

УДК 004.932

ФОРМИРОВАНИЕ ОБУЧАЮЩЕЙ ВЫБОРКИ ДЛЯ НЕЙРОННЫХ СЕТЕЙ ПРИ ПРОСТРАНСТВЕННО-ВРЕМЕННОМ ОБНАРУЖЕНИИ ДЫМА НА ВИДЕО



Е.Р. Адамовский

Старший преподаватель кафедры вычислительных систем и сетей Полоцкого государственного университета имени Евфросинии Полоцкой
e.adamovsky@psu.by



Р.П. Богуш

Заведующий кафедрой вычислительных систем и сетей Полоцкого государственного университета имени Евфросинии Полоцкой, доктор технических наук, профессор
r.bogush@psu.by

Е.Р. Адамовский

Окончил Полоцкий государственный университет имени Евфросинии Полоцкой. Область научных интересов: компьютерное зрение, машинное обучение, обработка изображений и видео.

Р.П. Богуш

Область научных интересов: обработка изображений и видео, интеллектуальные системы, машинное обучение.

Аннотация. В работе представлена технология формирования набора данных для обучения и тестирования моделей искусственных нейронных сетей, выполняющих пространственно-временной анализ кадров при детектировании дыма на видеопоследовательности. Предложена модификация формата представления последовательностей изображений и метаданных для данной задачи, обеспечивающая совместимость с форматом СОСО. Отличием предложенного подхода является также формирование синтетической выборки путем совмещения системы частиц и фона с автоматизированным аннотированием объектов.

Ключевые слова: детектирование дыма, компьютерное зрение, синтезированные изображения, машинное обучение.

Введение. Развитие искусственного интеллекта и увеличение производительности вычислительных устройств позволяет на практике применять искусственные нейронные сети (ИНС) для задач, которые ранее решались другими способами менее эффективно. Так, с целью детектирования пожара на ранней стадии и, в частности, дыма, являющегося одним из его признаков, используются физические датчики, которые обладают рядом недостатков и ограниченной областью применения, например, они неэффективны на открытых пространствах. При наличии высокопроизводительного аппаратного обеспечения, анализ видеозаписей может выполняться с помощью моделей ИНС. Их преимущество заключается в способности автоматически выделять из анализируемых данных наиболее релевантные признаки, необходимые для эффективного детектирования дыма в различных условиях съемки.

Известно, что важным фактором при обучении моделей ИНС является объем и разнообразие обучающей выборки. В таблице 1 приведены датасеты, которые содержат видеозаписи или последовательности кадров с областями дыма и их характеристики: количество видеозаписей N , их разрешение, средняя длительность t_{cp} , кадровая частота

(frames per second, FPS), а также содержание. Указаны способы аннотирования: классификация, локализация (ограничивающие рамки объектов); frame-level (аннотации на уровне отдельных кадров), video-level (аннотации на уровне видеозаписей).

Таблица 1. Наборы данных с областями дыма для обучения моделей ИНС

Датасет	N	Разрешение, пк	t_{cp} , с	FPS	Аннотирование	Содержание
FiSmo (FireVid) [1]	27	от 320×240 до 600×336	120	от 15 до 30	frame-level классификация	записи пожаров
RESCUER [1]	61	от 320×240 до Full HD	20	от 15 до 30		учения пожарных
VisiFire [2]	33	от 320×240 до 720×576	90	от 7 до 30	-	записи пожаров
D-Fire [3]	100	от HD до FullHD	30	2	-	городские пейзажи, записи пожаров
Mivia Smoke [4]	149	292×240	800	от 8 до 17	video-level классификация	дым, облака, солнце
FireSense [5]	49	от 320×240 до 640×360	35	от 10 до 30	video-level классификация	записи пожаров
FURG [6]	23	до HD	40	30	frame-level локализация	записи пожаров
HPWREN [7]	470	до 4K	-	-	frame-level классификация	кадры системы отслеживания пожаров
RISE [8]	12 567	180×180	360	~0,1	video-level классификация	записи промышленных выбросов
KMU FIRE & SMOKE [9]	38	320×240	115	24	video-level классификация	записи пожаров
M4SFWD [10]	28	от 776×452 до 1480×684	95	30	frame-level локализация	синтезированный дым, лесные пожары

Анализ таблицы 1 свидетельствует, что значительная часть датасетов не содержит поккадровых аннотаций. За исключением HPWREN, разрешение видеозаписей не превышает 1920 пк по ширине, а в большинстве случаев составляет меньшие значения, порядка 320×240 пк и ниже, до 180×180 пк в RISE. Почти все датасеты основаны на реальных видеозаписях, но также представлен M4SFWD с синтезированными изображениями дыма и фона. В FiSmo и FURG дым объединяют с пламенем в общий класс «пожар» (fire). Форматы аннотирования в датасетах различны, и требуют приведения к единой форме с целью возможности их совместного использования.

