

Алгоритмы компьютерного зрения для обнаружения дыма на видеопоследовательностях

Е.Р. Адамовский¹, Р.П. Богуш¹, С.В. Абламейко²

¹ Полоцкий государственный университет имени Евфросинии Полоцкой,
211440, Республика Беларусь, г. Новополоцк, ул. Блохина, д. 29;

² Белорусский государственный университет, 220030, Республика Беларусь, г. Минск, пр. Независимости, д. 4

Аннотация

Представлен аналитический обзор методов и алгоритмов детектирования дыма на последовательностях изображений для раннего обнаружения пожаров с применением систем видеонаблюдения. Рассмотрены визуальные признаки дыма, включая его статические и динамические особенности, которые позволяют отличать его от схожих явлений. Также представлены ограничивающие и мешающие факторы, влияющие на отображение визуальных признаков дыма в реальном мире на видеокадрах. Рассмотрены метрики для оценки эффективности детектирования и локализации областей дыма. Анализируются подходы, основанные на классических алгоритмах компьютерного зрения и на машинном обучении, включая сверточные нейронные сети и трансформеры. Алгоритмы, основанные на машинном обучении, систематизированы с учетом вычислительных затрат. Представлены существующие наборы данных машинного обучения моделей обнаружения дыма, приведены варианты аугментации изображений для расширения обучающей выборки.

Ключевые слова: детектирование дыма, пространственно-временные признаки дыма, системы видеонаблюдения, компьютерное зрение, нейронные сети.

Цитирование: Адамовский, Е.Р. Алгоритмы компьютерного зрения для обнаружения дыма на видеопоследовательностях / Е.Р. Адамовский, Р.П. Богуш, С.В. Абламейко // Компьютерная оптика. – 2026. – Т. 50, № 4. – 1827. – doi: 10.18287/COJ1827.

Citation: Adamovskiy YR, Bohush RP, Ablameyko SV. Computer vision algorithms for detecting smoke in video sequences. Computer Optics 2026; 50(4): 1827. doi: 10.18287/COJ1827.

Введение

Развитие средств вычислительной техники и широкое распространение систем видеонаблюдения (СВН) в настоящее время позволяет во многих случаях использовать визуальный анализ на основе алгоритмов компьютерного зрения для широкого спектра задач, ранее решаемых другими способами. Среди них одной из важных является обнаружение пожара на ранней стадии. В его распространении выделяют три фазы, при этом начальная, при которой область возгорания распространяется линейно вдоль горючего материала, длится до 10 мин. Затем пожар переходит в фазу объемного развития, и первичные средства тушения становятся малоэффективными. Таким образом, обнаружение пожара на ранней стадии является важной и актуальной задачей, решение которой обладает значительным экономическим эффектом за счет повышения уровня пожарной безопасности и снижения размера возможного ущерба.

Для обнаружения пожара на ранней стадии традиционно используются датчики, основанные на измерении температуры или содержании в воздухе продуктов возгорания или горючих газов, фиксации наличия открытого пламени или комбинированные [1]. Такие подходы имеют ограниченную область применения и различную задержку времени срабатывания, которое определяется как время, прошедшее от момента появления пожара до момента его обнаружения. В среднем температурные датчики характеризуются дальностью обнаружения до 37 м и задержкой до 10 с, дымовые датчики выполняют детектирование на расстоянии до 74 м и имеют время срабатывания до 58 с [2]. В помещениях большого объема, с высокими потолками или слабой циркуляцией воздуха эффективность традиционных средств детектирования дыма снижается. Изменение условий окружающей среды также может приводить к ошибкам обнаружений [3], например, изменение температуры и влажности. На открытых пространствах, включая леса и свалки, крупных инфраструктурных объектах, включая ангары, туннели и выставочные здания, традиционные датчики неэффективны [4], а также не способны установить точное местоположение источника возгорания [5]. Решения на основе анализа данных СВН более универсальны. Они могут быть развернуты с учетом возможностей существующей инфраструктуры без значительных системных изменений и финансовых затрат [6] при условии обеспечения визуального доступа к потенциальным объектам возгорания.

Визуально пожар представляет собой сочетание двух явлений различной природы: пламени и дыма. Дым образован газообразной дисперсной системой (аэрозоль), состоящей из мелких твердых частиц во взвешенном состоянии, которая формируется в процессе горения. Наличие дыма на ранней стадии пожара определяется типом возгорания, который, в свою очередь, зависит от материала горения. Основными компонентами дыма являются CO₂, CO, HCN, HCl, сажа, водяной пар, летучие органические соединения и различные оксиды [7]. При

тлении и пиролизе визуально наблюдается только дым, при пламенном горении область задымления может присутствовать [8]. Важными визуальными факторами пожара являются размер и динамика движения дыма, который способен занимать значительно больший объем по сравнению с пламенем и распространяться на значительные расстояния, поскольку его существование не требует непосредственного прилегания к высокотемпературной области горения. Таким образом, наличие дыма может быть зафиксировано даже при условии, что источник возгорания визуально недоступен. Дым является важной характеристикой пожара и присутствует как визуальный признак в большинстве случаев на ранней стадии [9]. Следовательно, надежное детектирование дыма представляет собой важную задачу при обнаружении пожара с использованием СВН согласно определенному алгоритму компьютерного зрения, который должен принимать решение о наличии и позиции дыма в кадре на основе выделенных признаков. Именно алгоритм является ключевым элементом, определяющим эффективность обнаружения дыма на видеоданных. Однако дым на ранней стадии пожара характеризуется областями малых размеров, а дисперсия цвета, текстуры дыма и помех велики, что делает его обнаружение очень трудным [10]. При этом присутствуют дополнительные потери признаков за счет преобразования трехмерного изображения в двухмерное, формируемое видеокамерами систем наблюдения.

В настоящее время существует ряд методов и алгоритмов для выделения дыма на видеоданных, которые используют комбинации признаков и подходы для их выявления, соответственно, характеризуются различной эффективностью и областью применения. В данной работе выполнен анализ таких алгоритмов. На основе анализа приведено заключение об эффективности рассмотренных алгоритмов и перспективе разработки новых подходов для решения задачи детектирования дыма на основе видеопоследовательностей.

1. Представление дыма на видеопоследовательности

1.1. Основные визуальные признаки дыма

Область задымления является частью пространства, примыкающей к зоне горения, с высокой концентрацией снижающих видимость продуктов горения [11]. Таким образом, визуальным дым представляет собой газообразную полупрозрачную структуру, возникающую вокруг области горения. Физический и химический состав дыма, а также его агрегатное состояние ограничивают диапазон характерных визуальных признаков, учет которых позволяет выделить дым на фоне других объектов. К таким признакам относятся следующие визуальные характеристики.

1. Преобладание вертикальной составляющей движения частиц дыма снизу вверх, которое обусловлено восходящими потоками горячего воздуха в зоне горения [12]. По мере распространения дыма движение его частиц приобретает хаотичный характер и в большей степени определяется такими факторами, как ветер, осадки, температура и влажность воздуха [13]. На удалении от источника дым становится похож на другие встречающиеся в природе системы взвешенных частиц: облака, пар (рис. 1а), смог, туман или пыль. Однако дым характеризуется большей динамикой, т.е. скоростью протекающих в нем процессов за счет сообщенной ему источником горения энергии.

2. Цветовой диапазон, ограниченный преимущественно оттенками серого [14], что объясняется рассеянием света на мелких частицах продуктов горения. При этом его яркость варьируется от белого до черного, а конкретный оттенок определяется материалом и активностью горения [13]. К примеру, низовой лесной пожар характеризуется светло-серым цветом дыма, а верховой пожар в лесу имеет более темный оттенок. При горении полимерных материалов цвет дыма близок к черному [7]. Дым, в отличие от пламени, не излучает в оптическом диапазоне, поэтому его итоговый цветовой оттенок в значительной степени определяется освещением сцены и может характеризоваться высокой насыщенностью цвета (рис. 1б). Следовательно, при выделении признаков дыма необходимо учитывать не только объекты окружающей сцены, но и особенности освещения сцены.

3. Статичность источника. Пожар, как правило, распространяется со скоростью до нескольких метров в минуту. В таком случае на интервале времени, измеряемом десятками секунд, что является характерным временем срабатывания традиционных датчиков [2], следует ожидать устойчивого наличия источника дыма в одном месте или небольшого смещения, что позволяет не учитывать при временном анализе дымоподобные объекты со слишком высокой скоростью перемещения [15].

4. Полупрозрачность дыма определяется как процентная мера того, насколько он блокирует собой видимый свет [16]. Отличная от нуля оптическая плотность реального дыма приводит к частичному перекрытию объектов фона, вид которых при этом теряет резкость граней [7] и насыщенность цветов. Полупрозрачность дыма может быть неравномерной, что приводит к значительным визуальным различиям в областях задымления [17]. Отсутствие резких контуров способно снизить точность локализации [18, 19], поскольку возникает сложность в определении содержащей дым области пространства из-за невозможности установить строгий ограничивающий критерий. Этот факт делает неэффективным применение методов сегментации, предполагающих для корректной работы наличие у объектов четких контуров [19]. Дым рассеивается по мере распространения, поскольку одно и то же количество частиц занимает увеличивающийся объем. Таким образом, чем ближе источник, тем плотнее дым, что важно для локализации источника возгорания. По мере

распространения дыма влияние объектов фона, которые могут иметь различный цвет и форму (рис. 1в), усиливается.

5. Хаотично изменяющаяся произвольная форма дыма [7]. Его границы могут быть четкими (рис. 1б), но чаще размытыми (рис. 1в), особенно на ранней стадии пожара.

6. Характерная текстура и форма краев дыма [20], содержащих динамические завихрения [21], которые могут быть оценены по фрактальной размерности, как в [22], или геометрическими способами, как в [6].

7. Мерцание области задымления из-за конвективного потока горячей парогазовой смеси продуктов горения, которое проявляет наибольшую выраженность на ее краях. Частота мерцания дыма составляет 1 – 3 Гц [23].

На визуальное представление дыма влияет его взаимодействие с другими объектами (рис. 1б), включая дымоподобные сущности, т.е. наличие пространственного контекста. Пример негативного влияния контекста в виде облаков на фоне дыма, что не позволяет визуально разделить их, показан на рис. 1г.



Рис. 1. Особенности визуального представления дыма:

(а) визуальная схожесть дыма (SM) и пара (ST) [10], (б) влияние освещения на цвет дыма, четкие границы, перекрытие объектами, (в) высокая прозрачность дыма на сложном динамическом фоне, размытые границы, (г) дым на фоне облаков

Таким образом, дым характеризуется следующими визуальными признаками: преобладающее вверх движение при статичности его источника, преимущественно серый цвет, полупрозрачность, произвольная форма, мерцание, фрактальная текстура. Дым визуально похож на другие явления, но отличается от них динамическими признаками: характер движения, изменение формы и прозрачности. На ранней стадии пожара дым находится в непосредственной близости от источника, что позволяет рассчитывать на более сильную выраженность его признаков. Значительное влияние на дым могут оказывать объекты фона, переднего плана и освещение сцены.

1.2. Потери информации и искажения при представлении дыма на видео

Выходные видеопоследовательности СВН представляют собой упорядоченный набор отдельных изображений $\{I_1, I_2, \dots\}$, которые отражают состояние объектов наблюдаемой сцены в различные моменты времени $\{t_1, t_2, \dots\}$. Следовательно, дым при его наличии визуализируется на видеокадрах как полупрозрачный или полностью непрозрачный объект с характерной текстурой и динамической формой, являясь продолжением двухмерной текстуры во временной области [7]. Каждый кадр размером $W \times H$ пикселей (пк) представляется в виде трехканального изображения $I \in \mathbb{R}^{W \times H \times 3}$, как правило, в цветовой модели RGB. Существует ряд факторов, приводящих к потере информации на изображениях систем видеонаблюдения, которые следует учитывать при построении алгоритмов обнаружения дыма:

1. Преобразование трехмерной сцены в двухмерное представление. Потеря информации о глубине приводит к невозможности однозначно выполнить сортировку объектов по их удаленности от точки наблюдения, определить расстояние, истинный размер объектов, их взаимное расположение.

2. Дискретность цифрового сигнала в виде ограниченного разрешения матрицы видеокамеры и конечная кадровая частота сокращают объем полезной информации, необходимой для выделения признаков дыма. На открытом пространстве достаточно высока вероятность возникновения ситуации, при которой местоположение пожара окажется далеко от точки наблюдения и проекция области задымления на матрицу видеокамеры будет иметь малую площадь, что затруднит обнаружение дыма. В рамках задачи обнаружения признаков пожара на ранней стадии, когда объект дыма еще занимает небольшой объем пространства, разрешение видеокамеры является достаточно значимым фактором. Стандартная скорость видеопотока составляет 25 – 30 к/с, что эквивалентно 33–40 мс на один кадр. Размер изображения в существующих системах видеонаблюдения чаще всего составляет 1920×1080 пк (формат FullHD, 1080p), но для современных реализаций на важных инфраструктурных объектах может достигать разрешения 3840×2160 пк (формат UltraHD, 4K). В отдельных случаях могут использоваться данные с меньшим разрешением и кадровой частотой. Качество данных ограничивается возможностями оборудования и пропускной способностью канала связи.

3. Ограниченная светочувствительность камеры усложняет детектирование дыма в темное время суток или в помещении без источников освещения. На небольшом расстоянии от точки наблюдения недостаток может быть скомпенсирован инфракрасной (ИК) подсветкой, если камера оснащена соответствующим модулем, однако для сцен большого масштаба такой подход бесполезен. Кроме того, на степень непрозрачности дыма в ИК-диапазоне

влияет его состав и условия наблюдения. СВН не содержат дополнительных датчиков, помимо видеокамеры, поэтому информация о температуре и составе воздуха недоступна для анализа. Коммерческие решения СВН с тепловизором существуют, однако имеют высокую стоимость [6] и значительно менее распространены. Также освещение сцены способно влиять на автоматически регулируемые настройки видеокамеры и, следовательно, представление дыма, искажая его истинный цвет.

4. Помехи, связанные с аппаратным несовершенством видеокамер. К ним относятся цифровые шумы, искажение перспективы, ограниченная цветопередача и различные эффекты оптической системы: блики, засветы, расфокус. Все это способно ухудшить надежность системы обнаружения. К примеру, в [24] отмечено, что алгоритм оптического потока, применяемый для получения направления движения дыма на последовательности видеок кадров, чувствителен к шуму камер видеонаблюдения, что может повлиять на эффективность детектирования дыма.

5. Сложные метеорологические условия, включая осадки, ухудшают видимость дыма и вносят дополнительные шумы на его изображение.

6. Алгоритмы сжатия данных для передачи видеопотока и соответствующие транспортные протоколы вносят существенные искажения. Наиболее популярный кодек H.264 (AVC) [25] использует чередующиеся опорные и разностные кадры, а также применяет межкадровое сжатие, что способно приводить к появлению графических артефактов на объектах дыма, особенно в условиях ухудшения качества связи. Используемое в H.264 энтропийное кодирование плохо справляется с изображениями дыма по причине его неструктурированной текстуры и непредсказуемости, из-за чего на видеок кадрах, передающихся с ограниченным битрейтом, возникают эффекты блочности, рябь и размытость в областях задымления.

