

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»

Кафедра математической кибернетики и компьютерных наук

ТАРГЕТИРОВАНИЕ ОБЪЕКТОВ НА ФОТОГРАФИЯХ ПРИ ПОМОЩИ
ГЛУБИННОГО ОБУЧЕНИЯ НА ПРИМЕРЕ FASHION-ИНДУСТРИИ
АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

студента 2 курса 273 группы
направления 01.04.02 — Прикладная математика и информатика
факультета КНиИТ
Тугушева Рушана Харрясовича

Научный руководитель
доцент, к. ф.-м. н.

А. А. Кузнецов

Заведующий кафедрой
к. ф.-м. н.

С. В. Миронов

Саратов 2019

АКТУАЛЬНОСТЬ ТЕМЫ

Машинное обучение стало основой для многих сервисов в современной ИТ индустрии. Глубинное обучение является частью широкого семейства методов машинного обучения. Одним из главных приложений глубинного обучения является распознавание изображений. В последние года распознавание изображений стало настолько точным, что иногда даёт результат более точный, чем люди-участники испытаний. Такие результаты позволили появиться рекомендательным системам нового поколения, применимым в коммерческом использовании. Как пример, их можно использовать для классификации товаров моды, что можно использовать либо для тегирования товаров, чтобы при поиске получатель с большей вероятностью получал товар, который соответствует его поисковому запросу, либо для обнаружения товаров, которые ранее были вручную размечены, но тегирование было выполнено неверно. Наиболее часто такие системы можно применять в сфере fashion-индустрии, а именно в интернет-магазинах одежды, т.к. это позволяет избежать потери возможных сделок между покупателем и продавцом. Целью магистерской работы являются:

- построение нейронной сети, пригодной для классификации предметов одежды;
- увеличение её показателей accuracy, precision, recall.

Поставленные цели определили следующие задачи:

- обзор глубинного обучения;
- обзор практических инструментов;
- построить модель нейронной сети;
- увеличить показатели эффективности сети при помощи подбора алгоритма обучения;
- увеличить показатели эффективности сети при помощи аугментации;
- увеличить показатели эффективности сети при помощи извлечения масок используя ключевые точки.

Методологические основы

Методологические основы таргетирования объектов на фотографиях при помощи глубинного обучения на примере fashion-индустрии представлены в работах *Imagenet classification with deep convolutional neural networks* /

A. Krizhevsky, I. Sutskever, G. E. Hinton; Deep learning in neural networks: An overview / J. Schmidhuber // Neural Networks, Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position / K. Fukushima, Tensorflow: Large-scale machine learning on heterogeneous distributed systems / M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado et al. , Gradient based learning applied to document recognition, Yann LeCun, Leon Bottou

Практическая значимость

Практическая роль работы состоит в том, с её помощью была построена модель, которая может быть использована для использования в e-commerce для проверки правильного тегирования товаров и построения рекомендательных систем.

Структура и объём работы.

Магистерская работа состоит из введения, 3 разделов, заключения, списка использованных источников и 8 приложений. Общий объем работы – 95 страниц, из них 64 страницы – основное содержание, включая 42 рисунков, список использованных источников информации – 121 наименование.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

1 Использование глубинного обучения при таргетировании изображений

1.1 Глубинное обучение

Глубинное обучение это подмножество более широкого множества методов машинного обучения, которые основываются на использовании данных для обучения, в противовес алгоритмам, направленных на конкретные задачи. Обучение есть с подкреплением, без подкрепления и с частичным подкреплением. Модели глубинного обучения некоторым образом связаны с обработкой информации и паттернами обмена информацией в биологической нервной системе. Такие архитектуры как глубокие нейронные сети, сети глубокого убеждения и рекуррентные нейронные сети применяется в отраслях таких как противомошеннические системы, распознавание речи, компьютерное зрение, естественная обработка человеческой речи, распознавание аудио, фильтрация социальных сетей, машинный перевод, биоинформатика и создание лекарственных средств, где они совершили результаты сравнимые и в некоторых случаях превосходящие человеческие.