Для рассмотренных датасетов произведена оценка разнообразия сцен, при этом отдельной сценой считалась уникальная комбинация места, времени и условий съемки, которая визуально отличалась от других видеозаписей в датасете. Анализ показал, что крупные наборы характеризуются низким разнообразием, например, в Mivia Smoke видеозаписи сняты в одном месте, в RISE представлено всего три локации, а HPWREN содержит сцены природных пожаров в одном регионе с однообразным ландшафтом. В наборах с менее чем 100 видеозаписями от 50% до 80% сцен являются уникальными, при этом выборка может быть ограничена собственной тематикой, например, в RESCUER представлены только видеозаписи тушения промышленных объектов. Суммарное

количество кадров на датасет, с учетом длительности видеозаписей и их кадровой частоты, составляет десятки и сотни тысяч, но малое количество уникальных сцен в датасетах нивелирует большой объем данных, поскольку важную роль при обучении моделей ИНС имеет разнообразие обучающих примеров. Таким образом, сложность задачи формирования обучающей выборки с объектами дыма обусловлена рядом причин.

1. Дефицит видеозаписей и наборов данных. Известно, что для достижения высокой обобщающей способности моделей ИНС, тренировочных образцов должно быть больше, чем обучаемых параметров, в современных архитектурах составляющих десятки миллионов. Количество и качество доступных в открытых источниках видеозаписей, содержащих области дыма, не обеспечивают необходимый объем обучающей выборки. Большинство видеозаписей имеют низкое разрешение и получены в условиях отсутствия жесткой фиксации камеры, что затрудняет точное аннотирование областей интереса. Пожар часто фиксируется уже на поздних стадиях, когда дым занимает большую площадь кадра. В открытом доступе обнаружено ограниченное количество датасетов, которые характеризуются небольшим объемом и низким уровнем уникальности сцен, а также имеют различные форматы представления. На практике высокие требования к объему обучающей выборки могут быть снижены за счет аугментирования данных и трансферного обучения. Так, в [11] показано, что предобученные модели с 30–40 млн параметров успешно обучаются на датасете с высоким разнообразием (2731 уникальная сцена) из 12 586 кадров с пламенем и дымом. В [12] модель на основе Deformable DETR (37 млн параметров) была обучена на датасете из 10 250 изображений с дымом, и достигла точности 49,7% mAP. В [13] модель на основе RT-DETR (36 млн параметров) была обучена на датасете из 12 564 изображений, и достигла точности 96,1% AP₅₀. Как правило, для извлечения первичных признаков применяются сверточные нейронные сети, предобученные на датасетах ImageNet: ImageNet-21k (14 млн изображений, 21 000 классов) или ImageNet-1k (1,3 млн изображений, 1 000 классов). Для моделей, выполняющих пространственно-временной анализ, часто используется датасет Kinetics-400 (246 тысяч видеозаписей, 400 классов). Однако следует учитывать, что, при разработке новых архитектур, использование предобученных моделей невозможно или в значительной степени ограничено.

2. Визуальное представление. Дым является полупрозрачным объектом без четкой формы и границ, что затрудняет его локализацию. Движение дыма является ключевым признаком, позволяющим отличать его от объектов, обладающих схожим визуальным представлением, но другими динамическими свойствами: туман, облака или пыль. Как следствие, обучающая выборка должна предоставлять возможность для отслеживания динамики областей дыма во времени. Решением является объединение кадров в связанные последовательности (серии), где последний кадр является опорным, аннотирован информацией о координатах ограничивающих рамок и содержит ссылки на предыдущие кадры серии в порядке их следования. Таким образом, формирование обучающей выборки для моделей ИНС, решающих задачу детектирования дыма на видеопоследовательности на основе пространственных и временных его особенностей, требует компиляции большого количества уникальных записей и аннотирования исходных данных в едином формате, который позволяет хранить информацию о сериях кадров.