Наличие упорядоченной во времени последовательности видеок кадров СВН позволяет проводить с помощью алгоритмов компьютерного зрения пространственно-временной анализ. Пространственный анализ включает выделение признаков дыма, ассоциированных с конкретными областями (зонами или регионами) на видеок кадрах. Временной анализ представляет собой отслеживание изменения признаков дыма в течение определенного количества кадров. Пространственно-временной анализ позволяет отслеживать динамику изменения признаков с привязкой к их положению на изображении и выделять неочевидные следы движения дыма [10].

2. Метрики оценки эффективности алгоритмов обнаружения дыма на видеопоследовательностях

Для оценки качества алгоритмов обнаружения дыма на видеопоследовательности используются метрики, которые оценивают точность классификации и локализации. Классификация позволяет выявить наличие дыма в кадре. Следует отметить, что классификация дыма является сложной задачей даже для человека. К примеру, датасет RISE [26] содержит 12567 вручную классифицированных видеозаписей промышленных выбросов с паром и дымом, из которых 193 клипа были размечены противоположным образом различными участниками процесса.

Метрики точности классификации применяются в том случае, если целью является определение наличия или отсутствия дыма в кадре, т.е. выполняется бинарная классификация «дым обнаружен» или «дым не обнаружен». В отдельных случаях метрики точности классификации приводятся и для методов, выполняющих локализацию дыма [27, 28], если производится оценка точности только модуля классификации, который обрабатывает выделенные на предыдущих этапах алгоритма области интереса (Region of Interest, RoI), а не изображение целиком. Также эти метрики применяются при классификации ячеек регулярной сетки (патчей), на которую разбивается каждый входной видеок кадр. Таким способом выполняется локализация дыма с кратно меньшим пространственным разрешением относительно исходного размера изображений, однако для задачи обнаружения дыма подобный компромисс является приемлемым. Например, в [13] размер патчей составлял 32×32 пк, в [29] ячейки сетки имели размер 224×224 пк, а в [30] – примерно 100×200 пк, в [4] использовались блоки до 30×30 пк.

К метрикам точности классификации относятся следующие показатели.

1. Доля истинно положительных (True-Positive, *TP*) предсказаний (*TP Rate*, *TPR*) или полнота (recall, *R*):

$$R = TPR = TP / (TP + FN). \quad (1)$$

Истинноположительным предсказанием является обнаружение дыма в кадре, на котором он действительно присутствует. Метрика полноты чувствительна к ложноотрицательной (False-Negative, *FN*) ошибке, которая в задаче детектирования дыма эквивалентна отсутствию реакции. В некоторых работах используется метрика Hit Rate (*HR*), которая аналогична *R*, например в [27].

2. Доля ложноотрицательных предсказаний (*FN Rate*, *FNR*) показывает относительное количество пропусков дыма при его фактическом наличии:

$$FNR = 1 - TPR = FN / (TP + FN). \quad (2)$$

3. Доля истинноотрицательных предсказаний (True-Negative Rate, *TNR*):

$$TNR = TN / (TN + FP). \quad (3)$$

Истинноотрицательным предсказанием является классификация кадра как не содержащего дым, на котором он отсутствует.

4. Доля ложноположительных предсказаний (False-Positive Rate, FPR):

$$FPR = FP / (FP + TN). \quad (4)$$

Большое значение FPR на практике даже при высоком истинноположительном предсказании приводит к тому, что система обнаружения дыма может быть практически непригодной к эксплуатации из-за частых ложных тревог, так как операторы быстро потеряют доверие к такой системе, соответственно, срабатывание на реальную угрозу может быть проигнорировано. В случае полной автоматизации ложное срабатывание может привести к остановке технологических процессов предприятий, эвакуации персонала и вызову пожарных.

5. Общая точность (Accuracy, AC), иногда также встречается обозначение Frame Accuracy (FA) [31]:

$$AC = (TP + TN) / (TP + TN + FP + FN). \quad (5)$$

Метрика AC используется, если классы сбалансированы по количественному соотношению, поэтому в задачах детектирования дыма ее применение ограничено или используется в составе комплексной оценки.

6. Точность отбора (Precision, P):

$$P = TP / (TP + FP). \quad (6)$$

Метрика P актуальна в случаях, когда цена ложноположительной ошибки велика, что не соответствует задаче обнаружения дыма [3], однако точность отбора используется для вычисления метрики F_1 .

7. F-мера (F_1 -Score), или среднее гармоническое значение, объединяет метрики P и R , что позволяет комплексно оценить качество детектирования, учитывая баланс между ложноположительными и ложноотрицательными ошибками системы обнаружения дыма:

$$F_1 = (2 \times P \times R) / (P + R). \quad (7)$$

8. Коэффициент корреляции Мэттьюза (Matthews Correlation Coefficient, MCC), метрика бинарной классификации, которая учитывает все четыре варианта результатов предсказаний (confusion matrix): TP , FP , TN и FN [32]:

$$MCC = (TP \times TN - FP \times FN) / \sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}. \quad (8)$$

Метрика MCC позволяет произвести более адекватное оценивание точности при несбалансированности классов [6, 30], чем F_1 , однако используется значительно реже по причине более сложной интерпретации и отсутствует в стандартных тестовых системах. Метрика может приобретать значения в диапазоне от -1 до 1.

Таким образом, метрики классификации могут быть использованы в качестве оценки точности обнаружения дыма, также для сравнения новых методов с существующими.

Во многих случаях необходимо учитывать качество предсказания не только наличия дыма на изображении, но и точность определения его положения, размера и формы. Усложнение в виде локализации дыма дополнительно увеличивает фактор неопределенности. В [33] отмечено, что полупрозрачность дыма создает трудности при разметке данных, а одновременное обнаружение пламени и дыма, объекты которых в значительной степени пересекаются, способно запутать систему обнаружения.

Локализация дыма является задачей установления его точного положения в кадре и выполняется одним из двух способов: при помощи ограничивающих рамок (bounding boxes) или путем сегментации изображения. Ограничивающие рамки представляют собой границы прямоугольника, внутри которых заключена соответствующая левому областю изображения. Рамки характеризуются координатами центра, шириной и высотой, или координатами левого верхнего угла, шириной и высотой. В большинстве случаев грани рамок выровнены относительно граней изображения [3, 6, 17, 18, 27, 28, 31, 34, 35, 36], однако могут использоваться и наклонные прямоугольники для адаптации к объектам дыма сложной формы, как в [37]. Сегментация является попиксельной классификацией объектов на отдельных изображениях или видеопоследовательности и представляется в виде маски [38, 39, 40].

В качестве метрики точности локализации дыма используется показатель Intersection over Union (IoU), который измеряет перекрытие двух прямоугольников, истинной ограничивающей рамки дыма S_{true} и предсказанной ограничивающей рамки S_{pred} .

$$IoU = |S_{true} \cap S_{pred}| / |S_{true} \cup S_{pred}|. \quad (9)$$

Метрика IoU может быть применена к областям любой формы, например, в случае использования сегментации, если существует возможность вычисления их площадей.

Наиболее популярной метрикой обнаружения дыма на последовательности видеок кадров является средняя точность (Average Precision, AP) по всем C классам (Mean AP , mAP).

$$mAP = (1/C) \times \sum_{c=1}^C AP_c. \quad (10)$$

AP вычисляется как площадь под кривой $P(R)$ (Precision-Recall Curve, PRC), которая на практике аппроксимируется, например, через 11 дискретных точек на оси R от 0 до 1 с шагом 0,1:

$$AP = (1/11) \times \sum_{R \in \{0, \dots, 1\}} \max(P(R)). \quad (11)$$

В случае детектирования только объектов дыма, т.е. при использовании лишь одного класса «smoke», метрика AP полностью эквивалентна метрике mAP ($C=1$). В популярных бенчмарках, например COCO [41], AP каждого класса усредняется для наборов PRC , полученных с различными пороговыми значениями IoU, как правило, от 0,5 до 0,95 с шагом 0,05. Таким образом, mAP является ключевой метрикой для оценки точности системы детектирования дыма с локализациями, поскольку объединяет в себе значения P и R аналогично F_1 .

Метрика AP имеет ряд расширений: по конкретному порогу IoU: AP_{50} ($IoU \geq 0,5$) и AP_{75} ($IoU \geq 0,75$), а также по размеру дыма AP_S , AP_M и AP_L . Значение AP_{50} показывает качество грубой локализации, когда общая площадь рамок составляет более 50 %. Значение AP_{75} отражает способность системы более точно определять местонахождение дыма. AP_S (small) вычисляется как AP для объектов с площадью меньше 32×32 пк, AP_M (medium) для объектов с площадью от 32×32 пк до 96×96 пк, а AP_L (large) для всех остальных объектов большей площади. На практике большинство работ ограничивается метриками mAP и AP_{50} , поскольку очень точная локализация областей задымления, с пересечением 75 % и выше, остается крайне сложной задачей и не является целью исследований.

Также для оценки используется средняя полнота (Average Recall, AR) в виде двух вариантов: по размеру дыма AR_S , AR_M и AR_L , а также по количеству предсказаний AR_1 , AR_{10} и AR_{100} . Эта метрика особенно важна для моделей, способных генерировать несколько предсказаний на один объект с разной степенью уверенности.

В некоторых работах, например, в [26], используется алгоритм Grad-CAM [42], который визуализирует карты внимания (attention maps) отдельных слоев искусственных нейронных сетей (ИНС) при обработке данных. Это позволяет с крайне низким пространственным разрешением локализовать участки изображения, содержащие дым. В работах, посвященных обнаружению дыма, Grad-CAM не используется с целью непосредственного вывода данных, а служит только для проверки корректности работы нейросетевых моделей, поскольку связанные с результатами работы этого алгоритма метрики точности локализации в литературе отсутствуют.

Также важной метрикой для системы детектирования дыма является производительность используемого алгоритма, которая выражается в скорости обработки данных и измеряется в количестве обрабатываемых кадров в секунду или Frame Per Second (FPS). Если алгоритм выполняет обработку со скоростью поступления видеок кадров или быстрее, он может считаться работающим в режиме реального времени. На практике может анализироваться не каждый поступающий кадр, а с предопределенным шагом, что кратно повышает запас времени на обработку данных.

Таким образом, сочетание метрик AP и AR позволяет комплексно оценить качество детектирования с локализациями дыма, поскольку аналогично метрикам классификации высокие значения метрик типа AP показывают точность истинноположительных прогнозов, но не отражают склонность системы к ложным срабатываниям.

3. Методы для обнаружения дыма на последовательностях изображений

3.1. Выделение признаков дыма

При рассмотрении способов выделения признаков дыма могут быть рассмотрены две категории подходов: основанные на классических алгоритмах компьютерного зрения и основанные на машинном обучении [34]. На практике оба подхода могут комбинироваться, как в [13, 27, 38, 39, 43, 44], в таких случаях реализации являются гибридными. Понятие алгоритма формально включает принцип функционирования ИНС, однако по отношению к их реализациям чаще применяются термины «модель» или «архитектура».

3.1.1. Классические методы

В подходах первой категории используются детерминированные алгоритмы, которые выделяют конкретные признаки на основе установленных правил, также именуемые как rule-based методы. Для обнаружения дыма часто используются следующие алгоритмы: детектирование движения (вычитание фона и построение его модели [4, 9, 12, 13], фильтр Калмана [6, 27]), вычисление оптического потока (optical flow) [38, 26, 45, 46], преобразование цветовых пространств (например, RGB в HSV [9, 13, 6, 27, 47, 48, 46] или YUV [43]), вейвлет-преобразование для частотного или временного анализа [4, 15, 49], анализ геометрических признаков [6], кластеризация [39], локальные бинарные шаблоны (Local Binary Patterns, LBP) [20, 46]. Также алгоритмические подходы для обнаружения дыма часто включают методы препроцессинга и постобработки: медианная фильтрация и морфологические операции для подавления шума [12], логические операции для объединения битовых масок признаков дыма, удаление на кадрах тумана или дымки [34, 38] с целью повышения качества изображения, выравнивание баланса между цветовыми каналами для компенсации преобладающего цвета в

сцене [50]. Таким образом, с помощью rule-based методов выделяются признаки: движение и его направление, цветовые характеристики, форма, положение в кадре, влияние на объекты заднего плана; выполняется препроцессинг и постобработка данных.

При наличии пространственно-временного анализа, rule-based методы, как правило, в качестве входных данных требуют видеопоследовательность, полученную при помощи стационарной камеры, которая в процессе съемки не изменяет свое положение и угол. Это ограничивает область их применения.

В первую очередь, методы алгоритмического типа производят выделение движения путем сравнения значений пикселей нескольких кадров (вычитание фона), а также выполняют построение модели фона [13], которая не содержит движущихся объектов, как алгоритмы MOG и MOG2 [51] в [9, 12]. С целью получения областей с движущимися объектами в [6, 27] использовался фильтр Калмана. Алгоритм оптического потока использует два кадра для вычисления направленности движения различных областей изображения в виде векторного поля с вертикальными и горизонтальными составляющими. Наиболее широко применяется реализация метода Фарнбэка (Farneback) [52]. Например, в [12] на последнем этапе алгоритма обнаружения дыма области-кандидаты анализировались при помощи оптического потока. Вычислялся средний угол и амплитуда полученных векторов движения, затем отмечались области с определенным диапазоном углов и достаточно высоким значением, как показано на рис. 2.

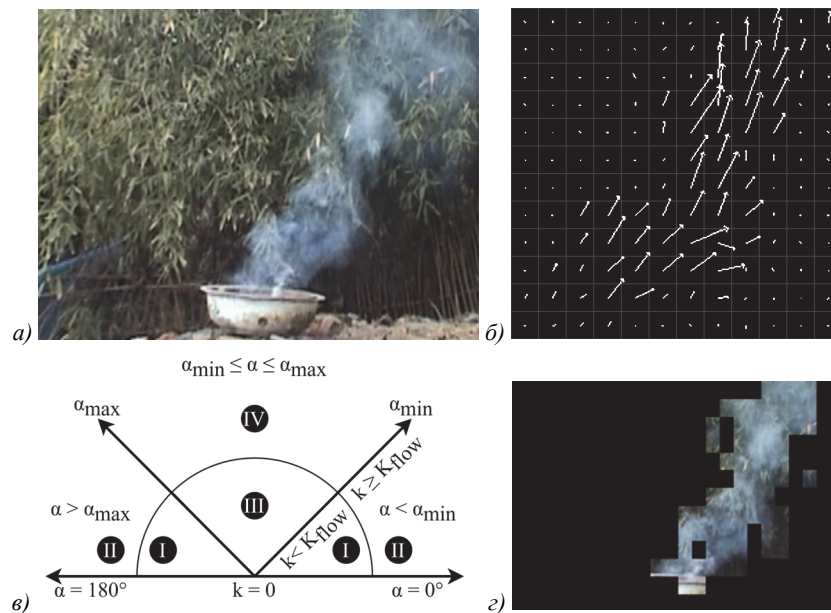


Рис. 2. Выбор областей задымления по направлению движения:

(а) кадр, (б) векторы движения блоков, (в) диапазон допустимых значений векторов (зона IV), (г) блоки-кандидаты на задымление [12]

Для выделения признаков дыма по его цветовым характеристикам применяется метод цветовой сегментации. Исходя из предположения, что дым имеет серый оттенок, обработка данных в модели RGB большинству исследователей представляется неудобной. В [27] выполнялось преобразование кадров в цветовую модель HSV, которая разбивает изображение на три канала: цвет (hue, H), насыщенность (saturation, S) и яркость (value, V). Затем формировалась маска дыма по пороговому значению канала S, которое подбиралось экспериментальным путем. В [43] использовалось преобразование в цветовое пространство YUV для минимизации влияния освещения, где Y – канал яркости, а U и V – цветоразностные каналы, на основе которых производилось формирование маски дыма по пороговым значениям. Другие цветовые модели используются значительно реже. Например, в [14] применялась цветовая модель HSI.