1.1.1 Определение

Глубинное обучение это класс алгоритмов машинного обучения который:

- использует каскадные или множественные слои нелинейной обработки юнитов для извлечения признаков и трансформации. Каждый последующий слой использует выход из предыдущего слоя как вход
- обученный с учителем/без учителя/комбинируя оба подхода.
- изучает множественные слои представлений который связаны с разными уровнями абстракции; уровни формируют иерархию концепций.

1.1.2 Обзор

Современные способы глубинного обучения основываются на построении искусственных нейронных сетей.

Слово «глубинный» в «глубинном обучении» ссылается на число слоёв через которые данные трансформируются. Точнее, системы глубинного обучения имеют значительную глубину CAP (путь передачи ответственности(англ. credit assignment path, CAP)). CAP это цепь преобразований от входа до выхода. CAP описывают потенциальные причинные связи между

входом и входом. Для сетей прямого распространения глубина САР это число скрытых слоёв плюс 1 (т.к. вывод тоже параметризован). Для рекуррентных нейронных сетей, в который сигнал может проходить более 1 раза, глубина САР потенциально неограничена. Универсальной договорённости о пороге, который разделяет поверхностное обучение и глубинное обучение не существует, но большинство исследователей согласны, что глубинное обучение включает САР глубиной > 2 .

1.1.3 Свёрточные нейронные сети

Сверточная нейронная сеть (CNN или ConvNet) представляет собой класс нейронных сетей, наиболее часто применяемых для анализа визуальных образов.

Они также известны как инвариантные или пространственно-инвариантные искусственные нейронные сети (SIANN), основанные на их общей архитектуре весов и характеристиках трансляционной инвариантности.

Они имеют приложения в распознавании изображений и видео, рекомендательных системах, классификации изображений, медицинском анализе изображений и обработке естественного языка.

Устройство свёрточной нейронной сети Сверточная нейронная сеть состоит из входного и выходного слоев, а также нескольких скрытых слоев. Скрытые уровни CNN обычно состоят из сверточных слоев, слоя с RELU, то есть с функцией активации, pooling уровней, полносвязных слоёв и слоёв нормализации.

Субдискретизация Операция субдискретизации (англ. subsampling, англ. pooling, также переводимая как «операция подвыборки» или операция объединения), выполняет уменьшение размерности сформированных карт признаков. В данной архитектуре сети считается, что информация о факте наличия искомого признака важнее точного знания его координат, поэтому из нескольких соседних нейронов карты признаков выбирается максимальный и принимается за один нейрон уплотнённой карты признаков меньшей размерности. За счёт данной операции, помимо ускорения дальнейших вычислений, сеть становится более инвариантной к масштабу входного изображения.

Слой свёртки Слой свёртки (англ. convolutional layer) — это основной блок свёрточной нейронной сети. Слой свёртки включает в себя для каждого канала свой фильтр, ядро свёртки которого обрабатывает предыдущий слой по фрагментам (суммируя результаты матричного произведения для каждого фрагмента). Весовые коэффициенты ядра свёртки (небольшой матрицы) неизвестны и устанавливаются в процессе обучения.

Слой активации Скалярный результат каждой свёртки попадает на функцию активации, которая представляет собой некую нелинейную функцию. Слой активации обычно логически объединяют со слоем свёртки (считают, что функция активации встроена в слой свёртки). Функция нелинейности может быть любой по выбору исследователя.

Пулинг или слой субдискретизации Слой пулинга (иначе подвыборки, субдискретизации) представляет собой нелинейное уплотнение карты признаков, при этом группа пикселей (обычно размера 2x2) уплотняется до одного пикселя, проходя нелинейное преобразование. Наиболее употребительна при этом функция максимума.

1.2 Практические инструменты

Примеры кода будут на Python. Этот язык стал дефакто стандартом в современном машинном обучении и обработке данных, хотя в нейронных сетях есть и другие примеры — например, библиотека Torch предлагает описывать структуру сетей в виде скриптов на языке Lua 3. Заметим, что это не означает, что сами вычисления в библиотеках целиком написаны на Python — конечно, для оптимизации низкоуровневых вычислений пишут на C и обращаются к еще более низкоуровневым библиотекам — драйверам видеокарты и особенно библиотеке cuDNN.

Основная абстракция, которая потребуется нам для того, чтобы понять все, что происходит в коде таких библиотек, как TensorFlow или Theano, — это граф вычислений.

