

Шамало В.В.

магистрант кафедры информатики и вычислительной техники

Кубанский государственный технологический университет

Россия, г. Краснодар

РАЗРАБОТКА ДАТАСЕТА И МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

ДЛЯ ДЕТЕКЦИИ ЛИЦ НА ФОТОГРАФИЯХ

Аннотация: качество детекторов лиц ограничивается не столько архитектурой модели, сколько качеством обучающих данных. К примеру, значительная часть общедоступных снимков из интернета имеет низкое разрешение, а метод интерполяции не восполняет утраченные детали. Цель работы — собрать датасет не менее чем из 20 000 размеченных изображений и обучить на нём детекторы семейства YOLO. Предложен конвейер подготовки данных: ручная разметка, апскейл сетью Real-ESRGAN вместо интерполяции и аугментация с моделированием шумов съёмки. Модели достигли точности 0,984 и 0,977 на подмножестве EASY набора WIDER FACE и были внедрены в действующий веб-сервис обезличивания фотографий.

Ключевые слова: компьютерное зрение, детекция лиц, глубокое обучение, подготовка данных, аугментация, суперразрешение.

Shamalo V.V.

Master's Student

Kuban State Technological University

Russia, Krasnodar

DEVELOPMENT OF A DATASET AND MACHINE LEARNING METHODS FOR FACE DETECTION IN PHOTOGRAPHS

Abstract: the quality of face detectors is limited not so much by model architecture as by the quality of training data: a large share of publicly available

images has low resolution, and interpolation does not restore lost detail. The aim is to compile a dataset of at least 20,000 annotated images and train YOLO-family detectors on it. A pipeline is proposed that combines manual annotation, upscaling with Real-ESRGAN instead of interpolation, and augmentation modelling shooting noise. The models reached an accuracy of 0.984 and 0.977 on the EASY subset of WIDER FACE and were deployed in a working anonymization service.

Keywords: computer vision, face detection, deep learning, data preparation, augmentation, super-resolution.

Введение

Детекция лиц лежит в основе систем видеонаблюдения, сервисов обработки фотографий и инструментов обезличивания персональных данных [1]. Качество детекторов ограничивается вариативностью условий съёмки: тени и контровый свет искажают черты лица, а мелкие и перекрытые лица на групповых снимках остаются наиболее сложным случаем.

Отдельная проблема — исходное разрешение. Значительная часть общедоступных изображений имеет размер 416×416 и подобный, а увеличение таких кадров билинейной или бикубической интерполяцией не добавляет информации, потому что кадр лишь становится крупнее и мягче. Цель работы — сформировать датасет не менее чем из 20 000 размеченных изображений лиц, обучить на нём детекторы семейства YOLO и встроить их в действующий веб-сервис обезличивания фотографий [2].

Методы и исследования

Данный корпус собран из фрагментов открытых наборов WIDER FACE и Fddb и изображений из открытых источников, которые были размечены вручную в среде Labellmg [3]; итоговый объём — 25 000 кадров. Ручная разметка выбрана осознанно: человек корректно

очерчивает границы перекрытого или сильно повёрнутого лица, где автоматическая предразметка ошибается и вносит в корпус шум.

Аннотации приведены к формату YOLO: каждая строка описывает один объект пятью числами — класс и нормированные по размеру кадра координаты центра рамки, её ширина и высота. Благодаря нормировке аннотации масштабируются автоматически при любом изменении размера изображения, что и позволяет свободно менять разрешение кадров.

Ключевое отличие предлагаемого конвейера от типовой практики — отказ от интерполяции в пользу генеративно-состязательной сети Real-ESRGAN [4]. Она восстанавливает правдоподобные высокочастотные детали, отсутствующие в исходном кадре, а не растягивает имеющиеся пиксели. Это позволило из одного размеченного корпуса получить два согласованных датасета — 640×640 и 1280×1280 пикселей.

Второе отличие — состав аугментаций. Помимо геометрических преобразований и изменения яркости в конвейер включены четыре типа шума («соль и перец», гауссовский, пуассоновский, размытие) и эластичные деформации; каждый воспроизводит конкретный артефакт съёмки — сбой сенсора, тепловой шум матрицы, низкую освещённость, смаз от движения. Такой набор редко применяется в детекции лиц, где обычно ограничиваются отражениями и яркостью [5].

