

DOI: 10.51790/2712-9942-2022-3-2-7

ИССЛЕДОВАНИЕ ВОЗМОЖНОСТЕЙ СОЗДАНИЯ НЕЙРОСЕТЕВОГО ПРОГРАММНО-АППАРАТНОГО КОМПЛЕКСА ДЛЯ РАННЕГО ВИДЕООБНАРУЖЕНИЯ ВОЗГОРАНИЯ В РЕЖИМЕ РЕАЛЬНОГО ВРЕМЕНИ

И. М. Клемышев^{1,a}, С. С. Лебедев^{1,b}, С. О. Старков^{2,c}

¹ ООО «КомпВи», г. Обнинск, Российская Федерация

² Обнинский институт атомной энергетики, Национальный исследовательский ядерный
университет «МИФИ», г. Обнинск, Российская Федерация

^a ORCID: <http://orcid.org/0000-0001-7831-9474>

^b ORCID: <http://orcid.org/0000-0002-9626-0223>, compvi19@mail.ru

^c ORCID: <http://orcid.org/0000-0002-0420-7856>, sergeystarkov56@mail.ru

Аннотация: в данной статье представлен метод, направленный на решение проблемы раннего обнаружения возгорания на видеоизображениях с камер наблюдения, основанный на временном анализе подозрительной области. Представленный метод позволяет проводить сжатие рассматриваемой области видеоизображения до временного ряда, который классифицируется рекуррентной нейросетью. Временной ряд собирается из метрик, снимаемых с самой области и окружения, которые отражают ее «мерцание». Анализ «мерцания» позволяет определить, содержит ли область огонь. Предложенный алгоритм позволяет значительно снизить количество ложных срабатываний при обнаружении огня благодаря анализу видеоизображения по времени. Способ сжатия исходных данных до временного ряда из набора характеристик области и окружения позволяет использовать небольшую рекуррентную нейросеть для классификации подозрительной области вне зависимости от ее размера. Эти особенности позволяют применить данную модель при создании автономного детектора для раннего обнаружения возгорания на основе одноплатного компьютера и видеокамеры. В статье приводится описание предложенной модели и обучения нейросети, оценки качества обучения и результаты экспериментов, а также примеры работы, выполненные на одноплатном компьютере Jetson Nano от NVIDIA.

Ключевые слова: нейросетевой программно-аппаратный комплекс, видеообнаружение, возгорание, пожар.

Для цитирования: Клемышев И. М., Лебедев С. С., Старков С. О. Исследование возможностей создания нейросетевого программно-аппаратного комплекса для раннего видеообнаружения возгорания в режиме реального времени. *Успехи кибернетики*. 2022;3(2):47–59. DOI: 10.51790/2712-9942-2022-3-2-7.

DEVELOPING A NEURAL NETWORK-BASED HARDWARE AND SOFTWARE SYSTEM FOR EARLY VIDEO DETECTION OF FIRES IN REAL TIME

I. M. Klemyshev^{1,a}, S. S. Lebedev^{1,b}, S. O. Starkov^{2,c}

¹ CompV, ООО, Obninsk, Russian Federation

² Obninsk Institute For Nuclear Power Engineering, Obninsk, Russian Federation

^a ORCID: <http://orcid.org/0000-0001-7831-9474>

^b ORCID: <http://orcid.org/0000-0002-9626-0223>, compvi19@mail.ru

^c ORCID: <http://orcid.org/0000-0002-0420-7856>, sergeystarkov56@mail.ru

Abstract: this article presents a method for early fire detection using video images from surveillance cameras, based on a temporal analysis of the suspicious area. The method makes it possible to compress the area of the video image to a time series, which is classified by a recurrent neural network. The time series contains metrics measured in the area and its surroundings to consider for its flicker. The flicker analysis determines whether the area is on fire. The proposed algorithm significantly reduces the number of false fire alarms as the video image are analyzed over time. The source data compression to a time series containing the area and surrounding characteristics allows using a small recurrent neural network to classify a suspicious area regardless of its size. These features make this model applicable for building an autonomous fire detector based on a single-board computer and a video camera. The paper describes the

proposed model and the neural network training, training quality evaluation and the results of experiments, as well the results generated with the Jetson Nano single-board computer from NVIDIA.

Keywords: neural network-based software and hardware system, video detection, ignition, fire.

Cite this article: Klemyshev I. M., Lebedev S. S., Starkov S. O. Developing a Neural Network-Based Hardware and Software System for Early Video Detection of Fires in Real Time. *Russian Journal of Cybernetics*. 2022;3(2):47–59. DOI: 10.51790/2712-9942-2022-3-2-7.

Введение

Пожары до сих пор являются угрозой для жизни и имущества. В современном мире, кроме традиционных способов обнаружения возгораний с помощью различных датчиков, реагирующих на изменение температуры или химического состава воздуха, предлагаются решения на основе компьютерного зрения и нейросетевых алгоритмов. Очевидное теоретическое преимущество решений на базе компьютерного зрения перед классическими датчиками — это возможность раннего обнаружения возгорания, когда до датчиков дыма или температуры просто не доходит воздействие источника горения. А возможность раннего обнаружения возгорания — это возможность спасти жизни и имущество от пожара.

Одновременно с развитием компьютерного зрения и машинного обучения на основе нейросетей активное развитие получили технологии граничных вычислений, и на рынок вышли целые семейства одноплатных компьютеров от различных технологических гигантов. Мощности одноплатных компьютеров растут благодаря использованию графических ускорителей (GPU), а в последнее время — и специальных тензорных процессоров (TPU) для работы с многомерными матрицами.

В данной работе внимание сосредоточено на задаче создания автономного детектора на основе одноплатного компьютера, который сможет произвести раннее обнаружение возгорания внутри и снаружи помещений по видеоизображению с камер наблюдения. Один из таких компьютеров — Jetson Nano от компании NVIDIA показан на рисунке 1.

Наблюдая возможности и успехи развития сферы машинного обучения, компьютерного зрения, а также расцвет интернета вещей (IoT) и граничных вычислительных устройств (edge computing), можно было бы предположить, что задача раннего обнаружения возгорания по видеоизображению давно уже под силу разработчикам охранных систем. Однако практика применения существующих на рынке решений для обнаружения огня с помощью интеллектуального анализа видео, а также обзор предлагаемых в нашей стране и за рубежом новых методов, дает понимание о том, что универсального решения нет, задача сложнее, чем кажется на первый взгляд и решить ее для наибольшего количества сценариев пока не удается.

