

СТАТЬИ

УДК 004.855

**ГЛУБОКОЕ ОБУЧЕНИЕ ДЛЯ КЛАССИФИКАЦИИ ИЗОБРАЖЕНИЙ
В РАЗЛИЧНЫХ ЦВЕТОВЫХ СИСТЕМАХ****Денисенко А.А.***ЧП Денисенко, Ирпень, e-mail: alexey.denisenko.work@gmail.com*

Классификация изображений является одним из наиболее фундаментальных приложений в области компьютерного зрения и по-прежнему вызывает огромный интерес. Люди способны распознавать большое количество объектов на изображениях без особых усилий, несмотря на то, что ряд характеристик объектов может меняться. Объекты могут распознаваться даже тогда, когда их обнаружение частично затруднено. В то же время алгоритмическое описание задачи распознавания для выполнения ее компьютерной реализации по-прежнему остается актуальной проблемой. Существующие методы ее решения эффективны только для частных случаев (например, для геометрических объектов, человеческих лиц, дорожных знаков, печатных или рукописных символов) и только в определенных условиях. Модель, предназначенная для идентификации и классификации объектов, должна уметь находить их местоположение, а также выделять различные признаки объектов, такие как края, углы, цветовые различия и т.д. Глубокие модели сверточных нейронных сетей показали наибольшую производительность при работе с изображениями, которая иногда превышает возможности человеческого зрения. Тем не менее даже при этом значительном улучшении все еще приходится сталкиваться с некоторыми проблемами, связанными с переобучением и вымыванием (затуханием) градиента. Для их решения используются некоторые известные методы: аугментация данных, пакетная нормализация и дропаут; современные модели, предназначенные для классификации, не выполняют преобразование цветового пространства исходных цветных изображений, представленных в формате RGB. Изучение использования различных цветовых пространств в задаче классификации изображений относится к актуальным проблемам глубокого обучения и определяет актуальность настоящего исследования. Решение этой задачи позволит улучшить производительность используемых моделей.

Ключевые слова: цветовое пространство, сверточные нейронные сети, глубокое обучение, классификация изображений, проблема затухания градиента

**DEEP LEARNING FOR IMAGE CLASSIFICATION
IN DIFFERENT COLOR SYSTEMS****Denisenko A.A.***PE Denysenko, Irpin, e-mail: alexey.denisenko.work@gmail.com*

Image classification is one of the most fundamental applications in the field of computer vision, which continues to receive attention. People recognize a large number of objects in images with little effort, even though a number of their characteristics can change. Objects can be recognized even when their detection is partially difficult. At the same time, the algorithmic description of the recognition problem for performing its computer implementation is still an urgent problem. The existing methods for solving it are effective only for special cases (for example, for geometric objects, human faces, road signs, printed or handwritten symbols) and only under certain conditions. A model designed to identify and classify objects must be able to find their location, as well as highlight various features of objects, such as edges, corners, color differences, etc. Deep models of convolutional neural networks have shown the greatest performance when working with images, which sometimes exceeds the capabilities of human vision. However, even with this significant improvement, there are still some issues with overfitting and gradient fading. To solve them, some well-known methods are used: data augmentation, batch normalization and dropout; modern models for classification do not perform color space conversion on original RGB images. The use of different color spaces in the task of image classification is one of the topical problems of deep learning and determines the relevance of this study. Solving this problem will improve the performance of the models used.

Keywords: color space, convolutional neural networks, deep learning, image classification, vanishing gradient

Методы глубокого обучения превратили приложения, которые ранее требовали экспертных знаний в области компьютерного и машинного зрения, в инженерные задачи. Глубокое обучение позволило перенести нагрузку от создателей приложений, которые разрабатывают и пишут сценарии алгоритмов, основанных на правилах, на инженеров, обучающих системы. Оно также открыло новые возможности для решений задач, которые никогда не выполнялись без человека. Таким образом, глубокое обучение дало возможность существенно облегчить работу в области машинного зрения,

расширяя границы возможностей компьютера и камеры, то есть того, что они могут точным образом изучить.