На рис. 3 показано представление дыма на исходном RGB-кадре в цветовых пространствах HSV и YUV. На рис. 3б представлены каналы изображения, показанного на рис. 3а, в цветовом пространстве HSV, сверху вниз: тон (H), насыщенность (S) и яркость (V). Заметно, что в канале насыщенности область задымления контрастирует с остальными объектами сцены. Рис. 3в представляет каналы в цветовом пространстве YUV, сверху вниз: канал яркости и цветоразностные каналы. Анализ обоих цветоразностных каналов позволяет выделить дым на фоне дороги и поля.

Частотный анализ позволяет оценить влияние дыма на объекты фона, поскольку в области перекрытия амплитуды высокочастотных составляющих изображения должны иметь сниженные значения относительно модели фона, как это было выполнено в [15]. В [4] применялось вейвлет-преобразование на основе вейвлета Хаара [23] после этапа вычитания фона с целью обнаружения резких переходов. В [49] вычислялся двухмерный пространственный вейвлет-анализ для определения полупрозрачных объектов на изображении, на которых присутствует дым, и определения мерцания его границ.

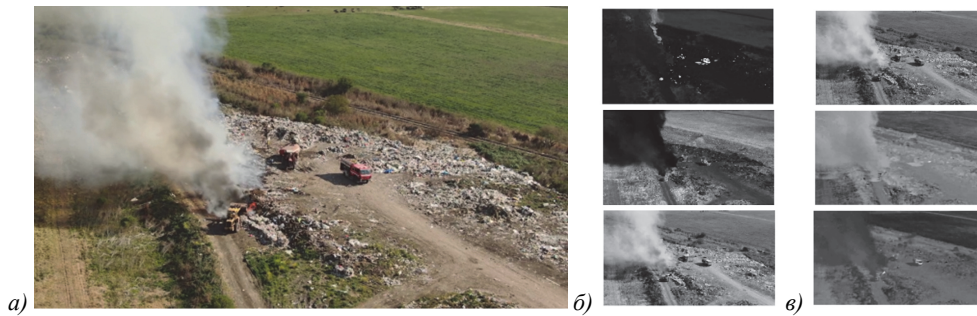


Рис. 3. Представление дыма в цветовых пространствах: (а) исходный кадр, (б) HSV, (в) YUV

Геометрический анализ включает в себя методы, которые оперируют двухмерными сущностями на изображении, включая контуры объектов дыма в виде замкнутых кривых и связные области пикселей (блобы). В [27] вычислялись площади и соотношения сторон ограничивающих рамок областей-кандидатов дыма, после чего удалялись регионы с площадью ниже пороговой и вытянутые по одной из осей, исходя из предположения, что слишком мелкие области являются шумом, а длинные не характерны для объектов дыма. В [6] выделенные на более ранних этапах блобы анализировались на основе их геометрических признаков: турбулентность, распространенность и выпуклость.

Турбулентность T представляет собой отношение квадрата периметра блоба P_{blob} к его площади S_{blob} :

$$T = P_{blob}^2 / S_{blob}. \quad (12)$$

Для области дыма, согласно [6], значение T должно быть высоким из-за его сложной формы.

Распространенность E выражает степень заполнения блобом описанной ограничивающей рамки S_{box} :

$$E = S_{blob} / S_{box}. \quad (13)$$

В [6] предполагается, что дым будет занимать от 20 % до 95 % площади ограничивающей рамки.

Выпуклость Cv представляет собой отношение площади блоба S_{blob} к минимальной площади выпуклой фигуры S_{cv} , описывающей этот блоб:

$$Cv = S_{blob} / S_{cv}. \quad (14)$$

Из-за нечеткой формы дыма ожидается, что значение Cv для дыма окажется выше в сравнении с другими объектами [6].

Для обнаружения дыма ограниченно используется кластеризация. Кластеризация представляет собой процесс группировки пикселей или областей изображения на основе их схожести, т.е. расстояния в пространстве признаков. Наиболее популярные методы: Simple Linear Iterative Clustering (SLIC) [53] и Density Based Spatial Clustering of Application with Noise (DBSCAN) [54]. Отличием от сегментации является то, что при кластеризации не используется семантический анализ, за счет чего достигается высокая скорость работы алгоритмов. В качестве признаков выступают: компоненты цвета в цветовом пространстве CIELAB, координаты пикселя в SLIC или более абстрактные признаки, задаваемые пользователем, в DBSCAN. В [19] была выполнена попытка применить SLIC с целью выделения регионов дыма (region-level segmentation), но в итоге этот этап был исключен из-за низкой результирующей точности, поскольку дым не имеет четких границ. При этом в [39] технология SLIC была успешно применена в сочетании с DBSCAN, но дым анализировался совместно с пламенем.

Общим недостатком rule-based подходов является необходимость подбора параметров, которые могут оказаться эффективными только в определенных сценариях. Особенно остро эта проблема проявляется для методов цветовой сегментации по причине отсутствия у них возможностей для автоматической адаптации под различные варианты освещения сцены, которое может придавать пикселям объекта дыма значительный уровень насыщенности. В качестве другого примера, фильтр Калмана [6, 27], несмотря на низкую вычислительную сложность, чувствителен к шуму, предполагает линейную модель движения и требует ручной настройки. Преимуществом алгоритмических методов является возможность простого изменения параметров, легкость объединения модулей для многокритериального анализа, низкие требования к мощности вычислительных устройств, интерпретируемость выходных данных.

3.1.2. Методы выделения признаков дыма с помощью нейронных сетей

Вторая категория методов представлена моделями глубокого обучения. Глубокое обучение (Deep Learning, DL) подразумевает использование различных типов архитектур ИНС с большим количеством слоев, включая: сверточные нейронные сети (CHC, также Convolutional Neural Networks, CNN) [27, 55, 43, 19, 39, 30, 46] и архитектуры на их основе, например, YOLO [28, 18, 37, 34, 47, 18, 36], LFNNet [17], TANet [40], FCN [38], 3D-CNN (Inception-v1 [26]), трансформерные модели (ViT [29], DETR [31, 35], RT-DETR [3]), метод опорных векторов

(Support Vector Machine, SVM) [9] и случайного леса [4]. При наличии временного анализа также применяется рекуррентная архитектура Long Short-Term Memory (LSTM) совместно с другими моделями [29].

Распространена практика добавления модулей внимания в существующие архитектуры с целью повышения точности обнаружения. Механизм внимания (attention) [56] позволяет моделям фокусироваться на релевантных частях входных данных за счет специальной архитектуры, которая оперирует компонентами ключей и запросов. Механизм внимания реализован в моделях типа «трансформер» и является вычислительно затратным, поэтому иногда от него отказываются в пользу других методов, как в [17].

Архитектуры ИНС, которые полностью самостоятельно выполняют анализ видеок кадров, называются сквозными (end-to-end). К преимуществу такого подхода относится отсутствие необходимости вручную настраивать параметры. Выделение информационных признаков в моделях ИНС происходит автоматически в процессе обучения за счет настройки внутренних весовых коэффициентов модели. Однако архитектура сети предопределяет типы извлекаемых свойств объектов. Например, модели CNN ориентированы на последовательное построение иерархии признаков, начиная с низкоуровневых (края, текстуры) и заканчивая высокоуровневыми абстракциями (формы, объекты). Семейство моделей YOLO (You Look Only Once) [57] использует выделенные признаки для предсказания ограничивающих рамок объектов на регулярной сетке. Трансформеры выполняют построение связей между элементами входных данных с помощью механизма внимания. В ходе обучения модели адаптируют эти представления под специфику задачи распознавания дыма.

Комбинации нейросетей используются для улучшения точности детектирования дыма, например, CNN, LSTM и ViT в [29], как показано на рис. 4.

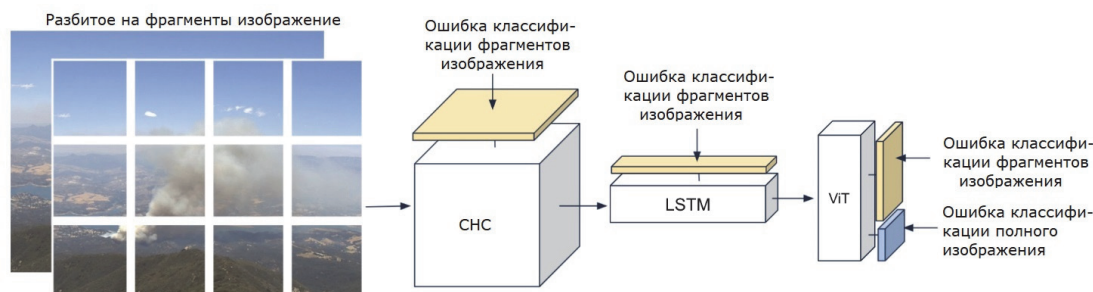


Рис. 4. Комбинирование архитектур СНС, LSTM и ViT для обнаружения дыма [29]

Однако увеличение числа модулей снижает общее быстродействие системы по причине увеличения количества вычислений и преобразований данных между моделями.

Для моделей DL актуальна проблема организации эффективного обучения, которая включает в себя подбор и разметку датасета, настройку параметров обучения, что значительно влияет на точность детектирования дыма. При этом качественно обученные модели способны обнаруживать дым в разнообразных условиях за счет способности к обобщению информации. Использование сегментации для задачи обнаружения и локализации дыма с использованием моделей ИНС сопряжено с рядом проблем, таких как высокая сложность разметки датасета и большой объем вычислительных ресурсов, требуемых для предсказывающих моделей, по сравнению с локализацией рамками, поэтому сегментация используется редко.

При использовании стандартных архитектур ИНС доступны предобученные модели в виде набора их весовых коэффициентов, которые могут быть дообучены на новых примерах для задачи детектирования дыма. Такой подход способен значительно сократить время и необходимый размер данных для обучения. При разработке новых архитектур или модификации существующих такая возможность отсутствует [13].

ИНС нуждаются в большом объеме данных для обучения и требовательны к вычислительным ресурсам, однако принцип их работы позволяет проводить большой объем операций независимо друг от друга. Таким образом, современные архитектуры способны работать с приемлемой скоростью обработки данных только на устройствах, оснащенных мощными графическими процессорами (Graphics Processing Unit, GPU), которые выполняют параллелизацию вычислений.

3.1.3. Гибридные методы выделения признаков дыма

Гибридные реализации используют алгоритмические подходы на первом этапе, а модели DL – на втором [44]. Основная идея такой структуры заключается в снижении вычислительной нагрузки на модули ИНС, характеризующиеся низкой скоростью обработки данных, путем выделения предварительных RoI относительно быстрыми rule-based методами.

Так, в [9, 13, 27] с помощью обучаемого классификатора анализировались области, выделенные по признакам движения и цвета. В [38] модель FCN выполняла обработку результата вычисления оптического потока. В [43] модель WSDD-Net осуществляла классификацию областей-кандидатов, полученных в результате цветовой сегментации в цветовом пространстве YUV. В [39] изображения разбивались на кластеры, а затем каждый из них классифицировался с помощью CNN. В [47] изображения из обучающего датасета преобразовывались в HSV до загрузки в модель,

предполагая, что каналы этой цветовой модели более информативны по сравнению с RGB. В [46] формировалось изображение LBP-motion на основе двух последовательных кадров путем вычисления и комбинирования данных оптического потока и LBP, а затем CHC MobileNetV2 выполняла его классификацию (рис. 5).

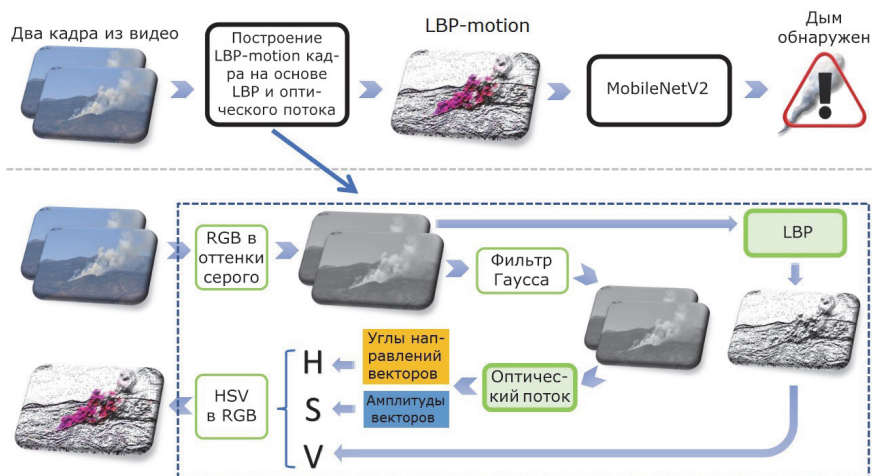


Рис. 5. Архитектура на основе MobileNetV2 с анализом LBP [46]

3.2. Учет временного контекста видеопоследовательности

Временной контекст отражает количество кадров, которые учитывает алгоритм или модель, для принятия решения о наличии или отсутствии дыма. При этом учитывается, что системы обнаружения дыма применяются в режиме реального времени путем анализа видеопотока, т.е. они должны успевать обрабатывать каждый кадр до поступления следующего.

Множество работ, посвященных обнаружению дыма на видеопоследовательности, ограничиваются только достижением необходимой скорости предлагаемых решений, но не используют временной анализ с целью повышения быстродействия и снижения вычислительных затрат [58]. Особенно очевидна эта особенность по отношению к методам на основе ИНС. В исследованиях часто используются существующие архитектуры, модификация которых с целью возможности производить временной анализ имеет слишком высокую трудоемкость. Даже при разработке принципиально новых моделей учет временного контекста выполняется редко и ограниченно, что связано с общей высокой сложностью данной задачи [37, 31, 18, 3, 35]. Анализ целой видеозаписи или ее фрагмента используется в случаях, когда архитектура системы обнаружения дыма, например, 3D-CNN, предполагает обработку серии кадров как единого блока данных.

В отдельных случаях учет временного контекста может быть избыточной операцией или плохо имплементироваться в выбранную архитектуру. Например, в [26] модель RGB-TC на основе 3D-CNN архитектуры Inception-v1, используя многомасштабные 1D-свертки по времени, анализировала сразу 36 кадров. Сделан вывод об ограниченной применимости временного анализа по ряду причин: дым появляется редко и нерегулярно, приводя к дисбалансу классов; погода, пар и облака и создают ложные движения; пространственные характеристики дыма важнее, чем его движение. Последний тезис в [26] подтверждается тем, что перемешивание кадров почти не влияло на точность модели (падение F_1 на 2 %), а использование оптического потока значительно ухудшило точность, на 23,9 %. Неэффективность применения временного анализа, возможно, была следствием неудачного выбора модели.