Существующие решения для определения огня на видеоизображении, полученном без использования специального оборудования (например, тепловизоров), а с помощью обычной цветной видеокамеры, часто ограничиваются применением сверточных нейросетей с некоторой предобработкой видеопотока. Точность таких детекторов, то есть способность находить в кадре вещи, похожие на огонь, на сегодняшний день достаточно высока и в некоторых случаях достигает близких к 100% значений. Одновременно с этим большой проблемой всех решений по огню является высокая ошибка первого рода, то есть неспособность детектора различать огонь и похожие на него объекты. Большая величина этой ошибки означает, что детектор будет давать много ложных срабатываний, замечая огонь там, где его нет. Максимальное снижение ошибки первого рода — это один из основных моментов, позволяющий перейти от просто модели, распознающей огонь на видеоизображении, к устройству, готовому к применению на практике.

Работы по теме

За последние годы глубокое обучение сделало огромный шаг вперед, во многом — в сфере компьютерного зрения и ориентирования в пространстве, благодаря новым подходам и архитектурам сверточных сетей. В ряде работ эти методы и подходы были применены к задаче раннего обнаружения возгораний на видеоизображениях. Несмотря на высокие показатели точности в нахождении огня на изображениях, эти модели не могут верифицировать подозрительную область во времени, что приводит к частым ошибкам первого рода, то есть ложным срабатываниям.

Одновременно с этим активно развивается и аппаратная часть вычислительных машин, что позволяет разрабатывать автономные устройства для построения умной и безопасной окружающей

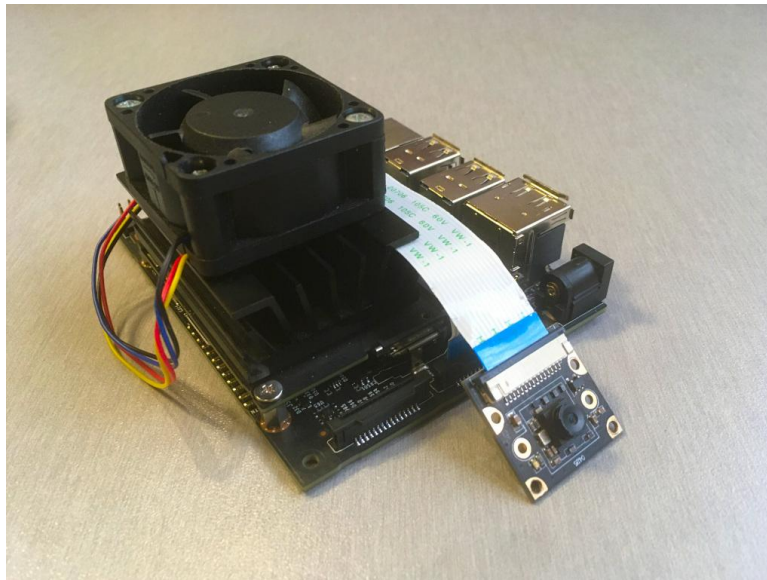


Рис. 1. Одноплатный компьютер Jetson Nano и RaspberryPI-камера (сенсор Sony IMX477)

среды. В работах [1, 2, 3] команды исследователей из разных стран демонстрируют возможности создания автономных детекторов для раннего обнаружения возгораний по видеоизображениям на базе одноплатных компьютеров. Детекторы в работах [1, 2] построены на базе одноплатного компьютера Jetson Nano, в работе [3] — на базе Banana Pi. В основе метода работы детекторов в работах [1, 2, 3, 4] — поиск источников возгорания с помощью сверточных нейросетей без анализа по времени. Однако, как было сказано выше, отсутствие верификации по времени не позволяет избежать ложных срабатываний, даже несмотря на высокие показатели точности распознавания этих моделей.

Одна из успешных работ, показывающих пространственно-временное обнаружение возгорания и задымления на видеопоследовательности, — работа китайской исследовательской команды [5]. В процессе обработки видеоизображения идет накопление кадров подозрительной области в течение некоторого времени, в результате которой данные этой области образуют трехмерный массив цветных изображений. Далее этот массив проходит классификацию с помощью нейросети на содержание дыма/огня. Такая проверка позволяет существенно уменьшить количество ложных срабатываний, но набирается большой объем данных для классификации каждой подозрительной области изображения. Это делает невозможным создание детектора на компьютерах вроде Jetson Nano, который вел бы мониторинг территории в режиме реального времени.

В 2010 году еще одна исследовательская команда [6] также прорабатывала вопрос раннего обнаружения возгорания по видео, отталкиваясь от поведения огня во времени. Их подход базируется на анализе мерцания, которое создает огонь в процессе горения. Они не анализируют область целиком как изображение, однако снимают некоторую характеристику с области и затем, анализируя изменение этой характеристики во времени, делают вывод о наличии/отсутствии огня в этой области.

Результаты работы китайских исследователей [5] показывают, что верификация подозрительных областей по времени позволяет существенно снизить ошибку первого рода при детекции огня и дыма на видеоизображениях. А работа 2010 года [6] говорит о том, что возможно существенно сократить объем данных, необходимых для того, чтобы различить огонь в поведении некоторой области изображения, отталкиваясь от его поведения как объекта.

Еще одна современная работа, демонстрирующая пространственно-временное обнаружение задымлений и возгораний [7], основана на последовательном выделении пространственных признаков подозрительных областей при помощи сверточной модели нейросети с архитектурой Faster-RCNN и классификации временного ряда собранных признаков с помощью рекуррентной нейросети с LSTM-архитектурой (Long-Short-Term-Memory). Предлагаемый подход во многом аналогичен данной работе в целом, и основное отличие — иной способ извлечения и накопления пространственных признаков для классификации области.

Другая интересная работа [8] представляет применение стереозрения для определения местопо-

ложения возгорания в трехмерном пространстве и использование расстояния от камеры до источника горения в качестве критерия для снижения уровня ложноположительных срабатываний. В перспективе любой автономный детектор возгораний и/или задымлений по видеоизображению может быть оборудован стереозрением и дополнен функционалом определения местоположения источника горения в трехмерном пространстве. Это может быть использовано для интеграции детектора раннего обнаружения возгораний, скажем, с адресной системой пожаротушения.

В предлагаемом методе предлагается сделать упор на то, чтобы верифицировать подозрительные области обязательно по времени, так как это способно уменьшить ошибку первого рода, и при этом использовать как можно меньше вычислительных мощностей для создания автономного детектора на компьютере Jetson Nano.