Разработка формата представления данных. Анализ показывает, что для аннотирования объектов на видеозаписях не существует единого общепринятого формата, а имеющиеся реализации специализированы, например, только для трекинга. В исследованиях применяются модификации с учетом специфики решаемых задач на основе распространенных форматов для аннотирования отдельных изображений: COCO (формат JSON), Pascal VOC (формат XML), YOLO (формат TXT), которые не эффективны для видеопоследовательностей. Поэтому предлагается модификация формата COCO, как показано на рисунке 1.

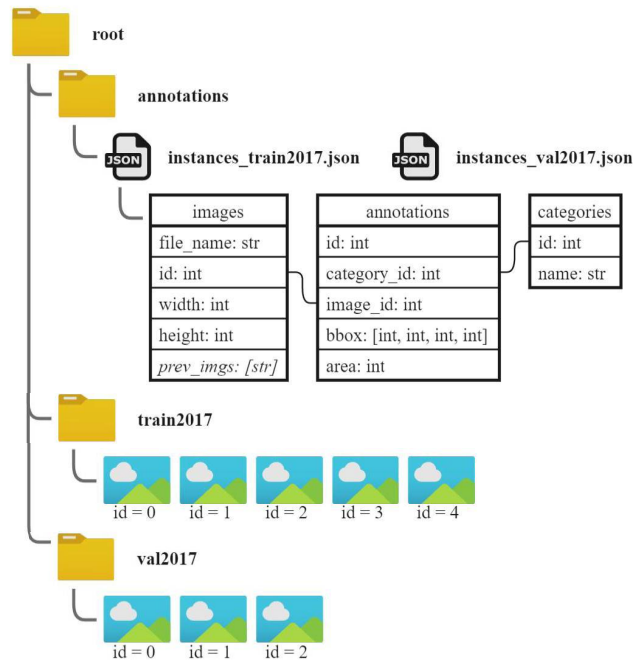


Рисунок 1. Структура формата представления кадров и метаданных в обучающем наборе

Формат предполагает наличие директории «annotations» для хранения двух JSON-файлов с аннотациями для тренировочной и валидационной выборки, «instances_train2017.json» и «instances_val2017.json» соответственно. Файлы аннотаций содержат разделы: «images» для перечисления кадров в соответствующей директории «train2017» или «val2017», «annotations» для аннотирования кадров, «categories» для определения списка детектируемых классов. В раздел «images» включено поле «prev_imgs», которое содержит относительные пути к предыдущим кадрам серии в порядке убывания их порядкового номера в серии. Предложенный подход допускает перекрытие между сериями кадров и различный размер серии T , включая вариант $T < T_m$, где T_m – предполагаемый фиксированный размер серии, который может быть обусловлен архитектурой конкретной модели ИНС. Предлагаемая стратегия заключается в циклическом заполнении серии до размера T_m . При $T = 1$ выполняется удаление всех аннотаций текущей серии, что обусловлено отсутствием в ней временных признаков объектов дыма. Современные программные реализации моделей ИНС содержат модули загрузки данных, которые процедурно, в рамках конвейера обработки, добавляют к изображениям аугментации, что позволяет снизить занимаемый датасетом объем памяти на носителе информации. В качестве примера может быть рассмотрен модуль загрузки COCO в реализации архитектуры DETR [14], который требует лишь небольшой модификации для возможности обработки серий кадров. Следовательно, предварительная аугментация датасета не требуется, однако может включать отражение, изменение яркости и контраста, масштабирование, поворот, наложение шума и размытие.

Генерация обучающей выборки. Покадровое аннотирование видеозаписей оператором имеет высокую трудоемкость, поэтому таким способом может быть подготовлен лишь ограниченный объем данных. В ряде исследований предлагается синтезировать сцены с объектами дыма при помощи инструментов трехмерного моделирования и визуализации: Unity и Unreal Engine [10], NVIDIA Omniverse [15], Blender [16]. Окружение в моделируемых сценах предлагается формировать на основе трехмерных объектов [10], реальных фотографий [17] и кадров из видеоигр с фотореалистичной графикой [16]. Совмещение дыма и фона может производиться сразу, при генерации синтезированного дыма, либо отложено, используя альфа-канал полученных изображений или технологию хромакей (green screen) [18, 19].