1.3 Источники данных

Ключевую роль в обучении играют данные, т.к. обучение с учителем позволяет получить намного более быстрые и правильные результаты. Так

как данная работа связана с распознаванием изображений для Fashion индустрии, то необходимы большие объёмы данных изображений. Для данной цели предлагается использовать такие вебсайты как:

1. www.instagram.com
2. www.twitter.com
3. www.pinterest.co.uk
4. www.lamoda.ru
5. www.wildberries.ru

Планируется с помощью поиска по хэштегам получать изображения с помощью специально написанной для этого программы. При этом хэштеги, по которым планируется сделать поиск по данным ресурсам, будут использованы как метки, для того, чтобы можно было их использовать при обучении с учителем.

2 Построение модели

2.1 Обзор моделей

Одно из преимуществ использования фреймворков таких как TensorFlow состоит в том, что они идут со множеством моделей, например таких как Inception-v3, которые уже обучены и способны распознавать тысячи признаков. Inception-v3 натренирована для соревнований ImageNet Large Visual Recognition Challenge используя данные для 2012г. Это стандартная задача в компьютерном зрении, где модели пытаются классифицировать все изображения из заданных 1000 классов, например «Зебра», «Долматинец» или «Посудомоечная машина». Тем не менее, Inception-v3 не тренирована для наших задач (классификации одежды и обуви в e-commerce, и как результат, не имеет интересующих нас атрибутов, таких как «рубашка-поло», например. Однако, мы можем получить преимущество из других аспектов моделей, таких как уже обученные слои векторизации и т.д.. Всё, что нам нужно, это переобучить полносвязный слой, который классифицирует изображения чтобы распознавать наши метки. Цель данной главы предоставить обзор современных, достигших наиболее высоких показателей алгоритмов, основывающихся на классификации изображений наиболее популярного датасета, ImageNet. Будут описаны наиболее современные архитектуры, которые привели к значительному улучшению классификации объектов на изображениях. Указанные датасеты были протестированы на разных версиях датасетах (датасет ImageNet выпускается каждый год всё с большим количеством изображений и меток).

ImageNet База изображений ImageNet это результат сотрудничества между Стэнфордским университетом и Принстонским университетом и стала ссылкой в области компьютерного зрения. Она содержит около 14 миллионов изображения изначально маркированных с помощью Synsets. Изначально соревнование состояло из простой классификационной задачи, где каждое изображение принадлежало к одной категории из 1000, от конкретной породы собак до определения типа еды. В настоящее время ещё проводится валидация с исходными условиями, но в дальнейшем соревнование эволюционировало в мультиклассификационную задачу с сегментацией объектов.

Машинное обучение В состязании ImageNet традиционно имели дело с такими алгоритмами такими как SIFT до поздних 90-х. Тем не менее, рывок в качестве верификации был достигнут с использованием нейронных сетей. Вдохновленная Gradient-Based Learning Applied to Document Recognition, первая модель глубокого обучения была опубликована в ImageNet Classification with Deep Convolutional Neural Networks притянула внимание публики получением ошибки 15.3%, что было лучше чем предыдущий результат в 26.2% используя модель SIFT. Эта знаменитая модель была названа «AlexNet» может быть рассмотрена сегодня, как простая архитектура с 5 последовательными сверточными слоями, max-pool слоями и 3 полносвязными слоями.

Дальнейшее развитие нейронных сетей Начиная со 2012 года, исследователи начали увеличивать количество свёрточных слоёв. В 2014 Симоньян и Циссерман выпустили модель VGG16, состоящую из 16 свёрточных слоёв, нескольких max-pool слоёв и 3-х полносвязных финальных слоёв. Одна из его особенностей в связывании множественных свёрточных слоёв с функцией активацией ReLU, что позволяло создать нелинейные преобразования, т.к. введение нелинейности сделало возможным обучать более сложные шаблоны. Также, они ввели 3x3 фильтры для каждой свёртки (в противовес 11x11 фильтрам для модели AlexNet) и отметили они могут распознать те же паттерны, большие чем фильтры по размеру, при этом снижая количество параметров для обучения. Эти преобразования позволили достигнуть уровня ошибок 7.3% в 2014 ImageNet соревновании.

