

Классификация изображений оперения для определения видовой принадлежности птиц

А.В. Белько¹, К.С. Добратулин^{1,2}, А.В. Кузнецов^{1,3}

¹ Самарский национальный исследовательский университет имени академика С.П. Королёва, 443086, Россия, г. Самара, Московское шоссе, д. 34;

² Национальный исследовательский технологический университет «МИСиС», 119049, Россия, г. Москва, Ленинский проспект, д. 4

³ ИСОИ РАН – филиал ФНИЦ «Кристаллография и фотоника» РАН, 443001, Россия, г. Самара, ул. Молодогвардейская, д. 151

Аннотация

В работе исследуется возможность применения нейронных сетей для классификации изображений оперения с целью определения видовой принадлежности птиц. Таксономическая идентификация птиц по перу широко применяется в авиационной орнитологии для анализа столкновений с летательными аппаратами и разработки методов их предотвращения. В данной статье производится обучение на основе набора данных с фотографиями оперения птиц. Проводится сравнение классификаторов, обученных на четырех выборках из исходного набора данных. Предлагается метод идентификации птиц по изображениям с реальными данными на основе нейронных сетей YoloV4 и моделей группы DenseNet. Проведенная экспериментальная оценка показала, что предложенный метод позволяет определить видовую принадлежность птицы по фотографии отдельного пера с точностью до 81,03 % для точной классификации и с точностью 97,09 % для первых пяти предсказаний классификатора.

Ключевые слова: машинное зрение, распознавание образов, сверточные нейронные сети, авиационная орнитология.

Цитирование: Белько, А.В. Классификация изображений оперения для определения видовой принадлежности птиц / А.В. Белько, К.С. Добратулин, А.В. Кузнецов // Компьютерная оптика. – 2021. – Т. 45, № 5. – С. 749-755. – DOI: 10.18287/2412-6179-CO-836.

Citation: Belko AV, Dobratulin KS, Kuznetsov AV. Classification of plumage images for identifying bird species. Computer Optics 2021; 45(5): 749-755. DOI: 10.18287/2412-6179-CO-836.

Введение

Классификация видов птиц по фотографии пера применяется в авиационной орнитологии для определения вида птицы при столкновении с летательным аппаратом и разработки мер предотвращения будущих столкновений [1]. Для решения этой задачи могут быть применены методы ДНК-анализа и томографии [2], однако во многих аэропортах до сих пор используется исключительно визуальная классификация. Несмотря на значительное количество неструктурированных наборов изображений оперения птиц, ранее данная задача не решалась с помощью методов машинного обучения. В разное время были созданы мультимедийные определители по перьевым останкам птиц, а также коллекции, систематизированные по цветам или размерам. Недостатком таких определителей является то, что использование подобных материалов требует от пользователя специальных знаний и поиск в большом количестве вариантов, что затруднительно использовать в полевых условиях.

Классификация изображений оперения птиц осложнена большой вариативностью между маховыми, рулевыми и пуховыми перьями одного вида птиц, половым диморфизмом особей, различиями в окрасе молодых и взрослых особей, различиями в окрасе,

вызванными селекцией, а также территориальными различиями. Оперение также может иметь различные вариации, вызванные генетическими отклонениями или недостатком витаминов.

Обучение алгоритмов классификации производилось на ранее собранных данных об оперении птиц [3]. Полученный набор является самым большим из наборов изображений оперения птиц в открытом доступе.

В первом параграфе данной статьи приводится обзор существующих работ по теме классификации видов птиц. Второй параграф посвящён описанию метода формирования набора данных для исследований. В следующих параграфах приводится описание предлагаемого алгоритма классификации и отдельных его этапов. Далее приводятся результаты экспериментальных исследований – точности классификации. Последний параграф работы посвящён выводам.

1. Обзор существующих работ

Классификация видов птиц является стандартной задачей машинного обучения [4, 5]. Эта задача относится к классу Fine-Grained Visual Classification (FGVC), которая является подзадачей машинного зрения. FGVC заключается в том, чтобы классифицировать объекты единого глобального класса по подклассам. Подобными работами можно считать решение задач классификации объектов других доменов:

листьев растений [6], моделей автомобилей [7], видов овощей [8] и т.д.

Задачи этого класса представляют трудность большую, чем задачи стандартной классификации изображений, так как отличаются высокой степенью схожести объектов разного класса. В то же время может встречаться внутриклассовая изменчивость из-за различий в позе, углах съемки и освещении.

Для решения задачи FGVC могут применяться те же методы, что и для задачи стандартной классификации изображений, такие как сверточные нейронные сети. Существует несколько распространенных методов для улучшения точности классификации этого типа задач.

Одним из методов улучшения точности классификации является использование выделения важных для классификации частей изображений, на основе которых производится принятие решения о классе изображения [9]. Метод используется для придания равного веса деталям изображения, которые могут иметь разный размер, но одинаково значимы для классификации [10]. Также это лучше адаптирует алгоритм к деформациям объектов, так как изменение отдельных деталей объекта существенно меньше, чем изменение объекта целиком, особенно при классификации живых объектов, принимающих разные позы. Например, при классификации птиц производится разметка таких частей, как голова птицы, хвост и кончики крыльев [5]. Данный подход может быть применен при обучении с учителем или без учителя.

