УДК 004.032.26

ИССЛЕДОВАНИЕ СПОСОБНОСТИ К TRANSFER LEARNING СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ, ОБУЧЕННЫХ НА IMAGENET

Богатырева А.А., Виноградова А.Р., Тихомирова С.А.

Национальный исследовательский ядерный университет «МИФИ», Москва,
e-mail: A2Bog@list.ru, vinogradova.a.rom@gmail.ru, tikhomirova3112@yandex.ru

В настоящее время нейронные сети применимы к различным задачам. Этих задач настолько много, что при решении каждой новой задачи построение новой нейросетевой модели для этой задачи утрачивает смысл. Поэтому в последнее время исследования в области переноса знаний (англ. transfer learning) стали снова набирать популярность. Более того, с каждым годом появляются все более и более сложные модели нейронных сетей, которые хорошо решают существующие задачи, и при решении новой задачи довольно сложно сразу понять, какая модель сети будет решать эту задачу наиболее качественно. Для того, чтобы проводить исследования в данном направлении, необходимо сначала понять, насколько применима концепция transfer learning к различным моделям нейронных сетей. В данной работе рассматривается концепция transfer learning применительно к различным задачам, в том числе к задачам классификации изображений, а также проводится обзор нескольких архитектур сверточных нейронных сетей (СНС), которые были заранее обучены на выборке изображений ImageNet. Было проведено исследование, на основании которого сделаны выводы о преимуществах обучения с использованием методов transfer learning по сравнению с полным обучением нейросети.

Ключевые слова: машинное обучение, классификация изображений, transfer learning, сверточные нейронные сети, предобученные нейронные сети

STUDY OF THE ABILITY OF CONLUTIONAL NEURAL NETWORKS PRETRAINED ON IMAGENET TO TRANSFER LEARNING

Bogatyreva A.A., Vinogradova A.R., Tikhomirova S.A.

National Research Nuclear University «MEPhI» (Moscow Engineering Physics Institute), Moscow,
e-mail: A2Bog@list.ru, vinogradova.a.rom@gmail.ru, tikhomirova3112@yandex.ru

Currently, neural networks are applicable to various tasks. There are so many of these tasks that when solving each new task, building a new neural network model for this task loses its meaning. Therefore, in recent years, research in the field of transfer learning has begun to gain popularity again. Moreover, every year more and more complex models of neural networks appear that solve existing problems well, and when solving a new problem it is rather difficult to immediately understand which network model will solve this problem most qualitatively. In order to conduct research in this direction, you must first understand how applicable the concept of transfer learning to different models of neural networks. In this work the concept of transfer learning applies to various tasks and, in particular, to image classification tasks, is considered, and a review of several architectures of convolutional neural networks (CNN), previously trained (pretrained) on a sample of images ImageNet. A study was conducted and conclusions about benefits of training CNN using transfer learning versus full training this CNN were made.

Keywords: machine learning, image classification, transfer learning, convolutional neural network, pretrained neural networks

Самое раннее упоминание о концепции переноса знаний (transfer learning) в машинном обучении датируется 1993 г. в работе [1], однако более подробно оно было рассмотрено в 1997 г. в журнале Machine Learning. Концепция заключается в передаче знаний, полученных в одной или нескольких исходных задачах, и использовании их для улучшения обучения в текущей задаче. Методы, обеспечивающие передачу знаний, направлены на то, чтобы процесс машинного обучения был таким же эффективным, как и обучение человека. В результате стало возможным переобучить СНС, обученную на одной выборке данных, для выполнения задач на новом множестве данных, что существенно ускорило процесс обучения сети.

В настоящее время проводится много исследований в направлении transfer learning. Например, ведутся работы по переносу знаний между текстами и изображениями [2], осуществляется перенос знаний из неразмеченных данных [3] и другие. Стоит заметить, что данная концепция успешно используется и на практике. Например, на различных соревнованиях, где любой исследователь может опробовать свои алгоритмы и модели анализа данных на серьезных актуальных практических задачах, участники стали все чаще использовать методы transfer learning для улучшения качества обучения своих моделей нейронных сетей. Так, для задачи распознавания диабетической ретинопатии, некоторые участники использовали transfer learning [4–5].