В [10] для обнаружения дыма на видеопоследовательности разработана модель архитектуры STCNet (рис. 6), которая представляла собой архитектуру, состоящую из двух параллельных блоков: блок выделения пространственных признаков использовал CNN для извлечения текстурных признаков дыма из текущего кадра, а блок выделения временных признаков с аналогичной CNN обрабатывал остаточные изображения, т.е. разницу между соседними кадрами. Таким образом, размер временного контекста составил два кадра. Результаты работы обоих блоков суммировались, после чего следовала полносвязная сеть для классификации.

Два последовательных кадра для вычисления оптического потока с целью дальнейшей обработки использовались в [38, 26, 59, 45, 46]. В [29] применялся временной анализ с использованием архитектуры LSTM и размером временного контекста в два кадра. Авторы экспериментировали с его увеличением до трех кадров, но значительного улучшения точности не получили. При этом модуль LSTM получал на вход признаки, выделенные сверточной сетью ResNet-34 из двух последовательных изображений, а выходные данные передавал в модель ViT для интегрирования пространственной информации между кадрами во времени.

При наличии временного анализа системы детектирования дыма функционируют одним из двух возможных способов: серия кадров подается на вход обработчика одновременно или же последовательно, по мере поступления новых данных, а обработчик выполняет накопление и обработку признаков.

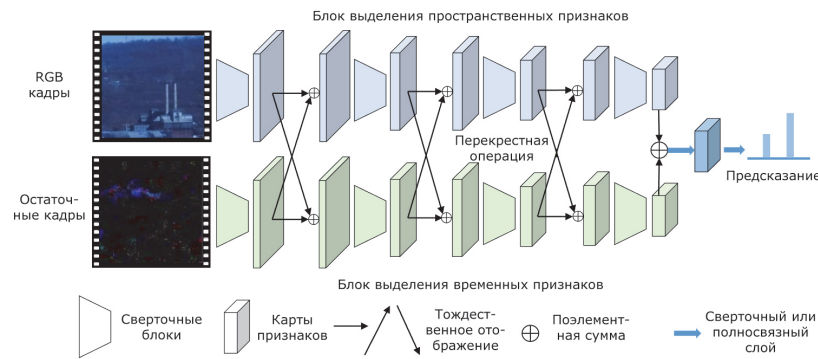


Рис. 6. Архитектура STCNet [10]

Первый способ характерен для rule-based методов и был реализован в [13, 6, 27], где применялся счетчик, который работал следующим образом. Выделялись пространственные признаки дыма, затем подсчитывалось количество кадров, в течение которых они непрерывно присутствовали в одной и той же пространственной области или были зафиксированы на всем кадре. При превышении заданного порога времени дым считался обнаруженным. Следовательно, размер временного контекста в указанных работах был численно равен порогу срабатывания счетчика признаков видеок кадров.

Второй способ реализуется в архитектурах ИНС, которые анализируют набор видеок кадров в виде единого блока данных. Например, в [16] архитектура на основе трехмерных сверточных слоев и модуля LSTM использовала 40 видеок кадров, однако в процессе обработки делала из них выборку, снижая объем анализируемой информации для классификации дыма в несколько раз. Аналогичный подход с использованием моделей на основе архитектуры I3D-ConvNet и временным контекстом в 36 кадров применялся в [26].

3.3. Вычислительная эффективность алгоритмов обнаружения дыма на основе машинного обучения

3.3.1. Алгоритмы с низкой вычислительной сложностью

Многие системы обнаружения дыма проектируются для одноплатных компьютеров (Single-Board Computer, SBC) с относительно низкой производительностью: серия устройств Raspberry Pi (RPi) и Jetson Nano (JN). Такой подход предполагает, что данные устройства будут функционировать в рамках СБН удаленно, совместно с видеок камерой, и выполнять предобработку данных с целью снижения нагрузки на центральный узел, агрегирующий видеопотоки множества точек наблюдения. Для таких условий подходят алгоритмы с низкой вычислительной сложностью, также используется термин «легковесные» или «облегченные» (lightweight). Для них характерны следующие особенности: работа в режиме реального времени, низкие требования программной реализации к вычислительным мощностям устройства и небольшой занимаемый объем памяти. Метрики точности облегченных алгоритмов могут быть ниже по сравнению с реализациями без ограничений на производительность и объем занимаемой памяти.

К примеру, в [55] разработана система обнаружения пожаров на трубопроводах с использованием легковесной трехслойной CNN, основанной на архитектуре FireNet [60] и предлагаемой для запуска на RPi 4. Обучение производилось на датасете FireNet [61] в виде 500 аннотированных в формате XML-изображений с пожаром, дополнительно масштабируемых до размера 64×64 пк с целью снижения вычислительной нагрузки на модуль CNN. Ключевая адаптация для задачи обнаружения дыма заключалась в смещении домена путем разбавления используемого набора данных изображениями пожаров на трубопроводах, т.е. под конкретный тип сцены, для повышения точности классификации кадров в специфических условиях.

В [28] использовалась упрощенная модель YOLOv2, занимающая объем памяти 7,1 МБ и обученная на 400 собранных из различных источников изображениях дыма. Изменения касались снижения размера входного слоя до 128×128 пк и уменьшения количества сверточных слоев. Модели архитектуры YOLO для детектирования объектов используют якорные рамки (anchor boxes), прямоугольные области различных размеров и ориентации. С целью адаптации к задаче обнаружения дыма якорные рамки YOLOv2 были скорректированы с использованием кластеризации методом k-средних (k-means) на основе предварительно размеченных данных, чтобы наилучшим образом соответствовать форме и размеру типичных областей дыма.

В [27] разработан алгоритм, который объединяет традиционные методы компьютерного зрения, включая фильтр Калмана [62] для детектирования движения и цветовую сегментацию с помощью конвертации кадров в цветовое пространство HSV (Hue-Saturation-Value), с моделью CNN, как показано на рис. 7.

На первом этапе выделялись области-кандидаты в виде блоков, которые потенциально могли содержать дым, а затем фрагменты исходного изображения масштабировались до 50×50 пк и подавались на вход CNN для классификации и последующего накопления с целью снижения количества ложных тревог. Обучение производилось на нескольких датасетах: набор видеозаписей FireSense [63], закрытый набор AdvISED с 30 записями дыма, а также датасет из 14270 изображений [51]. Таким образом, для обнаружения дыма использовался

многоэтапный подход, при котором выделялись признаки движения, цвета, а затем области-кандидаты обрабатывались обучаемым классификатором. Объем занимаемой памяти модели составил 126 КБ.

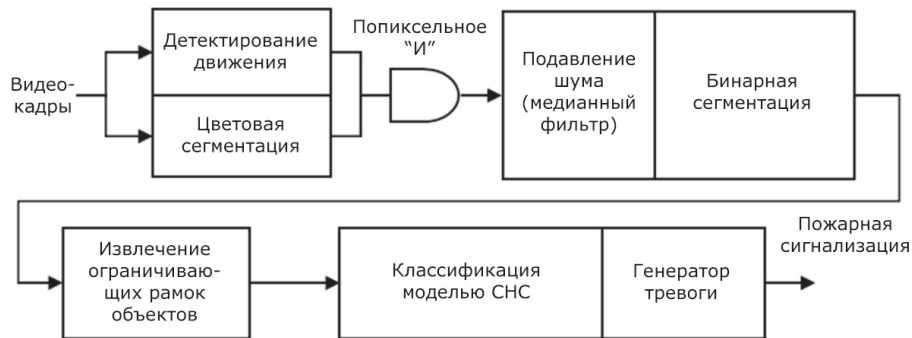


Рис. 7. Гибридная архитектура для обнаружения дыма [27]

В [6] предложен алгоритм, который выполнял аналогичную [27] предобработку, но на втором этапе вместо CNN использовались алгоритмические методы: извлечение геометрических признаков из блоков, выделение краев с помощью фильтра Собеля, пространственно-временной анализ путем накопления признаков. Представленный подход показал сопоставимую по сравнению с [27] точность и скорость обработки при тестировании на RPi 3.

В табл. 1 собрана информация по рассмотренным решениям для детектирования дыма для устройств с низкой вычислительной сложностью.

Табл. 1. Метрики и параметры легковесных алгоритмов обнаружения дыма

Алгоритм	FPS	Устройство	Размер кадров, пк	Метрики точности, %					
				FPR	FNR	AC	R	P	F ₁
Легковесная CNN [55]	24	RPi 4	64×64	1,9	0,8	97,9	97,2	98,3	97,9
Облегченная YOLOv2 [28]	21	JN	128×128	3,4	2,9	96,8	97,0	97,0	95,4
Фильтр Калмана, цветовая сегментация, CNN [27]	9	RPi 3	50×50	-	-	84,4	-	86,4	88,4
	55	JN							
AdViSED [6]	10	RPi 3	320×240	-	-	85,0	-	88,9	87,5

Можно заключить, что в легковесных решениях для SBC преобладают вычислительно дешевые алгоритмические операции, а использование обучаемых моделей ограничено вариациями архитектуры CNN и моделями на их основе. С целью снижения вычислительной нагрузки часто выполняется предварительное выделение RoI, а затем полученные фрагменты подаются на вход классификатора. Входные данные масштабируются для сохранения высокой скорости обработки данных в ущерб точности. Размер датасетов, используемых для обучения, обычно не превышает тысячи изображений. Временной анализ в большинстве случаев не выполняется или ограничивается простым накоплением пространственных признаков, поэтому модели для анализа временных рядов, такие как LSTM, не используются.

3.3.2. Алгоритмы со средней вычислительной сложностью

По относительно небольшому количеству обучаемых параметров в случае использования моделей ИНС, порядка нескольких миллионов, можно выделить категорию алгоритмов средней вычислительной сложности для обнаружения дыма. Как правило, для тестирования таких реализаций подходят видеокарты NVIDIA серии 10, реализующие проприетарную технологию Compute Unified Device Architecture (CUDA) [64] для управления параллельными вычислениями.

Например, в [47] для обнаружения дыма предлагалось в беспилотных летательных аппаратах (БПЛА) использовать модель на основе YOLOv10 с интегрированным модулем внимания. Архитектура обрабатывала изображения размером 1280×1280 пк, достигая скорости обработки 9 к/с при точности 79,3% mAP на устройстве с GPU NVIDIA GeForce RTX 1080Ti.

Также для применения на БПЛА в [36] описан алгоритм обнаружения дыма от лесных пожаров на основе YOLOv8. Модификации включали в себя облегченную сверточную сеть GSConv [65] и двухэтапный модуль внимания BiFormer [66] с целью улучшения способности модели выделять дым на фоне схожих объектов. Для обучения использовался набор данных из 6000 изображений, половина которых содержала дым. В результате модель достигла точности 79,4% mAP при скорости обработки свыше 100 к/с на устройстве с GPU NVIDIA GeForce GTX 1080 Ti.

Схожий подход представлен в [67], где описана легковесная модель LUFFD-YOLO на основе YOLOv8n с модификациями, направленными на ускорение обработки. Использовались датасеты: M4SFW [68] с 4000 изображений и FLAME [69] с 1000 изображений, а также собственный набор данных. Достигнутая точность

составила 86,3 % mAP, однако тестирование производилось в основном на синтезированных данных небольшого объема, что может не в полной мере отражать качество модели.

Подход с применением механизма внимания на основе YOLOX-L с 4 млн параметров предложен в [34], обучение производилось на 2200 изображениях с дымом. Результирующая точность облегченной модели составила 87,5% mAP со скоростью обработки 26 к/с. Важной особенностью предложенного подхода являлся алгоритм удаления тумана (дефоггинг) методом темного канала (Dark Channel Prior, DCP):

$$J = A + (I - A) / \max(t, t_0), \quad (15)$$

где J – изображение без тумана, A – уровень яркости тумана, t – прозрачность тумана, t_0 – минимальный порог значения прозрачности, константа.

DCP основан на атмосферной модели рассеяния (Atmospheric Scattering Model, ASM) [17, 40]:

$$I_f = J \times t + A \times (1 - t), \quad (16)$$

где I_f – изображение с туманом.

Удаление тумана позволило улучшить точность модели на 2,46 %. Однако у метода есть ограничения: он неприменим для ярких участков, например, неба; требует точной оценки A ; может привести к артефактам в областях с низкой контрастностью. Тем не менее DCP является одним из немногих алгоритмических методов, которые позволяют на основе цветовой информации отделить дым от другого дымоподобного объекта.

Схожая задача удаления дымки (haze) на основе атмосферной модели решалась при помощи DCP с целью улучшения качества детектирования дыма в [38]. Для этого использовалась гибридная модель, включающая оптический поток и FCN (Fully Convolutional Network), тестирование которой производилось на устройстве с GPU NVIDIA GTX 1080 Ti, а результирующая точность составила 95,4 % F_1 при скорости обработки 240 к/с.

В [43] архитектура WSDD-Net на основе DenseNet, которая занимала объем памяти около 40 МБ, достигла 99,3 % F_1 , выполняя обработку кадров со скоростью 100 к/с на GPU NVIDIA GeForce RTX 1080 Ti. Высокая точность была достигнута за счет предварительной сегментации объектов по признаку цвета.

Производятся попытки создания облегченных моделей путем конструирования собственных эффективных архитектур, например: объединение модулей 3D-CNN и LSTM в [16] с 4,6 млн параметров; STCNet [10] с 3,7 млн параметров; LNet-v2 [17], которая объединяет CNN LNet с модулями YOLOv3; TANet [40] с 3,2 млн параметров и механизмом внимания, которая фокусируется на анализе текстуры дыма; облегченная модель EfficientNet-B0 [30] за счет мозаичной обработки 3×3 с 5,4 млн параметров.

Таким образом, реализации со средней вычислительной сложностью подходят для применения в персональных устройствах, однако многие из них так же ограничены использованием технологии CUDA. Наиболее популярная архитектура YOLO и ее модификации. Широко используются авторские модели на основе CNN.

3.3.3. Алгоритмы с высокой вычислительной сложностью

Алгоритмы с высокой вычислительной сложностью, «высокопроизводительные» (high-performance) реализации, обеспечивают превосходящую точность, но доступны на устройствах с большим объемом оперативной памяти и оснащенных мощными GPU. Как правило, в исследованиях используются видеокарты NVIDIA серии 20 и выше, а количество обучаемых параметров в моделях ИНС составляет десятки миллионов.

Так, в [37] разработана модель на основе YOLOv5 MVMNet с 23 млн параметров, обученная на датасете из 7000 изображений. Ее точность составила 78,9 % mAP при обработке изображений размером 640×640 пк со скоростью 122 к/с на устройстве с GPU NVIDIA GeForce RTX 3090. Разработанная архитектура с помощью механизма внимания учитывала такие признаки, как цвет дыма, его текстуру и пространственные характеристики, для чего использовалось многонаправленное обнаружение через обратную регрессию якорных рамок. При этом данные обрабатывались в различных направлениях с целью получения большего количества пространственных признаков. Для получения адекватной оценки эффективности модели выбирались кадры, поступающие на вход каждые 4 с, однако временной анализ не применялся. Таким образом, в качестве адаптации к задаче обнаружения дыма реализован учет наклонных объектов и раздельное извлечение признаков в горизонтальной и вертикальной плоскостях. Следует отметить уникальный подход увеличения количества каналов входных RGB-изображений с 3 до 12 за счет уменьшения их пространственного размера. Авторы [37] предположили, что данный метод повышает эффективность извлечения текстурных признаков.

