

ИНФОРМАЦИОННЫЕ  
ТЕХНОЛОГИИ  
И ТЕЛЕКОММУНИКАЦИИ

INFORMATION  
TECHNOLOGIES  
AND TELECOMMUNICATIONS

УДК 551.501.816; 551.510.411

DOI: 10.17212/2782-2001-2023-3-69-86

## **Применение нейронных сетей для классификации объектов в видеоданных в задачах автоматического мониторинга состояния биологических объектов\***

**Е.А. ЧЕРКАШИН**

*199178, г. Санкт-Петербург, 14-я линия В.О., 39, Федеральное государственное  
бюджетное учреждение науки «Санкт-Петербургский федеральный исследова-  
тельский центр Российской академии наук»*

*egor@cea7.ru*

Устойчивая тенденция к развитию технологий в быту и промышленности ведет к их постоянному усложнению, к внедрению новых разработок и методов, появляются новые способы получения и обработки информации. Неотъемлемой частью повседневной жизни стали системы видеонаблюдения. Камеры установлены на многоквартирных домах, в общественных местах, даже в квартирах и частных домах. Одной из наиболее важных областей, где компьютерное зрение нашло значительное применение, является наблюдение за биологическими объектами и анализ их поведения и состояния, особенно в контексте охраны природы и автоматизации исследований поведения животных. С ростом качества и количества видеоматериалов становится актуальной проблема точных и эффективных методов классификации объектов в реальном времени, особенно когда речь идет об автоматическом мониторинге состояния животных.

В настоящей работе описаны основные виды распознавания объектов, рассмотрены популярные современные архитектуры нейронных сетей, проведен сравнительный анализ некоторых из них применительно к решаемой задаче. В ходе работы с использованием методов дополнительной обработки и аннотаторов изображений SuperVisely и VGG Image Annotator сформирован датасет с более чем тысячей уникальных изображений для извлечения релевантных характеристик объектов, а также проведены экспериментальные исследования качества распознавания объектов на видео с помощью предобученных известных нейросетевых моделей. Определены и сформулированы требования к исходным данным для эффективного решения задачи автоматического мониторинга и прогнозирования состояния биологических объектов. Показано, что во избежание появления слепых зон в месте обитания животных необходимо использовать достаточно большое количество камер, размещенных с перекрытием исследуемого пространства, чтобы объекты мониторинга были постоянно в поле видимости. Это позволит

---

\* Статья получена 15 июня 2023.

Исследование выполнено за счет гранта Российского научного фонда No 23-19-20081, <https://rscf.ru/project/23-19-20081/> и Санкт-Петербургского научного фонда.

впоследствии составить общую картину высокого разрешения, составленную из изображений всех камер, а по полученной картине осуществлять классификацию объектов посредством искусственных нейронных сетей.

**Ключевые слова:** искусственные нейронные сети, компьютерное зрение, нейросетевое прогнозирование, датасет изображений, аугментация данных, классификация изображений, локализация объектов, сегментация областей, аннотирование видеоданных

## ВВЕДЕНИЕ

В настоящее время нейронные сети используются во многих отраслях промышленности для решения сложных задач, которые ранее было трудно или невозможно решить традиционными методами. Одной из областей, где нейросетевое управление показало большие перспективы, является область прогнозного моделирования, где искусственные нейронные сети (ИНС) используются для прогнозирования результатов работы сложных систем.

Одним из наиболее интересных применений ИНС является область компьютерного зрения, где нейронные сети используются для анализа изображений и видеоданных. В последние годы методы компьютерного зрения используются для прогнозирования состояния биологических объектов, таких как растения, животные и даже люди [1]. Например, компания Cattle Care использует методы компьютерного зрения и ИНС для отслеживания жизненного цикла КРС [2].

Существует множество преимуществ прогнозирования биологических состояний с помощью ИНС. Например, ИНС предоставляют исследователям и врачам информацию о состоянии пациента в режиме реального времени, что позволяет проводить более своевременное и целенаправленное лечение [3]. ИНС также можно использовать для выявления биомаркеров, которые могут быть использованы для предсказания начала заболевания до появления симптомов, что позволяет проводить более раннюю диагностику и лечение [4, 5].

В сельском хозяйстве ИНС находят применение для увеличения точности прогнозирования состояния урожая, для расчета необходимых удобрений и пестицидов [6, 7]. Необходимо отметить, что наиболее развито использование современных информационных технологий, включающих методы искусственного интеллекта и ИНС, в точном растениеводстве, а в точном животноводстве цифровизация отрасли несколько отстает, при этом возникают правовые проблемы использования методов искусственного интеллекта [8, 9].

Целью настоящей статьи является исследование точной и эффективной модели для классификации и прогнозирования здоровья крупного рогатого скота (КРС) в режиме реального времени при разработке интеллектуальной системы видеомониторинга. Большое внимание уделяется модели, способной классифицировать различные аспекты поведения крупного рогатого скота, такие как кормление, отдых или признаки дистресса. В будущем модель должна быть обучена предсказывать потенциальные проблемы со здоровьем КРС на основе конкретных признаков, загруженных в нейронную сеть, что позволит обнаруживать их на ранней стадии и вовремя принимать меры.

Анализ собранных данных позволит принимать обоснованные управляющие решения относительно стратегий кормления, корректировки окружающей среды или распределения ресурсов, что в конечном итоге может значительно

улучшить производительность и эффективность всего поголовья в молочных хозяйствах.

Для достижения целей исследования необходимо собирать видеоданные из молочных хозяйств с учетом особенностей реализации систем видеонаблюдения в них, фиксируя при этом различные физиологические и поведенческие состояния КРС. Эти данные должны быть подвергнуты методам предварительной обработки для извлечения соответствующих характеристик и аннотаций для обучения нейросетевых моделей. В процессе работы будут исследованы различные архитектуры нейросетей и методики их обучения, чтобы в результате получилась модель, максимально точно классифицирующая КРС. Будут представлены и проанализированы результаты экспериментов с оценкой производительности различных сетевых архитектур и эффективности включения конкретных признаков для прогнозирования физиологического состояния биологических объектов. Полученные результаты позволят определить потенциальные проблемы и направления дальнейших исследований.

