

РАЗРАБОТКА ИНФОРМАЦИОННОЙ СИСТЕМЫ АВТОМАТИЧЕСКОГО ОБНАРУЖЕНИЯ ОБЪЕКТОВ ОПРЕДЕЛЕННОГО ВИДА НА ПРОИЗВОЛЬНЫХ ИЗОБРАЖЕНИЯХ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

М. А. Широбокова

Удмуртский государственный университет, Ижевск, Россия

E-mail: shirobokova.margarita@mail.ru

Развитие рынка искусственного интеллекта происходит постоянно, компаниями разных сфер все чаще применяют новые методы и модели в рамках их основных и сопутствующих бизнес-процессов. Однако использование новых технологий, как правило, требует высокой квалификации от сотрудников компании, и осложнено тем, что в большей части компании отсутствуют специалисты по построению моделей машинного обучения. Для решения данной проблемы предлагается готовая информационная система обнаружения объектов определенного вида на произвольных изображениях с использованием методов машинного обучения. Предлагаемая платформа для обнаружения экземпляров семантических объектов позволяет оптимизировать деятельность организации за счет ускорения работы с потоковыми изображениями (видеорядом), где детекция объектов на изображениях может значительно снизить затраты и долю ручного труда сотрудников. Такой подход позволяет использовать уже готовую предобученную модель в рамках бизнес-процессов компании, а также дообучать модель на новые типы объектов в виде «единого» окна.

DEVELOPMENT OF INFORMATION SYSTEM FOR AUTOMATIC DETECTION OF OBJECTS OF A CERTAIN TYPE ON ARBITRARY IMAGES USING MACHINE LEARNING METHODS

M. A. Shirobokova

The development of the artificial intelligence market is ongoing, companies in various fields are increasingly using new methods and models as part of their main and related business processes. However, the use of new technologies, as a rule, requires high qualifications from the company's employees, and is complicated by the fact that in most of the company there are no specialists in building machine learning models. To solve this problem, a ready-made information system for detecting objects of a certain type on arbitrary images using machine learning methods is proposed. The proposed platform for detecting instances of semantic objects allows you to optimize the organization's activities by speeding up work with streaming images (video sequences), where the detection of objects in images can significantly reduce costs and the share of manual labor of employees. This approach allows using a ready-made pre-trained model within the company's business processes, as well as re-training the model for new types of objects in the form of a "single" window.

Российский рынок искусственного интеллекта в 2020 году достиг 291 млн. долларов. Такие данные приведены в IDC Worldwide Artificial Intelligence Spending Guide, о чем компания IDC Russia сообщила TAdviser 27 апреля 2021 года. Согласно данным IDC, расходы на решения с применением искусственного интеллекта в коммерческих и государственных организациях выросли на

22,4% по отношению к 2019 году, оставаясь значительными на протяжении всего 2020 года. Основные потребители продукта – компании, оказывающие услуги и потребительский сегмент. Наиболее интенсивно искусственного интеллекта используется в организациях финансового сектора (22.8%) и торговли (13%). В связи с имеющейся тенденцией производится разработка информационной системы автоматического обнаружения объектов определенного вида на произвольных изображениях с использованием методов машинного обучения.

Платформа представляет собой информационную систему, которая решает задачу обнаружения объектов определенного вида на произвольных изображениях. Ключевым отличием предлагаемого решения является тот факт, основные задачи для построения модели машинного обучения, а именно построение модели, разметка данных, построение отчетов по работе с модели, дообучение модели и ее дальнейшая интеграция в информационную систему пользователя, полностью автоматизированы и реализованы в виде «единого окна» – единого приложения. Критическим при обучении моделей искусственного интеллекта является тот факт, что модель не умеет работать с образцами, которые не присутствовали в обучающей выборке. Для решения этой задачи, а именно дообучения модели требуется участие специалиста по построению моделей машинного обучения (Data Scientist), что не всегда возможно для компании заказчика. Основной концепцией платформы является ее автономность от участия специалиста по построению моделей машинного обучения (Data Scientist) и полная автоматизация основных этапов построения модели машинного обучения.