Другим частым подходом для улучшения точности классификации является сегментация переднего плана, так как зачастую задний фон у изображений разных классов может быть похожим и удаление заднего фона приводит к значительному улучшению точности классификации [10]. Сегментация переднего плана может выполняться при помощи различных алгоритмов, таких как Graph-cut [11] и DPM [12].

В случае если классы в наборе данных могут быть представимы в виде ориентированного дерева, то применима иерархическая классификация. Данный метод находит применение при классификации данных в сферах ботаники и зоологии, так как объекты имеют строгую таксономическую иерархию.

При классификации данных с иерархической структурой возможна классификация не на концевых узлах графа. Это может быть применено при недостаточном уровне уверенности для более точного определения класса.

Иерархическая классификация может реализовываться при помощи локальных классификаторов. При применении локальных классификаторов на уровне узла для каждого неконцевого узла обучается отдельный классификатор, который используется для классификации только дочерних классов этого узла. Плюсом такого метода является высокая точность классификации, а также простое добавление новых клас-

сов в структуру, так как для этого требуется изменение отдельных, сравнительно небольших локальных классификаторов, а не всей структуры в целом. Основным недостатком этого метода является необходимость обучения значительного числа локальных классификаторов. Альтернативным методом является применение локальных классификаторов для каждого из уровней дерева. Локальные классификаторы такого типа классифицируют все узлы дерева, имеющие одинаковый уровень. Подход уменьшает количество классификаторов, но может привести к нарушению иерархической структуры из-за несоответствия классификации на разных уровнях [13].

2. Описание метода формирования набора данных

Так как в открытом доступе не было сформированного набора данных с достаточным для обучения алгоритма машинного обучения количеством изображений, возникла необходимость сбора данных и создания нового набора данных. Для начального сбора данных были выбраны источники с фотографиями перьев. В качестве источников выступали специализированные сайты с профессиональными коллекциями фотографий, а также любительские фотографии с интернет-площадок для продажи и покупки перьев среди коллекционеров. Несмотря на то, что любительские фотографии имели качество ниже, чем профессиональные фотографии, а также зачастую на изображении присутствовали посторонние предметы, использование любительских фотографий позволило добавить большее разнообразие в данные и сделать алгоритм адаптированным к фотографиям с различным освещением и фоном. Использование более зашумленных и разнообразных фотографий может повысить точность алгоритмов FGVC [14].

После выбора источников у авторов были запрошены разрешения на использование фотографий. Несколько источников имели на сайте упоминание открытой лицензии: GNU FDL License, Creative Commons BY 4.0, а также Copyleft, позволяющие использовать фотографии для исследовательских целей. От авторов фотографий без упоминания лицензии получены письменные разрешения на использование фотографий.

После получения разрешений производился автоматизированный сбор данных при помощи технологии веб-скрапинга [15]. Разработан алгоритм, производящий обход по всем необходимым страницам сайтов и сохраняющий изображения. Скорость работы алгоритма искусственно занижена для того, чтобы производить меньшую нагрузку на сайты. Поиск изображений производится с помощью парсинга HTML-кода страницы, для ряда сайтов изображения перьев имели особый формат записи названия файла, что позволяло очистить данные от лишних изображений из интерфейса сайта.

В ходе исследования был сформирован набор данных из 1565 изображений, каждое из которых в среднем содержит 18 перьев. Каждая фотография содержит перья только одного вида птиц. Изображения были рассортированы по папкам с четырехуровневой таксономической организацией по отряду, семейству, роду и виду птиц. Набор данных содержит фотографии оперения 595 видов птиц.

После сбора данные были размечены при помощи ограничивающих прямоугольников (bounding boxes) в программе Microsoft VOTT. Каждый ограничивающий прямоугольник содержит изображение единственного пера, иногда с фрагментами соседних перьев при наложении на фотографии. Часть перьев не были отмечены, если они не представляют ценность для задачи классификации. Такими перьями являются пуховые и покровные перья, так как они выглядят практически идентично у разных видов птиц, а также перья, которые на фотографии расположены внахлест и практически полностью перекрыты соседними перьями.

В результате разметки получены 28272 метки расположения пера на фотографии. После аннотации изображений при помощи ограничивающих прямоугольников координаты углов прямоугольников разметки были экспортированы в формате CSV. С помощью программы на языке Python из оригинальных изображений по заданным координатам вырезаны и сохранены фрагменты изображений, содержащие отдельные перья. Таким образом, помимо 1565 исходных фотографий, был создан набор данных из 28272 изображений 595 видов птиц, содержащих одно перо на фотографии, для обучения модели классификации. Изображения расположены в иерархической структуре, содержащей данные о латинском наименовании отряда, семейства и биологического вида птицы. Формат наименования файла: Семейство/Отряд/Вид/Отряд_Вид_XXXXX.jpg. Данные были опубликованы в открытом доступе в репозитории Github.

3. Описание метода решения задачи

В качестве решения для данной задачи выбран алгоритм двухэтапной последовательной классификации [16]. Общая схема процесса двухэтапной классификации приведена на рис. 1.