Генерация синтезированных данных позволяет сформировать разнообразную выборку и автоматизировать процесс аннотирования. Однако эффективность подхода ограничена визуальными отличиями синтезированного дыма от настоящего, который является сложной дисперсной системой [17], т.е. актуальна проблема смещения распределения данных (domain shift). Следовательно, доля синтезированных кадров в обучающей выборке должна быть ограничена. Предлагаемые соотношения реальных и синтезированных кадров варьируются от 3:1 до 1:1 [15, 20]. Таким образом, решением проблемы дефицита тренировочных выборок является подготовка гибридного датасета, который содержит видеозаписи с реальным дымом и синтезированные образцы, а часть кадров не содержит объектов дыма. Негативную выборку следует разделить: сцены произвольного содержания и сцены с объектами, похожими на дым, включая облака, туман, пар и смог. Генерация синтезированных данных выполнена с помощью игрового движка Unity Engine 5 и инструмента системы частиц (Particle System), параметры которого подбирались для каждой серии кадров так, чтобы дым выглядел правдоподобно, имел небольшой размер и плотность, имитируя начальную стадию возгорания. Управляющий скрипт изменял следующие параметры: форму эмиттера дыма, его плотность, скорость и направление частиц, влияние ветра, яркость и направление освещения. Для каждой серии фиксировалось определенное количество кадров, в которых отражалась динамика развития пожара от момента появления объекта дыма. Предполагается, что модели ИНС, реализующие пространственно-временной анализ, при обучении на таких данных будут вынуждены учитывать временной контекст, необходимый для разрешения неопределенности, вызванной недостаточной информативностью пространственных признаков. Автоматизированная разметка в Unity Engine выполнялась путем проекции координат частиц с непрозрачностью выше порогового значения в экранное пространство и вычислением ограничивающей рамки на основе минимальных и максимальных значений координат по обеим осям. Фоновые изображения для каждой серии загружались из подготовленного набора данных, собранного на основе изображений из сети Интернет с произвольным содержанием: пейзажи, архитектурные объекты, интерьеры. На рисунке 2 показаны примеры с синтезированными объектами дыма, изображения кадрированы для лучшего восприятия.



Рисунок 2. Автоматически аннотированные кадры с синтезированным дымом на основе системы частиц в Unity Engine

Для аннотирования видеозаписей оператором разработано программное средство в виде десктопного приложения «Smoke Label» для операционной системы Windows. Приложение предоставляет функционал разметки произвольного количества объектов класса «дым» (smoke) на видеозаписях, используя ключевые кадры и графический интерфейс для создания и редактирования ограничивающих рамок. Отличие от аналогичных open-source инструментов, включая CVAT, Label Studio и LabelMe, заключается в том, что «Smoke Label» позволяет формировать выборки в предложенном формате. Создание обучающей выборки происходило следующим образом. Видеозаписи загружались в «Smoke Label» и автоматически разбивались на отдельные видеокadres. На выбранных кадрах создавались опорные рамки вокруг объектов дыма (рисунок 3а, б), а затем производилось автоматическое вычисление рамок в промежуточных кадрах путем линейной интерполяции координат и размеров опорных рамок (рисунок 3в).



Рисунок 3. Аннотирование объекта дыма в приложении «Smoke Label»: а) построение рамки для кадра №12; б) построение рамки для кадра №29; в) интерполированная рамка в кадре №21

На рисунке 4 показаны примеры аннотированных с помощью «Smoke Label» кадров из датасета с реальными объектами дыма, которые отмечены красными прямоугольниками.



Рисунок 4. Аннотированные кадры с реальными объектами дыма

На рисунке 5 показаны примеры кадров из датасета, которые содержат объекты, по отдельным визуальным признакам обладающие сходством с дымом.

