

Идентификация камеры по фотографии с использованием нейронных сетей

А.Ю.Денисова

Самарский национальный исследовательский университет им. академика С.П. Королева

Самара, Россия

ИСОИ РАН – филиал ФНИЦ «Кристаллография и фотоника»,

denisova_ay@geosamara.ru

Аннотация — Идентификация устройства съемки - это одна из основных задач криминалистических экспертиз, заключающихся в проверке подлинности изображений. Задача идентификации состоит в определении модели камеры по изображению. В настоящее время широкое развитие получили нейросетевые методы решения данной задачи. Настоящая статья посвящена проверке и модификации одного из таких методов, основанного на использовании нейронной сети EfficientNetB5. Данный метод отличается значительной простотой реализации и заявленной высокой эффективностью классификации камер. Однако проведенные нами исследования на основе собственной реализации данного метода в среде TensorFlow 2.0 показали, что заявленная в оригинальной статье точность работы метода была существенно занижена. В связи с этим была предложена своя аналогичная реализация на основе нейронной сети BagNet9, существенно повышающая точность детектирования камер по сравнению с базовым методом. Экспериментальные исследования были произведены с использованием набора данных Forchheim Image Dataset.

Ключевые слова— идентификация камеры, классификация фотоизображений, EfficientNetB5, BagNet9.

1. ВВЕДЕНИЕ

В задачах криминалистических экспертиз часто необходимо решать задачу установления источника изображения. Одной из подзадач в таком случае является определение модели регистрирующего устройства, т.е. идентификация камеры.

Для решения данной задачи в последнее десятилетие было предложено много успешных методов [1]. Среди них методы, основанные на использовании глубоких нейронных сетей, являются, пожалуй, самыми успешными и перспективными. Однако зачастую существующие реализации нейронных сетей имеют весьма сложную структуру и ограниченную производительность на наборах данных с большим количеством различных моделей камер. В этой связи подход на основе переноса обучения для нейронной сети EfficientNetB5, предложенный в работе [2], показался нам наиболее привлекательным, так как обладал существенной простотой реализации и заявленной высокой эффективностью.

С целью развития подхода, предложенного в статье [2], мы провели эксперименты по замене основной сети на более эффективную с сохранением предложенной авторами в [2] технологии обучения сети. Однако повторная реализация базового алгоритма [2] показала существенно более низкое качество детектирования камеры, чем было заявлено в статье [2] для оригинального метода. Тем не менее, рассматриваемая нами нейронная сеть BagNet9, которая была отобрана для замены основной сети, показала существенно более

качественное детектирование камер, что показывает ее более высокий потенциал для решения данной задачи, чем у EfficientNetB5.

Далее описывается структура оригинального метода, суть предлагаемой модификации метода и полученные экспериментальные результаты на наборе данных Forchheim Image Dataset, предложенном и впервые описанном в [2].

2. ОРИГИНАЛЬНЫЙ МЕТОД И ЕГО МОДИФИКАЦИЯ

Оригинальный метод, предложенный Hadwiger и Riess в [2] состоит в дообучении сети EfficientNetB5 с весами, полученными для ImageNet, путем переноса обучения. Для этого авторы используют свой собственный открытый набор данных Forchheim Image Dataset, содержащий изображения города, снятые 27 смартфонами. Набор данных содержит более 23000 изображений для 143 сцен, разбитых авторами на обучающий, валидационный и тестовый наборы. Для экспериментов авторы использовали данные 25 камер, убрав из датасета две повторяющиеся камеры.

Авторы используют нейронную сеть EfficientNetB5 с минимальной модификацией, а именно без слоев классификации, заменяя их слоем глобального среднего пулинга (global average pooling) и полносвязным слоем с 25 нейронами (по числу классов). Соответственно, дообучение производится только для добавленных новых слоев сети. Обучение производится с использованием оптимизатора Adam с параметрами по умолчанию и с использованием расписания настройки скорости обучения: в течение первых двух итераций скорость обучения равна 0,001, на последующих итерациях 0,0005. Для останова процесса обучения используется процедура раннего останова по значению функции потерь на валидационной выборке. В качестве функции потерь выступает категориальная кроссэнтропия.

Наиболее интересен процесс подготовки входных данных для сети. Во-первых, каждое изображение набора данных разбивается на кластеры размера 256×256 . Для каждого кластера вычисляется эвристическая метрика качества кластера:

$$Q = \frac{1}{3} \sum_{c \in \{R, G, B\}} \alpha \beta (\mu_c - \mu_c^2) + (1 - \alpha) (1 - e^{\gamma \sigma_c}), \quad (1)$$

где μ_c и σ_c обозначают среднее и стандартное отклонение яркости для кластера, а $\alpha=0,7$; $\beta=4$; $\gamma=\ln(0,01)$. Величина Q характеризует «разнообразность» изображения в пределах кластера. Для обучения и классификации отбираются только 100 кластеров с лучшей метрикой качества. Затем из кластера выделяется фрагмент (патч) размера 64×64 ,

который подается на вход сети. Результирующая классификация производится в трёх режимах: по патчам, по кластерам и по изображению целиком. В двух последних случаях кластер и изображение относятся к конкретной камере по принципу максимального голосования по всем патчам данного кластера и изображения.