Гипотеза состоит в том, чтобы использовать в качестве данных определенные метрики регионов-кандидатов, собранные в течении короткого времени, и провести анализ собранных данных с помощью нейросети. Поскольку данные такого рода представляют собой временной ряд, или последовательность данных, то для задачи классификации такого ряда нужно решить задачу много-к-одному с помощью рекуррентной нейросети (RNN).

Предлагаемый метод

Предлагаемый метод направлен на создание детектора возгораний, работающего на малых вычислительных мощностях (edge computing) и при этом не совершающего таких ошибок, как, например, реагирование на статичное изображение огня в кадре, проходящего человека в красной куртке или разного рода блики, часто появляющиеся на разных поверхностях в кадре.

Общий механизм распознавания огня с помощью компьютерного зрения аналогичен тому, каким образом это делает человек: заметив что-то напоминающее огонь в поле своего зрения, переместит свое внимание в эту область для принятия решения, действительно ли был замечен огонь. Конвейер обработки видеоизображения для задачи компьютерного зрения, в целом, состоит из тех же шагов:

- 1) выделения регионов-кандидатов, то есть прямоугольных областей кадра, предположительно содержащих огонь;
- 2) анализ данных областей во времени для классификации на содержание источника горения.

Задача данного исследования — найти способ анализировать регионы-кандидаты по времени с минимальными вычислительными затратами, высокой точностью и низким значением ошибки.

Цель исследования — получить легкую нейросетевую модель для классификации регионов-кандидатов по времени на содержание источника горения (огня) в реальном времени, которая будет использоваться в конвейере автономного детектора возгораний на основе граничных вычислений.

Основное техническое новшество предлагаемой модели детектора возгораний — временная интерпретация видеоизображения, а точнее — предлагаемый способ извлечения признаков и их классификация.

Для работы с временными признаками и последовательностями данных в глубоком обучении используются рекуррентные нейросети. Это класс сетей, предназначенных для моделирования последовательных данных, таких как временные ряды или естественный язык. Задача обнаружения возгорания требует классификации временного ряда, отражающего поведение некоторой области видеоизображения во времени. Поэтому в архитектуре нейросети после рекуррентного слоя стоят два полносвязанных слоя (Dense Layer) для классификации на содержание/отсутствие огня.

Логично, что чем больше пространственных признаков каждого кадра видеоизображения будет использовано для формирования временной последовательности для классификации, тем точнее будет результат. Обратной стороной такой полноты исходных данных является, само собой, большая ресурсоемкость их обработки.

Временные признаки в общем смысле представляют собой последовательный набор данных, которые могут быть как отдельными числовыми характеристиками, так и, например, целым изображением или картой признаков либо иной структурой данных. Какой бы процесс ни описывала эта последовательность признаков, она будет иметь некоторое минимальное значение длины, ниже которого процесс предсказания не будет эффективным на этом наборе данных. Поэтому выбран более оптимальный путь минимизации размера самих величин, из которых набирается последовательность признаков, вместо сокращения длины ряда.

Предлагаемый способ получения признаков для некоторой прямоугольной области R кадра видеопоследовательности и формирования из них временного ряда состоит в следующем:

- данная область R кадра берется в качестве основной области (с масштабом $1x$);
- строятся две дополнительные прямоугольные области с центрами и ориентацией как у основной области в масштабах $2x$ и $1/2x$;
- область кадра, содержащая все три масштабные области, преобразуется в цветовое пространство HSV (hue-saturation-value);
- для трех масштабных областей находятся средние значения величин h , s и v , обозначающие компоненты оттенка, насыщенности и яркости цвета; итого — 9 числовых значений для одного региона-кандидата в кадре.

Полученные средние значения величин h , s , v для трех масштабных областей, построенных для одного региона-кандидата, берутся в качестве временного признака, и из таких наборов значений составляется ряд для последующей классификации. По объему данных один набор значений равен объему данных цветного изображения из трех пикселей. Такой способ формирования временных признаков делает процесс инвариантным к размеру региона-кандидата. Для большего региона модель будет затрачивать немного больше времени для расчета средних значений, но на выходе иметь для классификации точно такой же небольшой набор признаков, что несравнимо с исходными объемами данных. Остается вопрос, почему такая модель будет работать на распознавание огня и насколько эффективен предлагаемый метод.

Для ответа обратимся к рисунку 2, на котором показан процесс сбора временных признаков с некоторой фиксированной области изображения указанным выше способом. Для примера, показанного на рисунке 1, имеются 4 области: заключающая огонь и окружающее его пространство (1), заключающая только огонь (2) с небольшой внешней окантовкой, немного меньше огня (3), внутри огня (4). Видно, что и значения величин h , s , v , и динамика изменений этих значений отличаются для разных областей в зависимости от «масштаба» и того, что область в себе содержит. Отсюда гипотеза о том, что можно научить нейросеть с RNN-архитектурой классифицировать такие последовательности временных признаков на те, которые соответствуют наличию огня, от всех остальных.

Поэкспериментировав с масштабными областями на разных видеоизображениях, содержащих огонь, и оценив величины h , s , v , мы выбрали следующий способ расчета областей и их количества: три области, которые строятся путем масштабирования в обе стороны средней области, заключающей в себя только огонь. Выбор именно такого способа расчета масштабных областей и их количества основан на особенностях самого огня, его большой вариативности в разных условиях горения и его поведении в пространстве.

Наглядно процесс получения величин h , s , v для областей, как содержащих огонь, так и не содержащих его, показан на рис. 3–5.

На рисунке 6 изображен общий алгоритм верификации региона-кандидата на содержание огня:

Алгоритм состоит из следующих шагов:

- 1) выделение области предполагаемого нахождения цели (огня, по цвету/движению, с помощью сверточной нейросети и т.д.) и двух масштабных областей ($x2$ и $x0.5$);
- 2) преобразование изображения в цветовое пространство HSV (hue, saturation, value);
- 3) вычисление средних значений HSV для трех выделенных ранее масштабных областей;
- 4) повторение пунктов 2) и 3) для последовательности кадров (экспериментально — 32 кадра) и получение в результате входного вектора для рекуррентной нейросети из 9 сигналов (h , s , v для трех областей) по 32 значения;
- 5) инференс обученной рекуррентной нейросети для классификации выбранного участка видеопоследовательности;
- 6) полученная в результате инференса вероятностная оценка того, что верифицируемый регион-кандидат действительно содержит огонь.