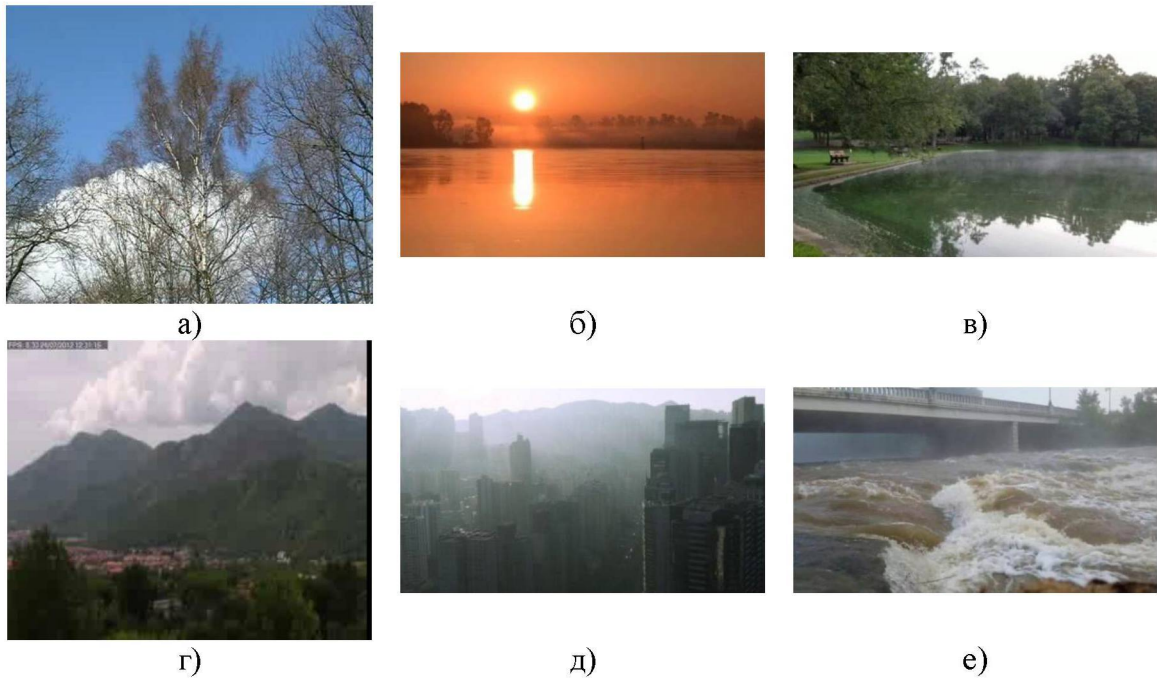


Рисунок 5. Примеры кадров с похожими на дым объектами: а) качающиеся на ветру деревья; б) солнечные блики; в) туман; г) облака; д) смог; е) водный поток и пар

Итоговый набор данных сформирован при помощи разработанных на языке Python наборов скриптов с заданными настройками, включая шаг выборки и максимальное количество кадров в серии, на основе датасетов, полученных ранее, как показано на рисунке 6.

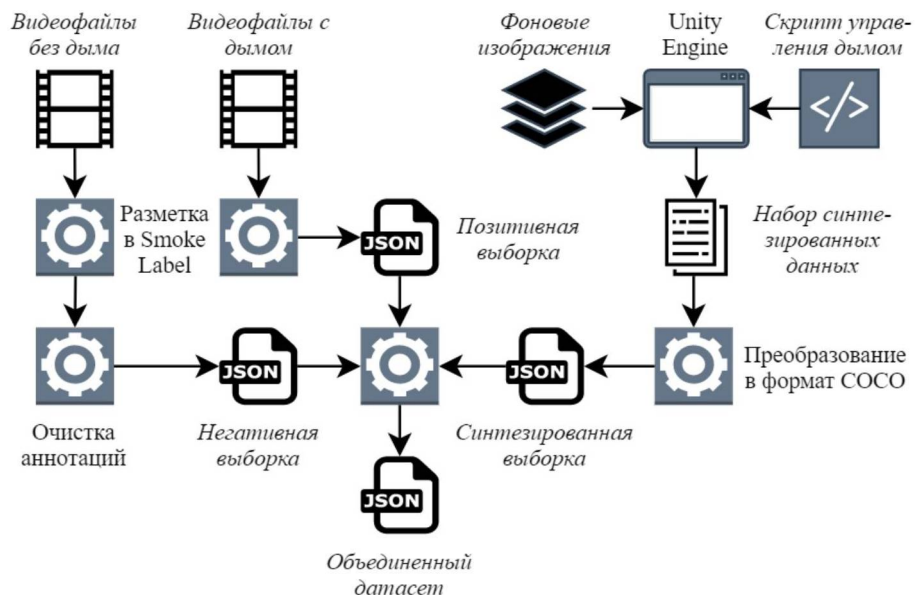


Рисунок 6. Сценарий формирования обучающей выборки

Тренировочная часть датасета (train) состоит из 36 360 серий (4 769 уникальных сцен), валидационная часть (val) состоит из 4 350 серий (475 сцен), всего 40 710 серий (5244 сцены). 9 000 серий (100 сцен), т.е. 22% от объема датасета, сформированы на основе рассмотренных в таблице 1 наборов, остальные выборки получены из собранных вручную