Стоит отметить и тот факт, что уже существует множество различных архитектур СНС, которые хорошо выполняют те или иные задачи, в том числе и задачи класси-

фикации изображений, поэтому при решении задач на новых данных зачастую проще и эффективнее выбрать одну из уже существующих нейронных сетей, а не строить её с нуля.

Для решения задачи классификации изображений в начале работы с новой выборкой данных, возникает вопрос о том, какую архитектуру нейронной сети использовать. Существуют два подхода к решению этой проблемы: построить свою архитектуру сети или выбрать одну из существующих. Первый подход имеет несколько существенных недостатков:

– зачастую необходимы большие затраты времени для построения корректно и быстро работающей сети;

– архитектуры сетей с каждым годом становятся все сложнее и многослойнее, также есть тенденция использования ансамблей сетей для обучения, что требует больших затрат ресурсов;

– для обучения требуются большие объемы данных.

Второй подход предполагает применение transfer learning, которое заключается в использовании существующей нейронной сети, обученной на какой-то задаче. На новой выборке данных эту сеть обучают уже не полностью, а только несколько последних слоев, предполагая, что при прогоне каждого изображения из этой выборки до этих слоев сеть выделила признаки, несущие всю необходимую информацию об изображении.

В данной работе ставится цель исследовать способность сверточных нейронных сетей (СНС) к transfer leaning между различными задачами классификации размеченных изображений.

Материалы и методы исследования

Обзор сверточных нейронных сетей. В задаче классификации нейронная сеть тем лучше, чем точнее она классифицирует объект, который ей подается на вход впервые. Классические полносвязные нейронные сети из-за большого количества связей между нейронами не подходят для такой задачи и необходимо подобрать другую архитектуру. Сверточная нейронная сеть – одна из самых эффективных архитектур для решения данной задачи. Это было, в частности, доказано на практике, когда на международном соревновании ImageNet в 2012 г. победила нейронная сеть AlexNet, в которой впервые была показана реализация сверточного слоя. Стоит заметить, что с тех пор во все последующие годы в сетях-победителях присутствуют сверточные архитектуры.

Сверточная нейронная сеть (СНС) представляет собой специализированный вид нейронной сети для обработки данных, имеющих сеточную топологию (grid cell topology). Архитектура состоит из следующих слоев:

– сверточный слой (convolutional operation). Слой получил свое название, так как основывается на операции свертки (convolutional), представляющей собой классическую математическую операцию свертки для вычисления определенных признаков. При свертке используется матрица весов небольшого размера (ядро свертки), которую двигают по всему обрабатываемому слою, формируя после каждого сдвига сигнал активации для нейрона следующего слоя с аналогичной позицией. В результате операции получаются карты активации (карты признаков);

– слой активации (detector stage), который представляет собой нелинейную функцию активации;

– слой субдискретизации (pooling function). Функция субдискретизации состоит в уменьшении размерности сформированных карт признаков. Выполняя данную операцию, исходят из идеи, что факт наличия искомого признака важнее, чем его местоположение на карте.

Комбинации этих слоев формируют иерархии усложняющихся признаков.

В данной работе рассматриваются следующие архитектуры СНС: AlexNet, VGG, ResNet и DenseNet с различным числом слоев.

AlexNet. AlexNet является сетью, включающей в себя сверточные слои (рис. 1). Как можно видеть, она представляет собой комбинацию сверточных слоев, операций субдискретизации и только последние три слоя являются полносвязными. Причем размеры ядра операции свертки сначала берут большие и уменьшают в процессе прохода по сети. Это можно объяснить тем, что пиксели входного изображения могут быть сильно коррелированы и рецепторную область можно брать большой, не боясь потерять информацию. Операция субдискретизации проходит после каждой свертки, увеличивая плотность некоррелированных участков. В итоге, после прохода по сети, получают класс, к которому относится входное изображение. Сеть содержит 650 тыс. нейронов с 60 млн параметров. Для того, чтобы итеративно научиться находить нужные веса, требуется очень много примеров и времени. Для достижения на соревнованиях ImageNet в 2012 г. (ILSVRC 2012) ошибки top-5 в 16,4% обучался ансамбль из семи сверточных сетей на двух GPU.