Архитектура нейросети

Задача классификации последовательности данных не является новой, как и архитектуры нейросетей, предложенные для ее решения. В данной работе использована архитектура, показанная на рисунке 7.

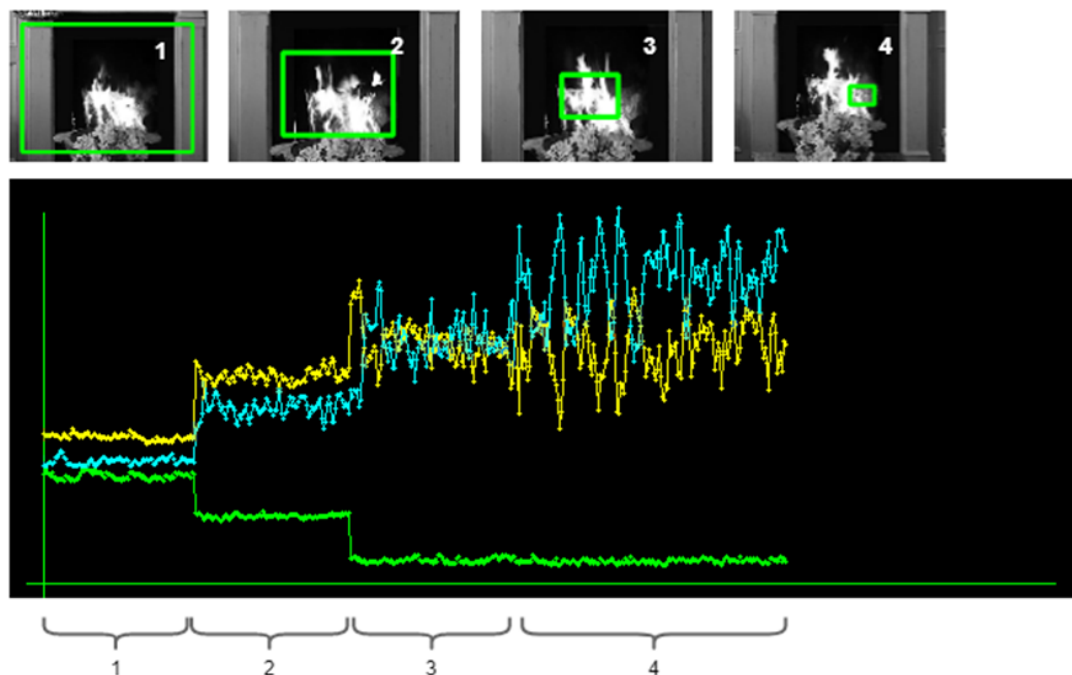


Рис. 2. Сбор данных средних величин h , s , v с областей разного масштаба, содержащих в себе огонь с разной степенью включения

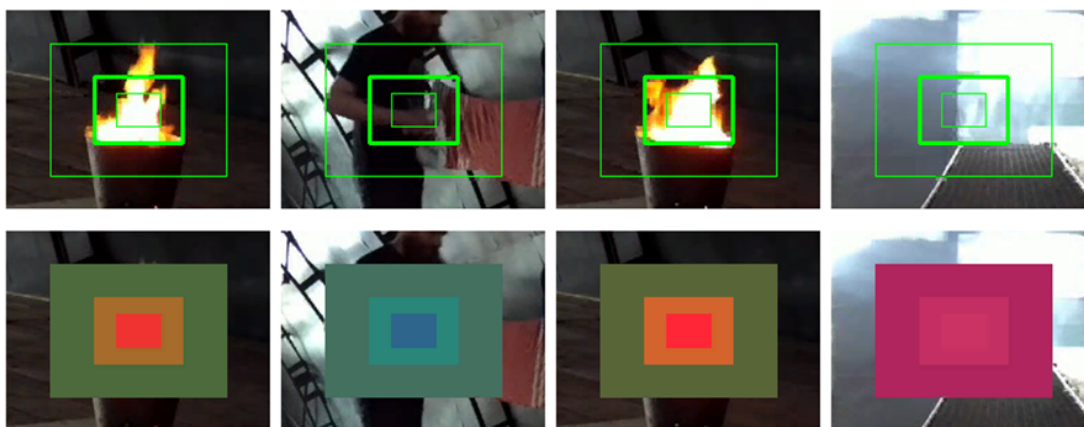


Рис. 3. Примеры масштабных областей и снимаемых с них сигналов средних величин Hue-Saturation-Value (показаны в нижнем ряду, преобразованы в пространство цвета RGB (Hue-Red, Sat-Green, Val-Blue))

Входной слой имеет размерность 32×9 , что означает 32 шага по времени и 9 значений на каждый шаг. Далее идет слой рекуррентной сети (LSTM или GRU), с входной размерностью 32×9 и восемью шагами на выходе. Для рекуррентного слоя заданы dropout и recurrent_dropout, которые способствуют лучшему обучению сети и помогают избежать переобучения за счет случайного отключения части узлов, соединяющих текущий и последующий слои. Выход рекуррентного слоя направлен на вход классификатора на основе двух полносвязанных (Dense) слоев с двумя байесовскими выходами нейросети, на которых получается результат классификации последовательности.

В процессе исследования и обучения модели произведена небольшая модификация архитектуры для сравнения результатов обучения и поиска более оптимальной конфигурации. В частности, модификации касались применения GRU-слоев вместо LSTM, а также использование dropout-ов как

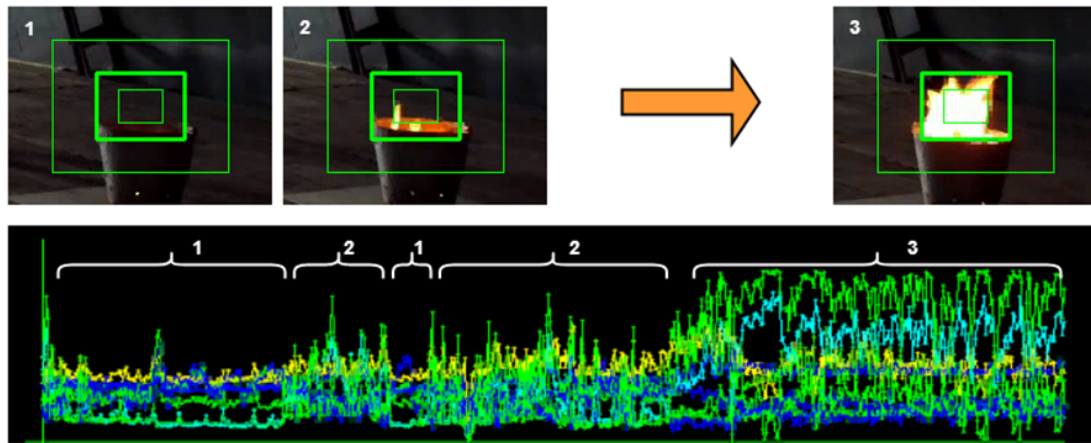


Рис. 4. Пример сигналов h , s , v , снимаемых с масштабных областей на видео, содержащих появление горения