Предлагаемая платформа для обнаружения экземпляров семантических объектов позволяет оптимизировать деятельность организации за счет ускорения работы с потоковыми изображениями (видеорядом), где детекция объектов на изображениях может значительно снизить затраты и долю ручного труда сотрудников. Разрабатываемое решение применимо в следующих сферах:

- ритейл: сканирование и подсчет товара, проверка актуальности ценников на полках магазинов;
- производство: обнаружение дефектов на деталях, проверка соблюдения техники безопасности на производстве – в т.ч. ношение касок\масок;
- медицина: диагностика отклонений от нормы\опухолей на снимках, обнаружение средств индивидуальной защиты - масок\халатов\перчаток;
- маркетинг: улучшение качества таргетированной рекламы за счет анализа частоты использования продуктов\брендов, обнаружение неприемлемого контента,
- документооборот: определение наличие печати/подписи/qr-кода/штрих-кода и других обязательных атрибутов на документе.

Основной задачей разработки автономной платформы для обнаружения экземпляров семантических объектов произвольного вида является обучение моделей машинного обучения хорошего качества, их автоматическая строгая валидация и формирование всего кода обучения в виде исполняемого файла.

Обучение модели базируется на технологии компьютерного зрения Object Detection (обнаружение объектов на изображении). Современные технологиче-

ские решения Object Detection основаны на использовании больших моделей нейронных сетей и алгоритмов их глубокого обучения. Подавляющее большинство текущих решений использует нейронные сети двух типов: R-CNN (Region-Based Convolutional Neural Network) и YOLO (You Only Look Once).

Однако согласно последним исследованиям, обнаружение объектов на открытом датасете COCO¹ производится с более высоким качеством при одновременном снижении времени ответа модели с помощью YOLO моделей.

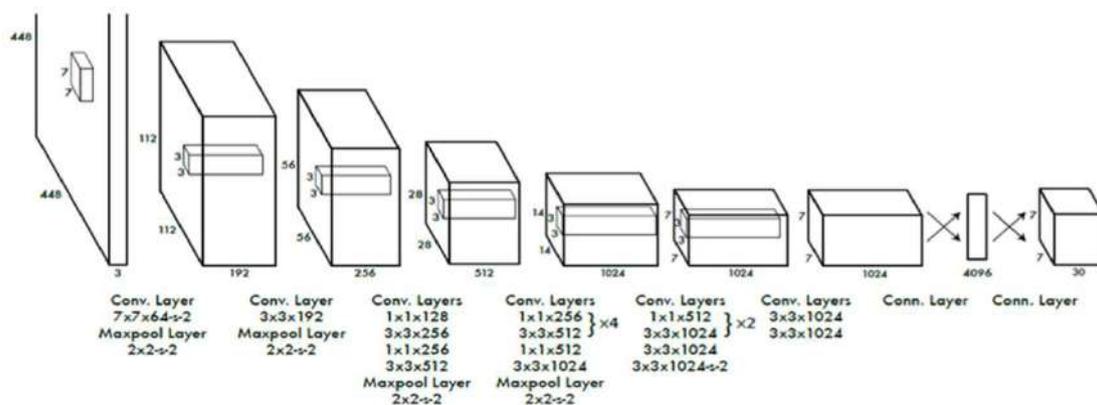


Рис. 1. Архитектура модели YOLO

Архитектура YOLO построена на основе модели GooLeNet для классификации изображений, как показано на рис. выше. Нейронная сеть имеет 24 сверточных слоя, за которыми следуют 2 полносвязных слоя.