VGG. Нейронные сети VGG уже глубже AlexNet – они состоят из 11–19 слоев. Стоит заметить, что размеры ядер операции свертки тут преимущественно равны 3*3 или 1*1, в отличие от предыдущей сети, что положительно сказывается на общем количестве параметров. Например, при свертках с ядрами 3*3 и 7*7 рецептивные поля одинаковые, однако число параметров меньше. Соответственно, обучение такой модели требует меньше памяти. Архитектура VGG-16 представлена на рис. 2.

Согласно данным из [6] на наборе данных ImageNet VGG-13 выдает ошибку 10,75%, VGG-16 – 9,62%, а VGG-19 – 9,12%. Из этого пытались сделать вывод, что чем глубже сеть, тем она будет более результативной, однако на практике выяснилось, что это не так: качество модели дошло до какого-то предела, а затем начало падать. Стоит отметить, что ансамбль моделей VGG-16 и VGG-19 стал победителем ImageNet в 2014 г. с результатом ошибки top-5 7,3%.

ResNet. Авторы Residual network смогли найти такую топологию, чтобы при увеличении глубины сети ее результативность, по крайней мере, не уменьшалась. Сеть состоит из блоков, реализующих residual-функции (так называемые residual-блоки), за

что и получила свое название. Подробнее специфика сети описана в работе [7]. Архитектура сети ResNet состоит из слоев, которые, в свою очередь, состоят из таких блоков. Последние два слоя – субдискретизация (average pooling) и полносвязный слой. Стоит отметить, что тут не использовался Dropout, что сохраняет больше информации. Отметим, что число параметров стало еще меньше, чем в предыдущих сетях. Ансамбль из шести сетей типа ResNet победил на ILSVRC 2015 с результатом ошибки top-5 3,57%, что превосходит результат человека.

DenseNet. Для улучшения информационного потока между слоями авторы работы [8] предложили немного отличную от ResNet схему: они ввели прямые соединения из любого слоя во все последующие. Благодаря такой плотной связи сеть и получила свое название. Также было уменьшено число карт признаков на переходных слоях. У данной сети обнаружена такая же особенность, как и у ResNet: при увеличении глубины улучшается точность модели. Если сравнить с ResNet, то DenseNet оказалась эффективнее. Результаты приведены в работе [8].

В данной работе рассматривается применение transfer learning при решении задач классификации, имеющих различные предметные области и параметры изображений.

MNIST. База изображений, представляющих собой рукописные числа. Число классов – 10 (от 0 до 9 включительно). Каждый объект – это серое полутоновое изображение с размерами 28*28 и одним цветовым каналом. В базе содержится 70 тыс. объектов: в обучающей выборке – 60 тыс. и в тестовой – 10 тыс.

Fashion-MNIST. База изображений, представляющая собой изображения одежды. Число классов – 10: футболка/топ, брюки, пуловер, платье, пальто, сандалия, рубашка, кед, сумка и ботильон. Этот датасет похож на MNIST и может служить прямой его заменой. Параметры изображений аналогичны.