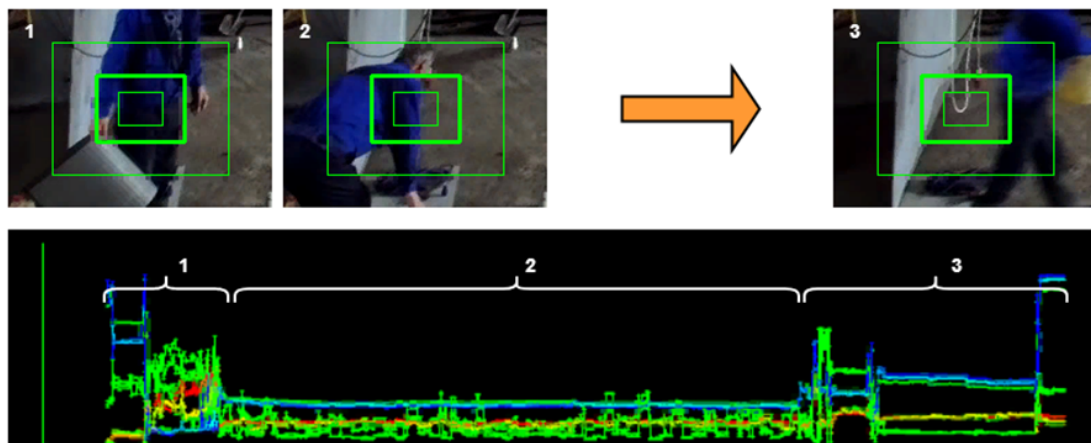


Рис. 5. Пример сигналов h , s , v , снимаемых с масштабных областей на видео, не содержащих источника горения (огня)

внутри рекуррентного слоя, так и отдельно в виде Dropout-слоя после рекуррентного.

Данные для обучения

Данные для обучения нейросети были подготовлены вручную с помощью небольших программ-утилит, написанных для этой задачи.

В лабораторных условиях были отсняты видеоролики, на которых присутствует:

- огонь на разном удалении от камеры;
- движущиеся люди;
- различные движущиеся предметы и механизмы;
- предметы, похожие на огонь, которые могут спровоцировать детектор на ошибку (естественное и искусственное освещение — дневной свет в окнах, лампы накаливания и газоразрядные).

Получить данные для обучения нейросети можно с помощью разметки видеороликов. Разметка в данном случае является преобразованием видеоизображения во временной ряд количественных характеристик, и назначение этому ряду соответствующего класса: огонь / не огонь. Для набора видеозаписей, содержащих огонь и снятых с фиксированного ракурса, человек выделяет в каждом ролике прямоугольную область, описывающую основную область пламени в кадре; при этом происходит расчет и запись усредненных значений HSV для масштабных областей (2x и 0,5x) в файл с данными. Аналогично для областей, не содержащих огонь, с той разницей, что на выходе для них ставится в

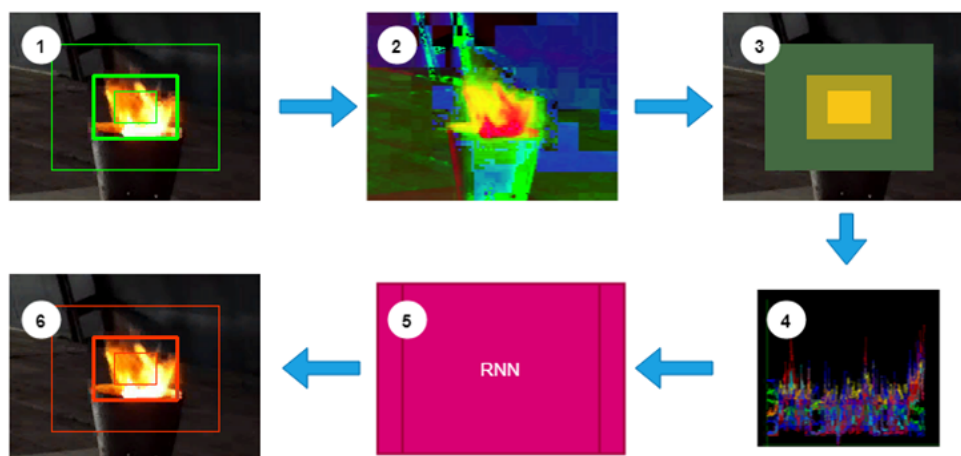


Рис. 6. Алгоритм верификации региона-кандидата на содержание огня: 1) выделение региона-кандидата и расчёт масштабных областей; 2) переход в цветовое пространство HSV; 3) расчёт средних значений Hue-Saturation-Value для масштабных областей; 4) набор временного ряда значений HSV для последующей 5) классификации с помощью рекуррентной нейросети и 6) получение результата огонь/не огонь