2.2 Первоначальная модель

В данной главе была построена первоначальная модель, используя фреймворк Keras.

3 Улучшение точности модели

В рамках работы над предыдущими отчётами пользуясь разработанной программой (Приложение ??) мы сумели достичь 69.7% ассигасу воспользовавшись техникой Transfer Learning , воспользовавшись моделью Inception v3 и обучив её на нашей выборке. Это является хорошим результатом, однако, т.к. целью данной работы является создание модели, применимой в production, это является недостаточно хорошим результатом, т.к. автоматическое тегирование товаров с ошибкой в $1/3$ случаев является недопустимым результатом. Существует множество направлений, в которых можно провести работы по улучшению существующего ассигасу.

3.1 Аугментация данных

В рамках данной главы были исследовано влияние на эффективность нейросети следующих способов аугментации:

- Отзеркаливание
- Вращения изображения
- Масштабирование
- Cutout regularization

3.2 Применение оптимизационных алгоритмов при обучении

Далее было исследовано влияние применения алгоритма обновления весов в ADAM. Алгоритм оптимизации ADAM является расширением стохастического градиентного спуска, который в последнее время получил широкое распространение для приложений глубокого обучения в области компьютерного зрения и обработки естественного языка. ADAM отличается от классического стохастического градиентного спуска. Стохастический градиентный спуск поддерживает единую скорость обучения (называемую альфа) для всех обновлений веса, и скорость обучения не изменяется во время тренировки. Скорость обучения поддерживается для каждого веса сети (параметра) и отдельно адаптируется по мере развития обучения. Метод ADAM вычисляет индивидуальные адаптивные скорости обучения для различных параметров из оценок первого и второго моментов градиентов. Авторы описывают ADAM как объединение преимуществ двух других расширений стохастического градиентного спуска. Его преимущества состоят в:

1. Прост для реализации.

2. Вычислительно эффективен.
3. Маленькие требования к памяти.
4. Инвариантен к изменению масштаба градиента по диагонали.
5. Хорошо подходит для задач, которые являются большими с точки зрения данных и / или параметров.
6. Подходит для нестационарных целей.
7. Подходит для задач с очень шумными / или редкими градиентами.
8. Гиперпараметры имеют интуитивно понятную интерпретацию и обычно требуют небольшой настройки.

3.3 Использование ключевых точек для улучшение точности

3.3.1 Ключевые точки

Извлечение эффективных признаков является ключевым шагом во многих алгоритмах машинного обучения и компьютерного зрения и их приложениях. В компьютерном зрении одна из форм выделения признаков связана с обнаружением и описанием важных областей изображения. Традиционно эти признаки извлекаются с использованием детекторов и дескрипторов, спроектированных вручную. Подходы, принимающие эту парадигму, обычно называют подходами, основанными на ключевых точках или признаках. В последнее время реинтродукция нейронных сетей во многих задачах компьютерного зрения в значительной степени заменила ручной подход, основанный на функциях. Подходы, основанные на нейронных сетях, обычно изучают извлечение признаков как часть сквозного конвейера. Хотя эти подходы показали большой успех в таких задачах, как распознавание сцены, обнаружение и классификация объектов, другие задачи, такие как структура движения, все еще зависят от вручную разработанных признаков, например, SIFT, для обнаружения и описания ключевых точек.

3.3.2 Оценка позы человека

Введение

Проблема оценки позы человека включает в себя определение местоположения ключевых точек тела, которое включает в себя основные части тела и суставы. Существуют различные приложения, связанные с этой проблемой, такие как классификация действий и прогнозирование движения тела. Определение ключевых точек тела оказалось сложной задачей из-за неболь-