Обучение проводилось на видеокарте Quadro RTX 5000 средствами PyTorch и Ultralytics; размер пакета — четыре изображения. Гиперпараметры взяты из документации моделей: начальная скорость обучения 0,01 для SGD (0,001 для Adam [6]), конечная — 1 % от начальной, моментум 0,937, регуляризация весов 0,0005, первые три эпохи — «разогрев». YOLOv5x6 обучалась на кадрах 1280×1280 , YOLOv8m — на 640×640 : обе модели видели один и тот же размеченный материал, что даёт корректную базу для сравнения архитектур.

Результаты

Точность оценивалась на тестовом наборе WIDER FACE, разделённом на подмножества EASY, MEDIUM и HARD [7]; дополнительно измерялось быстродействие на CPU и GPU. Результаты приведены в таблице 1.

Таблица 1 — Сравнение производительности моделей

Модель	Разрешение	GPU, fps	CPU, fps	EASY	MEDIUM	HARD
YOLOv5x6	1280×1280	8	0,337	0,984	0,964	0,913
YOLOv8m	640×640	54	4	0,977	0,949	0,860

YOLOv5x6 точнее на всех трёх уровнях, причём разрыв максимален на HARD (0,913 против 0,860) — там, где кадры содержат мелкие и размытые лица, то есть именно те, на которые нацелены апскейл и шумовые аугментации. YOLOv8m при вчетверо меньшем числе параметров уступает незначительно, но работает почти в семь раз быстрее на GPU и в десять раз быстрее на CPU.

Заключение

Сформированы два датасета по 25 000 размеченных изображений лиц в разрешениях 640×640 и 1280×1280, обучены детекторы YOLOv8m и YOLOv5x6. Предложенный конвейер подготовки данных обеспечил точность 0,984 / 0,964 / 0,913 для YOLOv5x6 и 0,977 / 0,949 / 0,860 для YOLOv8m на подмножествах EASY, MEDIUM и HARD.

Выбор модели определяется сценарием: для пакетной обработки архивов предпочтительна YOLOv5x6, для интерактивных сервисов и потокового видео — YOLOv8m. Обе модели внедрены в действующий веб-сервис [2] и опубликованы в открытых репозиториях [8, 9]; методика подготовки данных применима в смежных задачах детекции мелких объектов.

Использованные источники:

1. Majeed F., Khan F.Z., Iqbal M.J., Nazir M. Real-Time Surveillance System Based on Facial Recognition Using YOLOv5 // 2021 Mohammad Ali Jinnah University International Conference on Computing. 2021.
2. Сервис обезличивания лиц на фотографиях Hide Face [Электронный ресурс]. URL: <https://hide-face.com/> (дата обращения: 20.05.2023).
3. Dutta A., Zisserman A. The VIA Annotation Software for Images, Audio and Video // MM 2019: Proceedings of the 27th ACM International Conference on Multimedia. 2019.
4. Wang X., Yu K., Wu S., Gu J., Liu Y., Dong C., Qiao Y., Loy C.C. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks // Lecture Notes in Computer Science. 2019. Vol. 11133. P. 63–79.
5. Shorten C., Khoshgoftaar T.M. A Survey on Image Data Augmentation for Deep Learning // Journal of Big Data. 2019. Vol. 6. № 1.
6. Kingma D.P., Ba J.L. Adam: A Method for Stochastic Optimization // 3rd International Conference on Learning Representations. 2015.
7. Kim N., Kim J.H., Won C.S. FAFD: Fast and Accurate Face Detector // Electronics. 2022. Vol. 11. № 6.
8. Репозиторий модели YOLOv5x6 для детекции лиц [Электронный ресурс]. URL: <https://github.com/wadimAI3337/miem-yolov5x6-faceDetection> (дата обращения: 20.05.2023).
9. Репозиторий модели YOLOv8m для детекции лиц [Электронный ресурс]. URL: <https://github.com/wadimAI3337/miem-yolov8m-faceDetection> (дата обращения: 20.05.2023).