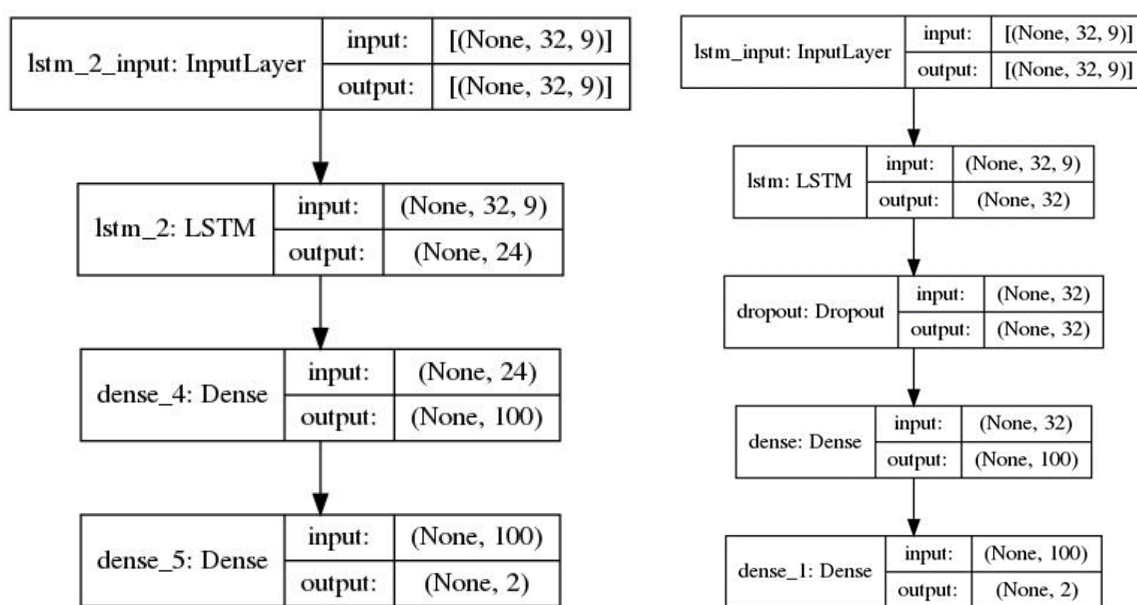


Рис. 7. Архитектура нейросети (с dropout внутри рекуррентного слоя (слева) и с отдельным Dropout слоем (справа)). Рекуррентный слой LSTM может быть заменен на GRU

соответствие класс «нет огня».

В итоге был получен следующий набор данных для обучающей и тестовой выборки:

30, 231 образцов для обучения;

10, 384 образца для тестирования;

Эксперименты и результаты

Обучение нейросети

Проведен ряд экспериментов с обучением нейросети, с небольшими изменениями как предложенной архитектуры, так и наборов данных для обучения, с целью определить наиболее подходящую конфигурацию модели и параметров обучения.

На собранных данных нейросеть хорошо обучается. Количество эпох обучения — 100. Результаты экспериментов относительно заданных условий представлены в таблице 1.

Таблица 1

Результаты обучения рекуррентной сети со слоями GRU (1.1–1.8) и LSTM (2.1–2.8). Для сравнения были составлены архитектуры с разным количеством ячеек памяти в рекуррентном слое из набора [8, 16, 24, 32, 40, 48, 56, 64] для каждого варианта рекуррентного слоя [GRU, LSTM]

Exp. №	RNN type	RNN units	Accuracy	Loss
1.1	GRU	8	0.940402	0.2399
1.2	GRU	16	0.959242	0.1842
1.3	GRU	24	0.982904	0.1022
1.4	GRU	32	0.972351	0.12024
1.5	GRU	40	0.963004	0.14234
1.6	GRU	48	0.94241	0.13233
1.7	GRU	56	0.941058	0.11323
1.8	GRU	64	0.956242	0.14234
2.1	LSTM	8	0.942045	0.2122
2.2	LSTM	16	0.95414	0.14234
2.3	LSTM	24	0.970502	0.10234
2.4	LSTM	32	0.990882	0.09742
2.5	LSTM	40	0.961082	0.10240
2.6	LSTM	48	0.932459	0.12234
2.7	LSTM	56	0.943529	0.13648
2.8	LSTM	64	0.955430	0.12234

Ниже приведены соответствующие графики истории обучения, где голубой тренд показывает параметр “accuracy”, оранжевый — параметр “loss”.

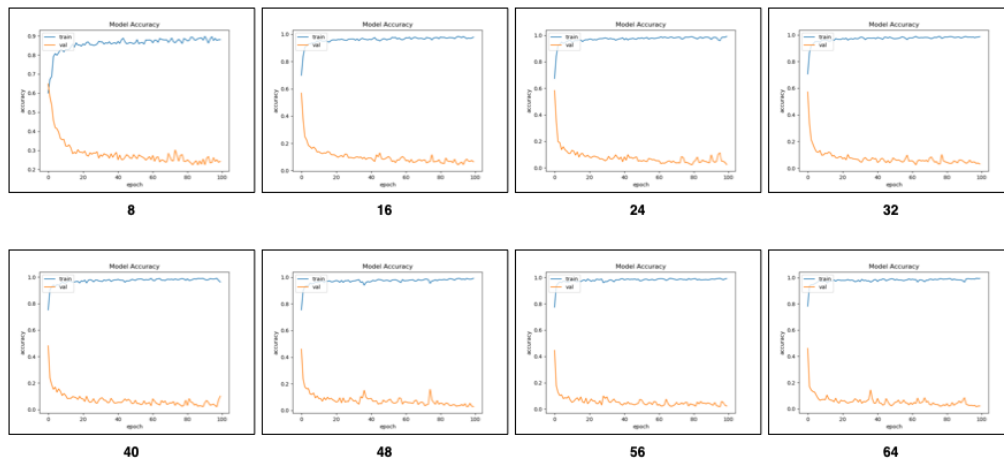


Рис. 8. Графики обучения рекуррентной нейросети со слоем LSTM с указанием количества ячеек слоя ([8, 16, 24, 32, 40, 48, 56, 64])

На фоне хорошего результата обучения для столь сильного сжатия исходных данных (степень сжатия равносильна сжатию региона-кандидата в каждом кадре до трех цветных пикселей — 9 значений средних величин пространства HSV) было решено провести дополнительный эксперимент с еще меньшим количеством входных данных, исключив из временного ряда величины hue (оттенок) для всех масштабных областей.

Если эксперимент будет удачным и сеть продолжит безошибочно отличать регионы-кандидаты, содержащие огонь, от всех других, то это будет означать, что нейросеть действительно способна отличать области, содержащие огонь по характеру того мерцания, который источник горения создает на видеоизображении, даже без учета цветовой составляющей огня. Это важный момент, т.к. горение

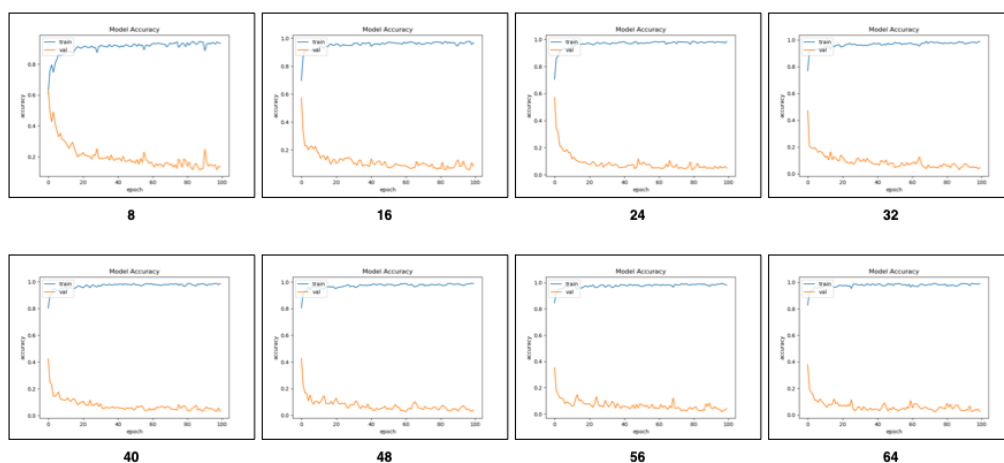


Рис. 9. Графики обучения рекуррентной нейросети со слоем *LSTM* с указанием количества ячеек слоя ([8, 16, 24, 32, 40, 48, 56, 64])

разных материалов порождает разную цветовую гамму языков пламени.