Классификация изображений является одним из наиболее фундаментальных приложений в области компьютерного зрения. Глубокие модели сверточных нейронных сетей показали наибольшую производительность при работе с изображениями, которая иногда превышает возможности человеческого зрения [1]. Тем не менее даже при этом значительном улучшении все еще приходится сталкиваться с некоторыми проблемами, связанными с пере-

обучением и вымыванием (затуханием) градиента. Для их решения используются некоторые известные методы: аугментация данных, пакетная нормализация и дропаут; современные модели, предназначенные для классификации, не выполняют преобразование цветового пространства исходных цветных изображений, представленных в формате RGB.

Возрастающая актуальность использования различных цветовых пространств для улучшения производительности определяет мотивацию эксперимента, представленного в данной работе.

Цель текущего исследования – изучить влияние цветового пространства в задаче классификации изображений. Для достижения цели набор данных CIFAR-10 будет преобразован в пять других цветовых пространств, а именно HLS, HSV, LUV, LAB, YUV, затем на каждом из них будут обучены две различные модели архитектуры глубокого обучения: ResNet20 [2] и CapsuleNet [3].

Материалы и методы исследования

В последнее время сверточные нейронные сети продемонстрировали существенное улучшение производительности в задачах классификации изображений [4]. Тем не менее в качестве входных данных они принимают наборы изображений в основном в пространстве RGB, хотя существует много других доступных цветовых пространств.

Для достижения цели исследования в работе была рассмотрена архитектура ResNet, занявшая первое место в международном соревновании по классификации 2015 г. [2]. Потребность в этой архитектуре была вызвана тем, что глубокие модели при добавлении слишком большого количества слоев могут показывать худшую точность в том случае, если сходимость сети была достигнута ранее. Таким образом, необходимо было обеспечить оптимальную глубокую сеть.

Архитектура ResNet (сокращение от residual network – рус. «остаточное обучение»). Здесь «остаточная» структура обучения подразумевает проверку соответствия добавляемых новых слоев некоторым «остаточным» слоям, смысл использования которых состоит в том, что, когда модель сети обучается, проще сбросить остаток слоев до нуля и обучить сеть нулю, чем согласовать обучение линейному преобразованию с помощью набора нелинейных слоев. То есть сеть обучается предсказывать функцию $F(x) - x$ вместо функции $F(x)$, а для компенсации этой разницы добавля-

ется так называемое «закрывающее соединение» между слоями «остаточного блока».

Применительно к задаче классификации изображений использование такой архитектуры позволяет решить проблему затухающего градиента, возникающей по причине того, что при дифференцировании по цепному правилу до глубоких слоев нейронной сети доходит небольшая величина градиента (из-за многократного домножения на небольшие величины на предыдущих слоях).

Вторая рассматриваемая архитектура – CapsNet – новая модель капсульной нейронной сети, написанная в результате экспериментов с фреймворком Keras [5], который представляет собой высокоуровневый API нейронных сетей, написанный на Python и способный работать поверх библиотеки глубокого обучения TensorFlow. Капсульная сеть обеспечивает механизм маршрутизации. Она может иметь много слоев, состоящих из капсул. Капсула – это группа нейронов, которые могут выполнять вычисления на своих входах, а затем вычислять выходные данные в форме вектора. Вычисления нейронов внутри капсулы могут представлять различные признаки изображения, такие как размер, положение, деформация, ориентация и т.д. объекта или части объекта, который присутствует на данном изображении. CapsNet использует длину выходного вектора для представления существования объекта. Длина выходного вектора капсулы не может превышать единицу благодаря применению нелинейной функции, которая оставляет направление вектора неизменным, но уменьшает его величину.

CapsNet предлагает включить механизм маршрутизации между двумя слоями капсул. Механизм маршрутизации создает капсулу в одном слое для связи с некоторыми или вообще всеми капсулами в следующем слое.

Следует отметить, что существуют и другие архитектуры, такие как VGG, GoogLeNet, и другие фреймворки, такие как Caffe, Pytorch. Текущий эксперимент сосредоточен на процессорных вычислениях.

Результаты исследования и их обсуждение

Для проведения эксперимента был использован набор данных CIFAR-10 [6], который состоит из 60000 цветных изображений 32x32 десяти классов (самолет, автомобиль, птица, кошка, олень, собака, лягушка, лошадь, корабль и грузовик), по 6000 изображений в каждом классе. Всего имеется 50000 изображений обучающей выборки и 10000 изображений тестовой выборки.

