучающей выборки	10
3.2.3	Процедура обучения	11
4	План дальнейших исследований	11

1 Цели обзора

В данном отчете представлены основные подходы для решения одной из задач машинного обучения — классификации и определения объектов на изображении. Рассматриваемые методы можно условно разделить на две группы: алгоритмы классификации с использованием признаков и классификаторы, которые выводят признаки в процессе обучения.

Главной целью обзора является разбор существующих систем решений, изучение их устройства и архитектуры, выбор оптимального подхода для решения задачи распознавания модели автомобиля по фотографии. В конце отчета предоставлен план дальнейших исследований и тестирования.

2 Существующие подходы определения и классификации объектов на изображении

2.1 Классификация по признакам

В данном пункте рассматриваются алгоритмы решения задачи распознавания автомобилей по цифровому изображению, полученному с камер дорожного наблюдения, авторегистраторов и т. д.. Перечисленные подходы основаны на обработке известной информации о структуре объектов, которые необходимо определить и классифицировать, а также использовании фактов о ракурсе съемке, погодных условиях и прочем для увеличения итоговой точности. Другими словами, эти методы работают с известными признаками объектов.

2.1.1 Выделение признаков по трехмерной структуре

Данный метод [1] основан на представлении автомобилей в виде полигональной модели, по которой происходит выделение признаков, передаваемых на вход нейросети. Трехмерная структура описания автомобиля разделена на восемь частей, в каждой из которых выделяется опорная вершина. В качестве признаков используется расстояние между каждой парой опорных вершин.

Дополнительно рассматриваются параметры колес автомобиля, их радиус и положение.

Авторы работ использовали трехслойную, полносвязную нейронную сеть с 30 входами (количество признаков) и 120 выходами (6×20 , 6 — количество классов, 20 — число выходных нейронов для каждого класса). Результат определялся поиском максимального значения в выходном слое.

При обучении сети использовался метод обратного распространения ошибки. В качестве функции ошибки, для улучшения сходимости, использовалась экспоненциальная функция вместо квадратичной.

Обучающая выборка состояла из 500 изображений с дорожных камер. Классификация производилась на большие/небольшие грузовики и автомобили. Тестовая выборка включала в себя еще 300 изображений, из которых правильно классифицировано 91%, ошибочно 4% и в 5% случаев система не смогла определить автомобиль на изображении.

2.1.2 Применение цепочки классификаторов

В работе [2] ставилась задача определения боковых стоек автомобиля. Решение задачи представляло собой многослойную схему классификации, от общего к частному. Такой подход называют усилением простых классификаторов (boosting classifier). Первые слои предназначены для обработки входного изображения, определения направления движения, колес, выделение боковой части автомобиля и только после этого производится поиск боковых стоек на основе гистограммы ориентированных градиентов и геометрических моделей.

Обучение системы производилось на изображениях автомобильного потока, в тестовую выборку также включены негативные примеры, не содержащие транспортных средств. Обучающее множество состоит из 100000 изображений без автомобилей и 4000 с автомобилями. Тестовая выборка содержит 1000 положительных и 200000 негативных примеров. Итоговая точность составила 90% для положительных и 99% — для негативных примеров.

2.1.3 Байесовские сети

Имея возможность выделения признаков для дальнейшего решения задачи классификации, можно воспользоваться байесовской сетью. Такой подход

использовался в работе выделения автомобилей на фотоаэроснимках низкого разрешения [3]. Основная проблема при решении этой задачи — высокое количество шумов: тени, солнечные блики, кроны деревьев и прочее.

Авторы предложили метод, основанный на использовании дополнительной известной информации о времени и месте съемки. Эти данные и выделенные признаки объектов вместе попадают в байесовскую сеть, которая производит отсев таких неправильно определенных кандидатов, как тени от настоящих автомобилей и объектов, находящихся вне дорожного полотна.