## 1. ЗАДАЧИ РАСПОЗНАВАНИЯ

Обеспечение качественного распознавания является фундаментальной задачей в компьютерном зрении и системах, обучаемых с помощью нейросетевых технологий. Точное распознавание различных жизненных аспектов КРС, таких как поведение и некие жизненные показатели, необходимо в вопросе обеспечения благосостояния всего поголовья. Существует несколько видов задач распознавания:

Классификация. Данную задачу можно рассмотреть как присвоение метки или категоризации объектам на полноценных изображениях или кадрах видео. В рассматриваемом проекте полезно безошибочно классифицировать не только вид животного (корова или лошадь, рис. 1), но и деятельность животного в рассматриваемый момент времени.



Рис. 1. Классификация объектов на фото с помощью предобученной нейросети: объект «лошадь» (слева), объект «корова» (справа)

Fig. 1. Classification of objects in the photo with a pre-trained neural network: a horse (left), a cow (right)

Локализация. Эта задача включает в себя определение точного местоположения в кадре или получение набора координат объекта. Точная локализация (рис. 2) позволит, к примеру, своевременно определить проблемы с ногами или позвоночником у КРС, ведь у больной коровы походка может сильно отличаться.



Рис. 2. Локализация коровы с помощью предобученной нейросети

Fig. 2. Localization of a cow using a pre-trained neural network

Сегментация – это задача присвоения меток не объекту в целом, а каждому пикселю, который принадлежит объекту. Это существенно повышает точность полученных из видео данных для дальнейшего анализа и принятия решений. Пример сегментации будет приведен далее по ходу работы (см. рассмотрение аннотатора SuperVisely).

Для выполнения поставленных задач на разных этапах понадобится решить все три задачи, так как для автоматического мониторинга состояния крупного рогатого скота недостаточно просто понимать, что в кадре находится КРС, необходимо отличать коров от предметов интерьера, от стен, людей и прочего, что может попасть в кадр. Также необходимо в реальном времени отслеживать жизнедеятельность крупного рогатого скота, что невозможно без детальной и четкой сегментации.

На данный момент существует несколько наиболее известных нейросетевых топологий, которые применяются для решения подобных задач (например, SegNet, DeepLab и другие) [10–12].

В работе [13] представлена модель SegNet, которая является архитектурой глубокого обучения, специально разработанной для семантической сегментации изображений. Модель SegNet состоит из структуры кодер–декодер, где кодер учится извлекать высокоуровневые характеристики из входного изображения, а декодер восстанавливает сегментированное изображение на основе этих характеристик.

Статья [14] описывает современную архитектуру для семантической сегментации изображений. По словам авторов статьи, DeepLab решает такие задачи, как улавливание мелких деталей, обработка больших рецептивных полей и сохранение пространственного разрешения. Авторы используют модифицированную версию популярной архитектуры CNN, такой как VGG-16 или ResNet, для сбора богатой контекстной информации. В дополнение к глубокому CNN и свертке Atrous авторы используют полностью связанные CRF для уточнения результатов сегментации.

В статье [15] представлена архитектура конволюционной нейронной сети (CNN), разработанной специально для задач сегментации биомедицинских изображений. Точная сегментация имеет решающее значение для различных приложений, включая диагностику заболеваний, планирование лечения и медицинские исследования.

Другой моделью, ориентированной на решение задач, связанных с выделением объектов на видеоданных, является YOLOv3 [16]. Проанализируем применимость моделей U-Net и YOLOv3 для решения поставленной задачи. Обе архитектуры легко обучаемы, достаточно легко адаптируются под разные данные и обеспечивают хорошее быстроедействие, они используются для задач компьютерного зрения, хотя направленность и области применения у них разнятся.

YOLOv3 – это нейросетевая архитектура, предназначенная для идентификации нескольких объектов на изображении или видео. Данная нейросеть делит входное изображение на сетку и назначает все ячейки данной сетки ответственными за обнаружение объектов. Данный метод позволяет обнаруживать объекты на видео в режиме реального времени.

U-Net, в свою очередь, предназначена для сегментации изображений, где целью является классификация и маркировка каждого пикселя в кадре. Данная архитектура используется в более точных областях, таких как компьютерное определение и сегментация опухолей.



Рис. 3. Сегментация областей в U-Net

Fig. 3. Segmentation of areas in the U-Net

Существенные различия между рассматриваемыми архитектурами.

- *Скорость работы.* YOLOv3 значительно быстрее оппонента в плане скорости, поэтому она является одним из лидеров в задачах обработки в режиме реального времени.

- *Точность работы.* U-Net, хоть и работает ощутимо медленнее, существенно превосходит конкурента в создании высокоточных сегментаций на кадрах.

- *Обучение.* Обе архитектуры требуют аннотированных обучающих данных, но формат этих данных существенно отличается, как и формат выходных данных. YOLOv3 требует аннотации ограничивающих рамок для обучения, тогда как U-Net требует аннотации пикселей.

- *Формат выходных данных.* YOLOv3 выводит координаты рамок вместе с вероятностью класса для всех обнаруженных объектов (объект «лошадь» на рис. 1), в то же время U-Net на выходе выдает сегментированное изображение, где все пиксели классифицируются по-разному.

## 2. АННОТИРОВАНИЕ ВИДЕОДАНЫХ

Для проработки на реальном видео были использованы предобученные модели с сайта ImageAI [17], однако результат на предобученных моделях не смог успешно определить даже малую часть коров на видео (рис. 4). Это показало, что предобученных моделей недостаточно для полноценного распознавания объектов.



