

В. А. ГАЛКИН, А. В. МАКАРЕНКО

Институт проблем управления им. В.А.Трапезникова РАН, Москва

РАЗРАБОТКА АЛГОРИТМА ОЦЕНИВАНИЯ МАССЫ НАБЛЮДАЕМЫХ ОБЪЕКТОВ ПО ИХ ВИДЕОИЗОБРАЖЕНИЯМ (НА ПРИМЕРЕ СЕЛЬСКОХОЗЯЙСТВЕННЫХ ЖИВОТНЫХ)

Представленная работа направлена на разработку подхода и алгоритма оценивания массы наблюдаемого объекта по его видеоизображению в видимом диапазоне длин волн. Алгоритм демонстрируется и анализируется на примере «взвешивания» сельскохозяйственных животных, располагающихся в производственной станке в условиях промышленной фермы. Подход основывается на комбинации методов сегментации экземпляров объектов и регрессионной модели зависимости между площадью сегментирующих масок и весом. Результирующая ошибка наилучшей модели составила 12 % на исходных данных и 4 % после гомогенизации данных.

Введение. Современное развитие науки и техники привело к развитию систем машинного зрения, основным алгоритмическим ядром которых стали глубокие искусственные нейронные сети. Становление нейронных сетей происходило поэтапно, от решения задачи классификации до решения задач локализации и сегментации экземпляров. Во множестве практических приложений желаемой целью является не чистый выход сети, а формируемые из него те или иные оценки, в частности метрические характеристики объектов, такие как: линейные размеры, объем, масса и т.п.

Классический «прямой» подход к оцениванию массы объекта состоит в использовании весов, который является наиболее точным. На практике, использование весов не всегда возможно, по причинам большой трудоемкости или невозможности проведения непосредственного взвешивания. По этим причинам возникает и актуализируется задача автоматического дистанционного «неконтактного» взвешивания объектов по их визуальным признакам. Одним из востребованных примеров прикладного применения подобных подходов является сельское хозяйство, где зачастую необходима оперативная оценка массы животных. Вес животного и его динамика в течение времени является не только фактором оптимизации в мясной промышленности, но и напрямую связана со здоровьем животного.

Целью данной работы является разработка алгоритма оценивания массы наблюдаемых объектов по их видеоизображениям, а демонстрация и анализ разработанного алгоритма проводятся на примере сельскохозяйственных животных, располагающихся в производственной станке в условиях промышленной фермы.

Обзор методов. Визуальная оценка веса животных стала направлением для множества научных работ. Превалирующая часть работ основана на эмпирических соотношениях между морфологией объекта и весом. Одним из методов получения морфологии объекта только по визуальным данным является фотограмметрия [1, 2], результатом которой является 3-D модель объекта с привязкой его размеров к абсолютной шкале. Составление модели с помощью фотограмметрии имеет множество ограничений на методики получения данных, например неподвижность объекта во время съемки с различных ракурсов, поэтому данный метод не применим к перемещающимся объектам и/или объектам меняющим форму.

В обзорной работе [3] было выделено 7 ключевых визуальных признаков, которые используют для оценки веса объектов, а именно: площадь тела (вид сверху), высота в холке, высота бедер, длина тела, ширина бедер, объем тела и обхват груди. Оценка некоторых из них является весьма трудоемкой, требующей составления сложной скелетной модели [4], поэтому в первом приближении можно ограничиться только площадью тела.

Базовый подход к решению задачи заключается в двух этапах: автоматическая оценка морфологических признаков, и далее аппроксимация регрессионными моделями эмпирической зависимости между массой и морфологией. Одним из примеров такого подхода является решение задачи сегментации поверхности объекта, с последующей оценкой зависимости площади и веса. Зачастую, для формирования морфологических оценок из «сырых» данных используют

глубокие нейронные сети. Алгоритмы глубокого обучения выделяются способностью автоматически выявлять наиболее релевантные признаки из данных. По этим причинам возможно решать сквозную задачу оценки веса, в котором входом алгоритма являются «сырые», необработанные изображения, а выходом – оценка веса. В контексте оценивания веса были предприняты попытки использовать глубокие сети [5, 6], однако они все ещё имеют ограничения, основным из которых является объём аннотированных данных.

Методология и эксперимент.