Суть метода заключается в том, чтобы до процесса распознавания провести предварительный анализ данных с целью выделения из них фрагментов, которые потенциально содержат интересующую информацию — Regions of Interest (ROI). В данном случае в качестве ROI были выделены объекты единственного класса «перо». Объем значимых данных по сравнению с первоначальным объемом данных оказывается небольшим. Затем по выделенным данным производится полный анализ с целью повышения качества классификации. Вычислительная сложность производимого анализа компенсируется малым числом анализируемых фрагментов.

Несмотря на то, что существуют алгоритмы, позволяющие выполнить обнаружение объекта и его классификацию одновременно, решено разделить этот процесс на отдельные этапы. Такой подход потенциально увеличивает точность обнаружения объектов, так как вместо малого количества объектов для каждого класса нейронная сеть обучена на большом количестве объектов одного класса, без необходимости обучения избыточным признакам различий разных классов перьев.

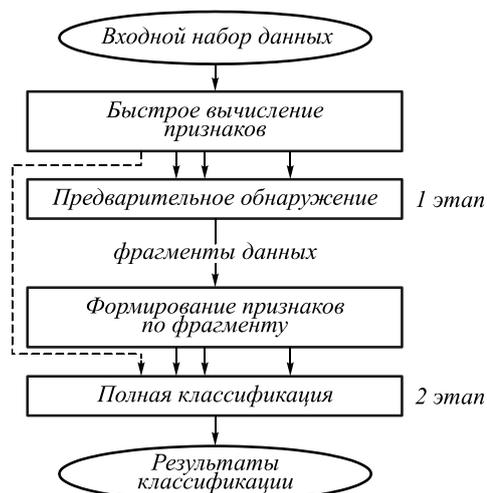


Рис. 1. Схема последовательной двухэтапной классификации

На этапе обнаружения все изображения, вне зависимости от класса, образуют обучающий набор для нейронной сети, выполняющей обнаружение объекта на изображении. После обнаружения объектов на фотографии происходит разделение изображения на отдельные изображения с единственным объектом на каждом из них и приведение всех изображений к единому размеру 40 на 240 пикселей. Это позволяет получить достаточно данных для дальнейшего обучения классификатора и придать больший вес не расположению и количеству перьев, а форме и цвету каждого отдельного пера. Далее применяется алгоритм классификации, который по обучающей выборке определяет набор признаков для осуществления классификации по ним. Схема процесса обработки изображения перед классификацией приведена на рис. 2.

Таким образом, было решено обучить две модели: модель обнаружения объектов класса «перо» и модель классификации по виду птицы. В итоговом виде на вход модели обнаружения объектов подаются изображения оперения, содержащие по одному или несколько перьев на каждом изображении, на выходе у модели обнаружения объектов — нормализованные изображения, каждое из которых содержит по одному перу. Полученные изображения находятся на входе у модели классификации, которая на их основе выдает предсказание вида птицы, которой эти перья принадлежат.

4. Процедура выделения регионов интереса

На первом этапе классификации было применено выделение значимых деталей изображения. В данном

случае требовалось выделение отдельных перьев на изображении в прямоугольники, что позволило снизить значимость фона изображения для классификации.

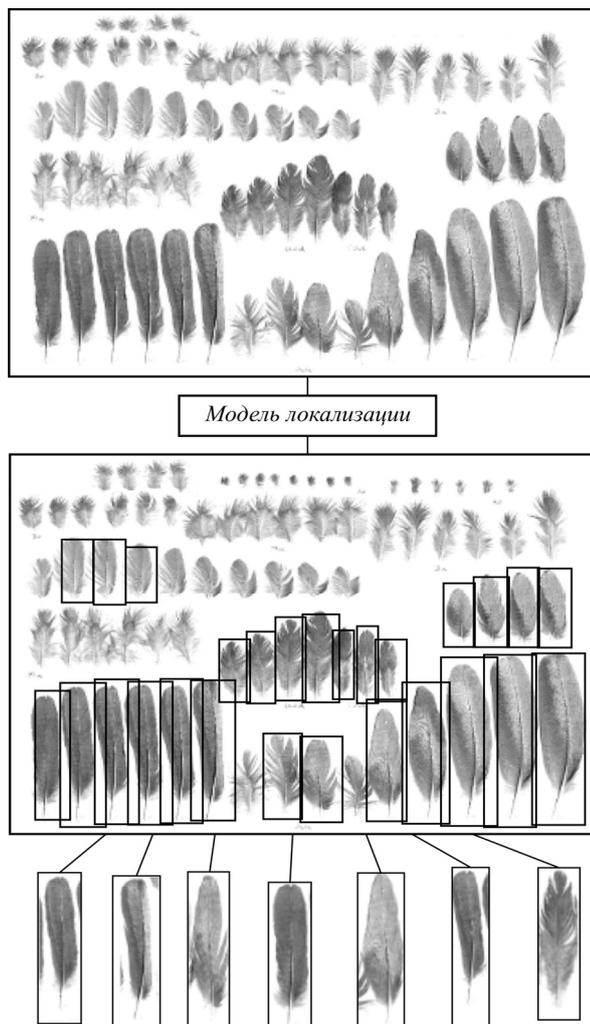


Рис. 2. Схема выделения значимых частей изображения перед классификацией

В качестве одного из наиболее современных методов обнаружения объектов на изображении была выбрана модель YOLOv4 [17] – архитектура семейства YOLO [18], демонстрирующая один из наиболее точных результатов для задачи обнаружения объектов на момент написания данной работы. Главная особенность этого семейства архитектур по сравнению с другими состоит в том, что большинство систем применяют свёрточную нейронную сеть (СНС) несколько раз к разным регионам изображения, в то время как в СНС YOLO применяется один раз ко всему изображению сразу, что отражено в названии – You Only Look Once. Сеть делит изображение на сетку, предсказывает bounding boxes и вероятности того, что в них есть искомым объект для каждого участка. Модель YOLOv4 оптимизирована для обучения на одном процессоре GPU, а также обеспечивает высокую скорость работы алгоритма, которую можно использовать для работы с видеопотоком.