В [18] разработана модель SF-YOLO с модулем внимания и 14 млн параметров на основе архитектуры YOLOv11 для обнаружения дыма и пламени на видеокдрах. Механизм внимания был добавлен для решения таких проблем, как перекрытие объектов дыма, их малый размер при сложном динамическом фоне на кадрах и обработка нечетких границ областей задымления. Модель обучалась на комбинированном датасете из реальных и синтезированных данных M4SFWD [68]. Обучение и тестирование производилось на устройстве с GPU NVIDIA RTX 3060, точность составила 45,85 % mAP.

Также механизм внимания для обнаружения дыма используют трансформерные модели на основе архитектуры DETR (Detection Transformer) [70], адаптирующей attention-подход для работы с изображениями.

Например, в [31] использовалась стандартная модель DETR с 41 млн параметров для анализа отдельных кадров видеопоследовательности с дымом на основе датасетов HPWREN [71] и AlertWildfire [72]. Достигнутая точность составила 43,2 % mAP при скорости обработки меньше 1 к/с. Была использована подготовленная система, в которой весовые коэффициенты трансформера взяты из предобученной модели, а модуль классификации заменен на одноклассовый, поскольку в оригинальной реализации DETR выполняет детектирование 80 классов COCO, среди которых нет дыма.

В [35] разработана модель обнаружения дыма от лесных пожаров на основе Deformable DETR [73], которая построена на основе базовой архитектуры DETR. Улучшения заключались в добавлении ряда модулей: блок извлечения многомасштабных локальных признаков со свертками различного размера и блок плотного пирамидального пулинга с картами признаков различного размера. Также был модифицирован выход модели путем итеративной коррекции предсказанных ограничивающих рамок объектов путем учета данных со всех слоев декодера. Для обучения собран закрытый датасет из 10250 изображений. Обученная модель достигла 49,7 % mAP, при этом метрика AP₅₀ составила 88,4 %, что выше, чем у оригинальных моделей DETR и Deformable DETR. Такое соотношение mAP и AP₅₀ отражает ситуацию, при которой объекты в целом обнаруживаются хорошо, но точность их локализации при этом относительно низкая. Скорость обработки составила 4 к/с на устройстве с GPU NVIDIA RTX 2080 Ti. Таким образом, в модели [35] реализованы модификации, направленные на улучшение точности детектирования дыма с учетом его сложной формы и небольшого размера: многомасштабная обработка признаков и алгоритм итеративного уточнения ограничивающих рамок.

В [3] в модели RT-DETR (Real-Time DETR) интегрированы модуль триплетного внимания, пирамидальная сеть слияния признаков и модуль Adown, адаптивно снижающий вычислительную сложность архитектуры за счет уменьшения разрешения карт признаков. Триплетное внимание выполняло обработку данных в пространственных и канальной плоскостях данных с целью наиболее эффективного извлечения признаков. Система запускалась на устройстве с GPU NVIDIA RTX 3080 и достигла скорости обработки 45 к/с при точности 96,1 % AP₅₀ даже без учета временных признаков.

В [29] совместно были применены архитектуры CNN (ResNet34), LSTM и ViT [74] (Visual Transformer) в модели SmokeNet, а обнаружение выполнялось для всего изображения и 45 блоков регулярной сетки (тайлов), на которые оно разбивалось в процессе первого этапа обработки. Общее количество обучаемых параметров модели составило 57 млн, а обучение производилось на датасете HPWREN [71]. Тайлы каждого кадра обрабатывались независимо на модулях CNN, а затем выделенные карты признаков для одинаково расположенных в пространстве тайлов из разных кадров объединялись и анализировались модулями LSTM и ViT. Способ имеет явный недостаток, связанный с небольшим размером патчей, которые могут быть неправильно классифицированы, т.к. они обрабатываются отдельно, без пространственного контекста. Как следствие, авторами [29] отмечено большое количество ложных срабатываний на низкие облака. Скорость обработки SmokeNet составила 19 к/с при точности обнаружения дыма 82,6 % F₁. Большое время на обработку одного кадра является следствием перегруженности алгоритма нейросетевыми моделями, которые частично дублируют друг друга. Так, архитектуры ViT и CNN функционально идентичны друг другу, поскольку обе выполняют классификацию изображений.

Таким образом, можно заключить, что модели, рассчитанные на использование в мощных вычислительных устройствах, используют механизмы внимания для повышения точности детектирования дыма, но, за исключением отдельных работ, не применяют временной анализ, полагаясь только на выделение и обработку пространственных признаков, включая rule-based методы в качестве предобработки. Использование технологии CUDA ограничивает их применение на низкопроизводительных встраиваемых устройствах или компьютерах без ее поддержки. Можно отметить, что модели на основе YOLO, даже с использованием механизмов внимания, способны демонстрировать высокую скорость обработки. При этом архитектуры на основе трансформеров характеризуются значительно меньшим быстродействием и требуют модификаций для использования в режиме реального времени. Комбинация архитектур позволяет достичь увеличения точности обнаружения дыма, но негативно влияет на скорость обработки данных.

В табл. 2 приведены показатели рассмотренных методов детектирования дыма за исключением алгоритмов с низкой вычислительной сложностью, указанных отдельно в табл. 1. Методы упорядочены по возрастанию следующим образом: по метрике mAP, затем, в случае ее отсутствия, по F₁, затем по AC. Следует отметить, что для обучения и тестирования представленных моделей и алгоритмов использовались разные наборы данных, включая закрытые, что не позволяет выполнить их строгое сравнение, однако дает возможность оценить в общем производительность и точность различных типов архитектур и подходов.

Анализ таблицы показывает, что по отношению к некоторым архитектурам представленные результаты имеют высокую достоверность. Например, модели DETR применялись в [31] и [35] разными авторами на различных данных, но значения их результирующих метрик для одной и той же задачи оказались очень близки друг к другу. Многие модели на основе YOLO также оказались сгруппированы по метрике mAP в единый кластер. Это свидетельствует о том, что модификации повышают качество оригинальных архитектур в определенных пределах, но не способны радикально улучшить эффективность обнаружения дыма.

Табл. 2. Показатели скорости и точности алгоритмов детектирования дыма

Метод	FPS	Метрики, %									
		TPR	FPR	F ₁	AC	mAP	AP ₅₀	AP ₇₅	AP _S	AP _M	AP _L
DETR [31]	1	-	3,0	96,9	96,8	43,2	79,0	-	38,6	67,6	84,1
DETR [35]	-	-	-	-	-	44,2	84,8	-	27,4	-	-
LFNet-v2 [17]	34	-	-	-	-	45,2	67,1	52,7	24,7	56,2	54,3
Deformable DETR [35]	-	-	-	-	-	45,5	85,8	-	33,5	-	-
SF-YOLO [18]	-	-	-	77,2	-	45,9	78,2	-	-	-	-
YoloV5m [33]	-	-	-	-	-	46,6	77,7	-	-	-	-
Deformable DETR-based [35]	4	-	-	-	-	49,7	88,4	-	36,9	-	-
YOLOv8 [36]	-	-	-	-	-	76,1	84,3	77,4	69,5	75,6	89,3
MVMNet [37]	122	-	-	-	-	78,9	88,1	81,2	-	-	-
YOLOv10-based [47]	9	-	-	76,1	-	79,3	-	-	-	-	-
YOLOv8-based [36]	100	-	-	-	-	79,4	87,1	82,4	71,3	78,5	92,6
YOLOv8n [67]	-	-	-	76,6	-	82,1	-	-	-	-	-
YOLOv8n-based [67]	-	-	-	81,0	-	86,3	-	-	-	-	-
YOLOX-L-Light [34]	26	-	-	-	-	87,5	-	-	-	-	-
Deformable DETR [3]	15	80,2	-	78,2	-	-	79,8	-	-	-	-
RGB-TC [26]	-	79,2	-	82,3	-	-	-	-	-	-	-
Оптический поток+FCN [38]	240	83,2	-	82,4	95,4	-	-	-	-	-	-
CNN+LSTM+ViT [29]	19	76,5	-	82,6	83,5	-	-	-	-	-	-
STCNet [10]	109	-	-	86,8	-	-	-	-	-	-	-
EfficientNet-B0 [30]	-	85,3	3,1	88,2	-	-	-	-	-	-	-
RT-DETR [3]	45	91,6	-	96,1	-	-	96,1	-	-	-	-
WSDD-Net [43]	100	99,7	0,34	99,3	99,2	-	-	-	-	-	-
YOLOv2 [28]	21	-	3,4	-	96,8	-	-	-	-	-	-
Движение+цвет+Inceptionv3 [13]	-	-	14,3	-	88,7	-	-	-	-	-	-

Значения mAP рассмотренных подходов варьируют в широком диапазоне значений, от 43 % до 87 % в различных рассмотренных моделях. Однако метрика AP₅₀ демонстрирует большую стабильность и для большинства решений находится в диапазоне 79 – 88 %. Это свидетельствует о том, что различные предложенные подходы примерно одинаково справляются с грубой локализацией дыма. Из табл. 2 также следует проблема обнаружения небольших объектов дыма, поскольку метрики APs в среднем имеют низкие значения.

Методы, представленные в табл. 1, используют метрики, аналогичные подходам, представленным в табл. 2, и отличаются только по области применения на низкопроизводительных устройствах, поэтому данные обеих таблиц могут быть сопоставлены между собой. Сравнение табл. 1 и табл. 2 показывает, что алгоритмы и модели как с низкой, так и с высокой вычислительной сложностью могут быть близки по точности.

Модели YOLO характеризуются большим быстродействием, чем архитектуры на основе DETR. Однако показатели быстродействия моделей во многом зависят от используемого оборудования.

3.4. Сценарии динамической сцены при обнаружении дыма

Задача обнаружения дыма на видеопоследовательности с помощью СВН является комплексной, поскольку обуславливается множеством факторов, влияющих на выбор программных инструментов и их параметров, которые должны в заданных условиях демонстрировать высокую эффективность. В исследовательских работах по тематике предложено большое количество сценариев, которые представляют собой сочетание значимых факторов, определяющих особенности реализаций систем обнаружения дыма.

В соответствии с классификацией, предложенной в [44], существующие методы обнаружения дыма могут быть отнесены к одному из четырех сценариев динамической сцены: малая дальность, низкая активность (Short Range Low Activity, SRLA), большая дальность, низкая активность (Long Range Low Activity, LRLA), малая дальность, высокая активность (Short Range High Activity, SRHA) и большая дальность, высокая активность (Large Range High Activity, LRHA).

В [6] пожары классифицируются как ближние и дальние, при этом ближними считаются области задымления на расстоянии менее 100 метров, что справедливо для закрытых помещений и отсеков внутри транспорта. Для каждого сценария предполагается собственный набор инструментов с целью обнаружения дыма на основе данных СВН.

В [44] предложена классификация пожаров по критерию удаленности объекта от точки наблюдения. Определено, что малая дальность соответствует разрешению объекта дыма более 50 пк/м, а большая дальность характеризуется меньшими значениями. Алгоритмы, разрабатываемые с учетом этого фактора, имеют ряд особенностей.

Так, при малой дальности, включая сценарии SRLA и SRHA, характерный размер входных изображений для систем обнаружения варьирует от 128×128 пк до 336×336 пк, т.е. данные имеют низкое разрешение, достаточное в рамках задачи.

При большой дальности, включая сценарии LRLA и LRHA, дым занимает слишком малую область кадра, поэтому выполняется предварительное выделение RoI, а размер входных изображений достигает 3072×2048 пк, поскольку высокое разрешение должно компенсировать небольшую площадь объектов дыма.

Согласно [44], активность фона оказывает значительное влияние на сложность обнаружения дыма и, в зависимости от удаленности точки наблюдения от пожара, требует применения различных подходов.

При малой активности на малом расстоянии (сценарий SRLA) существующие подходы, как правило, выполняют анализ отдельных изображений с помощью ограниченного числа алгоритмов для выделения признаков, поскольку увеличение их числа повышает вычислительную нагрузку без значительного улучшения производительности. Наиболее широко используются решения на основе CNN: VGG16 [19], DenseNet [43], EfficientDet, легковесные модели YOLO [33], а также алгоритмы компьютерного зрения, включая вычитание фона, оптический поток [59], суперпиксельную [39] и цветовую сегментацию [43]. Временной анализ не производится или ограничивается несколькими кадрами, как в [59] для вычисления оптического потока. С увеличением расстояния (сценарий LRLA) используются более сложные архитектуры: поздние модели YOLO и модели на их основе, например, YOLOv4 [75], YOLOv5 [37] и YOLOX-L [34], решения на базе DETR [31] или собственных трансформерных моделей [76]. В качестве входных данных используется временное окно из 2 – 3 видеок кадров [29], но чаще анализируется только одно изображение.

При высокой активности фона (сценарии SRHA и LRHA) в реализациях стараются учитывать временной контекст, поскольку на первый план выходит задача точного выделения детектируемых объектов. Для обработки временных рядов используется архитектура LSTM в связке с CNN [77] и алгоритмическими методами компьютерного зрения. При наличии временного анализа используется до нескольких десятков кадров, например 36 кадров для модели RGB-TC ConvNet [26] и до 30 кадров для [78]. В случаях, когда доступно лишь несколько пикселей RoI, необходимая для принятия решения информация может быть получена только с использованием длительного временного анализа. В этом случае применяются методы, использующие алгоритмы многокадровой локализации и распознавания дыма. Преобладают подходы, основанные на архитектуре YOLO: YOLOv5 [79] или гибридная 3D-CNN модель [80], основанная на YOLO. В [81] моделью ABi-LSTM учитывается 20 кадров.

На состав обучающих данных для моделей ИНС, параметры цветовой сегментации и применяемые меры борьбы с ложными срабатываниями в значительной мере влияет тип динамической сцены, определяемой наблюдаемой локацией. Например, для лесных пожаров в качестве дымоподобных объектов могут выступать туман и облака, а для городских условий наблюдения также должны учитываться машины, люди и многочисленные источники освещения.

Наибольшее число работ [13, 34, 38, 31, 47, 18, 43, 36, 17, 45] посвящено обнаружению дыма в условиях лесного пожара, где точка наблюдения, как правило, располагается на большом удалении от места возникновения пожара, а дым занимает малую область кадра и является единственным наблюдаемым визуальным признаком.

Работа [26] посвящена задаче обнаружения промышленных выбросов в виде дыма и пара. Датасет, сформированный в [26], выступает в качестве основных данных в [10]. Важной особенностью этой работы является попытка на основе незначительных различий пространственно-временных признаков выполнить разделение дыма и пара, которые выглядят практически идентично, и в контексте промышленных объектов могут часто присутствовать совместно в сцене. В [55] в качестве входных данных используются изображения пожаров на трубопроводах, поскольку они являются потенциальными источниками возгорания при слабом освещении, следовательно, обнаружение дыма в этих условиях представляет собой отдельную задачу повышенной сложности.