Набор данных. Исходными данными для решения поставленной задачи являются видеоролики производственного станка. Съёмка производилась с верхнего ракурса двумя широкоугольными камерами с перекрывающимися полями зрения, изображения с которых впоследствии сшивались в единую картину. Следует отметить, что в ходе экспериментов использовалось несколько станков, разрешающая способность сшитых кадров между которыми может варьироваться в силу различия параметров шивки.

Для получения аннотированного набора данных применялась оригинальная методика. Из производственного станка в порядке очереди по одному выводились животные с интервалом [0.5, 4] минуты, которые направлялись на весы для измерения массы. После того как станок оставался пустым, животные проходили через него по одному, задерживаясь в нем на [0.5, 2] мин. Общая длительность аннотированных данных более 12-ти часов записей. Максимальное число животных в производственном станке равно 44. Значение веса животных варьируется от 20 до 115 кг. Полученный набор обладает большой вариативностью и включает в себя данные как для оценки веса станка в целом, с различным количеством животных, так и для получения оценки веса отдельных особей на различной стадии развития.

Таким образом, размеченный набор данных представляет собой видеопоследовательности $\{x_{i,j}\}_{i=1}^N$ различной длины (j – счетчик по длине временного ряда) и соответствующие значения суммарного веса всех наблюдаемых животных в станке $\{y_i\}_{i=1}^N$, где количество роликов $N = 349$.

Метод оценивания веса. Для оценивания веса животных в производственном станке использовался подход на основе поиска зависимости между суммарной площадью масок, сегментирующих животных в кадре и суммарным весом станка. Сегментирующие маски получены путем решения задачи сегментации экземпляров. Используемый подход не зависит от метода решения задачи сегментации, поэтому, не теряя общности оставим конкретизацию подхода к сегментации за рамками данной работы.

Результирующими данными задачи сегментации являются временные ряды, каждый из которых содержит суммарную площадь масок наблюдаемых объектов на каждом кадре. Наблюдаемая сцена является динамической, поэтому возможны вариации площади масок от кадра к кадру, при неизменном суммарном весе объектов. Нормированная на вес медианная дисперсия площади в данных равна $\sigma = 8.02$, причем наблюдается высокий уровень дисперсии при малом количестве наблюдаемых объектов. Причина этого заключается в использовании статистической модели для получения сегментирующих масок. Статистическим моделям присущи ошибки, которые при немногочисленном числе объектов вносят в суммарную площадь ошибочный вклад одного порядка с корректными предсказаниями.

Агрегируем временной ряд в медианное значение площади. В таком случае, набор данных представляет собой пару векторов $X = \{\text{median } x_i\}_{i=1}^N$ и $Y = \{y_i\}_{i=1}^N$.

Для аппроксимации неизвестной функциональной зависимости между медианной площадью и суммарной массой используем три модели: RANSAC [7] и полиномиальную регрессию 2-го и 3-го порядков. Использование простых регрессионных моделей обуславливается высокой степенью коррелированности медианы площади и веса, коэффициент корреляции равен 0.98.

Для оценки качества обученных моделей были использованы следующие классические оценки качества регрессии: средняя абсолютная процентная ошибка (MAPE), медианная абсолютная процентная ошибка (median APE) и коэффициент детерминации (R^2). Обучение и тестирование моделей производилось на полной выборке данных, то есть без разбиения на обучающую и тестовую части. В силу простоты данных и моделей, а так же небольшого количества обучаемых параметров, в сравнении с размером выборки, полученные оценки будут объективными. С целью повышения устойчивости полученных результатов на каждую модель

производится 500 повторных запусков, из которых будет выбрана одна, наиболее точная (топ-1) по медианной абсолютной процентной ошибке модель.

Референсное решение. Как было сказано выше, в качестве моделей для решения задачи выступают: линейная регрессия RANSAC и полиномиальная регрессия 2-го и 3-го порядков. Точность референсных моделей на выборке данных приведена в табл. 1.

Таблица 1

Точность референсных моделей на выборке данных

| | MAPE | median APE | R^2 |
|------------------------------|-------|------------|-------|
| RANSAC | 0.216 | 0.188 | 0.918 |
| Полин. регрессия (2 порядок) | 0.139 | 0.118 | 0.976 |
| Полин. регрессия (3 порядок) | 0.591 | 0.128 | 0.976 |

Из полученных результатов можно заметить, что полиномиальная регрессия 2-го порядка имеет наилучшую точность. Полученная результирующая точность – достаточно низкая. Причина этого заключается в использовании нескольких производственных станков, с различными параметрами шивки и как следствие разрешением сшитых кадров, а также наблюдение объектов разных размеров и массы. Перечисленные причины, в свою очередь, приводят к неоднородности совмещенной выборки. Таким образом, встает вопрос о поиске преобразования выборок в однородные, с учетом возможности использования преобразования на новых данных. Одним из возможных вариантов построения такого преобразования является операция нормализации, откалиброванная посредством эталонных макетов животных.