Выбранная модель детектора была обучена с использованием платформы Google Colaboratory. Для обучения использована техника Transfer Learning [19], в качестве начальных весов модели использованы веса, полученные при обучении модели на наборе данных MS COCO [20]. Эта техника позволяет сократить время обучения модели. Примерное время обучения модели YOLOv4 на наборе данных оперения составила около 40 часов для 6000 эпох.

5. Процедура балансировки классов

Основной проблемой для дальнейшей классификации изображений является сильная несбалансированность классов (class imbalance). Количество изображений на класс может варьироваться между 2 и 620 изображениями, что может приводить к переобучению модели на более распространенных классах.

Были применены несколько методов для достижения баланса между классами. Были созданы две подвыборки основного набора данных, содержащие 50 и 100 классов с наибольшим количеством изображений, это позволило убрать классы, которые имели недостаточно данных для проведения точной классификации.

В качестве более сложного метода балансировки классов было решено воспользоваться иерархической структурой данных и объединить виды с недостаточным для обучения количеством изображений. Виды были объединены исключительно с другими видами из того же рода, таким образом проведена классификация до уровня рода птиц (макроклассы). Такие классы имеют в названии после названия рода сокращение *sp.* (от латинского *species*) – общепринятое в биологии обозначение того, что таксон определен с точностью до рода [21].

Классы, не имеющие достаточного количества данных после объединения, были исключены из выборки. В качестве минимального количества данных было выбрано 50 изображений. Максимальное количество данных было так же ограничено 300 изображениями, избыточные изображения были исключены. Таким образом, была создана еще одна подвыборка данных, обозначаемая далее как нормализованная (normalized).

Этот метод позволил добиться большего баланса классов, но при этом сохранить большее количество классов и изображений. График распределения данных по классам по всем четырем выборкам представлен на рис. 3а-г. Недостатком этого метода является то, что часть данных определяется с точностью до рода, не предоставляя информации о таксономическом виде.

6. Описание метода классификации изображений

Для решения задачи классификации было решено использовать СНС. В качестве моделей классификации изображений были выбраны три модели группы DenseNet: DenseNet121, DenseNet169 и DenseNet201 [22]. Архитектуры DenseNet показывают хорошие результаты для задач fine-grained классификации [23].