Рис. 4. Попытка определения объектов на видео с предобученной нейронной сетью

Fig. 4. An attempt to identify objects in a video with a pre-trained neural network

В дополнение к выбору нейросетевой архитектуры необходимо использовать инструменты аннотации, такие как SuperVisely (рис. 6) или VGG Image

Annotator – VIA (рис. 7). Эти инструменты представляют собой эффективные и удобные интерфейсы для аннотирования видеоданных.

Для аннотирования с помощью SuperVisely исходное видео было разбито на кадры с помощью утилиты ffmpeg (рис. 5).

```
karzhen@MacBook-Pro-Egor Downloads % /Users/karzhen/Downloads/ffmpeg -i short_vide
o.mp4 -r 3 frames/$image%03d.png
```

Рис. 5. Разбиение исходного видео на кадры

Fig. 5. Splitting the source video into frames

В SuperVisely можно покадрово разбивать видео на области с объектами, тем самым формируя базу данных для сегментации.

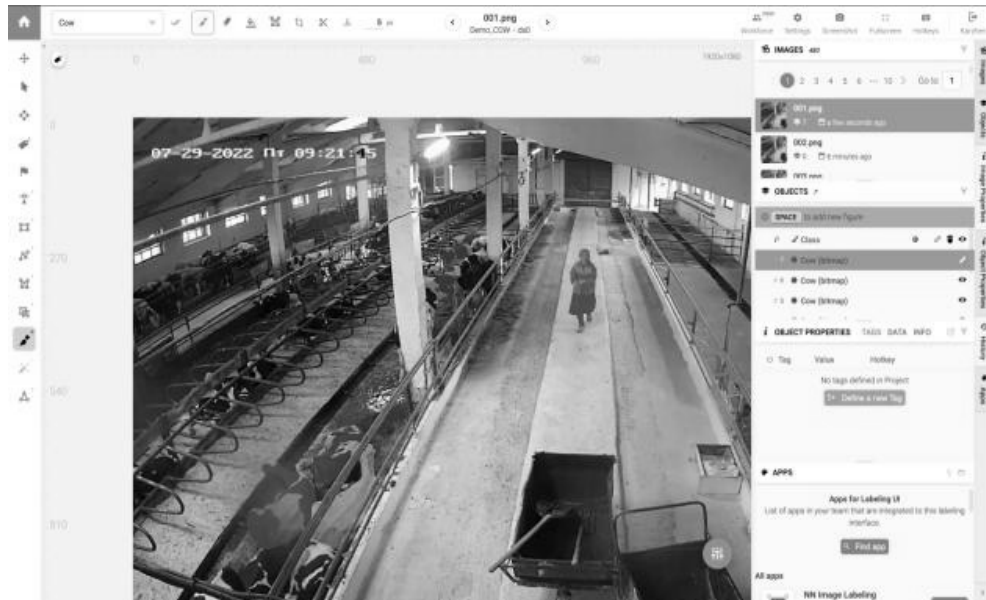


Рис. 6. Выделение зон (сегментов) на кадре в среде SuperVisely

Fig. 6. Selecting zones (segments) on the frame in the SuperVisely environment

При попытке более точной сегментации при масштабировании кадров видно, что области едва различимы даже для человеческого глаза, который не нужно тренировать для распознавания коров на видео (рис. 8). При этом возникает одна из проблем при решении задачи классификации биологических объектов в видео в задачах автоматического мониторинга их состояния: для качественного аннотирования изображений при дальнейшей классификации и сегментации объектов на видео необходимо использовать изображения с высоким разрешением, что, в свою очередь, требует больших вычислительных мощностей и увеличивает время обработки.

С помощью VIA можно довольно быстро выделять объекты в рамки для последующего пополнения датасета.



Рис. 7. Масштабирование видео для более точного сегментного аннотирования

Fig. 7. Scaling video for more accurate segment annotation



Рис. 8. Выделение зон (рамок) в VGG Image Annotator

Fig. 8. Selecting zones (frames) in the VGG Image Annotator



### 3. ЗАДАЧА ФОРМИРОВАНИЯ ДАТАСЕТА

На этапе формирования датасета необходимо использовать методы предварительной обработки для извлечения релевантных характеристик объектов. Перечислим основные методы.

- **Выборка кадров.** В некоторых случаях (рис. 11) далеко не все кадры имеют значимость при формировании датасета, так как кадры могут практически не отличаться. Если видео снято в разрешении 60 кадров в секунду, а животное стоит неподвижно, то сотни кадров могут быть не нужны, так как будут идентичными. Выборка кадров помогает сформировать множество кадров, которые будут сохранять временную характеристику, но при этом не будут использоваться дополнительные вычислительные мощности.

- **Изменение размера и масштабирование кадров.** Данный метод помогает стандартизировать датасет. Для более точного и быстрого обучения изображения должны иметь одинаковый формат.

- **Дополнение данных.** Данный метод применяется для увеличения разнообразия обучающих данных. Дополнение данных включает в себя переворачивание, зеркаливание, добавление шумов или избавление от них, высветление или затемнение изображения, изменение цветовой гаммы и прочее (рис. 9–12).

- **Вычитание фона.** Данный метод полезен в задаче сегментации объектов, где нужно отделить объекты переднего плана от элементов заднего.

Опишем подробнее метод дополнения данных.