2.1.4 Поиск выделенных признаков в базе данных

Car-Rec [4] — еще одно предложенное решение для задачи классификации автомобилей, заключающееся в выделении признаков и их обработке с использованием деревьев поиска по заготовленной базе. Для определения области поиска на изображении используется алгоритм SURF [5], основанный на интегральном представлении входного изображения. Разработанная система имеет точность более 90%.

2.2 Классификация без явного определения признаков

Не всегда имеется возможность выделения опорных признаков для классификации, так как в общем случае не понятно, по какому принципу их выбирать. Кроме этого, неизвестно, как с течением времени будет развиваться предметная область, и какие еще признаки необходимо зарезервировать для дополнительных классов.

При использовании нейросетей признаки для классификации автоматически вырабатываются при обучении модели. В случае необходимости добавить новые классы в классификатор, производят его переобучение, в результате чего выводятся новые признаки. Также возможно произвести более быструю настройку, скорректировав весовые коэффициенты без перестроения признаков (дообучение сети).

2.2.1 Глубокие сети доверия

Глубокие сети доверия [6] являются развитием идеи использования рекуррентных нейронных сетей, которые тяжело обучать из-за наличия обратных связей. Рассматривается архитектура, при которой внутри скрытых слоев используется ограниченная машина Больцмана.

Авторы рассматривают задачу определения автомобиля по фотографии сзади (как с камеры авторегистратора). Входной слой имеет размерность, соответствующую разрешению изображения (признаки — пиксели). Данные классифицируются по двум категориям: автомобиль и не автомобиль.

Тестирование происходило на 735 изображениях, рассматривались модели с одним, двумя и тремя скрытыми слоями. Наилучшая точность была получена при использовании двух скрытых слоев — более 96%. Также была произведена оценка точности для других алгоритмов: искусственных и сверточных нейросетей, k ближайших соседей и метода опорных векторов. Среди них наибольшая точность результата получилась при использовании сверточных нейронных сетей, почти 95%.

2.2.2 Сверточные нейронные сети

В работе [7] рассматривалась задача распознавания номерных табличек на домах. Использовалась традиционная архитектура сверточной нейросети: первый сверточный слой выделял 16 карт признаков со сверткой 5×5 , второй — 512, со сверткой 7×7 , два слоя классификации по 20 нейронов каждый. Применялся многоступенчатый подход выделения признаков (слоям классификации были доступны все предыдущие карты признаков). Описанная модель реализована на фреймворке EBLearn [8].

Был произведен подбор оптимального параметра степени свертки для алгоритма выделения признаков Lp-Pooling. Наилучшая точность получена при использовании коэффициента $p = 4$. Кроме этого, перед передачей изображения в сеть, вокруг него достраивалась рамка из двух пикселей с нулевым значением, чтобы граница попадала в операцию свертки рамкой 5×5 в разных положениях.

Набор данных состоял из обучающей, тестовой и дополнительной выборок. Дополнительная выборка состояла из большого количества легких для

классификации изображений, а обучающая содержала небольшое число более сложных примеров. Обучение происходило на 6000 примеров, выбранных случайным образом из обучающего множества ($2/3$) и из дополнительного ($1/3$). В результате при тестировании была получена точность более 95%.

2.2.3 Глубокие сверточные нейронные сети

Один из вопросов систем распознавания и классификации изображений — насколько большую предметную область можно охватить, используя автоматические методы. В рамках состязания ILSVRC [9] проводится попытка классифицировать огромный тестовый набор изображений (более миллиона) по тысяче категорий.

Авторы работы [10] занимались созданием одной из самых больших сверточных нейросетей для участия в ILSVRC-2012. Она состояла из 5 сверточных и 3 полносвязных слоев, включающих 60 миллионов параметров и 650 тысяч нейронов. В работе отмечена необходимость и важность большой размерности сети, так как при попытке избавиться хотя бы от одного слоя происходило сильное падение качества классификатора.