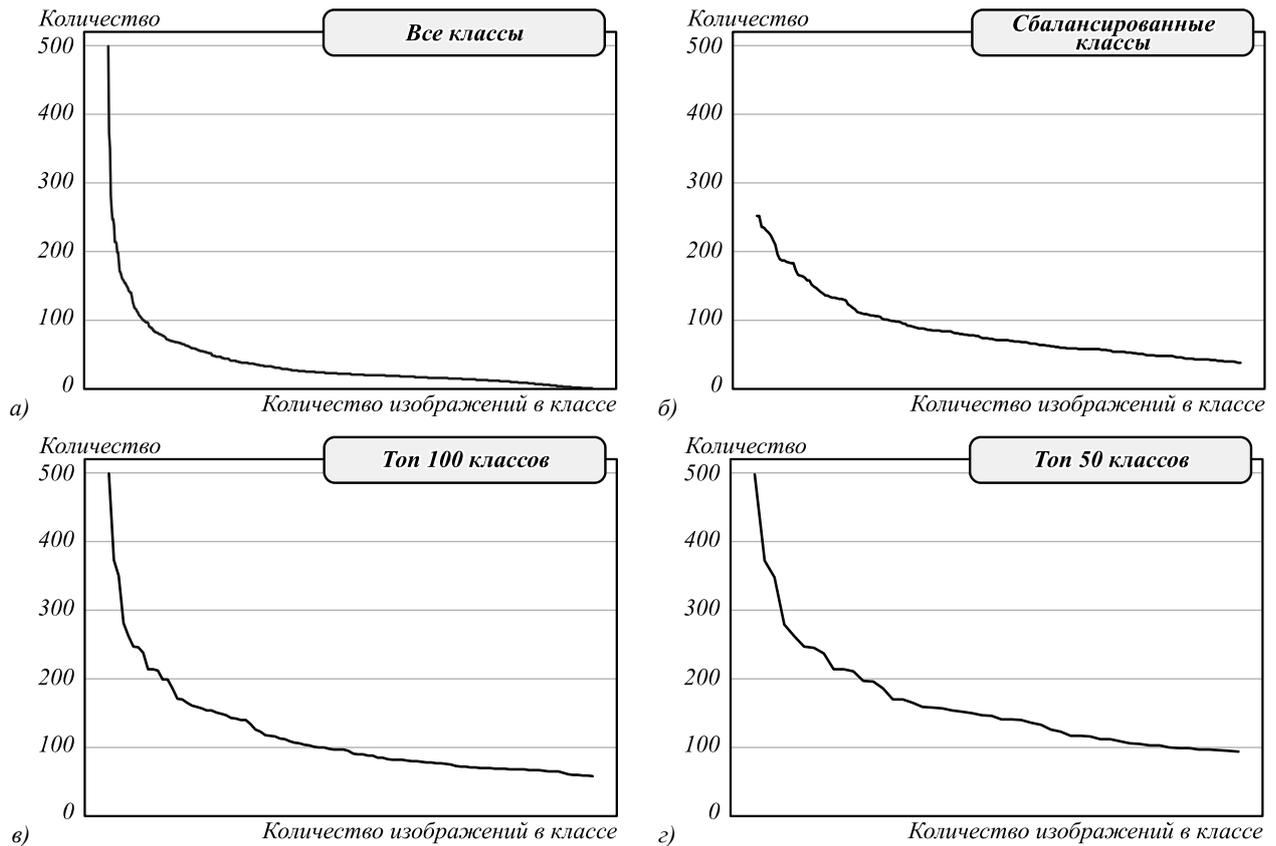


Рис. 3. Распределение данных по классам

Для обучения нейронной сети каждая из выборок изображений была разбита на обучающие и валидационные наборы данных в соотношении 80 на 20. Количество изображений в наборах данных и количество классов приведены в табл. 1.

Табл. 1. Распределение данных по классам и количество классов

Набор данных	Количество классов	Количество изображений
All	595	28272
Top-100	100	14941
Top-50	50	10584
Normalized	213	27582

Обучение производилось на базе платформы Google Colaboratory при помощи библиотеки Tensorflow. Для сравнения были обучены 3 модели на 4 наборах данных.

7. Анализ точности классификации разработанного алгоритма

Для определения точности алгоритма классификации моделей DenseNet была использована метрика mAP – усредненное значение метрики AP между всеми классами. Так как в данной задаче один класс, то метрика AP вычисляется для одного класса и не усредняется. Метрика AP рассчитывается по формуле:

$$AP = \frac{1}{11} \sum_{r \in 0.0, 0.1 \dots 1} \max_{\tilde{r} \geq r} p(\tilde{r}), \quad (1)$$

$$p = TP / (TP + FP),$$

$$r = TP / (TP).$$

Обозначения TP , FP и FN означают положительные (true positive), ложноположительные (false positive) и ложноотрицательные (false negative) предсказания соответственно.

При измерении точности полученных моделей максимальная точность по метрике AP составила 0,8736, 4620 меток были распознаны положительно, 490 меток распознаны ложноположительно и 986 – ложноотрицательно. Метки рассчитывались как верные при Intersection over Union (IoU) > 0,5. IoU определяется как площадь пересечения истинной метки с предсказанной, разделенная на площадь объединения данных меток, что приведено на формуле:

$$IoU = \frac{true \cap predicted}{true \cup predicted}. \quad (2)$$

Среднее значение IoU для модели составило 0,7522. Также для анализа точности моделей DenseNet была использована метрика $F1$ [24]:

$$F1 = 2 \frac{pr}{p+r}. \quad (3)$$

Точность по метрике $F1$ составила 0,86.

При анализе результатов была обнаружена зависимость между точностью классификации и распо-

ложением оперения на фотографии. На изображениях оперения, расположенного в форме «крыла» с имитацией натурального расположения оперения в раскрытом крыле, как правило, качество обнаружения объектов оказывается хуже из-за невозможности выделения единственного пера в ограничивающий прямоугольник без выделения соседних перьев. Для улучшения качества выделения рекомендуется располагать перья на фотографии без наложения.

В ходе экспериментов были получены 9 результатов для трёх моделей, обученных на четырех наборах данных. Для сравнения точности полученных алгоритмов были применены метрики Sparse Top-1 Categorical Accuracy и Sparse Top-5 Categorical Accuracy. Данные метрики вычисляют точность классификации модели. Для метрики Sparse Top-1 Categorical Accuracy предсказание считается верным при совпадении класса с наибольшим уровнем уверенности предсказания с точным классом. Для метрики Sparse Top-5 Categorical Accuracy предсказание считается верным при попадании точного класса в топ-5 классов, распознанных моделью. Результаты приведены в табл. 2.

Табл. 2. Результаты классификации

Набор данных	Архитектура модели	Sparse Top-1 Categorical Accuracy	Sparse Top-5 Categorical Accuracy
All	DenseNet121	0,7642	0,9482
	DenseNet169	0,7181	0,9360
	DenseNet201	0,7978	0,9586
Top-100	DenseNet121	0,7989	0,9709
	DenseNet169	0,7979	0,9695
	DenseNet201	0,7266	0,9491
Top-50	DenseNet121	0,6394	0,8871
	DenseNet169	0,6684	0,9186
	DenseNet201	0,5700	0,8740
Normalized	DenseNet121	0,7888	0,9603
	DenseNet169	0,8103	0,9608
	DenseNet201	0,7911	0,9616

Модели, обученные на полном и нормализованном наборах данных, показывают в среднем результаты лучше, несмотря на несбалансированные данные. Таким образом, добавление в набор данных редких видов птиц с малым количеством фотографий улучшило результаты набора данных [12].