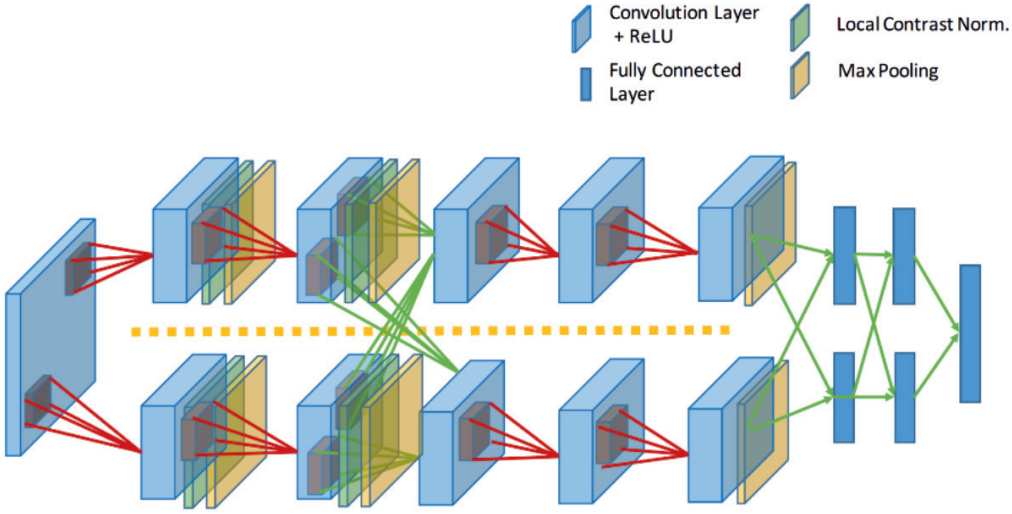

Рис. 1. Архитектура CHC AlexNet

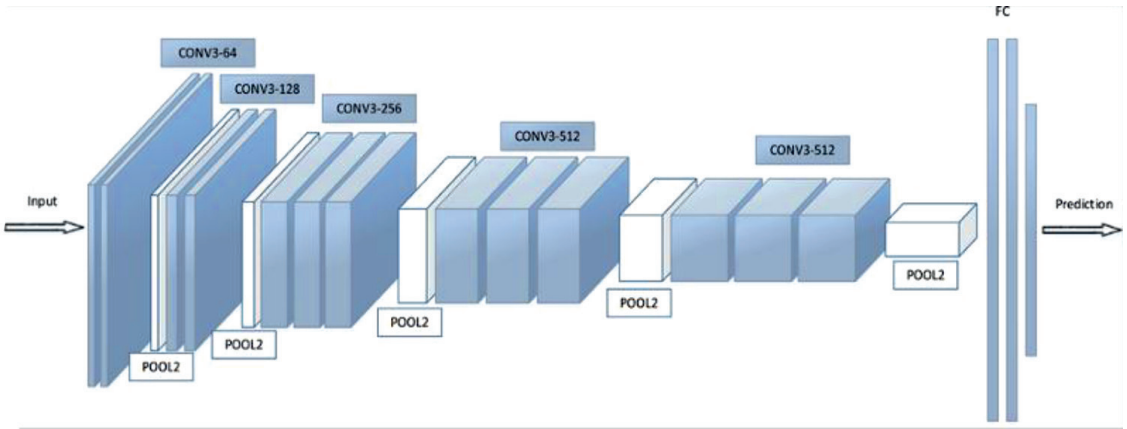

Рис. 2. Архитектура CHC VGG-16

CIFAR-10. База размеченных цветных изображений. Число классов – 10: самолет, автомобиль, птица, кошка, олень, собака, лягушка, лошадь, корабль, грузовик. Размеры каждого изображения – 32*32, 3 цветовых канала. Всего 60 тыс. изображений: по 6 тыс. на класс. Обучающая выборка состоит из 50 тыс., а тестовая – из 10 тыс.

CIFAR-100. Описание объектов похоже на предыдущее с отличием лишь в том, что в данном наборе 100 классов и, соответственно, на каждый класс приходится по 6 тыс. объектов. Стоит отметить особенность, что в данном датасете 20 суперклассов, к каждому из которых принадлежит по 5 классов. Отсюда следует тот факт, что у каждого изображения по 2 метки: «fine»-метка – класс, к которому принадлежит объект – и «coarse»-метка – обозначает суперкласс.

STL-10. Набор изображений для алгоритмов unsupervised learning и self-taught learning. Похож на CIFAR-10, однако имеет некоторые отличия:

– каждый класс имеет меньшее число примеров, чем в CIFAR 10;

– имеется большой набор (100 000) неразмеченных примеров для алгоритмов unsupervised learning. Стоит отметить, что среди этих примеров есть изображения, классы которых не представлены STL-10 (например, кролики, поезда и др.).

Число классов – 10: самолет, птица, машина, кошка, олень, собака, лошадь, обезьяна, корабль, грузовик. Размеры изображений – 96*96, 3 цветовых канала. 500 обучающих размеченных примеров на каждый класс (т.е. всего 5 тыс. изображений в обучающей выборке) и по 800 тестовых изображений на класс. Изображения были взяты из размеченных примеров ImageNet.