Для уменьшения вероятности переобучения использовался подход выбывающих нейронов [11]. В качестве активационной функции была выбрана модель параметрического выпрямляемого элемента ReLU [12]. При тестировании по мере оценки ошибки top-5 error (смотрится результат среди пяти самых вероятных классов, выделенных сетью) была получена точность 84.7%.

На соревнованиях ILSVRC-2014 была представлена еще большая сеть, включающая 27 слоев [13]. Команде GoogLeNet получилось добиться результатов в 6.67% ошибок.

2.2.4 Плиточные сверточные нейронные сети

Сверточные нейронные сети хорошо применяются для задач классификации и распознавания объектов, однако, подход использующий разделяемые веса может помешать в процессе обучения выделить некоторые инвариантные признаки. В работе [14] предлагался метод, в котором разделяемые веса чередовались с некоторым шагом k . Так, при $k = 1$ получалась обычная сверточная нейронная сеть, а при увеличении этого параметра строилась очеред-

ная модель, которая является компромиссом между способностью выявления инвариантности изображений и сложностью обучения системы.

Тестирование производилось на наборе данных CIFAR-10 [15], состоящем из 50000 обучающих изображений и 10000 тестовых, которые необходимо классифицировать по 10 категориям. Полученная точность метода при выбранном параметре $k = 2$ более 73%.

3 Выбранная модель

Согласно поставленным целям были рассмотрены существующие алгоритмы автоматической классификации объектов на изображении. Применительно для решения задачи НИРС определения модели автомобиля по фотографии был выбран подход использования нейронных сетей.

3.1 Обучение нейронных сетей

В этом пункте рассматриваются алгоритмы и их модификации для эффективного обучения нейронных сетей, разделение процесса обучения на независимые подзадачи и их вычисление на графическом процессоре управления. Кроме этого, приведен пример переобучения сети на более узкую предметную область без перестроения выделенных признаков, что говорит о возможности дообучения классификатора.

3.1.1 Эффективное обратное распространение ошибки

Обратное распространение ошибки — это один из классических методов обучения нейронных сетей, так как данный подход является простым при реализации, вычислительно не очень сложным и почти всегда работающим. Однако, его применение для получения действительно высоких результатов требует подбора большого числа параметров. В работе [16] приведены различные способы, применяемые исследователями для повышения результатов классификации при обучении сети, и их обоснование.

Основные проблемы алгоритма обратного распространения ошибки — это низкая скорость обучения глубоких сетей и попадание в точки локальных ми-

нимумов задачи оптимизации. Для ускорения без существенной потери точности можно воспользоваться стохастическим градиентным спуском. Кроме этого, рассматривались подходы нормализации входных переменных при обучении, перемешивания обучающей выборки, выбора активационной функции и способов начальной инициализации весов.

3.1.2 Использование графического процессора

Использование обучения без учителя позволяет обрабатывать огромные неразмеченные наборы данных и автоматически устанавливать взаимосвязи предметной области. Такой подход удобен тем, что не нужно собирать обучающую и тестовую выборки, которые приходится размечать вручную. С другой стороны, при таком обучении требуется выполнить намного больше вычислительных операций, из-за чего приходится отказываться от использования ЦПУ.

В статье [17] рассматривался пример обучения четырехслойной модели глубокой сети доверия со 100 миллионами настраиваемых параметров на графическом процессоре Nvidia GeForce GTX 280 с объемом оперативной памяти 1Гбайт. Программирование под CUDA SDK. Основная идея заключалась в минимизации числа операций обмена данными между ОЗУ и локальной памятью графического процессора. В итоге было получено увеличение скорости классификации в 5-15 раз по сравнению с вычислениями на ЦПУ.