Заключение

Несмотря на то, что у модели DenseNet121 на наборе данных Top-100 результаты классификации по метрике Sparse Top-5 Categorical Accuracy наиболее точные, версия модели DenseNet169, обученная на нормализованном наборе данных, показывает лучший результат по метрике Sparse Top-1 Categorical Accuracy и имеет незначительное отставание по метрике Sparse Top-5 Categorical Accuracy, из-за чего DenseNet169 можно считать наиболее полезной для

применения. Полученная модель показывает хороший результат классификации для большинства классов с учетом сложности классификации и несбалансированности данных. Благодаря иерархической структуре данных сохранено максимальное количество изображений и улучшен баланс классов, что привело к более точной классификации.

Литература

1. **Soldatini, C.** An ecological approach to birdstrike risk analysis / C. Soldatini, V. Georgalas, P. Torricelli, Y.V. Albores-Barajas // *European Journal of Wildlife Research*. – 2010. – Vol. 56, Issue 4. – P. 623-632.
2. **Yang, R.** Using DNA barcodes to identify a bird involved in a birdstrike at a Chinese airport / R. Yang, X.B. Wu, P. Yan, X.Q. Li // *Molecular Biology Reports*. – 2010. – Vol. 37, Issue 7. – P. 3517-3523.
3. **Belko, A.** Feathers dataset for fine-grained visual categorization / A. Belko, K. Dobratulin, A. Kuznetsov // *Proceedings of SPIE*. – 2020. – Vol. 11605. – 1160518. – DOI: 10.1117/12.2588386.
4. **Berg, T.** Birdsnap: Large-scale fine-grained visual categorization of birds / T. Berg, J. Liu, S.W. Lee, M.L. Alexander, D.W. Jacobs, P.N. Belhumeur // *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. – 2014. – P. 2019-2026.
5. **Fu, J.** Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition / J. Fu, H. Zheng, T. Mei // *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. – 2017. – P. 4476-4484.
6. **Araujo, V.M.** Fine-grained hierarchical classification of plant leaf images using fusion of deep models / V.M. Araujo, A.S. Britto, A.L. Brun, L.E.S. Oliveira, A.L. Koerich // *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*. – 2018. – P. 4476-4484.
7. **Yang, L.J.** A large-scale car dataset for fine-grained categorization and verification / L.J. Yang, P. Luo, C.C. Loy, X. Tang // *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* – 2015. – P. 3973-3981.
8. **Hou, S.H.** VegFru: A domain-specific dataset for fine-grained visual categorization / S.H. Hou, Y.S. Feng, Z.L. Wang // *2017 IEEE International Conference on Computer Vision (ICCV)*. – 2017. – P. 541-549.
9. **Dai, X.Y.** Efficient fine-grained classification and part localization using one compact network / X.Y. Dai, B. Southall, N. Trinh, B. Matei // *2017 IEEE International Conference on Computer Vision Workshops (ICCVW 2017)*. – 2017. – P. 996-1004.
10. **Zhao, B.** A survey on deep learning-based fine-grained object classification and semantic segmentation / B. Zhao, J.S. Feng, X. Wu, S.C. Yan // *International Journal of Automation and Computing*. – 2017. – Vol. 14, Issue 2. – P. 119-135.
11. **Priyadharshini, P.** Hyperspectral image classification using MLL and Graph cut methods / P. Priyadharshini, K. Thilagavathi // *Proceedings of 2016 Online International Conference on Green Engineering and Technologies (ICGET)*. – 2016. – P. 1-6.
12. **Pandey, M.** Scene recognition and weakly supervised object localization with deformable part-based models / M. Pandey, S. Lazebnik // *2011 IEEE International Conference on Computer Vision (ICCV)*. – 2011. – P. 1307-1314.

13. **Silla, C.N.** A survey of hierarchical classification across different application domains / C.N. Silla, A.A. Freitas // *Data Mining and Knowledge Discovery*. – 2011. – Vol. 22, Issues 1-2. – P. 31-72.
14. **Krause, J.** The unreasonable effectiveness of noisy data for fine-grained recognition / J. Krause, B. Leibe, J. Matas. – In: *Computer vision – ECCV 2016, Part III* / ed. by B. Leibe, J. Matas, N. Sebe, M. Welling. – Springer International Publishing AG, 2016. – P. 301-320.
15. **Mitchell, R.** Web scraping with Python: Collecting data from the Modern Web / R. Mitchell. – Sebastopol: O'Reilly Media, 2015. – 256 p.
16. **Глумов, Н.И.** Обнаружение и распознавание объектов на изображениях / Н.И. Глумов, В.В. Мясников, В.В. Сергеев. – Самара: Изд-во Самар. ун-та, 2010. – 141 с.
17. **Zhu, Q.F.** Study on the evaluation method of sound phase cloud maps based on an improved YOLOv4 algorithm / Q.F. Zhu, H.F. Zheng, Y.B. Wang, Y.G. Cao, S.X. Guo // *Sensors*. – 2020. – Vol. 20, Issue 15. – 4314.
18. **Redmon, J.** You only look once: Unified, real-time object detection / J. Redmon, S. Divvala, R. Girshick, A. Farhadi // *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. – 2016. – P. 779-788.
19. **Sarkar, D.** Hands-on transfer learning with Python / D. Sarkar, R. Bali, T. Ghosh. – Birmingham: Packt Publishing, 2018. – 440 p.
20. **Lin, T.-Y.** Microsoft COCO: common objects in context / T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, C.L. Zitnick. – In: *Computer vision – ECCV 2014, Part V* / ed. by D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars. – New York: Springer, 2014. – P. 740-755.
21. Международный кодекс зоологической номенклатуры, принятый XV Международным зоологическим конгрессом / пер. на рус. яз. Д.В. Обручева; Академия наук СССР, Отделение общей биологии. — М.-Л.: Наука, 1966. – 100 с.
22. **Huang, G.** Densely connected convolutional networks / G. Huang, Z. Liu, L. Van Der Maaten // *30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*. – 2017. – P. 2261-2269.
23. **Valev, K.** A systematic evaluation of recent deep learning architectures for fine-grained vehicle classification / K. Valev, A. Schumann, L. Sommer, J. Beyerer // *Proceedings of SPIE*. – 2018. – Vol. 10649. – 1064902.
24. **Everingham, M.** The PASCAL visual object classes (VOC) challenge / M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn, A. Zisserman // *International Journal of Computer Vision*. – 2010. – Vol. 88. – P. 303-338.