ASL Alphabet. Представляет собой выборку фотографий букв алфавита американского языка жестов. Число классов – 29: 26 представлены буквами A-Z и 3 – SPACE, DELETE и NOTHING. Последние три класса делают датасет приближенным к реальности, что является очень полезной практикой. Размеры каждого изображения – 200*200, 3 цветовых канала. Обучающая выборка состоит из 87 тыс., а тестовая – из 29 элементов.

Chest X-RAY Images. Набор данных представлен рентгеновскими снимками грудной клетки (передняя – задняя), разделенными по критерию наличия пневмонии. Число классов – 2: NORMAL или PNEUMONIA. Размеры снимков различны, но у каждого из них по 3 цветовых канала. Всего 5856 снимков, разделенных следующим образом: обучающая выборка состоит из 5216, валидационная – из 16, а тестовая – из 624 элементов.

10 Monkey Species. Представляет собой выборку фотографий обезьян десяти различных видов из Wikipedia's monkey cladogram. Число классов – 10, каждый класс соответствует какому-то одному виду обезьяны. Размеры каждого изображения – 400*300, 3 цветовых канала. Обучающая выборка состоит из 1097 тыс., а тестовая – из 272 элементов.

Transfer learning. Человеческий мозг способен к переносу знаний между задачами. Так, мы применяем соответствующие знания из своего накопленного опыта, когда сталкиваемся с новыми задачами, похожими на те, с которыми мы уже сталкивались ранее. Причем, чем теснее задача связана с предыдущим опытом, тем проще нам с ней справиться. Довольно популярным является пример про шашки и шахматы: умея играть в шашки, мы не учимся заново, с чистого листа, играть в шахматы, а только дообучаемся.

Основные алгоритмы машинного обучения (traditional machine learning), напротив, предназначены для отдельных конкретных задач: классификация кошек и собак, классификация чисел и др. Схема показана на рис. 3 (слева). Transfer learning, или явление переноса знаний – это попытка смоделировать эту особенность мозга человека в математическом аспекте. Концепция переноса знаний заключается в передаче знаний, полученных в одной или нескольких исходных задачах, и использования их для улучшения обучения в текущей задаче. Схема проиллюстрирована на рис. 3 (справа). Таким образом, машинное обучение может стать столь же эффективным, как и обучение человека.

Целью transfer learning является улучшение качества обучения в текущей (целевой) задаче, используя знания, полученные ранее из исходной задачи. Таким образом, перенос знаний может положительно повлиять на следующие показатели обучения:

– начальная производительность, достижимая в текущей задаче путем подбора начальных параметров модели, используя априорную информацию (переданные знания);

– время, которое занимает полное изучение текущей задачи с учетом переданных данных;

– конечный уровень производительности, достижимый в текущей задаче.

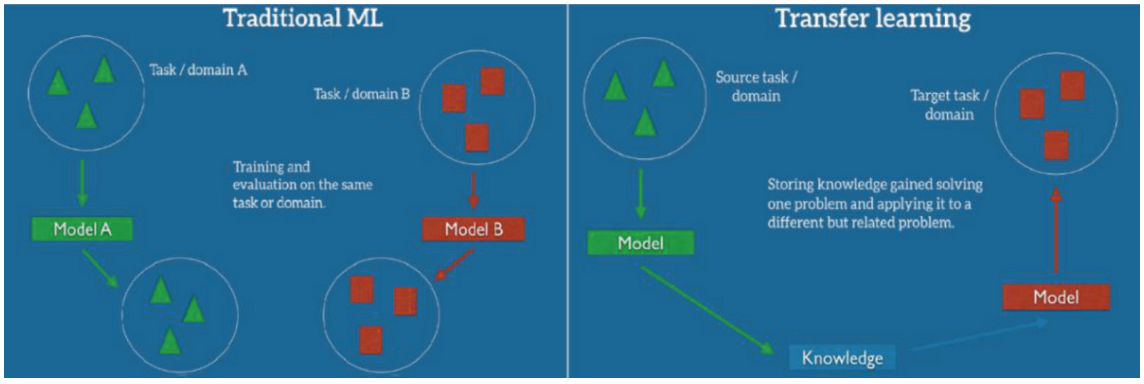

Рис. 3. Отличие классического машинного обучения (traditional ML) от обучения с применением transfer learning

Бывают случаи, когда исходная задача недостаточно тесно связана с текущей задачей или эта связь не учтена в полной мере при использовании метода переноса знаний. Тогда производительность может как остаться на том же уровне, так и уменьшиться. Такое явление называют отрицательный обмен (negative transfer). Таким образом, одна из задач исследователей в области transfer learning – получить положительный результат при переносе знаний между похожими задачами и устранить отрицательный обмен между малосвязанными между собой задачами.