3.1.3 Дообучение сети

Обучение нейросетей требует много времени на подготовку данных для обучающей выборки и тестирования, а также решение многомерной задачи оптимизации. При необходимости добавить в классификатор новый подкласс рассматриваемой предметной области не всегда есть возможность выполнить полное переобучение сети. Однако, можно попробовать произвести корректировку весов в полносвязных слоях без перестроения признаков.

Такой эксперимент был проведен в работе [18]. После обучения сети на широком классе задачи распознавания растений ее последний классификационный слой был инициализирован случайными весами, а затем произведено дообучение с применением алгоритма стохастического градиентного спуска

на более узком подклассе. Скорость обучения задавалась для каждого слоя в отдельности, наименьший шаг использовался для полносвязных слоев классификации.

С первых итераций корректировки весов наблюдался резкий подъем точности классификации, однако кривая обучения очень быстро вышла на плато в 60-65% для нового подкласса с использованием старых признаков. Сеть выделяла тысячу различных классов. Все вычисления производились с использованием фреймворка Caffe [19] на графическом процессоре NVIDIA Titan Z.

3.2 Определение параметров сети

При использовании нейронных сетей необходимо определить большое количество параметров их архитектуры. В этом разделе рассматриваются различные конфигурации, используемые для решения задач классификации, подобных поставленной НИРС.

3.2.1 Размер сети

Один из основных аспектов любого подхода машинного обучения заключается в балансировке системы между ее сложностью и качеством результата. Как уже говорилось в пункте 2.2.3, использование большого количества слоев необходимо для высокой точности классификации на большое количество классов. Таким образом, ограничение на размер сети сверху определяется временем, которое потребуется на ее обучение. С другой стороны, не всегда при увеличении количества скрытых слоев происходит рост точности, более того, чем больше нейронов и весовых коэффициентов, тем выше вероятность переобучения, когда в сети устанавливаются ложные взаимосвязи между элементами обучающей выборки [20].

Существуют два основных подхода к построению нейросетей — конструктивный и деструктивный. При первом из них вначале берется сеть минимального размера, и постепенно увеличивается до достижения требуемой точности. Деструктивным подход является полной противоположностью: из сети завышенного объема удаляются узлы и связи пока не будет получена модель заданного размера или точности.

В рассмотренных работах конкурса ILSVRC наилучшую точность показала сеть, состоящая из 27 слоев. Однако, стоит отметить, что решаемая задача классификации имела как очень близкие и похожие классы (например, сибирский хаски и чукотская лайка), так и совершенно разные (грибы и грузовые корабли). Предметная область НИРС не содержит таких резких перепадов между моделями автомобилей и больше схожа с работой по распознаванию лиц [21], в которой успешно использовалась сеть намного меньшей размерности — всего по два сверточных и объединяющих слоя, чередующихся друг с другом. Классификация происходила на сорок классов с применением самоорганизующихся карт Кохонена [22].

Используя идею дообучения можно воспользоваться готовыми моделями распознавания изображений с уже вычисленными картами признаков, например VGGNet [23]. Архитектура этой сети состоит из 16 слоев и имеет чрезвычайно однородную структуру, состоящую из поочередного применения операций свертки 3×3 и объединения 2×2 . Главная особенность заключается в том что по сравнению с другими моделями сетей (например GoogLeNet) VGGNet имеет большую обобщающую способность и лучше переносится на различные задачи классификации изображений. В частности, ее предобученная модель доступна для использования в фреймворке Caffe.

3.2.2 Формирование обучающей выборки

При подготовки данных для обучения необходимо подобрать достаточное количество разнообразных входных примеров, отражающих правила и закономерности, которые должны быть обнаружены и сохранены в нейросети [24]. Хорошее обучающего множество обладает следующими свойствами:

- достаточность — число примеров в обучающем множестве больше числа настраиваемых весов;
- разнообразие — в обучающей выборке присутствуют примеры всех выходных классов;
- равномерность представления классов — примеры различных классов представлены в обучающей выборке в одинаковых пропорциях [25];

Если не выполняется первое свойство, то вместо обобщения сеть просто запомнит данные и результат классификации будет неопределен для примеров, которые не вошли в обучающую выборку. Кроме этого, обучающее множество должно полностью покрывать пространство выходных классов. Требование равномерности является защитой от переобучения, когда класс с максимальным числом обучающих примеров будет определяться моделью как наиболее вероятный для любых входных данных.