ших суставов, окклюзий и необходимости учитывать контекст. В последнее время сверточные нейронные сети достигли значительных успехов в классификации изображений и проблем локализации объектов. Они очень похожи на обычные нейронные сети в том, что они составлены из нейронов с обучаемыми весами и уклонами. Однако нейронные сети плохо масштабируются для больших изображений. Каждый нейрон в слое полностью связан со всеми нейронами в предыдущем слое, таким образом, мы быстро генерируем огромное количество параметров и в итоге получаем переобучение на наборе данных. Сверточные нейронные сети используют тот факт, что входные данные состоят из изображений, поэтому они более разумно ограничивают архитектуру, что значительно уменьшает количество параметров. Сверточные нейронные сети привлекательны для оценки позы человека по двум причинам. Во-первых, нет необходимости явно проектировать представления элементов и детекторы для деталей, потому что модель и элементы извлекаются из данных. Во-вторых, изученная модель является целостной, где окончательные совместные оценки основаны на сложном нелинейном преобразовании полного изображения, в отличие от локальных детекторов, чьи рассуждения ограничены к одной части и может моделировать только небольшое подмножество действия между частями тела. Вот только некоторые из проблем в прогнозировании человеческих позных координат: ракурс конечностей, окклюзия конечностей, вращение и ориентация фигуры, а также наложение нескольких предметов.

Постановка задачи Наша задача состоит из двух задач: оценка позы человека и классификация действий. Для оценки позы сеть принимает в качестве входных данных необработанное изображение и выводит вектор координат ключевых точек тела. Мы стремимся идентифицировать координаты x-y пикселей для 14 суставов тела.

Для последней проблемы мы стремимся маркировать изображения на основе категории активности (20 классов, например, зимняя активность) и конкретной активности (410 классов, например, катание на горных лыжах).

Связанные задачи Свёрточные нейронные сети исторически использовались для задач классификации, но они все чаще применяются к проблемам локализации / обнаружения. Классификационные свёрточные нейронные сети могут быть преобразованы в локализационные свёрточные сети путем за-

мены окончательного классификационного слоя регрессионным слоем, для которого активации это вещественные значения и используя функции потерь регрессии. В качестве альтернативы Sermanet et al. предлагает комплексный подход к обнаружению, распознаванию и локализации объектов с помощью единой свёрточной нейронной сети. На высоком уровне абстракции деятельность человека часто может быть точно охарактеризована с точки зрения позы тела, движения и взаимодействия с объектами сцены [7]. Однако из-за сложного характера этой проблемы большинство современных моделей распознавания активности полагаются на целостные представления, которые извлекают элементы внешнего вида и движения из видео, из которого извлекаются изображения. Недавно Toshev et al. показали, что применение глубоких глубоких свёрточных нейронных сетей для оценки позы как проблемы регрессии имеет преимущество в том, что рассуждения о позе просты, но целостны. Они сформулировали оценку позы как проблему регрессии сустава тела, в которой местоположение каждого сустава прогнозируется с использованием 7-слойного CNN, в котором входом является полное изображение. Этот подход позволил получить современную оценку PСР (0,61) и был намного проще чем предыдущие методы, основанные на явно разработанных представлениях объектов и графических моделях.

Подходы к извлечению данных с помощью ключевых точек

Coco dataset Одна из главных целей компьютерного зрения это понимание визуальных сцен. Понимание сцен включает множественные задачи, включая понимание объектов, которые представлены на изображениях, локализация объектов в 2D и 3D, определение атрибутов объектов и сцен, характеристика отношений между объектами и предоставление семантического описания сцены.

Microsoft представила новый крупномасштабный набор данных, который направлен на решение ключевых исследовательских задач в понимании сцен: детектирование неконических изображений объектов, контекстуальный смысл между объектами и точной 2D локализации объектов.

Чтобы создать крупномасштабный набор данных который удовлетворяет этим трём целям была создана программа по сбору данных с помо-

пью обширного использования сервиса Amazon Mechanical Turk. Во-первых, Microsoft добыл большой набор изображений содержащих контекстуальные отношения и неканонические виды на объекты.

Далее, каждое изображение было размечено как содержащееся в какой-либо категории используя иерархическую структуру. Для каждой категории каждая индивидуальная копия промаркирована, проверена и сегментирована. Набор данных Microsoft Common Objects in Context (MS COCO) содержит 91 распространённую категорию объектов с 82 категориями имеющими более 5000 промаркированных изображений. В общей сложности датасет имеет 2 500 000 экземпляров в 328 000 изображениях. По сравнению с популярным ImageNet, COCO имеет меньше категорий, но больше экземпляров объектов на категорию. Это может помочь в обучении детализированных объектных моделей способных к точной 2D локализации изображений. COCO также значительно больше по количеству изображений на категорию чем PASCAL VOC или SUN наборы данных. Вдобавок, критическое отличие между COCO и другими наборами данных это количество промаркированных изображений на категорию что может помочь в обучении контекстной информации.