4. Обучение нейронных сетей и наборы изображений с дымом

Обучение моделей, выполняющих детектирование дыма, характеризуется рядом особенностей, продиктованных спецификой задачи или структурой входных данных.

В работах, посвященных обнаружению пожара на основе видеоданных или отдельных изображений, дым обнаруживается совместно с пламенем в качестве единого объекта «пожар» (fire) [33] или как отдельный класс «smoke» совместно с классом «пламя» [18] или без него. В некоторых случаях используется негативный класс «no fire», «no smoke» или аналогичные по сути классы с другим наименованием [27]. Элементы изображений, которые размечаются негативным классом, представляют собой визуально похожие на дым объекты, что позволяет более эффективно обучать модели и тестировать алгоритмы с низким уровнем ложных срабатываний в приближенных к реальным условиям [28]. Также иногда размечаются такие классы, как «лес» и «небо», наличие которых предполагается при эксплуатации системы детектирования [17]. Для балансировки классов многие изображения в данных для обучения не содержат никаких размеченных объектов [19].

В других случаях данные, содержащие дымоподобные объекты, вообще не размечаются, используются т.н. пустые метки (dummy annotations). В [31] пустые метки использовались для изображений без дыма, которые необходимы для предотвращения склонности модели обнаруживать дым там, где его нет. Однако авторы столкнулись с проблемой программной реализации, т.к. модель DETR не была способна их обрабатывать. В качестве решения на такие изображения добавлялась минимальная рамка 1×1 пк в случайном месте с классом «no smoke».

Наиболее значимой проблемой является недостаток реальных качественных наборов данных в открытом доступе [16]. Анализ популярных открытых датасетов общего назначения, рассчитанных на применение пространственно-временного анализа для детектирования объектов с локализацией, показал, что дым в них не используется как аннотированный класс никогда. Следовательно, они не могут использоваться для обучения. Большинство специальных наборов данных, содержащих объекты дыма, представляют собой неразмеченные фрагменты видеозаписей с классификационными метками или отдельные аннотированные рамками изображения, не образующие последовательности. В значительном количестве данных из проанализированных наборов пожар зафиксирован уже на поздних стадиях. Это является существенным недостатком, поскольку дым имеет разные визуальные признаки на разных стадиях горения [17]. Многие датасеты являются смешанными, т.е. содержат как дым, так и пламя. Таким образом, недостаток подходящих наборов данных представляет собой проблему при создании системы детектирования дыма на основе искусственных нейронных сетей. В качестве решения предлагается их синтезировать следующими путями: с помощью состязательных сетей [29]; моделировать с помощью графических или игровых движков, например NVIDIA Omniverse, Unreal Engine и Unity. В [16] отмечено, что оптимальное соотношение синтезированных и реальных данных составляет около 20 – 50 %. Качество синтезированных данных предлагается оценивать с помощью метрик SSIM (Structural Similarity Index) или FID (Frechet Inception Distance) [82]. Однако эффективность обучения ограничена разницей во внешнем виде между реальными и синтетическими образцами дыма, что может приводить к значительному снижению производительности [83, 40].

Другая особенность заключается в несбалансированности датасетов, которая может быть решена путем добавления объектов уточняющих и негативных классов или корректировкой коэффициентов при вычислении ошибки во время обучения. Так, в [17] использовался частично синтезированный датасет из 4000 изображений, который содержал классы «дым», «без дыма» «туман» и «дым + туман». В [37] объекты дыма на кадрах были разбиты на три отдельно детектируемых класса: «четкий дым», «рассеянный дым» и «смешанный дым». В [31] дым классифицировался по плотности как «низкая плотность», «средняя плотность» и «высокая плотность», однако мультиклассовый подход показал значительно худшие результаты в сравнении с одноклассовым по причине повышенной сложности задачи. В [40] выполнена сегментация изображений с градацией плотности дыма в диапазоне значений от 0 до 1. В [16] классифицировался 21 уровень непрозрачности дыма с шагом 5 %.

4.1. Обучение моделей ИНС для обнаружения дыма

Распространенной практикой является изменение способа вычисления ошибки при обучении. В [29] для оценки потерь использовалась взвешенная бинарная кросс-энтропия (Weighted Binary Cross-Entropy, *WBCE*).

$$WBCE = -(\omega_1 \times y_{true} \times \log(p) + \omega_0 \times (1 - y_{true}) \times \log(1 - p)), \quad (17)$$

где y – истинная метка, p – предсказанная вероятность, что изображение содержит дым, ω_0 , ω_1 – коэффициенты положительного и отрицательного классов.

Вес положительных примеров с дымом был увеличен в 5 раз, для обучения по тайлам в 40 раз, исходя из предпосылки, что файлов с негативными примерами в пропорции больше, чем для полных изображений. Аналогичная оценка была применена в [43].

В [36] использовалась метрика Wise-IoU ($W-IoUv3$) [84], с которой приоритет отдавался образцам среднего качества для избегания переобучения и влияния низкокачественных прогнозов. Значение потерь $L_{W-IoUv3}$ вычисляется следующим образом:

$$L_{W-IoUv3} = (\beta / (\delta \times \alpha^{\beta-\delta})) \times (e^{(d^2/(w^2+h^2))} \times L_{IoU}), \quad (18)$$

где α , β , δ – параметры коэффициента фокусировки, d – расстояние между центрами рамок, w и h – ширина и высота наименьшего описывающего обе рамки прямоугольника.

В [47] применялась метрика Shape-IoU ($S-IoU$) с целью адаптации системы обнаружения дыма к задаче локализации объектов с нестандартными формами и измененная мера ошибки на основе этой метрики L_{S-IoU} :

$$L_{S-IoU} = 1 - IoU + d_{shape} + 0,5 \times \Omega_{shape}, \quad (19)$$

где d_{shape} – взвешенное расстояние между центрами рамок, Ω_{shape} – штраф за различие формы рамок.

В [16] была задействована модифицированная метрика фокальных потерь (Focal Loss, FL), которая может быть полезна при дисбалансе классов. FL определяется следующим образом:

$$FL = -\alpha \times y_{true} \times (1 - y_{pred})^\gamma \times \log(y_{pred}), \quad (20)$$

где α , γ – гиперпараметры, контролирующие баланс между классами, y_{pred} – предсказанная метка.

Применение данной метрики позволило улучшить показатели точности модели на 5,2 % в [16].

В [37] использовалась метрика Poly-IoU ($P-IoU$) для оценки точности модели, которая предсказывает наклонные рамки для обнаруженных объектов дыма, попиксельно вычисляя область их пересечения $S_{intersec}$.

$$S_{intersec} = 0,5 \times \sum_{i=0}^{n-1} x[i] \times y[(i+1) \bmod n] - y[i] \times x[(i+1) \bmod n], \quad (21)$$

где $x[i]$, $y[i]$ – координаты вершин многоугольника пересечения, n – количество вершин многоугольника.

$P-IoU$ вычисляется следующим образом:

$$P-IoU = S_{intersec} / (S_{true} + S_{pred} - S_{intersec}). \quad (22)$$

В [3] модифицированная метрика Focaler-IoU ($F-IoU$) применялась для компенсации дисбаланса выборки:

$$F-IoU = \begin{cases} 0, & IoU < d, \\ (IoU - d)/(u - d), & d \ll IoU \ll u, \\ 1, & IoU > u, \end{cases} \quad (23)$$

где u , d – гиперпараметры, контролирующие баланс между классами.

В [18] предлагалось к внедрению сразу несколько модифицированных оценок, включая W-IoU [84] и SCYLLA-IoU [85], взвешенная сумма которых применялась в качестве результирующего значения ошибки модели. Применение этих оценок может снизить влияние выбросов и улучшить локализацию, учитывая угол, расстояние и соотношение сторон рамок.

4.2. Наборы данных для обучения моделей ИНС

Датасет FiSmo (FireVid) [86] (рис. 8) содержит 27 видеозаписей с разметкой в CSV-файлах, которые содержат координаты ограничивающих рамок для кадров с классами «fire», «no-fire» и «ignore». Разрешение файлов варьируется в диапазоне от 320×240 пк до 600×336 пк. Данные представляют записи пожаров, где присутствуют дым и пламя.



Рис. 8. Примеры кадров из датасета FiSmo

Датасет RESCUER [86] (рис. 9), как дополнение набора FiSmo, представляет собой аналогично аннотированный датасет из 62 видеофайлов, содержащих горящие промышленные объекты, однако значительная часть записей демонстрирует процесс тушения, что добавляет к пламени и дыму также объект рассеянной водяной струи. Разрешение видеофайлов RESCUER достигает разрешения FullHD, но также присутствуют записи в разрешении 320×240 пк. К недостатку рассмотренных наборов следует отнести нестандартную разметку, однообразие фонов и в случае RESCUER наличие струй воды, которые затрудняют интерпретацию данных.



Рис. 9. Примеры кадров из датасета RESCUER

Датасет VisiFire [87] (рис. 10) предлагает 17 видеозаписей без разметки, которые частично дублируют содержание FiSmo. Уникальные файлы набора данных имеют разрешение до 720×576 пк. Тематика видеозаписей включает лесные пожары и городской фон, дым и пламя присутствуют одновременно, на отдельных сценах с большим расстоянием от точки наблюдения пламя отсутствует. Использовался в [38].



Рис. 10. Примеры кадров из датасета VisiFire

Датасет D-Fire [88] (рис. 11) содержит 100 записей камер систем видеонаблюдения без разметки, однако многие из них сняты на одном и том же фоне, что снижает разнообразие данных. Записи имеют разрешение FullHD и HD. Кадровая частота большинства видеофайлов равна 2 к/с.

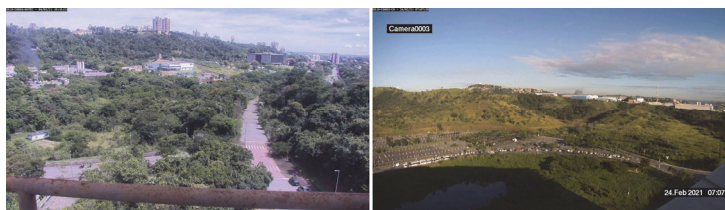


Рис. 11. Примеры кадров из датасета D-Fire

Датасет MIVIA Smoke [89] (рис. 12) разделен по категориям, включая негативные: облака – 19 записей, без дыма – 51 запись, блики – 1 запись, дым с бликами – 15 записей, дым – 59 записей, солнце – 4 записи. Следует отметить, что в рамках каждой категории видеофайлы единообразны, поскольку одни и те же объекты сняты статичной камерой в одинаковых условиях. Разметка по кадрам в датасете отсутствует. Все видеофайлы в наборе имеют разрешение 292×240 пк. Авторами [13] заявлена доработка датасета под названием MIVIA-SDD с попиксельной разметкой, однако в открытом доступе набор не был обнаружен.



Рис. 12. Примеры кадров из датасета MIVIA Smoke

Датасет FireSense [63] (рис. 13) предлагает 3 категории файлов без разметки: дым, пламя и их отсутствие. 13 видеозаписей с урбанистическими фонами и разрешением в диапазоне 320×240 пк до 640×360 пк содержат дым, снятый динамической камерой. Использовался в [27].



Рис. 13. Примеры кадров из датасета Fire Sense

Датасет FURG [90] (рис. 14), в отличие от предыдущих наборов, предоставляет разметку в собственном формате, который предлагает аннотации в формате XML к каждому из 23 видеофайлов. Разметка содержит координаты ограничивающих рамок для кадров с дымом. Все видеозаписи содержат как дым, так и пламя, а их разрешение достигает формата HD, но, как правило, ниже. Ролики сняты динамической камерой и фиксируют в основном крупные пожары либо макеты. Использовался в [33].



Рис. 14. Примеры кадров из датасета FURG

Датасет HPWREN [71] (рис. 15) представлен в виде 470 панорамных видеозаписей без разметки, которые разбиты на отдельные изображения и содержат сцены пожара на открытых локациях, включая дневные и ночные варианты. Кадры имеют высокое разрешение, вплоть до формата 4K. Использовался в [31, 29, 30].



Рис. 15. Примеры кадров из датасета HPWREN

Датасет RISE [26] (рис. 16) содержит 12567 видеозаписей промышленных объектов в течение всех времен года при различных погодных условиях, где каждый клип состоит из 36 кадров 180×180 пк, полученных с

интервалом 10 с. Особенность набора заключается в том, что его разметку выполняло множество людей, включая экспертов и энтузиастов. Разметка представлена меткой бинарной классификации о наличии дыма для всей записи. Датасет обладает двумя недостатками: отсутствие ночных сцен и всего три локации, для которых производилась съемка. Использовался в [10].



Рис. 16. Примеры кадров из датасета RISE

Датасет KMU FIRE & SMOKE [91] (рис. 17) состоит из набора 28 видеозаписей с огнем и дымом в разрешении 320×240 пк, включая сцены в помещении и на открытом воздухе при различных вариантах освещения, а также 10 роликов с негативными примерами, такими как облака, люди и туман. Записи сгруппированы по директориям, однако аннотирующая разметка отсутствует. Использовался в [38].

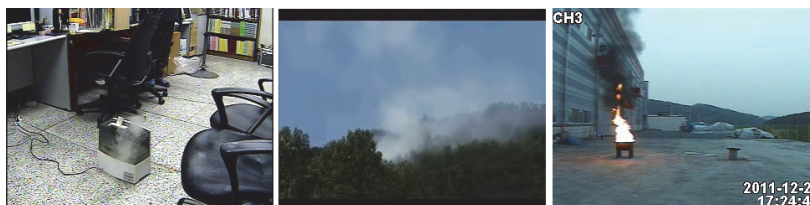


Рис. 17. Примеры кадров из датасета KMU FIRE & SMOKE

Датасет M4SFWD [68] (рис. 18) синтезирован на игровом движке Unreal Engine 5 и содержит 28 покaдрово аннотированных сцен лесных пожаров при различных условиях наблюдения, включая время суток и погодные условия, и при разном количестве источников дыма. Видеоролики имеют разрешение 1480×684 пк и продолжительность несколько минут. Использовался в [18, 67].



Рис. 18. Примеры кадров из датасета M4SFWD

Дополнительно следует отметить ресурс AlertWildfire [72] (рис. 19), который не содержит подготовленных наборов данных, однако предоставляет записи многочисленных камер, выполняющих мониторинг лесных пожаров на территории западной части США. Камеры имеют разрешение 2 Мпк и выполняют съемку с кадровой частотой менее 1 к/с. Данные AlertWildfire использовались в [31].



Рис. 19. Примеры кадров из датасета Alert Wildfire

Значимая часть классификационных моделей глубокого обучения, которые использовались для обнаружения дыма, были предобучены на наборе данных ImageNet [92], например [13, 38, 30, 26, 33].

Сравнение основных данных о датасетах представлено в табл. 3.

На основе анализа рассмотренных датасетов можно заключить, что большая их часть не содержит покaдровой разметки дыма или никакой разметки в принципе. В качестве исключений FiSmo и VisiFire предоставляют покaдровые аннотации в CSV файле, а FURG хранит разметку рамок в формате XML, RISE аннотирован

покадровыми JSON-файлами. Используемые форматы не являются общепринятыми, поэтому требуют создания дополнительной кодовой базы для возможности интерпретации и использования данных. Отмечено небольшое количество уникальных записей и низкое разнообразие сцен, что способно затруднить обучение моделей. Подавляющее большинство видеофайлов имеют низкое разрешение, что также снижает эффективность использования подобных датасетов.