Набор данных был преобразован в различные цветовые пространства так, как показано на рис. 1. Результат преобразования представлен в нижней части изображения.

Используемая в работе модель CIFAR10_ResNet состоит из двадцати слоев – и это

модификация модели ResNet20, сохраняющая ту же структуру и свойства. Вид этой модели и ее слоев приведен на рис. 2. Каждый слой ResNet состоит из блоков «обычных» слоев, при этом каждый блок имеет два или три уровня глубины.

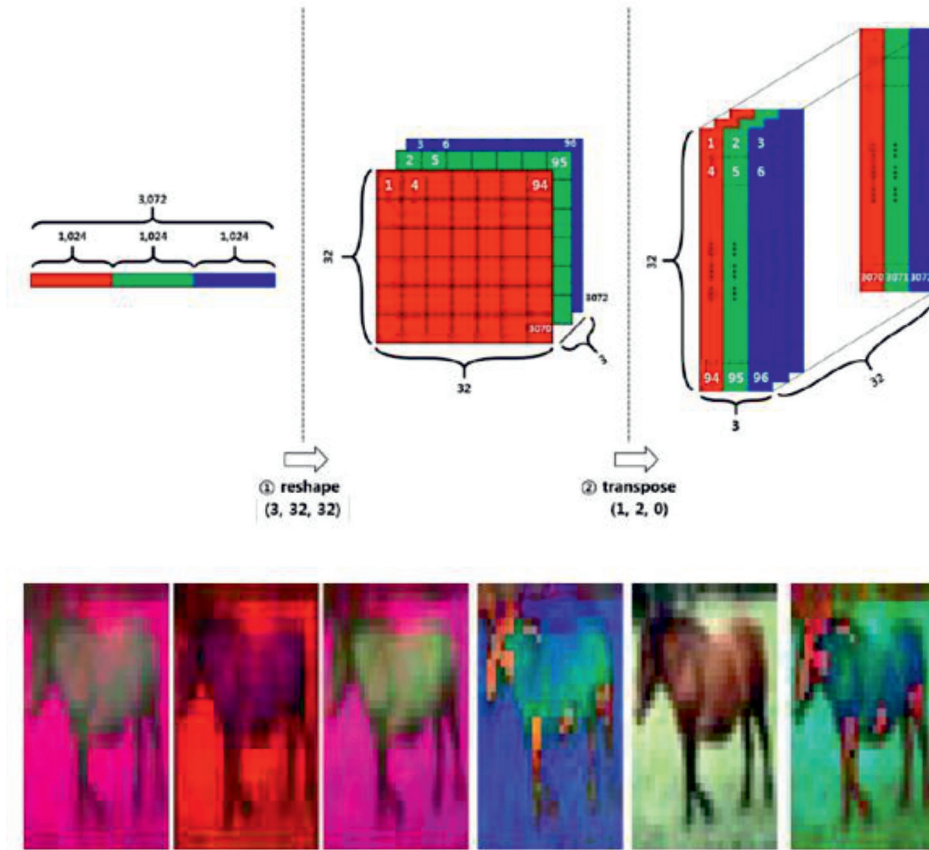


Рис. 1. Преобразование изображения в различные цветовые пространства

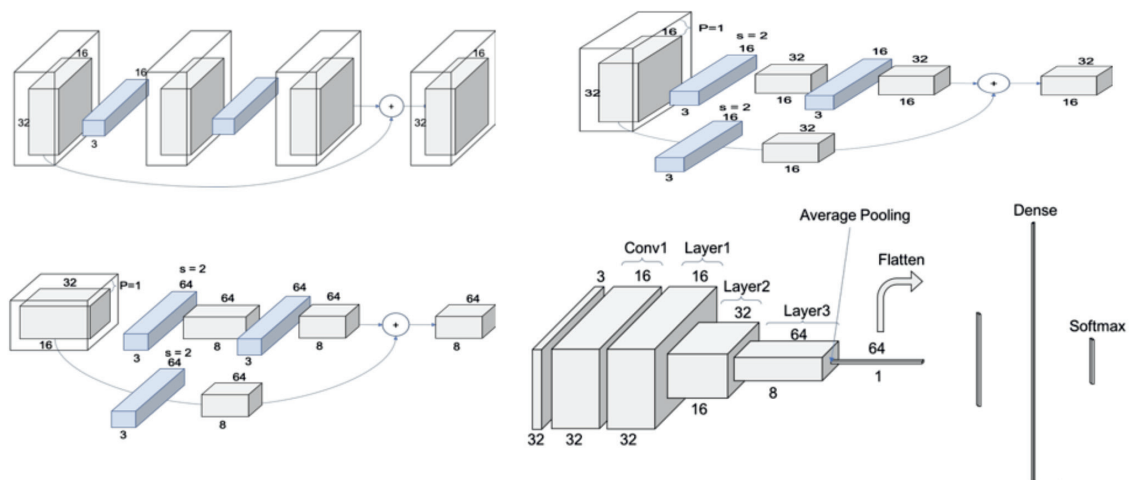


Рис. 2. Модификация ResNet

Структура сети ResNet для набора данных CIFAR-10 состоит из трех ResNet-слоев – здесь слои свертки и пулинга объединены в блоки, причем каждый блок обнаруживает схожее поведение. Слои выполняют свертку 3x3 с фиксированной размерностью карты признаков, минуя входные данные через каждые две свертки. Кроме того, параметры ширины и высоты фильтра свертки, а также каждой из карт признаков остаются постоянными в течение всего слоя. Уменьшение объема данных достигается за счет использования свертки.

В результате проведения экспериментов были получены результаты, показанные в табл. 1. Результаты демонстрируют, что есть небольшое изменение в точности в ResNet20, а также в CapsuleNet, но даже небольшой процент может иметь значение. С другой стороны, цветовое пространство LUV является неплохой альтернативой – оно показывает улучшение на 0,92% по сравнению с RGB в ResNet20 и 0,39% в CapsuleNet.