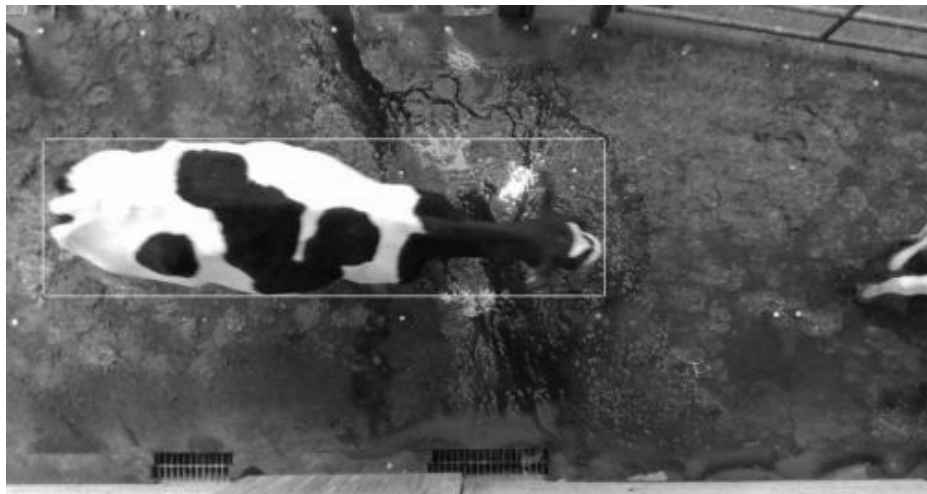


Рис. 9. Аннотирование КРС с помощью VGG Image Annotator

Fig. 9. Annotating cattle with the VGG Image Annotator

С помощью VIA аннотированы объекты «корова» на разных кадрах (рис. 9), затем с помощью скрипта на python для каждого изображения в исходной папке реализован метод дополнения данных (рис. 10), в результате которого для каждого кадра созданы перевернутое, отраженное, осветленное на 20 % и затемненное на 20 % изображения.

```

1 import os
2 import cv2
3
4 # Define the paths for the original dataset folder and the new dataset folder
5 original_folder = 'dataset'
6 new_folder = 'newDataset'
7
8 # Create the new dataset folder if it doesn't exist
9 if not os.path.exists(new_folder):
10     os.makedirs(new_folder)
11
12 # Define the data augmentation transformations
13 transformations = [
14     ('original', lambda img: img),
15     ('flipped', lambda img: cv2.flip(img, 1)),
16     ('mirrored', lambda img: cv2.flip(img, 0)),
17     ('brightened', lambda img: cv2.convertScaleAbs(img, alpha=1.2, beta=0)),
18     ('darkened', lambda img: cv2.convertScaleAbs(img, alpha=0.8, beta=0))
19 ]
20
21 # Loop through the files in the original dataset folder
22 for filename in os.listdir(original_folder):
23     # Load the original image
24     image_path = os.path.join(original_folder, filename)
25     img = cv2.imread(image_path)
26
27     # Apply data augmentation and save the images to the new dataset folder
28     for transform_name, transform_func in transformations:
29         transformed_img = transform_func(img)
30         new_filename = f'{os.path.splitext(filename)[0]}_{transform_name}.jpg'
31         new_image_path = os.path.join(new_folder, new_filename)
32         cv2.imwrite(new_image_path, transformed_img)

```

Рис. 10. Скрипт на языке python для дополнения данных

Fig. 10. A python script for data augmentation



Рис. 11. Исходный датасет с объектом «корова» в разных кадрах

Fig. 11. Source dataset with the “cow” object in different frames



Рис. 12. Датасет после аугментации данных

Fig. 12. Dataset after data augmentation

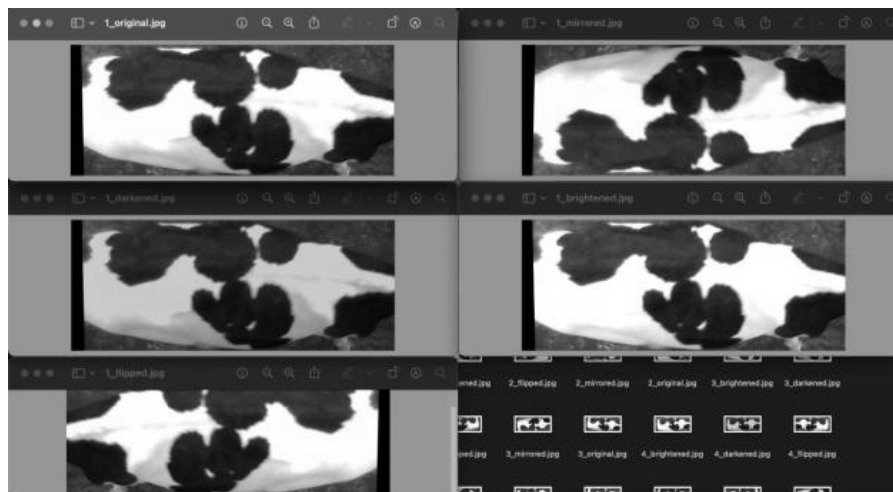


Рис. 13. Вариации одного изображения после применения скрипта

Fig. 13. Variations of a single image after applying the script

Важно отметить, что аугментация данных должна применяться только к обучающему набору данных. Наборы данных для проверки и тестирования не должны подвергаться увеличению данных, а должны содержать строго оригинальные изображения.

Для проверки работоспособности на предобученных моделях YOLOv3 и RetinaNET было взято небольшое видео, где человек четко будет видеть коров на видео. Существенная разница заметна в скорости выполнения программы с помощью разных моделей.

Результат работы YOLOv3 составил 42.914 секунды.

В свою очередь, время выполнения программы с RetinaNET заняло 218.175 секунды.

На рис. 14 и 15 представлены результаты для некоторых кадров из исходного видео.