Результаты эксперимента приведены в таблице 2.

Таблица 2

Результаты обучения рекуррентной сети со слоями *GRU* (1, 3) и *LSTM* (2, 4). Для сравнения были обучены архитектуры с разным количеством ячеек памяти в рекуррентном слое [16, 32] для каждого варианта рекуррентного слоя [*GRU*, *LSTM*]

Exp №	RNN type	RNN units	Accuracy	Loss
1	GRU	32	0.971279	0.10958
2	LSTM	32	0.97736	0.11395
3	GRU	16	0.952367	0.19234
4	LSTM	16	0.942547	0.20134

Примеры работы нейросети без HUE-величины с огнем разной природы, а соответственно, и разной цветовой палитрой пламени на изображении показаны на рисунке 10.

Нейросеть обучена и протестирована на устройстве Jetson Nano от NVIDIA — одноплатном компьютере с графическим ускорителем (128 ядер Maxwell), ориентированном на ИИ (нейросети, глубокое обучение, искусственный интеллект) и ресурсоемкие расчеты.

Примеры работы

На рисунках ниже показаны примеры классификации регионов-кандидатов (подозрительных областей) на содержание или отсутствие огня. Примеры сделаны с помощью программной утилиты, которая позволяет вручную выделить подозрительную область видеоизображения, образовать от нее масштабные области по описанному в статье методу и осуществить классификацию данной области видеоизображения по времени на наличие огня.

На рисунке 10 приведены примеры подтверждения наличия огня в выделенной области.

На рисунке 11 видно, что предложенная модель «отстраивается» от объектов, похожих на огонь, то есть не дает подтверждения его наличия в области. Это и есть заявленный результат по снижению ошибки первого рода с помощью временного анализа подозрительной области.

Оценка качества модели

Для того чтобы рассчитать вероятность ложноположительного срабатывания, был использован следующий подход:

- 1) собрана коллекция видеороликов, соответствующих следующим критериям:
 - а) фиксированное положение камеры при съемке;

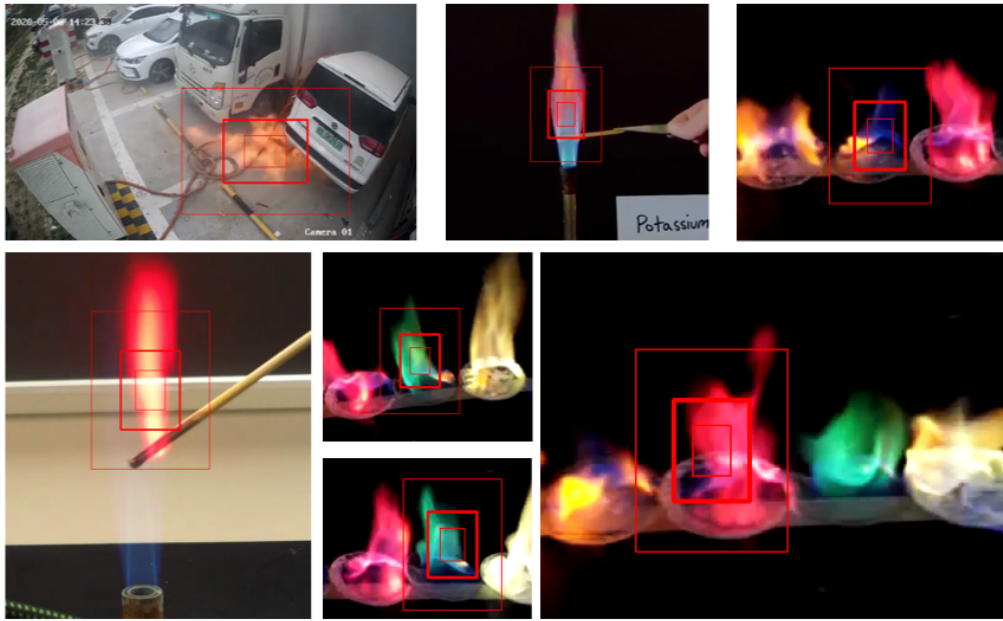


Рис. 10. Примеры работы сети, обученной без HUE-величин, с областями, содержащими огонь разной природы (что влияет на палитру пламени). На всех изображениях показано подтверждение наличия огня в подозрительной области

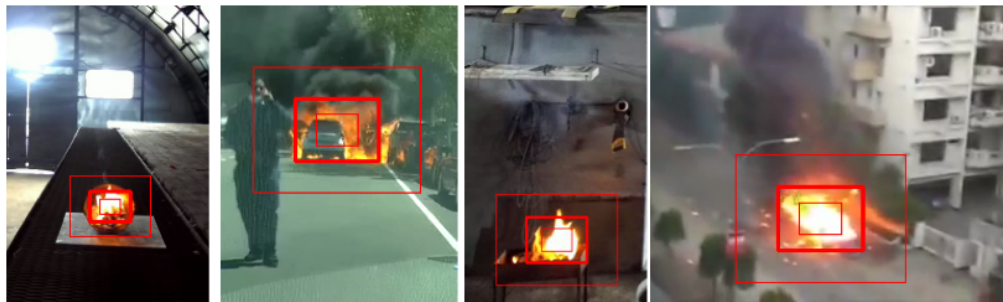


Рис. 11. Подтверждение огня в подозрительной области видеоизображения



Рис. 12. Примеры отсутствия ложных срабатываний на других объектах видеоизображений. Наличие огня в подозреваемой области не подтверждено

