

Применение ИИ в обработке видеоматериалов наблюдений за социальным поведением белогрудых ежей

Научный руководитель – Рutowская Марина Владимировна

Щербакова Юлия Сергеевна

Студент (бакалавр)

Российский государственный аграрный университет МСХА имени К.А. Тимирязева,
Зоотехнии и биологии, Зоологии, Москва, Россия

E-mail: yulischerbakova@yandex.ru

При анализе видеоматериалов с записью поведения белогрудых ежей (*Erinaceus roumanicus* Barrett-Hamilton, 1900) в открытых вольерах мы столкнулись с проблемой большого числа «пустых» видео и нехваткой времени на их просмотр. Для решения этой проблемы мы были вынуждены создать собственный продукт на базе нейросети с открытым исходным кодом YOLO [1].

Цель работы – разработка модели на базе технологии машинного обучения для распознавания белогрудых ежей на видеоматериалах и оценка её эффективности.

Для распознавания объектов на изображении с помощью технологии машинного обучения мы использовали свёрточную нейронную сеть от компании Ultralytics YOLOv11 [2]. Для первичного обучения мы отобрали 33 видеофайла записи ежей в разных вольерах, в разных ситуациях и в разное время суток. Далее видео загружали в программное обеспечение CVAT [3] и производили разметку датасета на наиболее «интересных» кадрах, а затем его выгрузку в формате YOLO. Дополнительно в датасет добавляли фоновые кадры. С помощью интерфейса Ultralytics запускали обучение нейронной сети (60 эпох) для предобученной модели YOLO11m. Для проверки модели использовали 781 видео с одного полигона. Нейросеть обрабатывала кадры, фиксируя координаты объектов. Результаты экспортировались в видеоредактор ShotCut для оценки и подсчёта активности ежей в Excel. Ошибки модели анализировали вручную: сложные кадры снова размечали в CVAT, объединяли с 30% исходных данных, и модель дообучалась повторно [3]. Для оценки эффективности каждой модели по отбору кадров в видеофайлах с ежами мы рассчитали точность работы ИИ по формуле $(TN/(TN+FP))*100\%$, затем – эффективность его работы по формуле $(TN/54)*100\%$ [1], где TN – число видео, которые ИИ распознал как «пустые», FP – не распознал как «пустые».

Мы получили работающую модель, использование которой позволит резко сократить время на просмотр видеоматериалов с записью поведения ежей в несколько раз. Анализ эффективности работы модели показал, что при использовании второй модели будет ошибка в 9% (не отмеченные ежи на видеозаписи), которой можно будет пренебречь. При сравнении трёх версий моделей, созданных с помощью ИИ, наиболее точной и эффективной оказалась вторая (табл. 1А): её точность работы выросла на 55%, а эффективность на 26%, по сравнению с первой. При этом третья версия оказалась ещё менее эффективной и точной, в ней встречались «ложные срабатывания» (табл. 1Б). Все версии хорошо детектируют ежей, но точность снижается при увеличении «шума» от ИИ.

Источники и литература

- 1) Chan A.H.H., Putra P., Schupp H., Köchling J., Straßheim J., Renner B., Kano F. YOLO-Behaviour: A simple, flexible framework to automatically quantify animal behaviours from videos // *Methods Ecol. Evol.* 2025. V. 16. P. 760–774.

- 2) COCO (Common Objects in Context): <https://cocodataset.org>
- 3) CVAT: <https://www.cvat.ai>

Иллюстрации

Таблица 1А

Показатели	1 версия ИИ	2 версия ИИ	3 версия ИИ
TN	6	17	4
FP	14	3	16
Общее число «пустых» видео	20	20	20
Общее число видео	54	54	54
Точность, %	30	85	20
Эффективность, %	11	37	7

Таблица 1Б

Тип ошибки	1 версия		2 версия		3 версия	
	Число видео с данным типом ошибки	Доля видео с ошибками от общего количества видео, %	Число видео с данным типом ошибки	Доля видео с ошибками от общего количества видео, %	Число видео с данным типом ошибки	Доля видео с ошибками от общего количества видео, %
Птицы	13	24,1	10	18,5	12	22,2
Ложное срабатывание	25	46,3	6	11,1	21	38,9
Ложная встреча	2	3,7	2	3,7	3	5,5
Неотмеченное взаимодействие	3	5,5	0	0	0	0
Неправильное количество ежей	0	0	1	1,9	0	0
Неотмеченные ежи	12	22,2	5	9	3	5,5
Общее число видео	54	-	54	-	54	-

Рис. : Таблица 1А: Сравнение трех версий ИИ (TN – истинно отрицательные, которые ИИ распознал как «пустые»*, FP – ложноположительные, ИИ не распознал как «пустые»)*Под «пустым» видео мы понимаем запись, на которой нет ни одного выхода ежей. Таблица 1Б: Сравнение ошибок, выдаваемых нейросетью, в каждой из версий