Табл. 3. Наборы данных для обучения моделей ИНС

Датасет	Количество записей	Разрешение записей, пк	Разметка (файл и тип разметки)	Классы	Тип записей
FiSmo [86]	27	от 320×240 до 600×336	CSV (классификация кадров)	fire, no-fire, ignore	реальные
RESCUER [86]	62	от 320×240 до 1920×1080			
VisiFire [87]	17	720×576	-	-	реальные
D-Fire [88]	100	от 1280×720 до 1920×1080	-	-	реальные
MIVIA Smoke [89]	149	292×240	классификация видеозаписей	дым, облака, блики, солнце, не дым	реальные
FireSense [63]	13	от 320×240 до 640×360	классификация видеозаписей	дым, пламя, не дым и не пламя	реальные
FURG [90]	23	до 1280×720	XML (локализация)	пожар	реальные
HPWREN [71]	470	до 4K	-	-	реальные
RISE [26]	12567	180×180	JSON (классификация видеозаписей)	дым	реальные
KMU FIRE & SMOKE [91]	38	320×240	классификация видеозаписей	дым, не дым	реальные
M4SFW [68]	28	1480×684	локализация	дым	синтезированные

4.3. Аугментация изображений в наборах данных с дымом

Аугментации являются способом увеличить разнообразие исходных данных путем их модификации, что особенно актуально в условиях недостатка датасетов. Произвольная форма дыма и широкий диапазон условий его наблюдения диктуют высокое разнообразие применяемых аугментаций.

В табл. 4 представлены используемые в рассмотренных работах аугментации, связанные с воздействием на исходное изображение или видеокadres. Аугментации упорядочены по возрастанию слева направо по количеству работ, в которых они были использованы. Работы упорядочены по убыванию сверху вниз по количеству применяемых аугментаций.

Табл. 4. Аугментации изображений, применяемые для обучения моделей, которые выполняют обнаружение дыма

Работа	Аугментации								
	Отражение	Мозаика	Изменение яркости и контраста	Кропирование	Изменение размера	Поворот	Шум	Сдвиг	Размытие
[3]	+	+	+	-	+	+	-	+	-
[67]	+	+	-	-	-	+	+	-	-
[17]	+	-	+	-	-	-	+	-	+
[71]	+	+	-	+	+	-	-	-	-
[10]	+	-	+	+	+	-	-	-	-
[18]	-	+	+	+	+	-	-	-	-
[30]	-	-	-	+	-	-	+	+	+
[16]	+	-	+	+	-	-	-	-	-
[47]	+	+	+	-	-	-	-	-	-
[36]	+	-	-	-	-	+	-	-	-
[34]	-	+	-	-	-	-	-	-	-
[31]	-	+	-	-	-	-	-	-	-

Анализ табл. 4 показывает, что наиболее часто используемые изменения включают зеркальное отражение изображений по горизонтали и вертикали, объединение изображений (мозаика), изменение яркости и контраста, а также кропирование. Чаще всего применяется 3 – 4 аугментации. Большее количество изменений, вероятно, негативно сказывается на способности моделей к обучению.

Таким образом, при обучении моделей используются различные стратегии: модификация оценки потерь, изменение весовых коэффициентов для различных классов, разнообразные способы учета негативных классов, введение уточняющих классов. Используемые в датасетах аугментации включают в себя: отражение

изображений, мозаику, изменение яркости и контраста, кропирование, размытие, шум. Количество, объем и качество существующих датасетов с объектами дыма являются неудовлетворительными для обучения крупных моделей глубокого обучения. Для решения проблемы предлагается синтезировать наборы данных.

Заключение

В работе представлен аналитический обзор методов и алгоритмов детектирования дыма на видеопоследовательностях для раннего обнаружения пожаров с применением систем видеонаблюдения. Выявлены визуальные признаки дыма, которые могут быть получены путем пространственно-временного анализа видеопоследовательностей и позволяют отличать его от схожих явлений. Представлены ограничивающие и мешающие факторы, влияющие на отображение визуальных признаков дыма реального мира на видеокадрах. Рассмотрены методы для выделения визуальных признаков дыма с учетом указанных ограничений, произведено их комплексное сравнение. Анализируются подходы, основанные на классических алгоритмах компьютерного зрения и на машинном обучении, включая сверточные нейронные сети и трансформеры. Алгоритмы, основанные на машинном обучении, систематизированы с учетом вычислительных затрат. Рассмотрены метрики для оценки эффективности детектирования и локализации областей дыма на кадрах, определены наиболее значимые из них. Представлены существующие наборы данных машинного обучения моделей обнаружения дыма, выявлены их недостатки, и рассмотрены способы их устранения. Приведены варианты аугментации изображений для расширения обучающей выборки.

Выполненный анализ показал, что в настоящее время для встраиваемых в системы видеонаблюдения с ограниченными вычислительными ресурсами модулей актуально применение классических алгоритмов компьютерного зрения и облегченных нейросетевых моделей. Для обеспечения высокой точности обнаружения дыма и снижения вероятности ложной тревоги за счет свойств адаптивности необходимо использовать методы с высокой вычислительной сложностью на основе машинного обучения. Показано, что наиболее популярными архитектурами являются модели YOLO и трансформеры на основе DETR, а распространенные модификации архитектур включают многомасштабную обработку и внедрение модулей внимания. Извлечение временных признаков на основе последовательности кадров для обнаружения дыма применяется ограниченно по причине увеличения вычислительной сложности, однако позволяет достигать высокой точности в сложных условиях.

К таким случаям относится сценарий, сочетающий большое расстояние до наблюдаемых объектов дыма и наличие большого количества помех в виде движущихся и похожих на дым объектов. Традиционно при анализе временных рядов используется архитектура LSTM и ее модификации в связке с другими моделями ИНС, также применяются трехмерные сверточные сети. Большие вычислительные затраты при таком подходе предполагают использование высокоскоростной аппаратной базы, включая графические процессоры или специализированные нейропроцессоры, что значительно увеличивает стоимость систем. Для снижения вычислительной сложности моделей первые этапы выделения областей дыма выполняются с помощью классических алгоритмов компьютерного зрения, таким образом реализуются гибридные системы. Поэтому актуальным является дальнейшее развитие алгоритмов обнаружения дыма с улучшенными качественными характеристиками и удовлетворительными временными затратами.

Также актуальным пока является формирование баз данных с изображениями дыма на ранних стадиях его возникновения в различных условиях съемки.

References

- [1] Fouziya Sulthana S, Wise A, Ravikumar C, Anbazhagan R, Idayachandran G, Pau G. Review study on recent developments in fire sensing methods. *IEEE Access* 2023; 11: 90269-90282. doi:10.1109/ACCESS.2023.3306812. [web:339]
- [2] Sokolov S, Zhilenkov A, Chernyi S, Nyrkov A. Assessment of the impact of destabilizing factors in the main engine shaft of the adaptive speed controller. *Procedia Comput Sci* 2018; 125: 420-426. doi:10.1016/j.procs.2017.12.055.
- [3] Yang L, Cheng Y, Xu F, Li B, Li X. Real-time smoke detection in surveillance videos using an enhanced RT-DETR framework with triplet attention and HS-FPN. *Fire* 2024; 7(11): 387. doi:10.3390/fire7110387.
- [4] Favorskaya M, Pyataeva A, Popov A. Spatio-temporal smoke clustering in outdoor scenes based on boosted random forests. *Procedia Comput Sci* 2016; 96: 762-771. doi:10.1016/j.procs.2016.08.231.
- [5] Pincott J, Tien PW, Wei S, Calautit JK. Indoor fire detection utilizing computer vision-based strategies. *J Build Eng* 2022; 61: 105154. doi:10.1016/j.jobe.2022.105154.
- [6] Gagliardi A, Saponara S. AdViSED: advanced video smoke detection for real-time measurements in antifire indoor and outdoor systems. *Energies* 2020; 13(8): 2098. doi:10.3390/en13082098.
- [7] Pyataeva AV. Research of methods and development of algorithms for smoke detection in open spaces using video sequences [PhD thesis, in Russian]. Krasnoyarsk: Siberian Federal University; 2016.
- [8] Yang Y, Fu T, Song F, Song X, Wang XL, Wang YZ. Wood-burning processes in variable oxygen atmospheres: thermolysis, fire, and smoke release behavior. *Polym Degrad Stab* 2022; 205: 110158. doi:10.1016/j.polymdegradstab.2022.110158.
- [9] Bohush RP, Ablameyko SV. Forest fire smoke detection algorithm in video image [in Russian]. *J Belarusian State Univ Math Inform* 2021; (1): 91-101. doi:10.33581/2520-6508-2021-1-91-101.
- [10] Cao Y, Tang Q, Lu X, Li F, Cao J. STCNet: spatio-temporal cross network for industrial smoke detection. *arXiv* 2020; arXiv:2011.04863. Available at: <https://arxiv.org/abs/2011.04863>.

- [11] GOST R 70861-2023. Forest protection from fire. Terms and definitions [in Russian]. Moscow: IPK Izdatelstvo Standartov; 2023.
- [12] Ye S, Adamovsky Y, Chen H, Bohush R, Ablameyko S. Real-time smoke detection in video based on two-step selection of regions of interest and directional movement analysis. *Pattern Recognit Image Anal* 2024; 34(4): 1323-1332. doi:10.1134/S1054661824701396. [web:340]
- [13] Carletti V, Greco A, Saggese A, Vento B. A smart visual sensor for smoke detection based on deep neural networks. *Sensors* 2024; 24(14): 4519. doi:10.3390/s24144519.
- [14] Yuan C, Liu Z, Zhang Y. Learning-based smoke detection for unmanned aerial vehicles applied to forest fire surveillance. *J Intell Robot Syst* 2018; 93(1-2): 337-349. doi:10.1007/s10846-018-0803-y.
- [15] Adamovskiy Y, Bohush R. Real-time algorithm for light gray smoke detection in video sequences. *Lect Notes Electr Eng* 2024; pp. 535-542. doi:10.1007/978-981-97-6934-6_64.
- [16] Seefried E. Smoke+: a video dataset for automated fine-grained assessment of smoke opacity [PhD thesis]. Fort Collins, CO: Colorado State University; 2024.
- [17] Wang C, Grau A, Guerra E, Shen Z, Hu J, Fan H. Semi-supervised wildfire smoke detection based on smoke-aware consistency. *Front Plant Sci* 2022; 13: 980425. doi:10.3389/fpls.2022.980425.
- [18] Zhao C, Zhao L, Zhang K, Ren Y, Chen H, Sheng Y. Smoke and fire – You only look once: a lightweight deep learning model for video smoke and flame detection in natural scenes. *Fire* 2025; 8(3): 104. doi:10.3390/fire8030104.
- [19] Xu G, Zhang Y, Zhang Q, Lin G, Wang Z, Jia Y, et al. Video smoke detection based on deep saliency network. *Fire Saf J* 2019; 105: 277-285. doi:10.1016/j.firesaf.2019.03.012.
- [20] Favorskaya M, Pyataeva A, Popov A. Verification of smoke detection in video sequences based on spatio-temporal local binary patterns. *Procedia Comput Sci* 2015; 60: 671-680. doi:10.1016/j.procs.2015.08.205.
- [21] Wang Y. Smoke recognition based on machine vision. In: *Proc Int Symp Comput Consum Control (IS3C)*; 2016. pp. 668-671. doi:10.1109/IS3C.2016.172.
- [22] Del-Pozo-Velázquez J, Aguiar-Pérez JM, Chamorro-Posada P, Pérez-Juárez MA, Wang X, Casaseca-de-la-Higuera P. Smoke detection in images through fractal dimension-based binary classification. *Digit Signal Process* 2025; 166: 105346. doi:10.1016/j.dsp.2024.105346.
- [23] Toreyin BU, Dedeoglu Y, Cetin AE. Wavelet-based real-time smoke detection in video. In: *Proc 13th Eur Signal Process Conf (EUSIPCO)*; 2005; Antalya, Turkey. pp. 1-4.
- [24] Gubenko DD. Study of image filtering algorithms in the problem of improving the quality of smoke detection in a video stream [in Russian]. *Izv Tula State Univ Tech Sci* 2022; (10): 23-29. doi:10.24412/2071-6168-2022-10-23-29.
- [25] Wiegand T, Sullivan GJ, Bjontegaard G, Luthra A. Overview of the H.264/AVC video coding standard. *IEEE Trans Circuits Syst Video Technol* 2003; 13(7): 560-576.
- [26] Hsu YC, Huang TH, Hu TY, Dille P, Prendi S, Hoffman R, et al. Project RISE: recognizing industrial smoke emissions. *Proc AAAI Conf Artif Intell* 2021; 35(17): 14813-14821.
- [27] Gagliardi A, de Gioia F, Saponara S. A real-time video smoke detection algorithm based on Kalman filter and CNN. *J Real-Time Image Process* 2021; 18(6): 1963-1975. doi:10.1007/s11554-021-01094-y.
- [28] Saponara S, Elhanashi A, Gagliardi A. Real-time video fire/smoke detection based on CNN in antifire surveillance systems. *J Real-Time Image Process* 2020; 18(3): 631-642. doi:10.1007/s11554-020-01044-0.
- [29] Dewangan A, Pande Y, Braun HW, Vernon FL, Perez I, Altintas I, et al. FlgLib & SmokeyNet: dataset and deep learning model for real-time wildland fire smoke detection. *Remote Sens* 2022; 14(4): 1007. doi:10.3390/rs14041007.
- [30] Fernandes AM, Utkin AB, Chaves P. Automatic early detection of wildfire smoke with visible light cameras using deep learning and visual explanation. *IEEE Access* 2022; 10: 12814-12828. doi:10.1109/ACCESS.2022.3145911.
- [31] Yazdi A, Qin H, Jordan CB, Yang L, Yan F. Nemo: an open-source transformer-supercharged benchmark for fine-grained wildfire smoke detection. *Remote Sens* 2022; 14(16): 3979. doi:10.3390/rs14163979.
- [32] Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* 2020; 21(1): 6. doi:10.1186/s12864-019-6413-7.
- [33] FireWarn: fire hazards detection using deep learning models. *IEEE Xplore* 2021. Available at: <https://ieeexplore.ieee.org/document/9750361>. doi:10.1109/CogMI52975.2021.00010.
- [34] Huang J, He Z, Guan Y, Zhang H. Real-time forest fire detection by ensemble lightweight YOLOX-L and defogging method. *Sensors* 2023; 23(4): 1894. doi:10.3390/s23041894.
- [35] Huang J, Zhou J, Yang H, Liu Y, Liu H. Small-target forest fire smoke detection model based on deformable transformer for end-to-end object detection. *Forests* 2023; 14(1): 162. doi:10.3390/f14010162.
- [36] Saydirasulovich SN, Mukhiddinov M, Djuraev O, Abdusalomov A, Cho YI. Improved wildfire smoke detection based on YOLOv8 and UAV images. *Sensors* 2023; 23(20): 8374. doi:10.3390/s23208374.
- [37] Hu Y, Zhan J, Zhou G, Chen A, Cai W, Guo K, et al. Fast forest fire smoke detection using MVMNet. *Knowl-Based Syst* 2022; 241: 108219. doi:10.1016/j.knosys.2022.108219.
- [38] Gupta T, Liu H, Bhanu B. Early wildfire smoke detection in videos. In: *Proc Int Conf Pattern Recognit (ICPR)*; 2021. pp. 8523-8530. doi:10.1109/ICPR48806.2021.9413231.
- [39] Sheng D, Deng J, Xiang J. Automatic smoke detection based on SLIC-DBSCAN enhanced convolutional neural network. *IEEE Access* 2021; 9: 63933-63942. doi:10.1109/ACCESS.2021.3075731.
- [40] Xia X, Peng Y, Li Z, Shi J, Fang Y. Texture-aware network for enhancing inner smoke representation in visual smoke density estimation. *IET Comput Vis* 2025; 19(1): 56-70. doi:10.1049/cvi2.70023.
- [41] Lin T.-Y., Maire M., Belongie S., Hays J., Perona P., Ramanan D., Dollár P., Zitnick C.L. Microsoft COCO: Common Objects in Context // *European Conference on Computer Vision (ECCV)*. – 2014. – P. 740–755. – DOI: 10.1007/978-3-319-10602-1_48.
- [42] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: *Proc IEEE Int Conf Comput Vis (ICCV)*; 2017. pp. 618-626. doi:10.1109/ICCV.2017.74.