Transfer learning имеет особенность, отличающую его от классического многозадачного обучения нейронных сетей: возможность передавать информацию строго в направлении от исходной задачи к текущей, в то время как в многозадачном обучении различные модели при решении задачи могут обмениваться информацией в любом направлении.

В данной статье рассматривается применение transfer learning к задачам классификации изображений, где обучение происходит на выборке данных, состоящих из пар «объект – метка».

Метод исследования. В процессе обучения СНС без использования transfer learning на каждой эпохе обучения происходит корректировка значений всех весов каждого слоя сети, т.е. происходит стандартный процесс обучения нейронной сети на выборке данных, представленных изображениями.

Обучение предобученных СНС с использованием transfer learning может быть произведено двумя способами [6]:

1) finetuning the convnet (тонкая настройка сети): вместо произвольной инициализации сети берут веса предобученной на большой выборке сети. В процессе обучения не только переобучается классификатор для нового набора данных, но также происходит точная настройка весов сети с помощью обратного распространения ошибки;

2) convent as fixed feature extractor: при инициализации предобученной сети «замораживают» веса всех слоев, за исключением последнего полносвязного слоя. Этот слой заменяется на новый со случайными весами, и только он обучается.

Таким образом, методика исследования заключается в загрузке предобученных сверточных нейронных сетей, рассмотренных ранее, и в их обучении с использованием методов transfer learning. Предлагается следующий алгоритм:

– загрузить обучающую и тестовую выборки изображений;

– загрузить предобученную модель СНС;

– зафиксировать («заморозить») параметры всех слоев сверточного блока нейронной сети;

– в блоке классификатора последний полносвязный слой заменить на новый с числом выходов равным числу классов в выборке, инициализировав его параметры случайными значениями, распределенными по нормальному закону;

– провести обучение такой СНС на заданной выборке.

В результате будет сделан вывод, к каким задачам классификации методы transfer learning применимы и применимы ли в общем случае.

Результаты исследования и их обсуждение

В табл. 1 приведены показатели точности классификации (англ. accuracy) некоторых СНС после обычного обучения.

Таблица 1

Точность (accuracy) классификации СНС изображений из тестовой выборки после классического обучения

	MNIST	CIFAR-10	CIFAR-100	Fashion MNIST	STL-10	ASL Alphabet	Chest X-RAY Images	10 Monkey Species
AlexNet	0,9534	0,9497	0,6350	0,8043	0,9366	0,9497	0,8562	0,9405
VGG 16	0,9687	0,9251	0,5913	0,7513	0,9436	0,8067	0,7949	0,9428
VGG 19	0,9686	0,9172	0,6199	0,7664	0,9487	0,8705	0,7865	0,9488

Таблица 2

Точность (accuracy) классификации СНС изображений из тестовой выборки после обучения с помощью transfer learning

	MNIST	CIFAR-10	CIFAR-100	Fashion MNIST	STL-10	ASL Alphabet	Chest X-RAY Images	10 Monkey Species
AlexNet	0,9479	0,8381	0,6206	0,8709	0,9016	0,9643	0,9016	0,9540
VGG 16	0,9230	0,8423	0,6240	0,8627	0,9528	0,8214	0,8188	0,9717
VGG 19	0,9180	0,8438	0,6300	0,8619	0,9520	0,9301	0,7938	0,9817
ResNet 50	0,9234	0,8180	0,6176	0,8463	0,9388	0,9643	0,8438	0,9960
ResNet 101	0,9148	0,8024	0,5908	0,8483	0,9381	0,9643	0,8438	0,9831
ResNet 152	0,9171	0,8032	0,5896	0,8495	0,9508	0,9485	0,8422	0,9931
DenseNet 121	0,9374	0,8257	0,6115	0,8509	0,9606	0,9716	0,8094	0,9896
DenseNet 161	0,946	0,8837	0,6723	0,8727	0,9644	0,9878	0,8797	0,9966
DenseNet 169	0,9483	0,8545	0,6946	0,8666	0,9671	0,9286	0,8391	0,9930

Критерием останова в процессе обучения является число эпох обучения. На такой процесс обучения было затрачено примерно от 5 до 8 ч.