Предлагаемая модификация данного метода заключается в простой замене сети EfficientNetB5 на другую, более эффективную для данной задачи.

3. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

В ходе экспериментальных исследований был реализован оригинальный алгоритм на основе EfficientNetB5 в пакете TensorFlow 2.0 для базового сценария обучения сети с использованием простейшей аугментации данных. На этапе обучения патч из кластера извлекается случайным образом и подвергается аугментации путем случайных отражений вдоль вертикальной и горизонтальной осей и поворотов на 90°. Для валидации используется случайное извлечение патча с фиксированным начальным значением датчика случайных чисел. Для тестирования используются по 16 неперекрывающихся патчей для каждого кластера.

В первом эксперименте были протестированы все возможные доступные реализации нейронных сетей для классификации из пакетов TensorFlow Applications и TF2CV. Всего было рассмотрено 230 архитектур сетей. Для ускорения работы, эксперимент по сравнению производительности сетей производился на 3 классах при размере кластера 128×128, и рассчитывалась только вероятность верной классификации камеры по патчам. Показатели точности классификации камеры для топ-5 нейронных сетей и EfficientNetB5 показаны в таблице 1.

Таблица I. Точность детектирования камеры на 3 классах

Нейронная сеть	Кол-во эпох	Точность классификации		
		Обучение	Валидация	Тестирование
BagNet9	22	0,9791	0,9828	0,9550
BagNet33	19	0,9943	0,9852	0,9465
BagNet	13	0,9804	0,9804	0,9365
SePreResNetbc26b	23	0,9335	0,9158	0,9158
PreResNetbc14b	54	0,9655	0,9508	0,9087
EfficientNetB5	18	0,825	0,7535	0,6579

Из таблицы 1 видно, что на трех классах наилучшие результаты 0,955 дала сеть BagNet9 по сравнению EfficientNetB5, у которой на тестовой выборке точность составила всего лишь 0,6579 вместо ожидаемых 0,95.

Во втором эксперименте была детально воспроизведена схема эксперимента с простой аугментацией из [2] и выполнено сравнение точности работы нейронных сетей EfficientNetB5 и BagNet9 при различных размерах патча для 25 классов. Для оригинального эксперимента размер патча составлял 64×64. Дополнительно мы рассмотрели патч размера 224×224, чтобы избежать необходимости изменения

размера изображения перед подачей в нейронную сеть BagNet9, для которой размер входа составляет 224×224. Как и в оригинальном эксперименте, точность классификации была рассчитана по патчам, по кластерам и по всему изображению. При извлечении патча 224×224 точность по патчам совпадает с точностью по кластерам, так как на один кластер приходится только 1 патч. Результаты данного эксперимента приведены в таблице 2 и 3.

Таблица II. Точность обнаружения камеры EfficientNetB5 и BagNet9 для патча 64×64

Нейронная сеть	Точность на тестовой выборке		
	По патчам	По кластерам	По целому изображению
BagNet9	0,5885	0,7261	0,8750
EfficientNetB5	0,2949	0,4584	0,6752

Таблица III. Точность обнаружения камеры EfficientNetB5 и BagNet9 для патча 224×224

Нейронная сеть	Точность на тестовой выборке		
	По патчам	По кластерам	По целому изображению
BagNet9	0,7779	0,7779	0,9253
EfficientNetB5	0,2013	0,2013	0,5043

Из таблиц 2-3 можно заметить, что BagNet9 во всех случаях превосходит по качеству EfficientNetB5. Однако при увеличении числа классов производительность обеих сетей существенно падает, что требует их более тонкой настройки. Также видно, что увеличение размера патча привело к росту производительности для BagNet9.

4. ЗАКЛЮЧЕНИЕ

В статье рассмотрена задача идентификации камеры по изображению. Для решения задачи использован подход на основе использования свёрточных нейронных сетей и переноса обучения. В статье сравниваются производительность двух сетей BagNet9 и EfficientNetB5 для решения данной задачи. Предлагаемая сеть BagNet9 показала более высокую точность по сравнению с EfficientNetB5, используемой в оригинальном методе. Однако достичь заявленной точности оригинального метода классификации так и не удалось. Тем не менее, произведенное сравнение показывает, что потенциал сети BagNet9 для решения данной задачи выше, чем для EfficientNetB5.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке гранта РФФИ № 19-29-09045.

ЛИТЕРАТУРА

- [1] Bernacki, J. A survey on digital camera identification methods / J. Bernacki // Forensic Science International: Digital Investigation. – 2020. – Vol. 34. – P. 300983.
- [2] Hadwiger, B. The Forchheim image database for camera identification in the wild/ B. Hadwiger, C. Riess // International Conference on Pattern Recognition. – 2021. – P. 500-515.