Если взглянуть на результаты точности распознавания по классам моделей CNN ResNet20 и CapsuleNet на тестовом наборе различных цветовых пространств в табл. 2, 3 и 4, можно заметить, что определенное цветовое пространство лучше подходит для конкретного класса конкретной модели, но эта особенность не является постоянной, поскольку ResNet20 меняется на CapsuleNet, или новая модель формируется путем аугментации данных со значительным изменением точности при переходе от одного цветового пространства к другому. Например, для класса корабля (ship) в ResNet20 без аугментации данных самая высокая точность 91,10% получается при использовании цветового пространства LAB; для ResNet20 с аугментацией самая высокая точность 92,50% получается с использованием цветового пространства YUV; для CapsuleNet с аугментацией данных самая высокая точность 91,50% получается с использованием цветового пространства RGB.

Таблица 1

Сравнение точности распознавания на валидационной выборке для разных цветовых пространств набора CIFAR-10 с ResNet20 и CapsuleNet

Цветовая модель	ResNet20 без аугментации, %	ResNet20 с аугментацией, %	CapsuleNet с аугментацией, %
RGB	74,76	85,02	83,15
HSL	73,48	83,73	81,65
HSV	75,12	84,81	82,56
LUV	75,59	85,94	83,54
YUV	75,96	85,88	83,31
LAB	76,52	85,66	82,94

Таблица 2

Сравнение точности по каждому классу набора CIFAR-10 с ResNet20 в различных цветовых пространствах (без аугментации данных)

Имя класса	RGB, %	HSL, %	HSV, %	LUV, %	YUV, %	LAB, %
airplane	78,70	72,20	80,90	77,30	75,40	76,60
automobile	90,90	84,00	85,50	86,90	78,60	90,00
bird	65,40	63,30	49,80	82,10	74,00	70,20
cat	36,80	69,90	64,70	47,20	54,30	57,10
deer	66,20	74,80	77,00	78,20	81,30	64,60
dog	79,40	59,80	68,30	67,40	66,70	63,60
frog	83,20	74,00	86,60	73,90	83,50	84,20
horse	72,30	66,40	68,40	81,30	74,80	74,60
ship	89,00	83,60	86,20	87,90	90,50	91,10
truck	85,70	86,80	83,80	73,70	80,50	90,20

Таблица 3

Сравнение точности по каждому классу набора CIFAR-10 с ResNet20 в различных цветовых пространствах (с аугментацией данных)