Первоначально размер исходного изображения изменяется до размера $L \times L$ (448×448 , как правило) и делится на сетки $S \times S$ ($S = 7$, поэтому каждое подизображение размером 64×64 принадлежит одной сетке). Каждая сетка предлагает B ограничивающих прямоугольников (на практике $B = 2$). Каждая ограничительная рамка соответствует 4 координатам (x_{center} , y_{center} , w , h) и одному доверительному значению. x_{center} и y_{center} — центральная координата ограничивающего прямоугольника, а w и h — вес и высота ограничивающего прямоугольника соответственно. Значение достоверности вычисляется по уравнению следующим образом:

$$Confidence = Pr(Object) \times IOU_{pre}^{truth}$$

Если в ячейке сетки присутствует истинный ограничивающий прямоугольник для семантического объекта, то $Pr(Object) = 1$, в противном случае — 0, а IOU_{pre}^{truth} — это значение метрики IOU между предсказанной ограничивающей рамкой и фактической истинной ограничивающей рамкой.

На рис. ниже фактической истинной ограничивающей рамкой обозначена зеленым цветом, предсказанная ограничивающая рамка — красным.

¹ cocodataset.org — крупномасштабный набор данных для обнаружения, сегментации и подписи объектов

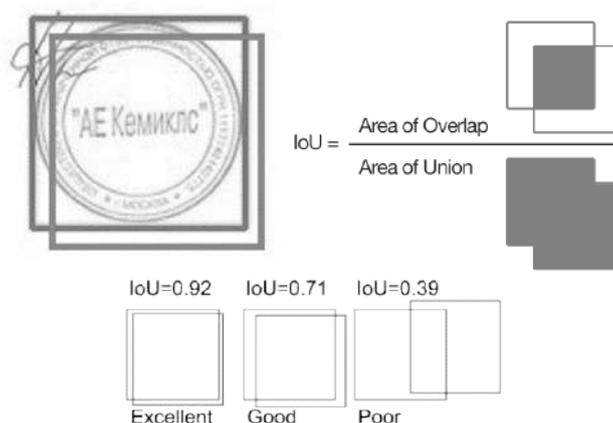


Рис. 2. Пример расчета метрики IOU

Кроме того, каждая сетка также генерирует C условных вероятностей класса $\Pr(Class_i|Object)$ (на практике $C = 20$). Таким образом, изображение, в итоге, производит $S \times S \times (B \times 5 + C)$ выходов (на практике $7 * 7 * (2 * 5 + 20) = 1470$ результатов). Во время тестирования YOLO умножает условные вероятности классов и прогнозы достоверности отдельных блоков,

$$\Pr(Class_i|Object) \times \Pr(Object) \times IOU_{pre}^{tru} = \Pr(Class_i) \times IOU_{pre}^{truth}$$

что дает оценки достоверности для каждой предсказанной ограничивающей рамки для каждого класса объектов.

Применимо к решаемой задаче скан-образы документов имеют структурированное изображение, где размеры и положение печатей и подписей детерминированы относительно размеров листа: то есть с высокой точностью можно предсказать геометрические размеры этих объектов, и они не будут значительно варьироваться. Это позволяет подобрать параметры структуры нейронной сети, которые минимизируют недостатки в локализации объекта и делает архитектуру YOLO наилучшим вариантом для обнаружения печатей и подписей на документе.

На текущий момент согласно заданной методике подготовлена модель машинного обучения, определяющая наличие объектов произвольного типа на документе:

- Типы объектов Печати, подписи, QR-коды, штрих-коды
- Объем обучающей выборки 6000 на каждый тип объекта
- Объем тестовой выборки 1500 на каждый тип объекта
- Скорость обучения модели 1.5ч / 1000 изображений на каждый тип объекта
- Скорость ответа модели ~ 1 сек. / изображение
- Объем памяти, занимаемый моделью ~ 85МБ
- Качество работы модели (mAP) ~ 0.98
- При использовании Yolo V5l
Google Colab (GPU K80)

Таким образом, текущее решение уже может применяться для обнаружения печатей и подписей на документах, а также при обучении модели по алгоритмам, собранным в виде единого программного кода, для поиска и распознавания ценников на товарах в магазинах, для подсчета количества товара на пол-

ках магазина, для обнаружения дефектов на деталях в масштабах промышленного производства и др.