видеозаписей или синтезированы. Ограниченное использование материалов из существующих датасетов обусловлено тем, что для каждого набора видеозаписи были сгруппированы по представленным в них сценам, и для каждой группы формировалась только одна выборка, за счет чего достигалось максимальное разнообразие данных без доминирования отдельных сцен. Формируемые выборки имели сопоставимый друг с другом объем (от 50 до 300 серий) с целью равномерной репрезентации всех возможных условий съемки объектов дыма.

Следует отметить, что в сформированном датасете количество серий эквивалентно числу кадров, поскольку каждый кадр гарантированно является опорным в одной из представленных серий, что продемонстрировано на рисунке 7.

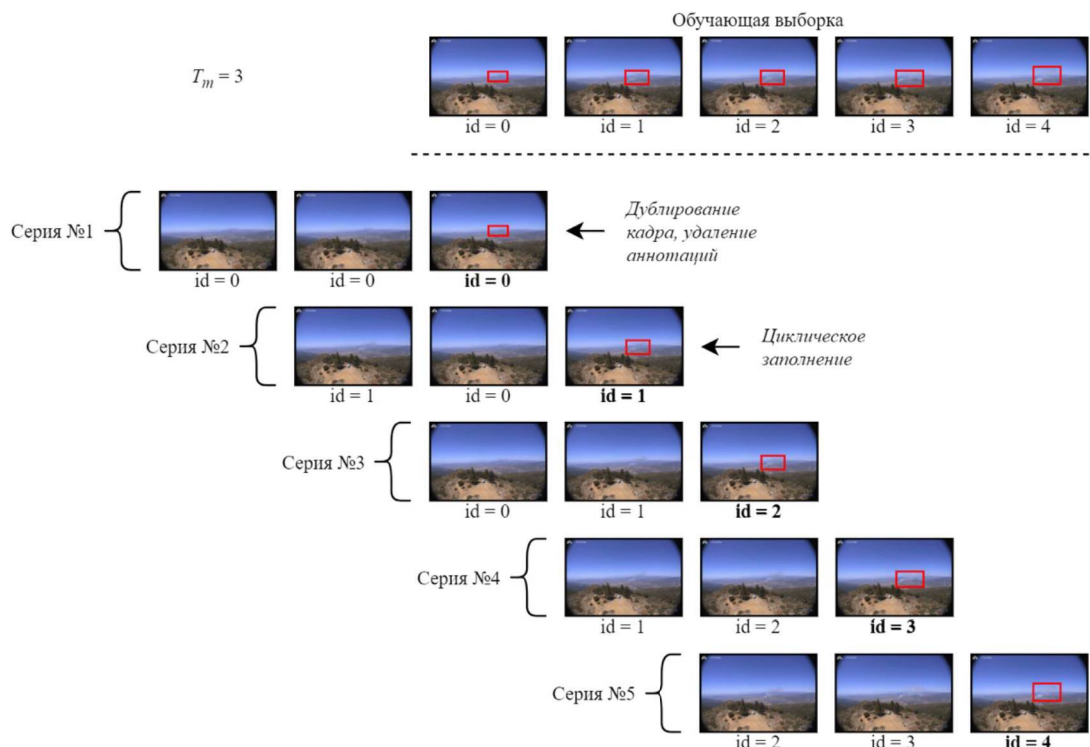


Рисунок 7. Общая схема формирования серий кадров

В тренировочных сериях 83% кадров содержат дым, выполняется соотношение реальных и синтезированных данных 3,3:1. Соотношение позитивных и негативных выборок составляет 3:1. Разделение тренировочной и валидационной выборки выполнено в пропорции 90/10, причем в валидационный набор не включены кадры с синтезированными объектами дыма. Объем занимаемой памяти обучающей выборки составил 5 ГБ.

Состав созданного датасета представлен в таблице 2.