- [43] Li T, Zhao E, Zhang J, Hu C. Detection of wildfire smoke images based on a densely dilated convolutional network. *Electronics* 2019; 8(10): 1131. doi:10.3390/electronics8101131.
- [44] Gragnaniello D, Greco A, Sansone C, Vento B. Fire and smoke detection from videos: a literature review under a novel taxonomy. *Expert Syst Appl* 2024; 255: 124783. doi:10.1016/j.eswa.2024.124783.
- [45] Xu M, Liu T, Zhu K, He Y. Smoke source localization using optical flow imagery in video-based detection. In: *Proc 14th Int Congr Image Signal Process Biomed Eng Informatics (CISP-BMEI)*; 2024. pp. 1-6.
- [46] Shi J, Wang W, Gao Y, Yu N. Optimal placement and intelligent smoke detection algorithm for wildfire-monitoring cameras. *IEEE Access* 2020; 8: 72326-72339. doi:10.1109/ACCESS.2020.2987991.
- [47] Fan X, Lei F, Yang K. Real-time detection of smoke and fire in the wild using unmanned aerial vehicle remote sensing imagery. *Forests* 2025; 16(2): 201. doi:10.3390/f16020201.
- [48] Song H, Chen Y. Video smoke detection method based on cell root-branch structure. *Signal Image Video Process* 2024; 18(5): 4851-4859. doi:10.1007/s11760-024-03121-4.
- [49] Ye S, Bai Z, Chen H, Bohush R, Ablameyko S. Effective algorithm to detect both smoke and flame using color and wavelet analysis. *Pattern Recognit Image Anal* 2017; 27(1): 131-138. doi:10.1134/S1054661817010138.
- [50] Bohush R, Adamovskiy Y, Chen H. Real-time smoke detection in video [in Russian]. *Herald Polotsk State Univ Ser Fundam Sci* 2023; 41(2): 2-8. doi:10.52928/2070-1624-2023-41-2-2-8.
- [51] Yuan F, Shi J, Xia X, Fang Y, Fang Z, Mei T. High-order local ternary patterns with locality preserving projection for smoke detection and image classification. *Inf Sci* 2016; 372: 225-240. doi:10.1016/j.ins.2016.08.040.
- [52] Sun Y. Describing motion in computer images' stream: optical flow. In: *Computer Vision for Multimedia Applications*; 2024. pp. 3-21. doi:10.1007/978-981-99-9191-4_2.
- [53] Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans Pattern Anal Mach Intell* 2012; 34(11): 2274-2282. doi:10.1109/TPAMI.2012.120.
- [54] Rehman SU, Asghar S, Fong S, Sarasvady S. DBSCAN: past, present and future. In: *Proc 5th Int Conf Appl Digit Inf Web Technol (ICADIWT)*; 2014. doi:10.1109/ICADIWT.2014.6814687.
- [55] Vaishnav P, Sankarasubramanian P. Realtime pipeline fire and smoke detection using a lightweight CNN model. In: *Proc Int Conf Mach Learn Appl Neural Technol (ICMLANT)*; 2021. doi:10.1109/ICMLANT53170.2021.9690545.
- [56] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *arXiv* 2017; arXiv:1706.03762. Available at: <https://arxiv.org/abs/1706.03762>.
- [57] Jiang P, Ergu D, Liu F, Cai Y, Ma B. Review of YOLO algorithm developments. *Procedia Comput Sci* 2022; 199: 1066-1073. doi:10.1016/j.procs.2022.01.135.
- [58] Özel B, Alam MS, Khan MU. Review of modern forest fire detection techniques: innovations in image processing and deep learning. *Information* 2024; 15(9): 538. doi:10.3390/info15090538.
- [59] Shi X, Lu N, Cui Z. Smoke detection based on dark channel and convolutional neural networks. In: *Proc Big Data Intell Anal Conf (BigDIA)*; 2019. pp. 23-28. doi:10.1109/BigDIA.2019.8802668.
- [60] Jadon A, Omama M, Varshney A, Ansari MS, Sharma R. FireNet: a specialized lightweight fire and smoke detection model for real-time IoT applications. *arXiv* 2019; arXiv:1905.11922. Available at: <https://arxiv.org/abs/1905.11922>.
- [61] FireNET: deep learning model for detecting fire in video and camera streams. Available at: <https://github.com/OlafenwaMoses/FireNET>.
- [62] Khodarahmi M, Maihami V. Review on Kalman filter models. *Arch Comput Methods Eng* 2022; 30: 727-747. doi:10.1007/s11831-022-09815-7.
- [63] FireSense dataset. Available at: <https://www.kaggle.com/datasets/chrisfilo/firesense>.
- [64] Filonenko A, Hernandez DC, Jo KH. Fast smoke detection for video surveillance using CUDA. *IEEE Trans Ind Inform* 2018; 14(2): 725-733. doi:10.1109/TII.2017.2757457.
- [65] Li H, Li J, Wei H, Liu Z, Zhan Z, Ren Q. Slim-neck by GSConv: a better design paradigm of detector architectures for autonomous vehicles. *J Real-Time Image Process* 2024; (in press). doi:10.1007/s11554-024-01436-6. Available at: <https://arxiv.org/abs/2206.02424>.
- [66] Zhu L, Wang X, Ke Z, Zhang W, Lau R. BiFormer: vision transformer with bi-level routing attention. In: *Proc IEEE Conf Comput Vis Pattern Recognit (CVPR)*; 2023. doi:10.1109/CVPR52729.2023.00995.
- [67] Han Y, Li B, Hu Y, et al. LUFFD-YOLO: A lightweight model for UAV remote sensing forest fire detection based on attention mechanism and multi-level feature fusion. *Remote Sens* 2024; 16(12): 2177. doi: 10.3390/rs16122177.
- [68] Wang G. Multiple scenarios, multiple weather conditions, multiple lighting conditions and multiple wildfire objects synthetic forest wildfire dataset (M4SFWD). *IEEE Dataport* 2024. doi:10.21227/m9kz-bw61. Available at: <https://iee-dataport.org/documents/multiple-scenarios-multiple-weather-conditions-multiple-lighting-conditions-and-multiple>.
- [69] Fire-Smoke-Dataset. Available at: <https://github.com/DeepQuestAI/Fire-Smoke-Dataset>.
- [70] Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S. End-to-end object detection with transformers. *arXiv* 2020; arXiv:2005.12872. Available at: <https://arxiv.org/abs/2005.12872>.
- [71] HPWREN FIgLib data. Available at: <https://cdn.hpwren.ucsd.edu/HPWREN-FIgLib-Data/index.html>.
- [72] ALERT Wildfire camera network. Available at: <https://www.alertwildfire.org/>.
- [73] Zhu X, Su W, Lu L, Li B, Wang X, Dai J. Deformable DETR: deformable transformers for end-to-end object detection. *arXiv* 2021; arXiv:2010.04159. Available at: <https://arxiv.org/abs/2010.04159>.
- [74] Yuan L, Chen Y, Wang T, Yu W, Shi Y, Francis J, et al. Tokens-to-token ViT: training vision transformers from scratch on ImageNet. *arXiv* 2021; arXiv:2101.11986. Available at: <https://arxiv.org/abs/2101.11986>.
- [75] Huo Y, Zhang Q, Jia Y, Liu D, Guan J, Lin G, et al. Deep separable convolutional neural network for multiscale image-based smoke detection. *Fire Technol* 2022; 58(3): 1445-1468. doi:10.1007/s10694-021-01199-7.
- [76] Yang C, Pan Y, Cao Y, Lu X. CNN-Transformer hybrid architecture for early fire detection. *Lect Notes Comput Sci* 2022; pp. 570-581. doi:10.1007/978-3-031-15937-4_48.

- [77] Tao H, Duan Q. Adaptive frame selection network with enhanced dilated convolution for video smoke recognition. *Expert Syst Appl* 2023; 225: 119371. doi:10.1016/j.eswa.2022.119371.
- [78] Chen X, An Q, Yu K, Ban Y. Novel fire identification algorithm based on improved color segmentation and enhanced feature data. *IEEE Trans Instrum Meas* 2021; 70: 1-15. doi:10.1109/TIM.2021.3075380.
- [79] Vinicius P, Campos RJ, Rezende TM, Lisboa AC, Barbosa AV. Hybrid method for fire detection based on spatial and temporal patterns. *Neural Comput Appl* 2023; 35(13): 9349-9361. doi:10.1007/s00521-023-08260-2.
- [80] Huo Y, Zhang Q, Zhang Y, Zhu J, Wang J. 3DVSD: an end-to-end 3D convolutional object detection network for video smoke detection. *Fire Saf J* 2022; 134: 103690. doi:10.1016/j.firesaf.2022.103690. [web:341][web:344]
- [81] Cao Y, Yang F, Tang Q, Lu X. Attention-enhanced bidirectional LSTM for early forest fire smoke recognition. *IEEE Access* 2019; 7: 154732-154742. doi:10.1109/ACCESS.2019.2946712.
- [82] Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. *arXiv* 2017; arXiv:1706.08500. Available at: <https://arxiv.org/abs/1706.08500>.
- [83] Xu G, Zhang Q, Liu D, Li G, Wang JJ, Zhang Y. Adversarial adaptation from synthesis to reality in fast detector for smoke detection. *IEEE Access* 2019; 7: 29471-29483. doi:10.1109/ACCESS.2019.2902606.
- [84] Tong Z, Chen Y, Xu Z, Yu R. Wise-IoU: bounding box regression loss with dynamic focusing mechanism. *arXiv* 2023; arXiv:2301.10051. Available at: <https://arxiv.org/abs/2301.10051>.
- [85] Gevorgyan Z. SIoU loss: more powerful learning for bounding box regression. *arXiv* 2022; arXiv:2205.12740. Available at: <https://arxiv.org/abs/2205.12740>.
- [86] dsw2017 smoke dataset. Available at: <https://github.com/mtcazzolato/dsw2017>.
- [87] VisiFire demo videos. Available at: <http://signal.ee.bilkent.edu.tr/VisiFire/Demo>.
- [88] D-Fire videos. Available at: <https://clk.li/BhNw>.
- [89] MIVIA smoke dataset. Available at: <https://mivia.unisa.it/datasets/video-analysis-datasets/smoke-detection-dataset/>.
- [90] furg-fire-dataset. Available at: <https://github.com/steffensbola/furg-fire-dataset>.
- [91] Computer Vision and Pattern Recognition Laboratory homepage. Available at: <https://cvpr.kmu.ac.kr/>.
- [92] ImageNet dataset. Available at: <https://image-net.org/>

Сведения об авторах

Адамовский Егор Русланович. Старший преподаватель кафедры вычислительных систем и сетей Полоцкого государственного университета имени Евфросинии Полоцкой. Научные интересы: компьютерное зрение, машинное обучение, обработка изображений и видео. E-mail: e.adamovsky@psu.by ORCID: 0000-0003-1044-8741

Богуш Рихард Петрович. Доктор технических наук, заведующий кафедрой вычислительных систем и сетей в Полоцком государственном университете имени Евфросинии Полоцкой. Научные интересы: обработка изображений и видео, интеллектуальные системы и машинное обучение. E-mail: bogushr@mail.ru ORCID: 0000-0002-6609-5810

Абламейко Сергей Владимирович. Академик Национальной академии наук Беларуси, доктор технических наук, профессор. Профессор механико-математического факультета Белорусского государственного университета. Научные интересы: искусственный интеллект, компьютерное зрение, системы, основанные на знаниях, географические информационные системы и медицинская визуализация. E-mail: ablameyko@bsu.by ORCID: 0000-0001-9404-1206

Поступила в редакцию 12 сентября 2025 г.



Окончательный вариант – 22 декабря 2025 г.

Computer vision algorithms for detecting smoke in video sequences

Y.R. Adamovskiy¹, R.P. Bohush¹, S.V. Ablameyko²

¹ Polotsk State University, Blokhina St., 29, Novopolotsk, 211440, Republic of Belarus;

² Belarusian State University, Prospekt Nezavisimosti 4, Minsk, 220030, Republic of Belarus

Abstract

This paper presents an analytical survey of methods and algorithms for smoke detection in image sequences, aimed at early fire detection using video surveillance systems. Static and dynamic visual features of smoke that enable its discrimination from visually similar phenomena are examined. Limiting and interfering factors that degrade the visibility of smoke visual features in real-world video frames are discussed. Criteria for evaluating the effectiveness of smoke detection and localization are presented. Approaches based on classical computer vision techniques, as well as machine learning methods, including convolutional neural networks and transformers, are analyzed. Machine learning-based algorithms are systematically categorized according to computational cost. Existing datasets for training smoke detection models are reviewed, along with image augmentation techniques employed to expand the training set.

Keywords: smoke detection, spatio-temporal smoke signatures, video surveillance systems, computer vision, neural networks.

Citation: Adamovskiy YR, Bohush RP, Ablameyko SV. Computer vision algorithms for detecting smoke in video sequences. *Computer Optics* 2026; 50(4): 1827. doi: 10.18287/COJ1827.

About authors

Yahor Adamovskiy. Senior Lecturer at the Department of Computer Systems and Networks, Polotsk State University. Research interests: computer vision, machine learning, image and video processing. E-mail: e.adamovsky@psu.by ORCID: 0000-0003-1044-8741

Rykhhard Bohush. Doctor of Technical Sciences, Head of Computer Systems and Networks department of Polotsk State University. Research Interests: image and video processing, intelligent systems, and machine learning. E-mail: bogushr@mail.ru ORCID: 0000-0002-6609-5810

Sergey Ablameyko. Academician of the National Academy of Sciences of Belarus, Doctor of Technical Sciences, Professor. Professor of the Faculty of Mechanics and Mathematics of the Belarusian State University. Research Interests: artificial intelligence, computer vision, knowledge based systems, geographical information systems, medical imaging. E-mail: ablameyko@bsu.by ORCID: 0000-0001-9404-1206

Received September 12, 2025. The final version – December 22, 2025.