3.3.3 Использование нейронных сетей для использования ключевых точек

Введение Компанией Microsoft была представлена новая архитектура, а именно High Resolution Net (HRNet), который способен поддерживать представление в высоком разрешении в течение всего процесса обучения. Обучение начинается из подсети с высоким разрешением в качестве первого этапа, постепенно добавляются подсети с высоким и низким разрешением по очереди, чтобы сформировать больше ступеней, и параллельно подключать подсети с несколькими разрешениями. Проводятся многократные разномасштабные слияния, обмениваясь информацией через подсети с множественным разрешением параллельно снова и снова в ходе всего процесса. Оценка ключевых точек происходит по представлениям высокого разрешения, выводимым данной сетью.

3.3.4 Обучение сети для извлечения ключевых точек

Для того, чтобы максимально корректно вырезать изображения, было решено использовать нахождение ключевых точек на фигуре, для того, чтобы получить область, которая будет содержать в себе минимальное количество фона и при этом содержание вырезанное по прямоугольнику захватит человека, чтобы не повлечь игнорирование признаков, что приведёт к некорректной классификации.

Для получения массива векторов состоящего из ключевых точек для получения координат по обрезке изображений была использована нейросеть HR-48, разработанная компанией Microsoft, о которой было рассказано выше. В данный момент она является лучшей сетью в мире по показателю точности предсказаний ключевых точек на фигуре человека, что и нужно нам при предобработке данных. Вышеуказанная нейросеть была обучена на нейросети Coco.

Ниже показана структура проекта:

Полный исходный код доступен по адресу github.com/haskellx/deep

После тренировки нейросети возможно перейти к оценке нашего набора данных на имеющемся наборе данных.

В коде, указанном в Приложении представлен модуль для работы с ключевыми точками. Ниже представлен модуль для работы с ключевыми точками.

После этого у нас имеется очищенный набор данных, который возможно использовать для улучшения предсказаний.

3.3.5 Использование обученной сети для увеличения точности предсказаний

Имея набор исправленный набор данных, стало возможно обучить нейросеть без лишних данных. По-прежнему используется тот же набор инструментов: Keras, TQDM, Python 3, Tensorflow. Так как используется технология transfer learning, была взята предобученная сеть Inception V3 с весами для набора ImageNet.

Результаты эксперимента можно признать успешными, так как обучение на масках позволило достичь повышение прироста эффективности предсказаний сети до 81%.

ЗАКЛЮЧЕНИЕ

В рамках магистерской работы была проведена обширная работа.

Была дан обзор технологии глубинного обучения, практических инструментов для работы с глубинным обучением, дан пример работы с низкоуровневой работы с библиотекой Tensorflow.

В дальнейшем, построена первоначальная модель для работы с классификацией изображений на основе технологии Transfer Learning, заключающейся в использовании предобученной сети, как правило обученной на огромном источнике данных, что как правило недоступно для небольших организаций, с последующим дообучением на последнем слое и соответственно изменением классификатора.

Далее проведены эксперименты по увеличению показателей полученной сети с использованием аугментации - были получены как отрицательные, так и отрицательные результаты (для различных видов аугментаций). Так же был изменён метод поиска оптимальных весов с классического градиентного спуска на более современный ADAM, был получен положительный результат.

В последней части работы были использованы маски для обучения, извлекая маски при помощи обучения сети HRNet и обучения на наборе изображений COCO. Получен значимый прирост эффективности.

Таким образом, поставленные цели были достигнуты.

ОСНОВНЫЕ ИСТОЧНИКИ ИНФОРМАЦИИ

1. Krizhevsky, A. Imagenet classification with deep convolutional neural networks / A. Krizhevsky, I. Sutskever, G. E. Hinton // Neural Information Processing Systems.
2. Schmidhuber, J. Deep learning in neural networks: An overview / J. Schmidhuber // Neural Networks.
3. Fukushima, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position / K. Fukushima // Biological cybernetics.
4. Tensorflow: Large-scale machine learning on heterogeneous distributed systems / M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado et al.

5. Yann LeCun Leon Bottou, Y. B. P. H. Gradient based learning applied to document recognition