Имя класса	RGB,%	HSL,%	HSV,%	LUV,%	YUV,%	LAB,%
airplane	83,10	92,70	91,60	89,40	89,40	87,50
automobile	95,00	88,90	96,20	87,20	97,80	78,50
bird	83,60	78,30	71,00	73,30	80,20	81,30
cat	65,10	78,40	67,90	75,60	77,80	76,70
deer	91,00	83,70	81,20	93,70	84,90	90,70
dog	74,00	66,80	79,20	78,50	65,40	82,90
frog	90,80	88,40	89,00	88,50	89,50	91,20
horse	87,20	89,60	93,80	88,10	94,80	79,50
ship	84,90	78,70	85,10	90,90	92,50	91,30
truck	95,50	91,80	93,10	94,20	86,50	97,00

Таблица 4

Сравнение точности по каждому классу набора CIFAR-10 с CapsuleNet в различных цветовых пространствах (с аугментацией данных)

Имя класса	RGB,%	HSL,%	HSV,%	LUV,%	YUV,%	LAB,%
airplane	86,40	84,80	83,30	81,00	81,90	86,40
automobile	90,42	92,50	90,30	92,50	94,20	89,50
bird	96,90	70,90	74,30	77,00	78,00	70,80
cat	67,40	61,70	68,50	71,20	64,40	68,00
deer	79,40	80,00	80,30	84,00	81,30	82,20
dog	70,20	81,90	71,90	75,10	77,10	71,50
frog	93,10	88,20	89,70	90,40	90,90	90,20
horse	84,90	81,80	87,00	86,00	86,30	87,90
ship	91,50	84,70	90,40	88,00	91,20	89,00
truck	87,50	90,00	89,90	90,20	87,80	93,90

При анализе точности классификации изображений набора данных CIFAR-10 для различных цветовых пространств, обученных с помощью ResNet без аугментации данных, можно заметить, что все модели имеют одинаковое поведение со склонностью к переобучению. В то же время при анализе зависимости величины ошибки от количества эпох обучения модели оказывается, что начиная с десятой эпохи потери все еще падают, но потери на обучающей выборке растут, что является явным признаком переобучения.

Для ResNet20 с аугментацией данных модели менее переобучены, при этом точность на тестовой выборке резко возрастает.

С другой стороны, можно отметить, что CapsuleNet с аугментацией хорошо справляется с задачей и точность на тестовой выборке плавно растет.

Заключение

Проведенный эксперимент показал, что перспективным направлением является использование моделей, которые обучались на наборе данных в RGB, но вместе с тем также способны принимать на вход и классифицировать аналогичное изображение в различных цветовых пространствах с соответствующей структурой данных, а затем переучивать самих себя для того, чтобы получить лучшие результаты распознавания.

Будущие исследования могут заключаться в том, что вместо того, чтобы преобразовывать изображения из цветового пространства sRGB в другое цветовое пространство, можно подавать на вход модели изображения в желаемом цветовом пространстве и использовать соответствующий

ющую структуру для хранения данных, чтобы предотвратить потерю информации в процессе преобразования из одного цветового пространства в другое. Это позволит существенно улучшить производительность используемой модели глубокого обучения, но при этом придется также столкнуться с проблемой доступности размеченного набора данных для проведения эксперимента, поскольку большая часть наборов данных с изображениями находится в RGB-пространстве.

Список литературы

1. Андреас М. Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными. М.: Альфа-книга, 2017. 487 с.
2. Trains a ResNet on the CIFAR-10 dataset. [Электронный ресурс]. URL: https://keras.io/examples/cifar10_resnet (дата обращения: 01.08.2020).
3. Train a simple CNN-Capsule Network on the CIFAR10. [Электронный ресурс]. URL: https://keras.io/examples/cifar10_cnn_capsule (дата обращения: 01.08.2020).
4. Шолле Ф. Глубокое обучение на Python. СПб.: Питер, 2018. 400 с.
5. Keras: The Python Deep Learning library. [Электронный ресурс]. URL: <https://keras.io> (дата обращения: 01.08.2020).
6. Фомин Я.А. Распознавание образов: теория и применения. М.: ФАЗИС, 2012. 429 с.