Модифицированное решение. Как было показано в предыдущем подразделе, для повышения точности методов оценивания веса с использованием площади объектов требуется поиск преобразования, устраняющего неоднородность в данных. По причине отсутствия данных, достаточных для формирования такого преобразования, найдем его эмпирически, для демонстрации потенциально возможной точности предложенного подхода. Результирующая точность моделей после преобразования данных приведена в табл. 2. Полученная ошибка по метрике median APE – не более 4 % веса от реально измеренного.

Таблица 2

Точность модифицированных моделей на выборке данных

| | MAPE | median APE | R^2 |
|------------------------------|-------|------------|-------|
| RANSAC | 0.243 | 0.091 | 0.991 |
| Полин. регрессия (2 порядок) | 0.061 | 0.042 | 0.991 |
| Полин. регрессия (3 порядок) | 0.892 | 0.132 | 0.983 |

Заключение. В представленной работе предложен алгоритм оценивания массы наблюдаемого объекта по его видеоизображению в видимом диапазоне длин волн. Демонстрация алгоритма проведена на примере «взвешивания» сельскохозяйственных животных, располагающихся в производственном станке в условиях промышленной фермы. Применяемый подход основывается на поиске зависимости между медианной площадью масок наблюдаемых объектов за некоторый промежуток времени и суммарной массой объектов. В качестве модели для получения из видеоданных площади масок наблюдаемых объектов выступала глубокая сверточная нейронная сеть. Для оценки веса было рассмотрено 3 модели, а именно линейная регрессия RANSAC и полиномиальные регрессии 2-го и 3-го порядков. Было сформировано референсное и модифицированное решение, в котором показана эффективно достижимая точность. Наиболее точной моделью оказалась полиномиальная регрессия 2-го порядка, ошибка которой по метрике median APE равна 0.118 и 0.042 для референсного и модифицированного решения соответственно. Для анализа моделей были построены зависимости ошибки оценки веса от различных характеристик данных. Наблюдаемые зависимости продемонстрировали стабилизацию ошибки моделей при наблюдении не менее 3-х объектов одновременно.

ЛИТЕРАТУРА

1. Jeremy A Bot и др. Using 3D photogrammetry to create open-access models of live animals: 2D and 3D software solutions. в: Grayburn et al., eds. D 3 (2019), с. 54–72.

2. Sakir Tasdemir и Ilker Ali Ozkan. «ANN approach for estimation of cow weight depending on photogrammetric body dimensions». *International Journal of Engineering and Geosciences* 4.1 (2019), с. 36–44.
3. Roel Dohmen, Cagatay Catal и Qingzhi Liu. «Computer vision-based weight estimation of livestock: a systematic literature review». *New Zealand Journal of Agricultural Research* 65.2-3 (2022), с. 227–247.
4. Steven Yik и др. «DIAT (Depth-Infrared Image Annotation Transfer) for Training a Depth-Based Pig-Pose Detector». *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2020, с. 2459–2466.
5. Dan B Jensen, Katarina N Dominiak и Lene J Pedersen. «Automatic estimation of slaughter pig live weight using convolutional neural networks». *II International Conference on Agro BigData and Decision Support Systems in Agriculture*. 2018.
6. Roel Dohmen, Cagatay Catal и Qingzhi Liu. «Image-based body mass prediction of heifers using deep neural networks». *Biosystems Engineering* 204 (2021), с. 283–293.
7. Martin A Fischler и Robert C Bolles. «Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography». *Communications of the ACM* 24.6 (1981), с. 381–395.

V.A.Galkin, A.V.Makarenko, (Institute of Control Sciences RAS, Moscow)

Algorithm development for the estimation of the weight of observed objects by their video images (by example on the animal husbandry)

The presented paper is aimed at developing an approach and an algorithm for estimating the weight of an observed object from its video images in the visible wavelength range. The algorithm is demonstrated and analyzed on the example of “weighing” livestock placed on industrial farm. The approach is based on a combination of methods for instances segmenting object and a regression model of the relationship between the area of segmenting masks and weight. The resulting error of the best model was 12 % on the source data and 4 % on the data after applying homogenization.