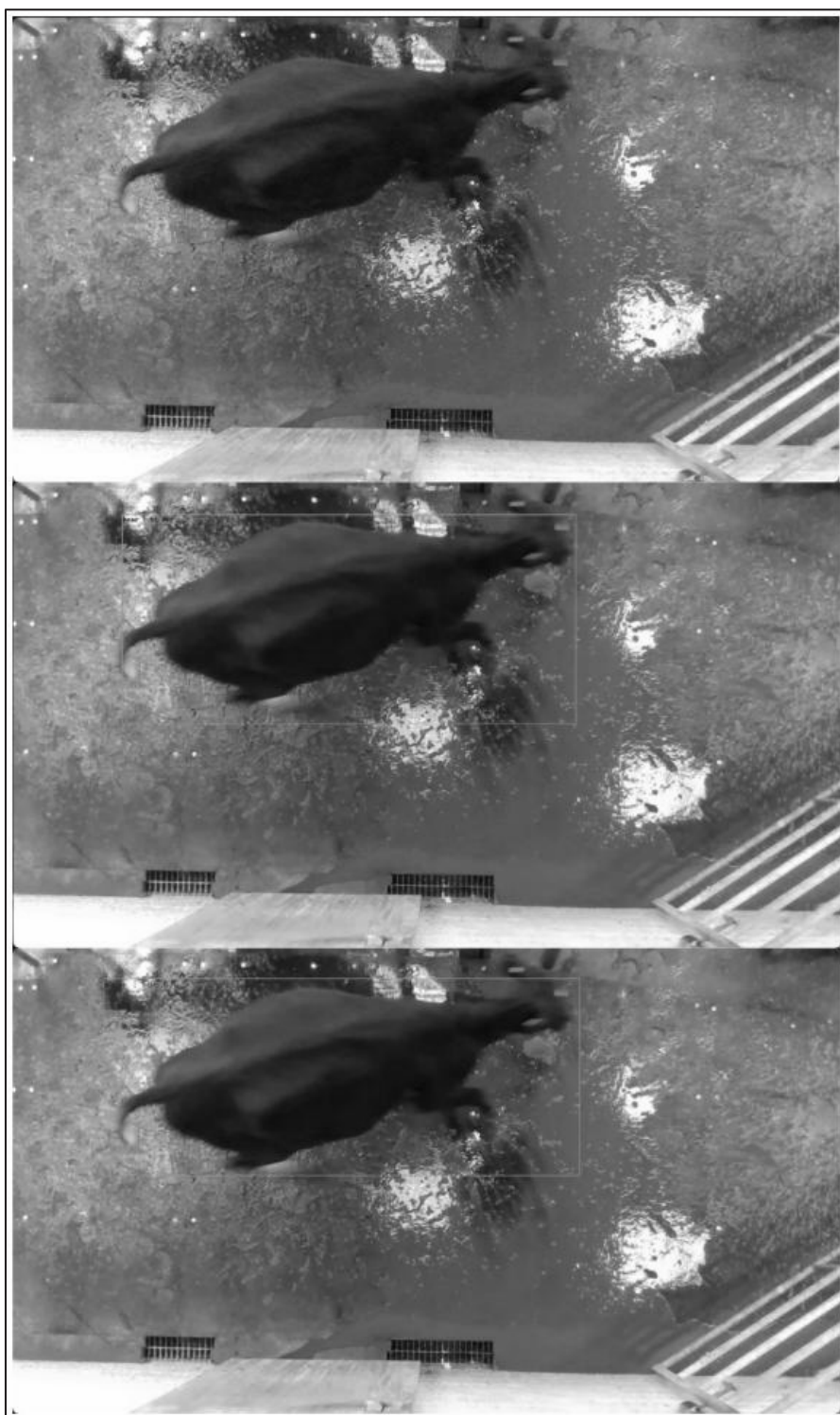


Рис. 14. Исходное изображение, RetinaNet (51,88 %), YOLOv3 (96,5 %)

Fig. 14. An original image, RetinaNet (51.88 %), YOLOv3 (96.5 %)