В табл. 2 приведены показатели точности (accuracy) классификации изображений предобученными нейронными сетями после обучения с использованием методов transfer learning. Критерием останова в процессе обучения является число эпох обучения. На такой процесс обучения было затрачено от 3 мин до 4 ч (для нейросетей с очень большим числом слоев). Стоит отметить, что эти показатели могут быть выше, но для этого необходим другой критерий останова, а также требуется больше времени и более мощные компьютерные ресурсы.

Исходя из результатов, приведенных в табл. 1–2, можно сказать, что применение в процессе обучения методов transfer learning применимо для различных задач классификации, причем результаты классификации изображений из тестовых выборок довольно высокие в обоих вариантах обучения, однако на обычное обучение тратится больше времени. Более того, не всегда возможно полное обучение многослойных сверточных моделей сетей, так как зачастую для этого требуются мощные компьютерные ресурсы.

Выводы

В данной статье была рассмотрена концепция transfer learning применительно к задачам классификации изображений, которые отличаются предметной областью и параметрами изображений, а также был проведен краткий обзор нескольких сверточных архитектур. Было проведено исследование, задача которого провести два различных процесса обучения (обычное обучение и обучение с использованием методов transfer learning) на нескольких выборках изображений и сравнить полученные показатели. Исходя из результатов, можно сделать вывод, что обучение СНС с применением методов transfer learning применимо к различным задачам классификации, причем обучение было более эффективно при значительно меньших временных затратах. Довольно важным является также объем компьютерных ресурсов, затрачиваемых на исследование, так как не всякую модель СНС можно полностью обучить, не имея при этом мощных процессоров. Transfer learning позволяет частично решить данную проблему, так как в процессе такого обучения изменяются параметры не всех, а только нескольких слоев.

Список литературы

1. Pratt L.Y. Discriminability-based transfer between neural networks. NIPS Conference: Advances in Neural Information Processing Systems. 1992. vol. 5. P. 204–211.

2. Yin Zhu, Yuqiang Chen, Zhongqi Lu, Sinno Jialin Pan, Gui-Rong Xue, Yong Yu, Qiang Yang. Heterogeneous Transfer Learning for Image Classification. Twenty-Fifth AAAI Conference on Artificial Intelligence. 2011. P. 1304–1309.

3. Rajat Raina, Alexis Battle, Honglak Lee, Benjamin Packer, Andrew Y. Ng. Self-taught Learning: Transfer Learning from Unlabeled Data. Proceedings of the 24th International conference on Machine learning. 2007. P. 767–774.

4. Lakshmi Govind, Dharmendra Kumar. Diabetic retinopathy detection using transfer learning. Journal for advanced research in applied science. 2017. vol. 4. P. 463–471.

5. Deep Learning: Transfer learning и тонкая настройка глубоких сверточных нейронных сетей // «Хабрахабр». 2016. [Электронный ресурс]. URL: https://habrahabr.ru/company/microsoft/blog/314934/ (дата обращения: 15.05.2019).

6. Torch Contributors, PyTorch Documentation // 2017. [Электронный ресурс]. URL: http://pytorch.org/docs/0.3.0/index.html (дата обращения: 15.05.2019).

7. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition // Cornell University. 2015. [Электронный ресурс]. URL: https://arxiv.org/abs/1512.03385v1 (дата обращения: 15.05.2019).

8. Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger. Densely Connected Convolutional Networks // Cornell University. 2018. [Электронный ресурс]. URL: https://arxiv.org/abs/1608.06993v5 (дата обращения: 15.05.2019).