Таблица 2. Состав датасета

№	Train		Val		Описание
	Серии	Сцены	Серии	Сцены	
1	19400	334	3600	96	Серии на основе видеозаписей с реальным дымом
2	5960	35	380	9	Серии на основе видеозаписей с похожими на дым объектами
3	4000	4000	370	370	Серии на основе видеозаписей без дыма и похожих объектов
4	7000	400	0	0	Серии с синтезированными объектами дыма

Заключение. Разработана технология формирования набора данных для обучения и тестирования моделей искусственных нейронных сетей, которые выполняют задачу детектирования объектов дыма на видеопоследовательности. Технология формирования обучающей выборки включает: сбор видеозаписей с дымом и объектами, обладающими схожими визуальными признаками; аннотирование видеозаписей с помощью специального программного средства «Smoke Label»; формирование синтетических объектов дыма на основе системы частиц в игровом движке Unity Engine; объединение полученных выборок. На основе разработанной технологии создан датасет в модифицированном формате COCO, состоящий из 40 710 серий кадров, включая выборки с реальным и синтезированным дымом, который может быть использован для обучения моделей искусственных нейронных сетей, реализующих пространственно-временной анализ видеок кадров.

Список литературы

- [1] Cazzolato, M. T. FiSmo: A Compilation of Datasets from Emergency Situations for Fire and Smoke Analysis / M. T. Cazzolato, L. P. S. Avalhais, D. Y. T. Chino [et al.] // SBBD 2017 – Proceedings of Satellite Events: Dataset Showcase Workshop (DSW). – São Paulo : SBC, 2017. – URL: <https://sbbd.org.br/2017/wp-content/uploads/sites/3/2017/10/proceedings-satellite-events-sbbd-2017.pdf>. – Дата доступа: 26.03.2026.
- [2] VisiFire [Электронный ресурс] // Bilkent University, Department of Electrical and Electronics Engineering. – URL: <http://signal.ee.bilkent.edu.tr/VisiFire/Demo>. – Дата доступа: 26.03.2026.
- [3] D-Fire Videos [Электронный ресурс]. – Режим доступа: <https://clk.li/BhNw>. – Дата доступа: 26.03.2026.
- [4] Fiore, E. M. et al. A video dataset for smoke detection based on deep learning / E. M. Fiore, G. Fenu, M. Marras, R. Pellicanò // Journal of Visual Communication and Image Representation. – 2020. – Vol. 72. – P. 102857.
- [5] chrisfilo. FireSense: Dataset for Fire and Smoke Detection [Электронный ресурс] // Kaggle. – URL: <https://www.kaggle.com/datasets/chrisfilo/firesense>. – Дата доступа: 26.03.2026.
- [6] Hogan, I. Fire Warn: Fire Hazards Detection Using Deep Learning Models / I. Hogan, D. Qiao, R. Luo [et al.] // 2021 IEEE Third International Conference on Cognitive Machine Intelligence (CogMI) : proceedings. – Atlanta, GA, USA : IEEE, 2021. – PP. 1-10. – DOI: 10.1109/CogMI52975.2021.00010.
- [7] Yazdi, A. Nemo: An Open-Source Transformer-Supercharged Benchmark for Fine-Grained Wildfire Smoke Detection / A. Yazdi, H. Qin, C. B. Jordan [et al.] // Remote Sensing. – 2022. – № 14(16). – Art. 3979. – DOI: 10.3390/rs14163979.
- [8] Hsu, Y. C. Project RISE: Recognizing Industrial Smoke Emissions / Y. C. Hsu, T. H. Huang, T. Y. Hu [et al.] // Proceedings of the AAAI Conference on Artificial Intelligence. – 2021. – № 35(17). – PP. 14813-14821. – DOI: 10.1609/aaai.v35i17.17777.
- [9] Gupta, T. Early Wildfire Smoke Detection in Videos / T. Gupta, H. Liu, B. Bhanu // 2020 International Conference on Pattern Recognition (ICPR) : proceedings. – 2021. – PP. 8523-8530. – DOI: 10.1109/ICPR48806.2021.9413231.
- [10] Wang, G. Multiple scenarios, multiple weather conditions, multiple lighting conditions and multiple wildfire objects: Synthetic Forest Wildfire Dataset (M4SFWD) [Электронный ресурс] / G. Wang // IEEE Dataport. – 2024. – DOI: 10.21227/m9kz-bw61. – Дата доступа: 26.03.2026.
- [11] Han, X. Benchmarking Multi-Scene Fire and Smoke Detection / X. Han, N. Pu, Z. Feng [et al.] // Chinese Conference on Pattern Recognition and Computer Vision (PRCV) : proceedings. – Singapore : Springer Nature Singapore, 2024. – PP. 203-218. – DOI: 10.1007/978-981-97-8795-1_14.
- [12] Huang, J. A Small-Target Forest Fire Smoke Detection Model Based on Deformable Transformer for End-to-End Object Detection / J. Huang, J. Zhou, H. Yang [et al.] // Forests. – 2023. – № 14(1). – Art. 162. – DOI: 10.3390/f14010162.
- [13] Yang, L. Real-Time Smoke Detection in Surveillance Videos Using an Enhanced RT-DETR Framework with Triplet Attention and HS-FPN / L. Yang, Y. Cheng, F. Xu [et al.] // Fire. – 2024. – № 7(11). – Art. 387. – DOI: 10.3390/fire7110387.
- [14] Carion, N. End-to-End Object Detection with Transformers / N. Carion, F. Massa, G. Synnaeve [et al.] // European Conference on Computer Vision (ECCV) : proceedings. – Cham : Springer, 2020. – PP. 213-229. – DOI: 10.1007/978-3-030-58452-8_13.
- [15] Seefried, E. Balancing Quality and Quantity: The Impact of Synthetic Data on Smoke Detection Accuracy in Computer Vision / E. Seefried, C. Jung, J. Fitzgerald [et al.] // CVPR 2024 Workshop SynData4CV : proceedings. – 2024. – URL: <https://openreview.net/forum?id=ckhhCcGq9n>. – Дата доступа: 26.03.2026.
- [16] Mao, J. Wildfire Smoke Classification Based on Synthetic Images and Pixel- and Feature-Level Domain Adaptation / J. Mao, C. Zheng, J. Yin [et al.] // Sensors. – 2021. – № 21(23). – Art. 7785. – DOI: 10.3390/s21237785.
- [17] Xu, G. Deep Domain Adaptation Based Video Smoke Detection using Synthetic Smoke Images / G. Xu, Y. Zhang, Q. Zhang [et al.] // Fire Safety Journal. – 2017. – Vol. 93. – DOI: 10.1016/j.firesaf.2017.08.004.