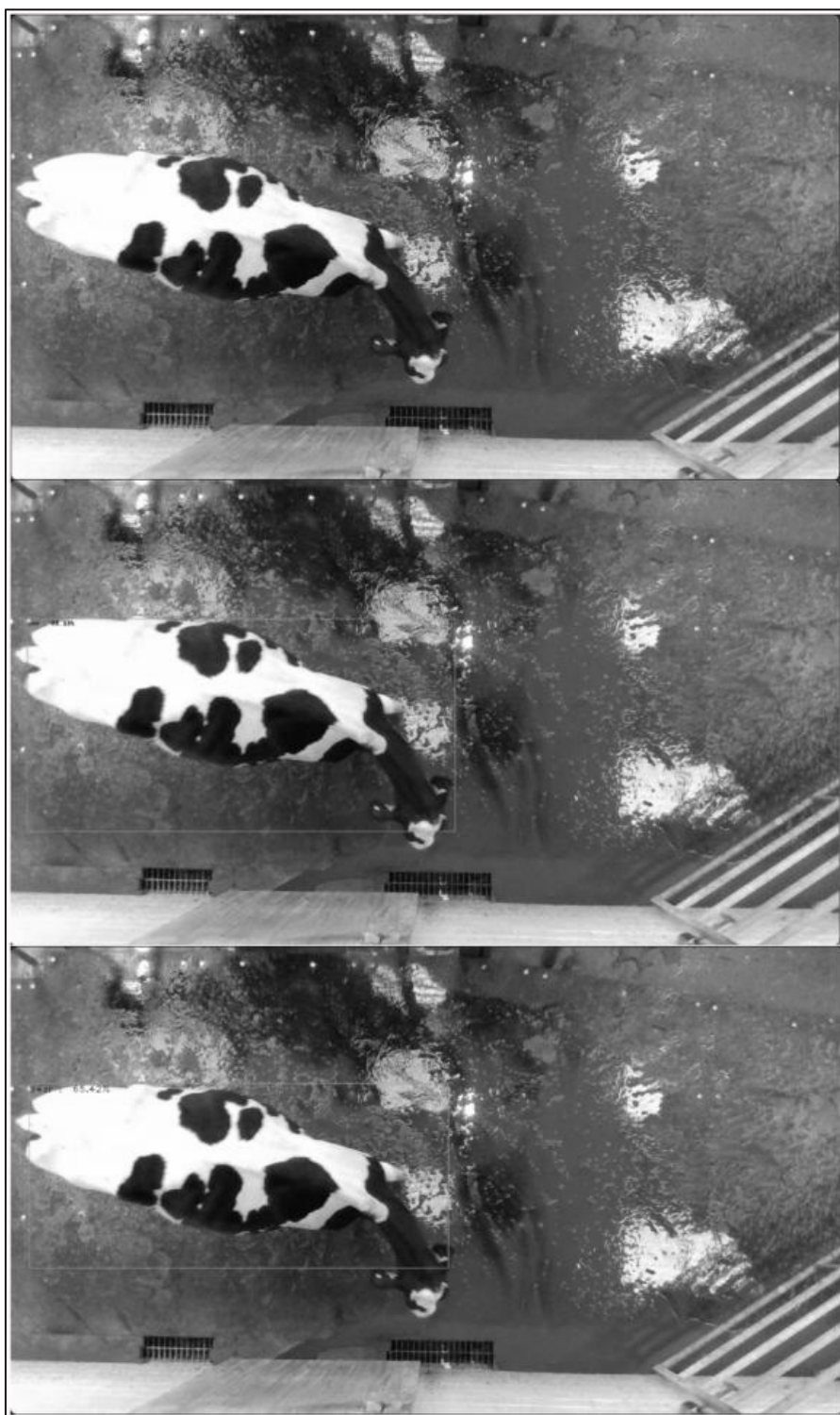


Рис. 15. Исходное изображение, RetinaNet (89,13 %), YOLOv3 (65,42 %)

Fig. 15. An original image, RetinaNet (89.13 %), YOLOv3 (65.42 %)

Сравнение рассмотренных моделей показало только существенную разницу по скорости, точность классификации не оказалась стабильно лучше у одной из моделей, поэтому без дополнительного обучения рано делать выводы и расставлять приоритеты использования.

В таблице приведен сравнительный анализ рассмотренных моделей для видео длительностью 5 секунд.

**Сравнительный анализ работы RetinaNet и YOLOv3**  
**Comparative analysis of RetinaNet and YOLOv3 performance**

Название модели	Время выполнения	Процент классификации в первом примере	Процент классификации во втором примере
RetinaNet	218.175 с	51.88 %	89.13 %
YOLOv3	42.914 с	96.5 %	65.42 %

### ЗАКЛЮЧЕНИЕ

В работе были рассмотрены задачи распознавания объектов на видео, описаны основные современные нейросетевые технологии, используемые для поставленных задач. Проведен сравнительный анализ наиболее релевантных для решаемой задачи архитектур U-Net и YOLOv3. При использовании каждой архитектуры выделены преимущества и недостатки. U-Net предназначена для высокоточных сегментаций в кадре, а YOLOv3 значительно быстрее по скорости, поэтому она больше подходит для использования в режиме реального времени. Во время работы использовались аннотаторы изображений SuperVisely и VGG Image Annotator. С их помощью был сформирован базовый датасет, который впоследствии значительно увеличился путем аугментации данных. Для проверки сверточных нейросетей в задаче классификации объектов на видео были использованы предобученные модели RetinaNet и YOLOv3.

Сформулированы следующие требования к исходным данным для эффективного решения поставленной задачи выделения и классификации биологических объектов в реальном времени:

- Необходимо использовать камеры как можно более высокого разрешения с учетом стоимости и экономической эффективности для четкого приближения объектов на этапе аннотирования.
- Во избежание появления слепых зон в месте обитания КРС необходимо использовать достаточно большое количество камер, размещенных с перекрытием исследуемого пространства, чтобы объекты мониторинга были постоянно в поле видимости. Это позволит впоследствии составить общую картину высокого разрешения, составленную из изображений всех камер, а по полученной картине осуществлять классификацию объектов посредством ИНС.
- Для достижения необходимой точности классификации с помощью исследованных моделей ИНС RetinaNet и YOLOv3 необходимо либо самостоятельно формировать датасет из реальных данных, либо брать за основу пред-

обученные модели, но дообучать их с помощью аннотированных вручную видеоданных.

Выявленные требования будут учтены при дальнейших исследованиях в рамках решения задачи автоматического мониторинга и прогнозирования состояния биологических объектов.