Сведения об авторах

Белько Алина Вадимовна, 1998 года рождения, является студентом Национального исследовательского университета имени академика С.П. Королёва по направлению магистратуры «Data Science». Область научных интересов: глубокое обучение, машинное обучение, компьютерное зрение, анализ данных. E-mail: alinabelko@gmail.com.

Добратулин Константин Сергеевич, 1998 года рождения, в 2020 году окончил Самарский национальный исследовательский университет имени академика С.П. Королёва по направлению «Прикладная математика и информатика», с 2020 года является студентом Национального исследовательского технологического университета (НИТУ) «МИСиС» по направлению «Информационные системы и технологии». Область научных интересов: глубокое обучение, машинное обучение, компьютерное зрение, анализ данных. E-mail: dobratulin@yahoo.com.

Кузнецов Андрей Владимирович, родился в 1987 году. В 2010 году окончил СГАУ с отличием по специальности «Прикладная математика и информатика». В 2010 поступил в аспирантуру СГАУ, в 2013 г. защитил диссертацию на соискание степени кандидата технических наук. В настоящее время работает старшим научным сотрудником НИЛ-97 СГАУ и научным сотрудником в ИСОИ РАН. Круг научных интересов включает обработку и анализ изображений, распознавание образов, обнаружение искажений изображений, геоинформатику. Имеет 37 публикаций, в том числе 18 научных статей и 1 монографию. E-mail: kuznetsoff.andrey@gmail.com.

ГРНТИ: 28.23.15

Поступила в редакцию 14 ноября 2020 г. Окончательный вариант – 02 июля 2021 г.

Classification of plumage images for identifying bird species

A.V.Belko¹, K.S. Dobratulin^{1,2}, A.V. Kuznetsov^{1,3}

¹ Samara National Research University, 443086, Samara, Russia, Moskovskoye Shosse 34;

² National University of Science and Technology "MISiS",
119049, Moscow, Russia, Leninsky Prospect 4;

³ IPSI RAS – Branch of the FSRC "Crystallography and Photonics" RAS,
443001, Samara, Russia, Molodogvardeyskaya 151

Abstract

This paper studies the possibility of using neural networks to classify plumage images in order to identify bird species. Taxonomic identification of bird plumage is widely used in aviation ornithology to analyze collisions with aircraft and develop methods for their prevention. This article provides a method for bird species identification based on a dataset made up in the previous research. A method for identifying birds from real-world images based on YoloV4 neural networks and DenseNet models is proposed. We present results of the feather classification task. We selected several deep learning architectures (DenseNet based) for a comparison of categorical crossentropy values on the provided dataset. The experimental evaluation has shown that the proposed method allows determining the bird species from a photo of an individual feather with an accuracy of up to 81.03% for accurate classification, and with an accuracy of 97.09% for the first five predictions.

Keywords: machine vision, pattern recognition, neural networks, aviation ornithology.

Citation: Belko AV, Dobratulin KS, Kuznetsov AV. Classification of plumage images for identifying bird species. *Computer Optics* 2021; 45(5): 749-755. DOI: 10.18287/2412-6179-CO-836.