3.2.3 Процедура обучения

4 План дальнейших исследований

Список литературы

- [1] W. Wu, Z. QiSen, W. Mingjun. A method of vehicle classification using models and neural networks. Vehicular Technology Conference. 2001.
- [2] R. Brehar, S. Nedevschi, L. Dăian. Pillars detection for side viewed vehicles. Intelligent Computer Communication and Processing (ICCP). 2010.
- [3] T. Zhao, R. Nevatia. Car detection in low resolution aerial images. Image and Vision Computing. 2003.
- [4] D. Jang, M. Turk. Car-Rec: A real time car recognition system. Workshop on applications of computer vision (WACV). 2011.
- [5] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool. Speeded-up robust features (SURF). Computer vision and image understanding. 2008.
- [6] H. Wang, Y. Cai, L. Chen. A vehicle detection algorithm based on deep belief network. The scientific world journal. 2014.
- [7] P. Sermanet, S. Chintala, Y. LeCun. Convolutional neural networks applied to house numbers digit classification. International Conference on Pattern Recognition (ICPR). 2012.
- [8] P. Sermanet, K. Kavukcuoglu, Y. LeCun. Eblearn: Open-source energy-based learning in c++. 21st International Conference on Tools with Artificial Intelligence, ICTAI'09. 2009.
- [9] Image Large Scale Visual Recognition Challenge. URL: <http://www.image-net.org/challenges/LSVRC/>.
- [10] A. Krizhevsky, I. Sutskever, G. Hinton. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems. 2012.
- [11] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research. 2014.

- [12] G. E. Dahl, T. N. Sainath, G. E. Hinton. Improving deep neural networks for LVCSR using rectified linear units and dropout. International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2013.
- [13] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich. Going deeper with convolutions. 2014.
- [14] J. Ngiam, Z. Chen, D. Chia, P. Koh, Q. Le, A. Ng. Tiled convolutional neural networks. Advances in Neural Information Processing Systems. 2010.
- [15] A. Krizhevsky, H. Krizhevsky, Geoffrey. Learning multiple layers of features from tiny images. 2009.
- [16] Y. A. LeCun, L. Bottou, G. B. Orr, K. R. Müller. Efficient backprop. Neural networks: Tricks of the trade. 2012.
- [17] R. Raina, A. Madhavan, N. Y. Andrew. Large-scale deep unsupervised learning using graphics processors. Proceedings of the 26th annual international conference on machine learning. 2009.
- [18] A. K. Reyes, J. C. Caicedo, J. E. Camargo. Fine-tuning deep convolutional networks for plant recognition. Working notes of CLEF 2015 conference. 2015.
- [19] Caffe: Convolutional Architecture for Fast Feature Embedding. URL: <http://caffe.berkeleyvision.org/>.
- [20] S. Lawrence, C. L. Giles, A. C. Tsoi. What size neural network gives optimal generalization? Convergence properties of backpropagation. 1998.
- [21] S. Lawrence. Face recognition: A convolutional neural-network approach. 1997.
- [22] T. Kohonen. The self-organizing map. 1990.
- [23] K. Simonyan, A. Zisserman. Very deep convolutional networks for large-scale image recognition. 2014.
- [24] R. Thirumalainambi, J. Bardina. Training data requirement for a neural network to predict aerodynamic coefficients. 2003.

- [25] H. Ian Witten, Eibe Frank and A. Mark. Hall Data Mining: Practical Machine Learning Tools and Techniques. 2011.