[18] Chen, S.-H. Real-Time Video Smoke Detection Based on Deep Domain Adaptation for Injection Molding Machines / S.-H. Chen, J.-H. Jang, M.-J. Youh [et al.] // *Mathematics*. – 2023. – № 11(17). – Art. 3728. – DOI: 10.3390/math11173728.

[19] Zhang, Q. Wildland Forest Fire Smoke Detection Based on Faster R-CNN using Synthetic Smoke Images / Q. Zhang, G. Lin, Y. Zhang [et al.] // *Procedia Engineering*. – 2018. – Vol. 211. – PP. 441-446. – DOI: 10.1016/j.proeng.2017.12.034.

[20] Seefried, E. Smoke+: A Video Dataset for Automated Fine-Grained Assessment of Smoke Opacity : thesis for the Candidate's degree in Technical Sciences / E. Seefried. – Fort Collins : Colorado State University, 2024.

Авторский вклад

Адамовский Егор Русланович – анализ предметной области, формирование обучающей выборки, проектирование специального программного обеспечения.

Богущ Рихард Петрович – постановка задачи исследования, анализ полученных результатов, корректировка рукописи статьи.

TRAINING DATASET FORMATION FOR NEURAL NETWORKS IN SPATIO-TEMPORAL SMOKE DETECTION FROM VIDEO

Y.R. Adamovskiy

*Senior Lecturer at the Department of
Computer Systems and Networks, Polotsk
State University*

R.P. Bohush

*Doctor of Technical Sciences, Professor, Head of
Computer Systems and Networks department of
Polotsk State University*

Abstract. This paper presents a dataset formation technology for training and testing neural network models performing spatio-temporal analysis of frames for smoke detection on video sequences. A modification of the image sequence and metadata representation format is proposed for this task, ensuring compatibility with the COCO format. A distinctive feature of the proposed approach is the formation of a synthetic dataset by combining a particle system with background images and automated object annotation.

Keywords: smoke detection, computer vision, synthesized images, machine learning.