References

- [1] Soldatini C, Georgalas V, Torricelli P, Albores-Barajas YV. An ecological approach to birdstrike risk analysis. *Eur J Wildl Res* 2010; 56(4): 623-632.
 - [2] Yang R, Wu XB, Yan P, Li XQ. Using DNA barcodes to identify a bird involved in a birdstrike at a Chinese airport. *Mol Biol Rep* 2010; 37(7): 3517-3523.
 - [3] Belko A, Dobratulin K, Kuznetsov A. Feathers dataset for fine-grained visual categorization. *Proc SPIE* 2020; 11605: 1160518. DOI: 10.1117/12.2588386.
 - [4] Berg T, Liu J, Lee SW, Alexander ML, Jacobs DW, Belhumeur PN. Birdsnap: Large-scale fine-grained visual categorization of birds. *IEEE Conf on Computer Vision and Pattern Recognition (CVPR)* 2014: 2019-2026.
 - [5] Fu J, Zheng H, Mei T. Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition. *2017 IEEE Conf on Computer Vision and Pattern Recognition (CVPR)* 2017: 4476-4484.
 - [6] Araujo VM, Britto AS, Brun AL, Oliveira LES, Koerich AL. Fine-grained hierarchical classification of plant leaf images using fusion of deep models. *IEEE 30th Int Conf on Tools with Artificial Intelligence (ICTAI)* 2018: 4476-4484.
 - [7] Yang LJ, Luo P, Loy CC, Tang X. A large-scale car dataset for fine-grained categorization and verification. *IEEE Conf on Computer Vision and Pattern Recognition (CVPR)* 2015: 3973-3981.
 - [8] Hou SH, Feng YS, Wang ZL. VegFru: A domain-specific dataset for fine-grained visual categorization. *IEEE Int Conf on Computer Vision (ICCV)* 2017: 541-549.
 - [9] Dai XY, Southall B, Trinh N, Matei B. Efficient fine-grained classification and part localization using one compact network. *IEEE Int Conf on Computer Vision Workshops (ICCVW)* 2017: 996-1004.
 - [10] Zhao B, Feng JS, Wu X, Yan SC. A survey on deep learning-based fine-grained object classification and semantic segmentation. *Int J Autom Comput* 2017; 14(2): 119-135.
 - [11] Priyadharshini P, Thilagavathi K. Hyperspectral image classification using MLL and Graph cut methods. *Proc 2016 Online Int Conf on Green Engineering and Technologies (IC-GET)* 2016: 1-6.
 - [12] Pandey M, Lazebnik S. Scene recognition and weakly supervised object localization with deformable part-based models. *IEEE Int Conf on Computer Vision (ICCV)* 2011: 1307-1314.
 - [13] Silla CN, Freitas AA. A survey of hierarchical classification across different application domains. *Data Min Knowl Discov* 2011; 22(1-2): 31-72.
 - [14] Krause J, Leibe B, Matas J. The unreasonable effectiveness of noisy data for fine-grained recognition. In Book: Leibe B, Matas J, Sebe N, Welling M, eds. *Computer vision – ECCV 2016, Pt III*. Springer International Publishing AG; 2016: 301-320.
 - [15] Mitchell R. *Web scraping with Python: Collecting data from the Modern Web*. Sebastopol: O'Reilly Media; 2015.
 - [16] Glumov NI, Myasnikov VV, Sergeev VV. Detection and recognition of objects in images [In Russian]. Samara: Samara University Publisher; 2010.
 - [17] Zhu QF, Zheng HF, Wang YB, Cao YG, Guo SX. Study on the evaluation method of sound phase cloud maps based on an improved YOLOv4 algorithm. *Sensors* 2020; 20(15): 4314.
 - [18] Redmon J, Divvala S, Girshick R, Farhadi A. You ONLY LOOK ONCE: Unified, real-time object detection. *IEEE Conf on Computer Vision and Pattern Recognition (CVPR)* 2016: 779-788.
 - [19] Sarkar D, Bali R, Ghosh T. *Hands-on transfer learning with Python*. Birmingham: Packt Publishing; 2018.
 - [20] Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollar P, Zitnick CL. Microsoft COCO: common objects in context. In Book: Fleet D, Pajdla T, Schiele B, Tuytelaars T, eds. *Computer vision – ECCV 2014, Part V*. New York: Springer; 2014: 740-755.
 - [21] *International code of zoological nomenclature*. London: I Natural History Museum; 1999.
 - [22] Huang G, Liu Z, Van Der Maaten L. Densely connected convolutional networks. *30th IEEE Conf on Computer Vision and Pattern Recognition (CVPR)* 2017: 2261-2269.
-

[23] Valev K, Schumann A, Sommer L, Beyerer J. A systematic evaluation of recent deep learning architectures for fine-grained vehicle classification. Proc SPIE 2018; 10649: 1064902.

[24] Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The PASCAL Visual Object Classes (VOC) Challenge. Int J Comput Vis 2010; 88: 303-338.

Authors' information

Alina Vadimovna Belko (b. 1998) is a student of Samara National Research University, majoring in Data Science. Research interests: deep learning, machine learning, computer vision, data science. E-mail: alinabelko@gmail.com.

Konstantin Sergeevich Dobratulin (b. 1998) graduated from Samara National Research University in 2020, majoring in Applied Mathematics and Informatics. Since 2020 is a student of National University of Science and Technology MISIS (NUST MISIS), majoring in Information Systems and Technologies. Research interests: deep learning, machine learning, computer vision, data science. E-mail: dobratulin@yahoo.com.

Andrey Vladimirovich Kuznetsov, born in 1987. Graduated from SSAU in 2010 with honors in "Applied Mathematics and Computer Science". In 2010, he entered the post-graduate school of SSAU, and in 2013, defended dissertation for the Ph.D. degree. Currently works as a senior researcher NIL-97 Samara State Aerospace University and a research fellow at the at IPSI RAS. Research interests include image processing and analysis, image recognition, image distortion detection, and geoinformatics. Has 37 publications, including 18 scientific articles and 1 monograph. E-mail: kuznetsoff.andrey@gmail.com.

Received November 14, 2020. The final version – July 2, 2021.