## СПИСОК ЛИТЕРАТУРЫ

1. Селянкин В.В. Компьютерное зрение. Анализ и обработка изображений. – СПб.: Лань, 2019. – 152 с.
2. Video analytics for dairy farm operations. – URL: <https://www.cattle-care.com/> (accessed: 30.08.2023).
3. Koç E., Türkoğlu M. Forecasting of medical equipment demand and outbreak spreading based on deep long short-term memory network: the COVID-19 pandemic in Turkey // *Signal, Image and Video Processing*. – 2022. – Vol. 16 (3). – P. 613–621. – DOI: 10.1007/s11760-020-01847-5.
4. Progress and opportunities to advance clinical cancer therapeutics using tumor dynamic models / R. Bruno, D. Bottino, D.P. de Alwis, A.T. Fojo, J. Guedj, C. Liu, K.R. Swanson, J. Zheng, Y. Zheng, J.Y. Jin // *Clinical Cancer Research*. – 2020. – Vol. 26 (8). – P. 1787–1795. DOI: 10.1158/1078-0432.CCR-19-0287.
5. Support to early clinical decisions in drug development and personalised medicine with checkpoint inhibitors using dynamic biomarker-overall survival models / R. Bruno, P. Chanu, M. Kågedal, F. Mercier, K. Yoshida, J. Guedj, C. Li, U. Beyer, J.Y. Jin // *British Journal of Cancer*. – 2023. – P. 1–6. – DOI: 10.1038/s41416-023-02190-5.
6. Применение искусственного интеллекта при оптимизации орошения и применении гербицидов / А.Ю. Федосов, А.М. Меньших, В.А. Фартуков, М.И. Зборовская, Д.М. Васильев // *Экономика строительства*. – 2023. – № 2. – С. 42–51.
7. Федосов А.Ю., Меньших А.М. Внедрение искусственного интеллекта в растениеводство для оптимизации орошения // *Сельскохозяйственные машины и технологии*. – 2022. – Т. 16, № 4. – С. 45–53.
8. Сурай Н.М., Кудинова М.Г. Внедрение цифровых технологий в молочном животноводстве // *Парадигма устойчивого развития агропромышленного комплекса в условиях современных реалий: материалы Международной научно-практической конференции, посвященной 70-летию создания ФГБОУ ВО Красноярский ГАУ*. – Красноярск, 2022. – С. 180–186.
9. Мохов А.Ю., Абезин Д.А. Правовые аспекты использования технологий искусственного интеллекта в целях обеспечения продовольственной безопасности страны // *Аграрное и земельное право*. – 2022. – № 12 (216). – С. 97–100.
10. Deep residual learning for image recognition / K. He, X. Zhang, S. Ren, J. Sun // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – Las Vegas, NV, USA, 2016. – P. 770–778. – DOI: 10.1109/CVPR.2016.90.
11. DeepHeart: semi-supervised sequence learning for cardiovascular risk prediction / B. Ballinger, J. Hsieh, A. Singh, N. Sohoni, J. Wang, G. Tison, G. Marcus, J. Sanchez, C. Maguire, J. Olgin, M. Pletcher // *Proceedings of the AAAI Conference on Artificial Intelligence*. – 2018. – Vol. 32 (1). – DOI: 10.1609/aaai.v32i1.11891.
12. Лукашевич М.М. Цифровая обработка изображений и распознавание образов. – Минск: БГУИР, 2023. – 72 с.
13. Badrinarayanan V., Kendall A., Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling. – 2015. – arXiv:1511.00561. – DOI: 10.48550/arXiv.1511.00561.
14. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs / L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2018. – Vol. 40 (4). – P. 834–848. – DOI: 10.1109/tpami.2017.2699184. – PMID: 28463186.
15. Ronneberger O., Fischer P., Brox T. U-Net: convolutional networks for biomedical image segmentation // *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. – Cham: Springer, 2015. – P. 234–241. – (Lecture Notes in Computer Science; vol. 9351). – DOI: 10.1007/978-3-319-24574-4\_28.

16. Redmon J., Farhadi A. YOLOv3: An incremental improvement: report. – University of Washington, 2018. – URL: <https://tethys.pnnl.gov/publications/yolov3-incremental-improvement> (accessed: 30.08.2023).
17. Olafenwa M. ImageAI: Object Detection. – URL: <https://github.com/OlafenwaMoses/ImageAI/tree/master/imageai/Detection> (accessed: 30.08.2023).

*Черкашин Егор Александрович*, аспирант, младший научный сотрудник лаборатории автоматизации научных исследований Федерального государственного бюджетного учреждения науки «Санкт-Петербургский федеральный исследовательский центр Российской академии наук». Основное направление научных исследований – методы нейросетевого управления и прогнозирования биологических объектов, методы компьютерного зрения. Имеет 3 печатные работы. E-mail: [egor@cea7.ru](mailto:egor@cea7.ru)

*Cherkashin Egor A.*, a PhD student, junior researcher at the laboratory of research automation in the St. Petersburg Federal Research Center of the Russian Academy of Sciences. The main field of his scientific research is methods of neural network control and prediction of biological objects, methods of computer vision. He has 3 publications. E-mail: [egor@cea7.ru](mailto:egor@cea7.ru)

DOI: 10.17212/2782-2001-2023-3-69-86

### ***Application of neural networks for the classification of objects in video data in the tasks of automatic monitoring of the condition of biological objects\****

*E.A. CHERKASHIN*

*St. Petersburg Federal Research Center of the Russian Academy of Sciences, 39 V.O., 14th line, 199178, Saint-Petersburg, Russian Federation*

*[egor@cea7.ru](mailto:egor@cea7.ru)*

#### **Abstract**

A steady trend towards the development of technologies in everyday life and industry leads to their constant complication, to the introduction of new developments and methods, new ways of acquiring and processing information. Video surveillance systems have become an integral part of everyday life. Cameras are installed in apartment buildings, public places, even in apartments and private houses. However, one of the most important areas where computer vision has found significant application is the observation of biological objects and analysis of their behavior and state, especially in the context of nature protection and automation of animal behavior research. With the growth of quality and quantity of video materials, the problem of accurate and efficient methods for classifying objects in real time becomes relevant, especially when it comes to automatic monitoring of animal condition. There can be no mistakes here, so it is necessary to use only the latest technology. This paper describes the main types of object recognition, considers popular modern neural network architectures, and performs a comparative analysis of some of them in relation to the problem to be solved. In the course of work using methods of additional processing and image annotators SuperVisely and VGG Image Annotator, a dataset with more than a thousand unique images for extraction of relevant characteristics of objects was formed, and experimental studies of the quality of object recognition on video using pre-trained known neural network models were carried out. The requirements to the initial data for effective solution of the problem of automatic monitoring and prediction of the condition of biological objects have

---

\* Received 15 June 2023.

The research is supported by the grant of the Russian Science Foundation No. 23-19-20081, <https://rscf.ru/en/project/23-19-20081/>, and St. Petersburg Science Foundation.



been defined and formulated. It is shown that in order to avoid the appearance of blind zones in the animal habitat it is necessary to use a sufficiently large number of cameras placed with overlapping of the investigated space so that the monitoring objects were constantly in the field of visibility. This will subsequently make it possible to compile an overall high-resolution picture made up of images from all cameras and on the basis of the obtained picture to classify objects using artificial neural networks.