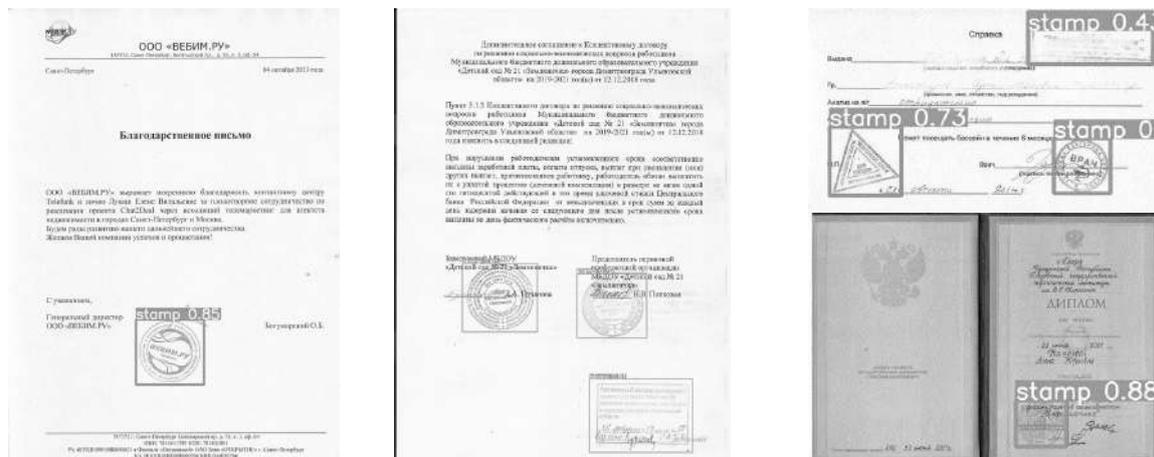


Рис. 3 Пример работы платформы с подключенной моделью определения печатей на разных типах документах²

Программное средство «Платформа для обнаружения экземпляров семантических объектов произвольного вида» зарегистрировано в Реестре программ для ЭВМ Федеральной службы по интеллектуальной собственности под номером 2022610966.

СПИСОК ЛИТЕРАТУРЫ

1. *Adebayo A. A.* MACE correlation filter algorithm for face verification in surveillance scenario // *Journal of Computer Science and Engineering*. 2013. Vol. 13. Iss. 1. P. 6-15.
2. *Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A.* CoRR. 2014.
3. *Redmon J., Divvala S., Girshick R., Farhadi A.* You Only Look Once: Unified, Real-Time Object Detection. [Электронный ресурс]. URL: <https://arxiv.org/abs/1506.02640> (дата обращения: 15.09.2022).
4. *Peiyuan J., Daji E., Fangyao L., Ying C., Bo M.* A Review of Yolo Algorithm Developments // *Procedia Computer Science*. 2022. Vol. 199. P. 1066-1073.
5. Будущее сервиса: чего ожидают клиенты в РФ, Беларуси и Казахстане. Исследование. [Электронный ресурс]. URL: <https://rusability.ru/articles/buduschee-servisa-chego-ozhidayut-klienti-v-rf-belarusi-i-kazahstane-issledovanie/5fd294ac2dda593c3483d4d2> (дата обращения: 20.09.2022).
6. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных. [Электронный ресурс]. URL: <http://www.machinelearning.ru> (дата обращения: 18.09.2022).
7. *Широбокова М. А., Приказчиков П. А.* Разработка алгоритма оценки работы персонала по аудио и видео данным с использованием искусственного интеллекта // XXII Всероссийская конф. молодых ученых по математическому моделированию и информационным технологиям, тезисы докладов. 2021.

² Представленные здесь и далее примеры документов найдены в сети Интернет.