- б) если в ролике присутствует огонь, то он не меняет своего местоположения в кадре;
 - с) независимо от наличия или отсутствия огня в роликах присутствует движение людей, машин и других объектов;
 - д) ролики содержат искусственные источники освещения, свет, проникающий в окна, объекты, напоминающие огонь по цвету, блики и поверхности, отражающие огонь на воспроизведении;
- 2) каждый ролик из коллекции размечается так, чтобы был известен регион возгорания на данном

ролике (прямоугольная область, описанная вокруг области горения в кадре);

3) программная утилита прогоняет коллекцию роликов, обнаруживая в каждом подозрительные области (регионы-кандидаты на содержание огня); два программных процесса занимаются обнаружением регионов-кандидатов: первый выделяет области, по цвету напоминающие огонь, второй — все области, где замечено движение (эти области отслеживаются встроенным трекером для того, чтобы была возможность набрать необходимое количество кадров); подозрительные области подвергаются классификации с помощью RNN-сети по описанному методу;

4) результат классификации интерпретируется следующим образом:

- a) если область классифицирована как содержащая огонь и находится внутри вручную указанной области огня для данного ролика (или же совпадает с указанной областью на 95 %, коэффициент IoU = 95 %), то результат положительный; (TP — true positive);
- b) если область классифицирована как содержащая огонь и находится вне вручную указанной области огня для данного ролика (или же совпадает с указанной областью менее чем на 95 %, коэффициент IoU < 95 %), то результат ложноположительный; (FP — false positive);
- c) если область классифицирована как область без огня и находится внутри вручную указанной области огня для данного ролика (или же совпадает с указанной областью на 95 %, коэффициент IoU = 95 %), то результат ложноотрицательный (FN — false negative);
- d) если область классифицирована как область без огня и находится вне вручную указанной области огня для данного ролика (или же совпадает с указанной областью менее чем на 95 %, коэффициент IoU < 95 %), то результат отрицательный (TN — true negative).

Искомые метрики модели: True Positive Rate (TPR) и False Positive Rate (FPR) — рассчитываются следующим образом:

$$TPR = \frac{TP}{TP + FN}; \quad FPR = \frac{FP}{FP + TN}.$$

В результате эксперимента получены следующие значения величин:

- TP = 284;
- TN = 945;
- FP = 7;
- FN = 4.

Соответственно, вероятности работы модели:

$$TPR = 284 / (284 + 4) = 0.986;$$

$$FPR = 7 / (7 + 945) = 0.0074;$$

Точность классификации составляет 98.6%, а ошибка первого рода — 0.7 %.

Выводы

В статье предложен метод раннего обнаружения огня на видеопоследовательности, который позволяет значительно сократить количество ложных срабатываний (ошибку первого рода), что является одним из ключевых моментов на пути к созданию автономного детектора, способного к самостоятельной работе в реальных условиях.

Метод предлагает верификацию найденных регионов-кандидатов путем анализа временного ряда характеристик с помощью рекуррентной нейросети. Основное новшество предложенной модели — способ получения временной последовательности, отражающей временное изменение проверяемой области и пространства вокруг нее с помощью косвенных величин. Способ заключается в выделении трех масштабных областей, названных так по способу образования и пространственного расположения, и представляющих собой в некотором виде «пирамиду признаков», часто используемую в архитектуре сверточных нейросетей. В данном случае признаки — это средние значения величин h , s , v , рассчитанные для каждой масштабной области в цветовом пространстве HSV (h — оттенок, s — насыщенность, v — интенсивность). Три области отражают видимые во времени изменения пространства внутри и снаружи предполагаемой области горения. «Колебания» величин h , s , v в этих областях образуют собой некоторое представление, которое практически безошибочно классифицируется рекуррентной нейросетью.

Таким образом, регионы-кандидаты, предположительно содержащие область возгорания (или горения), проходят окончательную проверку по времени и по видимым изменениям пространства внутри, снаружи и в пределах подозрительной области.

Полученные для предложенной модели метрики качества работы: точность классификации 98.6% и ошибка первого рода, равная 0.7 %, позволяют говорить о высоком достигнутом результате и перспективном использовании данного метода.

Кроме уменьшения ложноположительных срабатываний модели, метод позволяет значительно сократить количество данных для классификации региона-кандидата, что делает эту модель пригодной для использования в устройствах граничных вычислений, вроде одноплатных компьютеров семейства Jetson от NVIDIA или других аналогичных, для создания автономных детекторов раннего обнаружения возгораний по видеоизображениям.

Вся работа, описанная в статье, должна рассматриваться не как уже найденное решение задачи раннего обнаружения огня на видеоизображениях, а как перспективное направление дальнейшего исследования и основа для создания промышленных аппаратно-программных детекторов, которым человек сможет доверить охрану имущества от возможных угроз, связанных с огнем.

ЛИТЕРАТУРА

1. Saponara S., Elhanashi A., Gagliardi A. Real-Time Video Fire-Smoke Detection Based on CNN in Antifire Surveillance Systems. *J Real-Time Image Proc.* 2021;18:889–900.
2. Hongyi Pan, Diaa Badawi, Ahmet Enis Cetin. Computationally Efficient Wildfire Detection Method Using a Deep Convolutional Network Pruned via Fourier Analysis. *Sensors.* 2020;20:2891.
3. Abdusalomov A., Baratov N., Kutlimuratov A., Whangbo T. K. An Improvement of the Fire Detection and Classification Method Using YOLOv3 for Surveillance Systems. *Sensors.* 2021;21:6519.
4. Samarth G., Bhowmik N., Breckon T. P. Experimental Exploration of Compact Convolutional Neural Network Architectures for Non-Temporal Real-Time Fire Detection. *Proceedings of the 18th IEEE International Conference on Machine Learning and Applications ICMLA 2019.* Piscataway, NJ: IEEE. 2019:653–658.
5. Lin G., Zhang Y., Xu G., Zhang Q. Smoke Detection on Video Sequences Using 3D Convolutional Neural Networks. *Fire Technol.* 2019;55:1827–1847.
6. Günay O., Taşdemir K., Töreyn B. U., Çetin A. E. Fire Detection in Video Using LMS Based Active Learning. *Fire Technol.* 2010;46:551–577.
7. Kim B., Lee J. A Video-Based Fire Detection Using Deep Learning Models. *Appl. Sci.* 2019;9:2862.
8. Li Guodong, Lu Gang, Yan Yong. Fire Detection Using Stereoscopic Imaging and Image Processing Techniques. *Proceedings of IEEE International Conference on Imaging Systems and Techniques 2014 (IST2014).* 2014:28–32. DOI: 10.1109/IST.2014.6958440.