**Keywords:** artificial neural networks, computer vision, neural network prediction, image dataset, data augmentation, image classification, object localization, area segmentation, video data annotation

## REFERENCES

1. Selyankin V.V. *Komp'yuternoe zrenie. Analiz i obrabotka izobrazhenii* [Computer vision. Analysis and processing of images]. St. Petersburg, Lan' Publ., 2019. 152 p.
2. *Video analytics for dairy farm operations*. Available at: <https://www.cattle-care.com/> (accessed 30.08.2023).
3. Koç E., Türkoğlu M. Forecasting of medical equipment demand and outbreak spreading based on deep long short-term memory network: the COVID-19 pandemic in Turkey. *Signal, Image and Video Processing*, 2022, vol. 16 (3), pp. 613–621. DOI: 10.1007/s11760-020-01847-5.
4. Bruno R., Bottino D., Alwis D.P. de, Fojo A.T., Guedj J., Liu C., Swanson K.R., Zheng J., Zheng Y., Jin J.Y. Progress and opportunities to advance clinical cancer therapeutics using tumor dynamic models. *Clinical Cancer Research*, 2020, vol. 26 (8), pp. 1787–1795. DOI: 10.1158/1078-0432.CCR-19-0287.
5. Bruno R., Chanu P., Kågedal M., Mercier F., Yoshida K., Guedj J., Li C., Beyer U., Jin J.Y. Support to early clinical decisions in drug development and personalized medicine with checkpoint inhibitors using dynamic biomarker-overall survival models. *British Journal of Cancer*, 2023, pp. 1–6. DOI: 10.1038/s41416-023-02190-5.
6. Fedosov A.Y., Menshikh A.M., Fartukov V.A., Zborovskaya M.I., Vasiliev D.M. Primenenie iskusstvennogo intellekta pri optimizatsii orosheniya i primeneni gerbitsidov [The use of artificial intelligence in the optimization of irrigation and the use of herbicides]. *Ekonomika stroitel'stva = Economics of Construction*, 2023, vol. 2, pp. 42–51.
7. Fedosov A.Yu., Menshikh A.M. Vnedrenie iskusstvennogo intellekta v rastenievodstvo dlya optimizatsii orosheniya [Implementation of artificial intelligence in agriculture to optimize irrigation]. *Sel'skokhozyaistvennyye mashiny i tekhnologii = Agricultural Machines and Technologies*, 2022, vol. 16, no. 4, pp. 45–53.
8. Surai N.M., Kudinova M.G. [Introduction of digital technologies in dairy farming]. Materials of the International Scientific and Practical Conference dedicated to the 70th anniversary *Paradigma ustoychivogo razvitiya agropromyshlennogo kompleksa v usloviyakh sovremennykh realii* [Paradigm of sustainable development of agroindustrial complex in the conditions of modern realities]. Materials of the International Scientific and Practical Conference, Krasnoyarsk, 2022, pp. 180–186. (In Russian).
9. Mokhov A.Yu., Abezin D.A. Pravovye aspekty ispol'zovaniya tekhnologii iskusstvennogo intellekta v tselyakh obespecheniya prodovol'stvennoi bezopasnosti strany [Legal aspects of the use of artificial intelligence technologies to ensure the food security of the country]. *Agrarnoe i zemel'noe pravo = Agrarian and Land Law*, 2022, no. 12 (216), pp. 97–100. DOI: 10.47643/1815-1329\_2022\_12\_97.
10. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.
11. Ballinger B., Hsieh J., Singh A., Sohoni N., Wang J., Tison G., Marcus G., Sanchez J., Maguire C., Olgin J., Pletcher M. DeepHeart: semi-supervised sequence learning for cardiovascular risk prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018, vol. 32 (1). DOI: 10.1609/aaai.v32i1.11891.
12. Lukashovich M.M. *Tsifrovaya obrabotka izobrazhenii i raspoznavanie obrazov* [Digital image processing and pattern recognition: manual]. Minsk, BGUIR Publ., 2023. 72 p.
13. Badrinarayanan V., Kendall A., Cipolla R. *SegNet: a deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling*. 2015, arXiv:1511.00561. DOI: 10.48550/arXiv.1511.00561.

14. Chen L.C., Papandreou G., Kokkinos I., Murphy K., Yuille A.L. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, vol. 40 (4), pp. 834–848. DOI: 10.1109/tpami.2017.2699184. PMID: 28463186.
15. Ronneberger O., Fischer P., Brox T. U-Net: convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham, Springer, 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4\_28.
16. Redmon J., Farhadi A. *YOLOv3: An incremental improvement*. Report. University of Washington, 2018. Available at: <https://tethys.pnnl.gov/publications/yolov3-incremental-improvement> (accessed 30.08.2023).
17. Olafenwa M. *ImageAI: Object Detection*. Available at: <https://github.com/OlafenwaMoses/ImageAI/tree/master/imageai/Detection> (accessed 30.08.2023).

Для цитирования:

Черкашин Е.А. Применение нейронных сетей для классификации объектов в видеоданных в задачах автоматического мониторинга состояния биологических объектов // Системы анализа и обработки данных. – 2023. – № 3 (91). – С. 69–86. – DOI: 10.17212/2782-2001-2023-3-69-86.

For citation:

Cherkashin E.A. Primenenie neironnykh setei dlya klassifikatsii ob"ektov v videodannykh v zadachakh avtomaticheskogo monitoringa sostoyaniya biologicheskikh ob"ektov [Application of neural networks for the classification of objects in video data in the tasks of automatic monitoring of the condition of biological objects]. *Sistemy analiza i obrabotki dannykh = Analysis and Data Processing Systems*, 2023, no. 3 (91), pp. 69–86. DOI: 10.17212/2782-2001-2023-3-69-86.